Distributed Adaptive Signal Processing for Frequency Estimation

Shri Siva Sithan Kanagasabapathy
Department of Electrical and Electronic Engineering
Imperial College London

A thesis submitted for the degree of
Doctor of Philosophy (Ph.D.)
December 2016

To the people and events that made this possible.

Abstract

It is widely recognised that future smart grids will heavily rely upon intelligent communication and signal processing as enabling technologies for their operation. Traditional tools for power system analysis, which have been built from a circuit theory perspective, are a good match for balanced system conditions. However, the unprecedented changes that are imposed by smart grid requirements, are pushing the limits of these old paradigms.

To this end, we provide new signal processing perspectives to address, some fundamental operations in power systems such as frequency estimation, regulation and fault detection. Firstly, motivated by our finding that any excursion from nominal power system conditions results in a degree of non-circularity in the measured variables, we cast the frequency estimation problem into a distributed estimation framework for noncircular complex random variables. Next, we derive the required next generation widely linear, frequency estimators which incorporate the so-called augmented data statistics and cater for the noncircularity and a widely linear nature of system functions. Uniquely, we also show that by virtue of augmented complex statistics, it is possible to treat frequency tracking and fault detection in a unified way.

To address the ever shortening time-scales in future frequency regulation tasks, the developed distributed widely linear frequency estimators are equipped with the ability to compensate for the fewer available temporal voltage data by exploiting spatial diversity in wide area measurements. This contribution is further supported by new physically meaningful theoretical results on the statistical behaviour of distributed adaptive filters. Our approach avoids the current restrictive assumptions routinely employed to simplify the analysis by making use of the collaborative learning strategies of distributed agents. The efficacy of the proposed distributed frequency estimators over standard strictly linear and stand-alone algorithms is illustrated in case studies over synthetic and real-world three-phase measurements.

An overarching theme in this thesis is the elucidation of underlying commonalities between different methodologies employed in classical power engineering and signal processing. By revisiting fundamental power system ideas within the framework of augmented complex statistics, we provide a physically meaningful signal processing perspective of three-phase transforms and reveal their intimate connections with spatial discrete Fourier transform (DFT), optimal dimensionality reduction and frequency demodulation techniques. Moreover, under the widely linear framework, we also show that the two most widely used frequency estimators in the power grid are in fact special cases of frequency demodulation techniques.

Finally, revisiting classic estimation problems in power engineering through the lens of noncircular complex estimation has made it possible to develop a new self-stabilising adaptive three-phase transformation which enables algorithms designed for balanced operating conditions to be straightforwardly implemented in a variety of real-world unbalanced operating conditions. This thesis therefore aims to help bridge the gap between signal processing and power communities by providing power system designers with advanced estimation algorithms and modern physically meaningful interpretations of key power engineering paradigms in order to match the dynamic and decentralised nature of the smart grid.

Contents

List of Figures

List of Tables

Nomenclature

ACRONYMS

ACEKF	(Augmented) Complex extended Kalman filter
ACLMS	(Augmented) Complex least mean square
AR	Autoregressive
CRLB	Cramer-Rao lower bound
D-ACEKF	Diffusion-augmented complex extended Kalman filter
D-ACLMS	Diffusion-augmented complex least mean square
DFT	Discrete Fourier transform
FM	Frequency modulation
MLE	Maximum likelihood estimation
MSE	Mean square error
RoCoF	Rate of change of frequency
SLAR	Strictly linear autoregressive
TSO	Transmission system operator
WLAR	Widely linear autoregressive

SYMBOLS

$\mathbf{1}_N$	$N \times 1$ column vector of ones
$\mathbf{A}$	Matrix
$\boldsymbol{a}$	Vector
$\bar{\boldsymbol{x}}$	Augmented input vector
$\mathbb{C}$	Field of complex numbers
$\mathbb{R}$	Field of real numbers
ρ	Circularity quotient
a	Scalar
j	Imaginary unit, $\sqrt{-1}$
$\mathbf{I}_M$	$M \times M$ Identity matrix

OPERATORS

$(\cdot)^*$	Complex conjugation
$(\cdot)^{\mathsf{H}}$	Vector/Matrix conjugate transposition
$(\cdot)^{\mathsf{T}}$	Vector/Matrix transposition
$\stackrel{\text{def}}{=}$	Defined as
$\mathbb{E}\{\cdot\}$	Statistical expectation
$\mathsf{Im}\{\cdot\}$	Imaginary part
$\otimes$	Kronecker product
$\mathsf{Re}\{\cdot\}$	Real part
$\mathsf{blkdiag}(\cdot)$	Creates a block-diagonal matrix of its arguments
$\mathsf{col}(\cdot)$	Creates a column vector of its arguments
$\mathsf{diag}(\cdot)$	Creates a diagonal matrix of its arguments
$\arg[\cdot]$	Argument (extracts angle of a complex-valued scalar)
$\mathsf{Tr}[\cdot]$	Trace of a square matrix
$\varrho(\cdot)$	Spectral radius of a square matrix

Statement of Originality

I declare that the work presented in this thesis is the outcome of my research under the supervision of Prof. Danilo P. Mandic. Ideas and results from the work of others are appropriately acknowledged with the standard referencing practices employed in this discipline. The material in this thesis has not been submitted for any degree at any other academic or professional institution.

Copyright Declaration

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Danilo Mandic who gave me the opportunity to undertake this PhD and for his support during the tough times. During the last four years, through his encouragement and guidance, he has given me countless opportunities to develop as a person and a scholar. There is still a lot left to learn from him but I hope that I was able to extract as much knowledge as a student during this period.

Next, I am thankful to Prof. Anthony Constantinides, Prof. Scott Douglas, Prof. Ron Hui, Dr. Sudhin Thayyil and Dr. Yili Xia for their enlightening advice and comments which greatly benefited this thesis and its related publications. Furthermore, I thank Dr. Adrià Junyent-Ferré and Dr. Yousef Pipelzadeh for their time, effort and invaluable insights on the research regarding distributed synthetic inertia provision. I am extremely thankful to my examiners, Prof. Lajos Hanzo and Dr. András György for their valuable input which improved the rigour and clarity of this thesis.

I am also grateful to my former colleagues Dahir Dini, Cyrus Jahanchahi and Felipe Tobar for showing me the ropes during the early days. It is important to also thank my collaborators, Sayed Pouria Talebi and Min Xiang, with whom I have had multiple interesting experiences and discussions throughout the years (especially towards the last few days before a deadline)! I also had the pleasure working with Alex Stott, Ilia Kisil and Thayne Thanthawaritthisai who have a great future ahead of them.

I would like to thank my friend and colleague, John Murray-Bruce who introduced me to Prony's early work which formed an integral part in my understanding of the frequency estimation problem and Richard Stanton for being a good friend during my time at Imperial. I also thank Rob Tod for being a kind friend over the years and his lovely family for all the memorable times in Llanddewi Rhydderch.

Nevertheless, I am grateful to the CSP group administrator Ms. Charlotte Grady, and graduate tutor, Dr. Imad Jaimoukha, for all their help during my PhD. My experiences in the CSP group has been enriched by the lunch/social chats with Valentin, David, Thanos, Ali, Wilhelm, Alicia, Takashi, Tricia, Apit, Theerasak, Steph, Hamdi, Michael and Charlie, where conversations ranged from the mundane to the bizarre. Thank you all.

I would also like to acknowledge my aunt who has been extremely supportive in my progress over the last two years and for understanding my unsociable working hours. Last but not least, my heartfelt gratitude goes to my mother and father who provided me their love, support and guidance on crucial issues while also giving me a lot of freedom to explore, question and discover. I would not be here without them.

Sithan Kanna
December 2016

Publications

Submitted Journals

[J1] S. Kanna and D. P. Mandic, "Modern view of three phase transforms," *Submitted to IEEE Power Electronics Magazine*, 2016.

[J2] S. Kanna, A. G. Constantinides, and D. P. Mandic, "The equivalence of FM demodulation and recursive DFT for frequency tracking in power systems," *Submitted to International Journal of Electrical Power & Energy Systems*, 2016.

[J3] Y. Xia, S. Kanna, and D. Mandic, "Maximum likelihood frequency estimation of unbalanced three-phase power signals, *Submitted to IEEE Transactions on Power Systems*, 2016.

Published Peer-Reviewed Journals

[J4] S. Kanna and D. P. Mandic, "Steady-state behavior of general complex-valued diffusion LMS strategies," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 722–726, May 2016.

[J5] D. P. Mandic, S. Kanna, and A. G. Constantinides, "On the intrinsic relationship between the least mean square and Kalman filters [Lecture Notes]," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 117122, Nov 2015.

[J6] S. Kanna, D. H. Dini, Y. Xia, S. Y. Hui, and D. P. Mandic, "Distributed widely linear Kalman filtering for frequency estimation in power networks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 1, no. 1, pp. 45–57, Mar 2015.

[J7] C. Jahanchahi, S. Kanna, and D.P. Mandic, "Complex dual channel estimation: Cost effective widely linear adaptive filtering," *Signal Processing*, 104:3342, Nov 2014.

Journals in Preparation

[J8] S. Kanna, Y. Pipelzadeh, A. Junyent-Ferre, T. Green, and D. Mandic, "Aggregating synthetic inertia using wind turbines: A rapid frequency response solution for low inertia grids," *In preparation for IEEE Transactions on Power Systems*, 2016.

Published Peer-Reviewed Conference Proceedings

[C1] S. Kanna, M. Xiang, and D. P. Mandic, "Real-time detection of rectilinear sources for wireless communication signals," in *Proc. of the 2015 International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2015, pp. 541–545.

[C2] D.P. Mandic, S. Kanna, and S.C. Douglas, "Mean square analysis of the CLMS and ACLMS for non-circular signals: The approximate uncorrelating transform approach," in *Proc. of the IEEE Intl. Conf. on Acoust. Speech and Signal Process. (ICASSP)*, Apr 2015, pp. 3531–3535.

[C3] S. Kanna, S.C. Douglas, and D.P. Mandic, "A real time tracker of complex circularity," in *Proc. of the 8th IEEE Sensor Array and Multichannel Signal Process. Workshop (SAM)*, Jun 2014, pp. 129–132.

[C4] S. Kanna, S.P. Talebi, and D.P. Mandic, "Diffusion widely linear adaptive estimation of system frequency in distributed power grids," in *Proc. of the IEEE Intl. Energy Conf. (ENERGYCON)*, May 2014, pp. 772–778.

Chapter 1

Introduction

> It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment.
>
> *Carl Friedrich Gauss*

1.1 Perspective

Modern electricity grids are undergoing unprecedented changes to meet several ambitious goals. These goals include the need to reduce our reliance on fossil fuels and better management of the current ageing infrastructure, while creating economic benefits for society [1]. For example, legally binding CO2 emissions targets require the UK to reduce emissions by 32% by 2020 and 60% by 2050 compared to the baseline emissions levels in 1990 [2]. Fulfilling each of these demands compounds the number of challenges and their impact on transmission system operators (TSOs), utilities, and end-consumers alike.

It is widely expected that these challenges are to be met by the next generation power network, commonly known as the "smart grid". The smart grid aims to make a paradigm shift from the top-down mode of operation of the current grid by incorporating distributed generation, typically from renewable sources (wind, solar), distributed storage from plug-in-vehicles and demand response technologies [1]. The backbone to this future grid is an intelligent real-time wide area monitoring system (WAMS) capable of estimating key indicators of the stability of the grid [3].

One of the most closely monitored indicator is the system frequency as its deviation from the nominal frequency of 50 Hz or 60 Hz indicates a mismatch between the supply and demand of electricity or a fault. Since electricity is a non-storable commodity, its generation and demand have to balanced in real time. If demand is greater than generation, the frequency drops, whereas if generation is greater than demand, the frequency rises. Therefore, accurate tracking of the system frequency is a prerequisite for the task of balancing the demand and generation of electricity [4]. Figure 1.1 shows a scene form the National Grid control centre where the system frequency is monitored continuously.

Current power system frequency tracking algorithms were designed in the early 1980s and would be adequate for the smart grid if the dynamics of the grid remain within the status quo [5]. However, future grids are expected to have vastly different dynamics including more rapid frequency deviations [6], together with a multitude of power quality issues (e.g. more noise, voltage sags, harmonics). To understand the evolution of the issues surrounding these problems, we shall briefly discuss the history of the electricity grid, the present frequency regulation

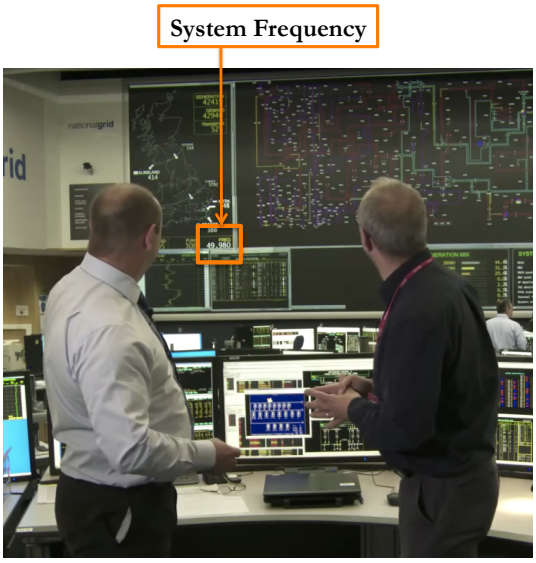

Figure 1.1: National Grid control centre. Scene taken from the show "Fully Charged", aired on May 2012.

landscape and the anticipated challenges that are transforming the requirements for frequency tracking algorithms in the future.

1.2 System Frequency: Past, Present and Future

1.2.1 Historical Context

The electrification of objects has been one of mankind's greatest triumphs, earning it the title "*the greatest engineering achievement of the 20th Century*" [7]. In its early years, the electric grid was far less homogeneous than it is today, where disparate electricity suppliers and equipment manufacturers adopted different standards which suited their own applications. Of these standards, the most important debate centred around the choice between alternating current (AC) and direct current (DC) in the so-called "War of Currents" between Westinghouse who favoured AC and Edison who favoured DC [8].

The AC system eventually "won" the war since it offered several advantages compared to DC system. Firstly, using AC allowed for the voltages to be stepped up and down with transformers which lowered the cost of transmission as cheaper copper wires could be used to supply the same amount of power with lower currents. Secondly, AC systems could cater for loads designed for different voltages with the same transmission lines through the use of transformers, unlike DC systems which required separate DC transmission lines for different load ratings.

Once the war of currents paved the way for the AC framework, the subsequent standardisation debate was regarding the AC system frequency. Prior to the standardisation of the frequency to 50/60 Hz, the AC supply around the world existed at arbitrary frequencies which ranged from $16\frac{2}{3}$ Hz to 133 Hz. The choice of the system frequency came from different (often

times opposing) needs of the various electrical equipments like incandescent bulbs, transformers, induction motors and synchronous converters [9]. Generally, the system frequency needed to be high enough to have less flickering in light-bulbs and for the flux rate in the transformers to be sufficiently high for an efficient operation. However, having a high frequency required generators to spin at higher rotations speeds and contain more poles – this was considered to be more hazardous than generators operating at lower speeds [10]. Besides that, transmitting power through long distance cables with a high AC frequency worsened the transmission losses in the cables through what is known as "skin effect" [11].

The 60 Hz frequency was standardised in the US since the 50 Hz frequency introduced considerable flickering in the arc lighting system which was widespread in the US in the late 19th century [12]. The reason behind the precise choice of 50 Hz and 60 Hz (as opposed to e.g. 52 Hz or 61 Hz) is less clear. Speculations range from the observation that 60 Hz fits into the units of time where there are 60 seconds in a minute and 60 minutes in an hour [12], while "50" was chosen since it was a commonly used value in industrial design applications [10].

1.2.2 Present Landscape

Transmission system operators (TSO) have a legal requirement to maintain the system frequency within a pre-determined range. For example, the National Grid in the U.K. is required to maintain the system frequency within 1% of the nominal frequency of 50 Hz. However, in the event of an unexpected large generator loss ($>$ 1.8 gigawatts (GW)), the frequency is allowed to deviate to a minimum of 49.2 Hz [13].

Frequency regulation actions can be broadly split into two categories: (i) under nominal operating conditions and (ii) during a contingency. During nominal operating conditions, the TSO plans and manages the supply-demand balance at different timescales:

- **Long term (timescale: 1 year to 10 years).** The long term planning required to ensure an adequate supply of power is conducted by the TSO. The National Grid, for example, releases the "Electricity 10-year Statement" outlining the electricity transmission planning cycle and the future needs of the transmission network [14].
- **Monthly to intra-day (timescale: several months to few hours ahead).** In this time scale, electricity generators and consumers trade power contracts to fulfil the demands in the market. Intra-day balancing is carried out by power schedulers who have a range of options that include increasing or curtailing generation to load shedding and exporting any deficits to neighbouring markets [15].

When a contingency occurs, the supply-demand balance is temporarily disrupted due to either a large generator loss (supply-side), or a sudden change in the load (demand-side). Responses of the grid during such contingencies typically resemble the following scenario, which is also graphically depicted in Figure 1.2 [13]:

- **Inertial response (timescale: instantaneous up to 10s).** *Synchronous electrical machines* are electrical motors and generators that rotate at a rate which is directly proportional to the system frequency (e.g. 50 Hz). Their rotational speed therefore changes when the system frequency changes. However, by virtue of the kinetic energy stored within them, these

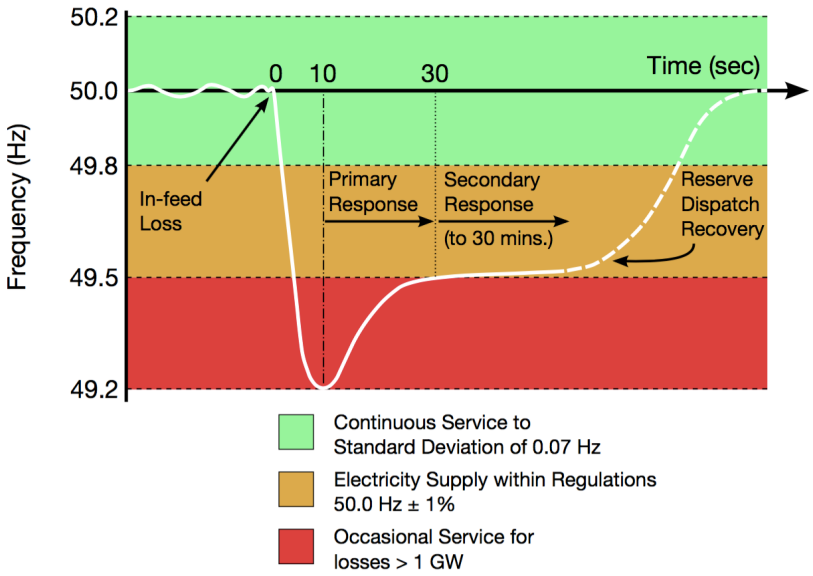

Figure 1.2: Illustration of the frequency response time-scale during a generation loss.

electrical machines resist any sudden changes to their state of motion and consequently regulate changes in the supply and demand of electricty in the very beginning of a contingency. The phenomenon which describes the resistance an electrical machine has to any change in its state of motion is referred to as *inertia*.

- **Primary response (timescale: from 10s up to 30s).** As soon as a contingency occurs, a primary response is automatically supplied by the governors within synchronous electrical machines which change the power output to stabilise the frequency.
- **Secondary response (timescale: from 30s up to 30mins).** Additional power is dispatched by the grid operators from the reserve power available in the grid which may come either from power plants that operate below their capacity or energy storage devices.

1.2.3 Future Grids: The Challenge

The key challenge for the frequency regulation scheme described in the previous section is the large penetration of intermittent renewable resources (wind, solar) which is expected to: (i) shorten the frequency regulation time-scales [16], and (ii) degrade the quality of the current/voltage signals used by conventional control and estimation algorithms [17].

Firstly, the frequency regulation time-scales are expected to be shortened in future grids, since replacing a large fraction of conventional generators by renewable energy sources is expected to reduce the overall inertia available in the grid [6]. The reason for this is that, unlike conventional power plants which are interfaced to the grid via synchronous machines, renewable sources are interfaced to the grid by power electronic inverters. These inverters are electronic circuits which do not provide any inertial support to the grid. In the United Kingdom,

the transmission system operator estimates a 40% – 50% reduction of the available inertia in the Great Britain (GB) network by 2035 [18], see Figure 1.3.

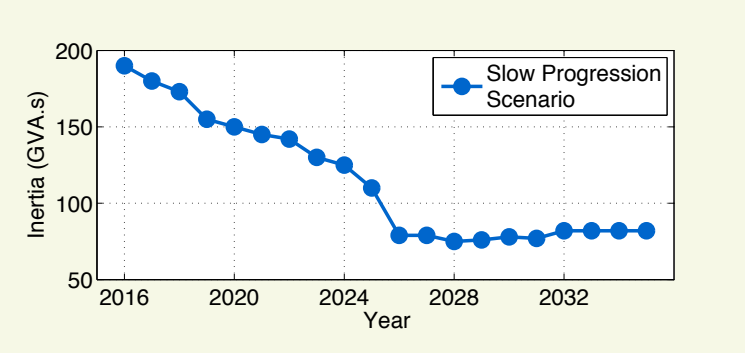

Figure 1.3: Forecast of available inertia in the grid, excluding embedded generation under the "slow progression" scenario, measured in Giga VoltAmpere seconds (GVA.s).

This is of great concern as the system inertia is a key factor in determining the rate of change of frequency (RoCoF) when a sudden power imbalance occurs (e.g. the loss of a generation unit) [16], see Figure 1.4. Systems with higher inertia have lower RoCoFs, and are considered to be more stable than low-inertia systems. Power systems with higher RoCoFs cause more mechanical stress for generators and exhibit more extreme frequency excursions, necessitating a rapid response from transmission system operators (TSO) [6].

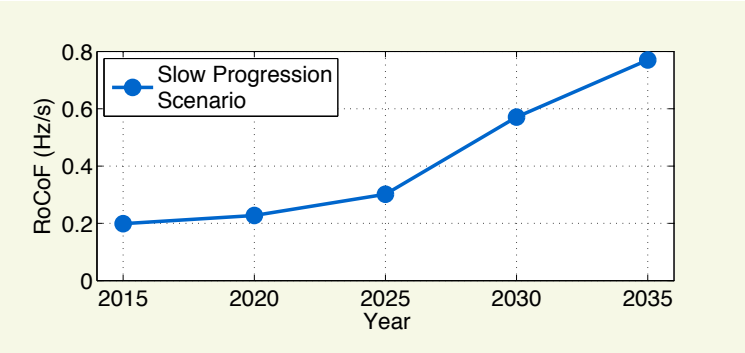

Figure 1.4: Forecast of typical RoCoF values occuring at 10% of the year, under the "slow progression" scenario.

The second challenge is concerned with the degradation of the signal quality (voltage and current measurements) used in estimation algorithms. This is attributed to the diverse profile of users who are plugging into and out of the grid at a more rapid pace. For example, uneven distribution of loads for across the three-phase voltages, as defined in (1.1), result in unbalanced system conditions almost all the time [19]. Besides that, the increased use of electronic inverter-based equipment (e.g. consumer electronics) introduces switching noise, and non-sinusoidal currents and harmonics [17], which are not well-represented by algorithms derived in standard linear estimation theory [20, 21, 22].

1.3 Problem Formulation

In light of the challenges facing the frequency regulation landscape in future low inertia grids, it is paramount that the frequency estimation and tracking problem is revisited. To illuminate this point, consider a general electricity grid represented by a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$. The N buses in the network are represented by a node set, $\mathcal{N} = \{1, 2, \ldots, N\}$, while the power lines between the buses (connections) are represented by an edge set, $\mathcal{E}$. The neighbourhood of a node i, denoted by $\mathcal{N}_i$, comprises all the nodes connected to node i including itself, that is $\mathcal{N}_i = \{j \,|\, (i,j) \in \mathcal{E}\}$ [23], see Figure 1.5.

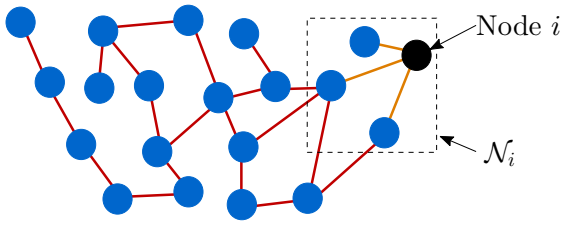

Figure 1.5: A distributed network with $N = 20$ nodes.

Each node (e.g. node i) has access to sampled three-phase voltage measurements, at the discrete time instant k, given by

$$\boldsymbol{s}_{i,k} = \begin{bmatrix} v_{a,i,k} \\ v_{b,i,k} \\ v_{c,i,k} \end{bmatrix} = \begin{bmatrix} V_{a,i}\cos(\omega k + \phi_{a,i}) \\ V_{b,i}\cos(\omega k + \phi_{b,i} - \frac{2\pi}{3}) \\ V_{c,i}\cos(\omega k + \phi_{c,i} + \frac{2\pi}{3}) \end{bmatrix}, \tag{1.1}$$

where the amplitude of the phase voltages $v_{a,i,k}$, $v_{b,i,k}$, $v_{c,i,k}$, are $V_{a,i}, V_{b,i}, V_{c,i}$ while the phase values are denoted by $\phi_{a,i}, \phi_{b,i}, \phi_{c,i}$. The angular frequency is $\omega = 2\pi f T$, with f the fundamental power system frequency which is identical throughout the network and T the sampling interval.

As addressed in Chapter 7, the three-phase representation of the $\boldsymbol{s}_{i,k}$ in (1.1) is over-parametrised and can be compactly represented as "two-phase" Clarke voltages, $v_{\alpha,i,k}$ and $v_{\beta,i,k}$ via a projection onto a new orthogonal basis using the so-called "Clarke transform", given by

$$\begin{bmatrix} v_{\alpha,i,k} \\ v_{\beta,i,k} \end{bmatrix} \stackrel{\text{def}}{=} \underbrace{\sqrt{\frac{2}{3}} \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix}}_{\text{Clarke matrix}} \begin{bmatrix} v_{a,i,k} \\ v_{b,i,k} \\ v_{c,i,k} \end{bmatrix}. \tag{1.2}$$

Moreover, the Clarke transform enables, $v_{\alpha,i,k}$ and $v_{\beta,i,k}$, to be conveniently represented jointly as a complex-valued scalar,

$$s_{i,k} \stackrel{\text{def}}{=} v_{\alpha,i,k} + j v_{\beta,i,k}. \tag{1.3}$$

Figure 1.6 illustrates the operation of the Clarke transform in (1.2).

It can be shown that the complex-valued Clarke voltage (also referred to as the $\alpha\beta$ voltage)

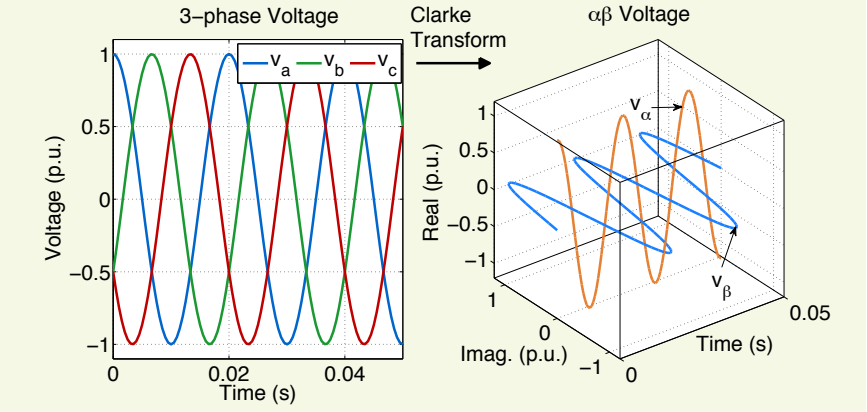

Figure 1.6: Geometric representation of the Clarke Transform. Observe that the Clarke voltages v_α and v_β are orthogonal and admit a convenient complex valued representation in the form $s_{i,k} = v_{\alpha,i,k} + jv_{\beta,i,k}$

.

in (1.3) takes the form

$$s_{i,k} = A_i e^{j\omega k} + B_i e^{-j\omega k}, \tag{1.4}$$

where the positive and negative sequence phasors A_i and B_i are given by[1] [24]

$$A_i = \frac{\sqrt{6}}{6}\left[V_{a,i}e^{j\phi_{a,i}} + V_{b,i}e^{j\phi_{b,i}} + V_{c,i}e^{j\phi_{c,i}}\right] \tag{1.5a}$$

$$B_i = \frac{\sqrt{6}}{6}\left[V_{a,i}e^{-j\phi_{a,i}} + V_{b,i}e^{-j\left(\phi_{b,i}+\frac{2\pi}{3}\right)} + V_{c,i}e^{-j\left(\phi_{c,i}-\frac{2\pi}{3}\right)}\right]. \tag{1.5b}$$

Our task is to estimate the system frequency, ω, given noisy observations of the Clarke voltage $s_{i,k}$ in (1.4). In particular, the noisy observations of $s_{i,k}$ can be expressed in the form

$$y_{i,k} \quad = \quad s_{i,k} + \eta_{i,k}, \tag{1.6}$$

where $\eta_{i,k}$, is a zero-mean complex-valued white Gaussian noise process, with variance $\sigma^2_{\eta_i} = \mathbb{E}\left\{|\eta_{i,k}|^2\right\}$. This thesis aims to answer, in a systematic way, several questions pertinent to the estimation problem in (1.6), the first of which is on the signal model suitable for the general Clark voltage $s_{i,k}$.

Notice from the noise-free signal in (1.5b) the negative sequence phasor vanishes, that is $B_i = 0$, for a balanced system (equal amplitudes, $V_{a,i} = V_{b,i} = V_{c,i}$, and uniform phase separation, $\phi_{a,i} = \phi_{b,i} = \phi_{c,i}$), thus yielding the balanced Clarke voltage

$$s_{i,k} = A_i e^{j\omega k}. \tag{1.7}$$

Figure 1.7 depicts the real and imaginary parts of the balanced signal in (1.7) which exhibits a circular trajectory. However, for unbalanced systems, $B_i \neq 0$, and therefore $s_{i,k}$ does not fol-

[1] A detailed explanation of the Clarke transform is given in Chapter 7.

low a uniform circular trajectory. During a fault in the voltage lines, the currents and voltages across the three-phases fail to maintain uniformity and can either experience a sag (under-voltage) or a swell (over-voltage). These sags and swells are classified into distinct categories denoted by alphabet symbols from A to G [25]. For example, a system with Type C and Type D voltage sags, as illustrated in the phasor diagram in Figure 1.8, exhibits noncircular Clarke voltage trajectories as explained by (1.4), see also Figure 1.7.

An unbalanced system condition therefore introduces a problem, as conventional complex-valued linear estimation theory does not cater for noncircular signals. In fact, it was recently shown that the standard *strictly linear* model for (1.4) is inadequate for unbalanced systems and a *widely linear* model is required [24, 26]. Widely linear estimation shall therefore be a cornerstone to this thesis, where we present several contributions, such as the design of computationally efficient widely linear adaptive filters and new results in the convergence analysis of widely linear adaptive filters, all within a general framework of power systems.

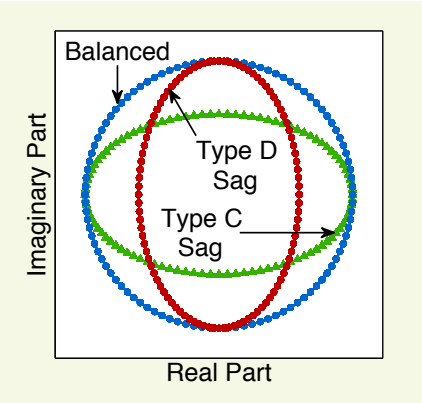

Figure 1.7: For a balanced system, characterised by $V_{a,i} = V_{b,i} = V_{c,i}$ and $\phi_{a,i} = \phi_{b,i} = \phi_{c,i}$, the trajectory of Clarke's voltage v_k is circular (blue line). For unbalanced systems, the voltage trajectories are noncircular (red and green lines).

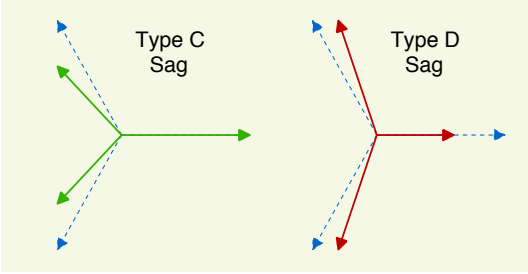

Figure 1.8: Phasor diagrams of voltage sags where the dashed blue arrows designate a set of balanced three-phase voltage phasors. Notice the change in magnitudes and phase separations during a fault (sag).

Furthermore, the multi-sensor nature of the estimation problem in (1.6), calls for the extension of standard singe-node frequency estimators to a distributed setting. In the particular application of frequency tracking in low-inertia grids, the distributed estimation algorithms are

crucial to exploit spatially diverse measurements across the grid, so as to compensate for the fewer temporal measurements available in the first few hundred milliseconds after a contingency. Although spatially diverse measurements, $y_{i,k}$, from (1.6) can be collected and processed in a fusion centre, the growth in the number of measurements from phasor measurements units (PMU) and smart-meters makes the network vulnerable to a single point of failure [27, 28, 29].

The research undertaken for this thesis is motivated by: (i) the need for faster and more robust frequency estimates in future low-inertia grids, together with (ii) growth in multi-sensor distributed measurements and (iii) latest developments in widely linear modelling of three-phase voltages motivated. The main research objective can therefore be summarised as *"to solve the frequency tracking problem in low-inertia grids using complex-valued distributed adaptive filters."* This thesis aims to achieve the goal from various angles which include formulation of new algorithms, uncovering their learning behaviour (stability, convergence), giving new perspectives to some well-established solutions and finally by testing the algorithms with real-world case studies.

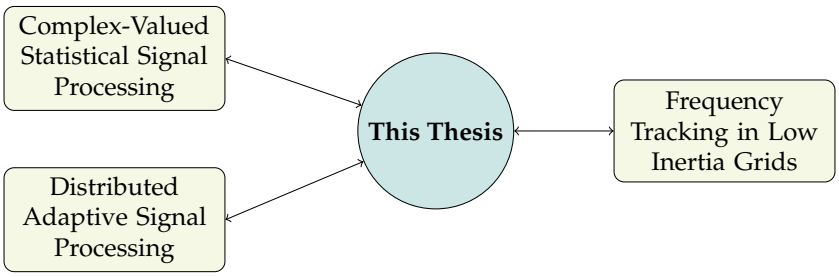

Figure 1.9: Subject areas of the thesis.

1.4 Contribution of the Thesis

The contributions of this thesis can be divided into four areas, this roughly corresponds to how this thesis is organised, as shown in Figure 1.10.

1. **Novel adaptive filters for widely-linear estimation**

   - Dual Channel Estimation: Low Complexity Adaptive Filters

     Complex-valued estimation algorithms are computationally intensive (e.g. one complex multiplication requires four real multiplications) while standard methods to reduce this computational complexity cast the problem in the real domain, where the physical meaning inherent to working in the complex-domain is lost. A new estimation method, referred to as "dual channel estimation" is proposed by decomposing the standard mean square error (MSE) into the MSE in estimating the real and imaginary parts. This allows for the development of computationally efficient complex-valued estimation algorithms while still remaining within the complex domain. This contribution is elaborated in **Chapter 3** and was introduced in [30].

     Relevant Publications:

[30] C. Jahanchahi, S. Kanna, and D.P. Mandic, "Complex dual channel estimation: Cost effective widely linear adaptive filtering," *Signal Processing*, 104:3342, Nov 2014.

- Adaptive Tracking of Complex Circularity

  The circularity quotient of a complex-valued signal acts as a "fingerprint" in many areas including communications, speech processing and power engineering. Since the circularity quotient is a function of the second order statistics of the signal, current methods that test the circularity are block-based. A structurally elegant adaptive algorithm to track the circularity quotient is derived using the MSE estimator of a complex-valued signal from its conjugate. This algorithm was proposed in [31] and was applied to track the circularity of communication signals in [32], see **Chapter 4**.

  Relevant Publications:

  [31] S. Kanna, S.C. Douglas, and D.P. Mandic, "A real time tracker of complex circularity," in *Proc. of the 8th IEEE Sensor Array and Multichannel Signal Process. Workshop (SAM)*, June 2014, pp. 129–132.

  [32] S. Kanna, M. Xiang, and D. P. Mandic, "Real-time detection of rectilinear sources for wireless communication signals," in *Proc. of the 2015 International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2015, pp. 541–545.

2. **Distributed Adaptive Filters with application to Frequency Tracking**

- Diffusion augmented complex least mean square (D-ACLMS) for frequency estimation

  The problem of estimating the frequency in grid is cast into the distributed learning setting by applying the D-ACLMS algorithm for tracking the frequency in unbalanced three-phase power systems. The proposed D-ACLMS is able to improve the steady-state estimation performance by allowing individual agents to share their estimates. This contribution was presented in [33] and shall be discussed in **Chapter 5** and **Chapter 8**.

  Relevant Publications:

  [33] S. Kanna, S.P. Talebi, and D.P. Mandic, "Diffusion widely linear adaptive estimation of system frequency in distributed power grids," in *Proc. of the IEEE Intl. Energy Conf. (ENERGYCON)*, May 2014, pp. 772–778.

- Distributed augmented complex extended Kalman filter (D-ACEKF)

  The augmented complex extended Kalman filter is proposed in a distributed setting for general complex-valued state estimation problems. Current methods for distributed Kalman filtering assume circular noise and strictly linear models which are both inadequate for a wide range of applications including frequency tracking in unbalanced smart grids. The proposed D-ACEKF is applied to track the frequency

of both synthetically generated and real-world measurements of unbalanced voltage signals. This work has been published in [34] and is outlined in **Chapter 6** and **Chapter 8**.

Relevant Publications:

[34] S. Kanna, D. H. Dini, Y. Xia, S. Y. Hui, and D. P. Mandic, "Distributed widely linear Kalman filtering for frequency estimation in power networks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 1, no. 1, pp. 45–57, Mar 2015.

3. **Learning Behaviour of Distributed Adaptive Filters**

- Mean square analysis of complex-valued LMS algorithms

  The mean square analysis of widely linear (augmented) least mean square algorithm is typically not addressed due to the difficulty in diagonalising the augmented covariance matrix. By exploiting new results in complex-valued linear algebra, elegant and physically meaningful bounds for the learning rate and steady-state misadjustment are derived as functions of the degree of non-circularity of the input signal. The analysis was first published in [35] and is included in the background section in **Chapter 2** and in the analysis in **Chapter 5**.

  Relevant Publications:

  [35] D.P. Mandic, S. Kanna, and S.C. Douglas, "Mean square analysis of the CLMS and ACLMS for non-circular signals: The approximate uncorrelating transform approach," in *Proc. of the IEEE Intl. Conf. on Acoust. Speech and Signal Process. (ICASSP)*, Apr 2015, pp. 3531–3535.

- Mean square analysis of distributed complex-valued LMS algorithms

  The mean square analysis of the ACLMS is next extended to bound the steady-state mean square performance of the diffusion complex least mean square (D-CLMS) and the diffusion augmented CLMS (D-ACLMS) algorithms. The distinguishing factor of the proposed analysis is the consideration of the second-order terms in the weight error covariance matrix recursion, without compromising the mathematical tractability of the problem. Unlike previous results, the proposed analysis is able to clearly show the relationship between data noncircularity and estimation performance of the D-CLMS and D-ACLMS. This result is introduced in [36] and is the main discussion point of **Chapter 5**.

  Relevant Publications:

  [36] S. Kanna and D. P. Mandic, "Steady-state behavior of general complex-valued diffusion LMS strategies," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 722–726, May 2016.

4. **New Interpretations of Existing Methods**

- Signal processing perspective of three-phase transforms

Although three-phase transforms are routinely used in signal processing applications in power engineering, they lack a comprehensive signal processing interpretation, a prerequisite for their use in the future smart grids. This thesis explores the three most common power system transforms: the symmetrical, Clarke and Park transforms. We show that the symmetrical transform can be derived as a spatial discrete Fourier transform (DFT), while the the Clarke transform is found to be the optimal dimensionality reduction algorithm for balanced three-phase systems. Finally, the Park transform is shown to be a frequency demodulation scheme in the complex-domain. These interpretations strengthen the physical meaning of wide-class of machine learning, signal processing and control algorithms that employ three-phase transforms. In addition, an adaptive real-time Clark and Park transformed is also proposed for dynamic power systems. This contribution is outlined in **Chapter 7**.

Relevant Publications:

[37] S. Kanna and D. P. Mandic, "Modern view of three phase transforms" *Submitted to IEEE Power Electronics Magazine*, 2016.

- Relationship between the Kalman filter and the LMS

  The Kalman Filter and the LMS are typically treated as separate entities, with the former as a realization of the optimal Bayesian estimator and the latter as a recursive solution to the optimal Wiener filtering problem. We reveal a joint perspective on Kalman filtering and LMS-type algorithms, achieved through analyzing the degrees of freedom necessary for optimal stochastic gradient descent adaptation in a system identification framework. This is included in **Appendix A.**

  Relevant Publications:

  [38] D. P. Mandic, S. Kanna, and A. G. Constantinides, "On the intrinsic relationship between the least mean square and Kalman filters [Lecture Notes]," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 117122, Nov 2015.

- Relationship Between the DFT and the FM Techniques

  We have provided a new, compact framework for the unification of the two best known frequency tracking algorithms in the context of power system frequency estimation. The *de facto* standard recursive discrete Fourier transform (DFT) based frequency tracker is shown to be a special case of the frequency demodulation techniques commonly applied in communication systems. This unification is an important contribution to the field of power system measurements as it provides deeper insights into the similarities and differences between the two algorithms. Moreover, our treatment of the recursive DFT and demodulation technique reveals new opportunities in the design of phasor and frequency tracking algorithms in phasor measurement units (PMU). This result is outlined in **Appendix B** and is under review.

  Relevant Publications:

  [39] S. Kanna, A. G. Constantinides, and D. P. Mandic, "The equivalence of FM

demodulation and recursive DFT for frequency tracking in power systems," *Submitted to International Journal of Electrical Power & Energy Systems*, 2016.

1.5 Thesis Organisation

Figure 1.10 shows an overview of structure of the thesis which is organised as follows:

- **Chapter 2:** Estimation in the complex-domain

  The aim of this chapter is to provide the reader with the necessary background for the treatment of unbalanced three-phase voltages and widely linear algorithms. This is achieved through a discussion on noncircularity of complex-valued signals and widely linear estimation theory. Recent developments in complex-valued matrix factorisations are also introduced, together with widely linear autoregressive (WLAR) modelling and augmented complex adaptive filters. The latter ideas are instrumental in developing frequency estimator for unbalanced power systems in Chapter 8.

The chapters outlining the contribution of the thesis are grouped into three parts.

Part I: Adaptive filters for widely linear estimation

- **Chapter 3:** Low-complexity complex-valued adaptive filters

  This chapter elaborates the "dual channel" estimation methodology which decomposes the MSE cost function into the MSE in estimating the real and imaginary parts of a signal so that separate estimators can be used to minimised the two cost functions independently. We proceed to show that the dual-channel estimation scheme allows for the necessary degrees of freedom to model general complex-valued signals, while enabling computationally efficient adaptive filters.

- **Chapter 4:** Adaptive tracking of complex circularity

  In this chapter, a new insight into the relationship between the complex circularity quotient and MSE estimation is presented. We show that MSE estimator of a complex random variable from its complex conjugate is the complex circularity quotient. A sequential (online) version of this estimator, suitable for real time tracking of the degree of complex circularity, is outlined based on the complex least mean square (CLMS) algorithm.

Part II: Distributed adaptive algorithms

- **Chapter 5:** Diffusion complex least mean square

  The diffusion augmented complex least mean square (D-ACLMS) is introduced in this chapter. Some background information on distributed learning over graphs and the diffusion adaptation scheme necessary for this chapter and Chapter 6 is included. The technical contribution of this chapter is the mean square analysis for the D-ACLMS algorithm. The analyses make use of the similarity assumption to bound the steady-state mean square performance. As the frequency estimation examples are reserved for Part III, we apply the D-ACLMS to synthetic data to verify the analysis.

- **Chapter 6:** Diffusion complex extended Kalman filter

  This chapter introduces the distributed complex and augmented complex extended Kalman filter (D-ACEKF) for general complex-valued state-space models, while also accounting for the correlated and noncircular natures of the observation noise. We give a unified account of existing distributed Kalman filtering approaches and show that the proposed distributed augmented complex extended Kalman filter is suitable for the generality of complex-valued signals.

Part III: Distributed frequency estimation in the smart grid

- **Chapter 7:** Modern view of three-phase transformations

  In this chapter, the focus goes back to electricity grid where we introduce the mathematical models for three-phase power systems. The symmetrical, Clark and Park transforms are derived from first principles and are connected to well-known signal processing concepts. Specifically, the symmetrical transform is shown to be a spatial DFT scheme, the Clark transform is derived as the optimal dimensionality reduction algorithm for balanced three-phase systems while the Park transform is shown to be a frequency demodulation technique. The outputs of the Clarke and Park transforms for systems operating under unbalanced conditions and frequency deviations are revealed to be non-ideal. To this end, we propose an adaptive Clarke and Park transform for general dynamic systems.

- **Chapter 8:** Distributed frequency tracking examples

  Using insights developed in the previous chapters, a distributed frequency tracking algorithm is introduced. Both synthetic and real world data are used to perform case studies over a range of power system conditions, to illustrate the theoretical and practical advantages of the proposed distributed frequency estimation methodology. Practical implementation issues are also included for the benefit of the reader.

- **Chapter 9:** Conclusions & future work

  Concluding remarks drawn from the thesis and ideas for future works are outlined in this chapter.

Appendices

- **Appendix A:** Intrinsic relationship between the Kalman filter and LMS

  This appendix provides a much needed link between stochastic gradient algorithms like the LMS and state-space estimators like the Kalman filter. This appendix can be seen as a link between the estimators proposed in Chapter 5 and Chapter 6. For notational simplicity and clarity of ideas, only the real-valued counterparts of the Kalman filter and LMS are used.

- **Appendix B:** Equivalence of classical frequency tracking techniques

  In a similar fashion to Appendix A, we show that two most well-known frequency estimators in the electricity grid can in fact be unified into a single framework. Using

this framework, the recursive discrete Fourier Transform based frequency tracker can be viewed as a special case of the frequency demodulation (FM) technique.

- **Appendix C:** Cramer-Rao lower bound and maximum likelihood estimation

  The Cramer-Rao lower bound (CRLB) is derived for the frequency estimation problem for general unbalanced three-phase voltages. Furthermore, we derive the maximum likelihood frequency estimation scheme and show that it differs from the standard periodogram maximiser when the voltage signal is unbalanced. This algorithm is referred to as the augmented periodogram maximiser. Part of this work has been submitted for publication [40].

- **Appendix D:** Rapid Frequency Response for Low Inertia Grids

  This appendix describes a potential application of the distributed frequency estimation schemes proposed in this thesis for the provision of a rapid frequency response in low inertia grids. We show that exploiting spatially diverse frequency estimates enables wind turbines to provide synthetic inertia without having to increase their power output levels to mimic the conventional synchronous generators. Simulations using a detailed 2850-bus model of the Great Britain system in the DIgSILENT PowerFactory software support the proposed ideas. This work is currently being prepared for publication [41].

> **Contribution Panel**
> Throughout the different chapters and appendices, a visual aid in the form of this panel will serve to highlight the contributions of the thesis.

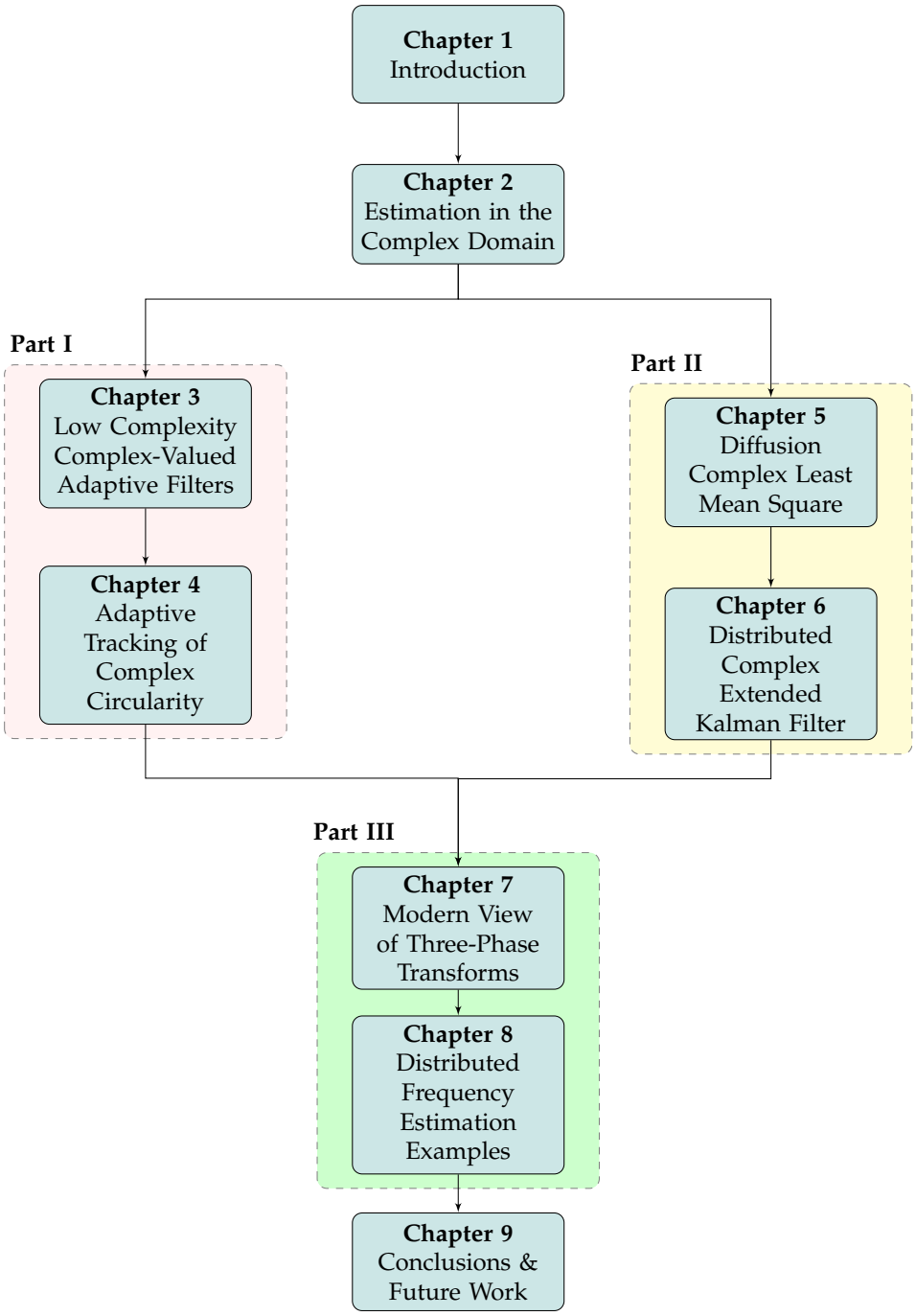

Figure 1.10: Thesis Structure

Chapter 2

Estimation in the Complex Domain

> The shortest path between two truths in the real domain passes through the complex domain.
>
> *Jacques Hadamard*

Chapter Overview

This chapter provides the required background in complex-valued statistical signal processing for the ideas presented in this thesis. This includes a brief introduction to noncircular random variables, widely linear estimation, including widely linear autoregressive (AR) models and adaptive filters. We also introduce the the notion of mean convergence and mean square stability in the analysis of adaptive algorithms. The analysis technique introduced in this chapter shall be instrumental to understanding the convergence proofs in Chapters 3 to 6.

2.1 Benefits of a Complex-Valued Representation

Complex-valued signals are the backbone of many engineering disciplines including communications, array processing, and power engineering as they are convenient representations of sinusoidal waveforms and bivariate real-valued signals [42, 43]. Although it may be natural to process complex-valued data by representing the real and imaginary parts as a bivariate signal in the real domain, any intuition and physical meaning inherent in processing in the complex domain would be largely obscured.

Specifically, working in the complex-domain provides conceptual and modelling advantages to describe certain phenomena (e.g. time-delay, phase shifts) which in turn benefits signal processing engineers at the algorithm design stage. Consider this example of a frequency estimation task, whereby the goal is to estimate the frequency, ω, from signals $x_k = \cos(\omega k)$ and $y_k = \sin(\omega k)$, which are nonlinear functions of ω with k denoting the discrete time index. Estimating the frequency of the signals x_k and y_k is made much simpler by defining a complex variable, $z_k \overset{\text{def}}{=} x_k + jy_k = \cos(\omega k) + j\sin(\omega k) = e^{j\omega k}$. Notice that, the complex-representation of the signal z_k can be expressed as

$$z_k = e^{j\omega k} = e^{j\omega(k-1+1)} = e^{j\omega}e^{j\omega(k-1)} = az_{k-1}, \tag{2.1}$$

where estimating the linear coefficient, $a = e^{j\omega}$, is sufficient to estimate the frequency, ω. This shows that frequency estimation, which originally presents itself as a nonlinear estimation problem, can be transformed into a linear problem, by virtue of working in the complex-domain.

Another example worth mentioning is the modelling of the "XOR" function in neural networks. When working in the real-domain, a multilayer perceptron is required to implement a classifier which takes two binary inputs and outputs a binary variable according the XOR rule. However, when the problem is represented in the complex-domain, a single linear neuron is able to accomplish the task of performing the XOR function. We refer the reader to [44], for a detailed explanation of this problem.

2.2 Circularity of Random Variables

Statistical tools for complex random variables are traditionally treated as generic extensions of their real-valued counterparts [26, 43]. However, this approach is valid only for second-order circular random variables. To illustrate this point, consider a zero-mean complex-valued scalar random variable (r.v.), $x \in \mathbb{C}$, expressed in its Cartesian form as

$$x = x_r + jx_i,$$

where $x_r = \mathsf{Re}\{x\} \in \mathbb{R}$ and $x_i = \mathsf{Im}\{x\} \in \mathbb{R}$. The covariance of x is then given by

$$\begin{aligned} r &\overset{\text{def}}{=} \mathbb{E}\left\{xx^{\mathsf{H}}\right\} = \mathbb{E}\left\{|x|^2\right\} \\ &= \mathbb{E}\left\{x_r^2\right\} + \mathbb{E}\left\{x_i^2\right\} \\ &= \sigma_r^2 + \sigma_i^2, \end{aligned} \tag{2.2}$$

and only reflects the total joint power within the real and imaginary components. To ensure a sufficient number of degrees of freedom in the processing of general complex-valued data, the pseudocovariance, p, defined as,

$$\begin{aligned} p &\overset{\text{def}}{=} \mathbb{E}\left\{xx^{\mathsf{T}}\right\} = \mathbb{E}\left\{x^2\right\} \\ &= \mathbb{E}\left\{x_r^2\right\} - \mathbb{E}\left\{x_i^2\right\} + 2j\mathbb{E}\left\{x_r x_i\right\} \\ &= \sigma_r^2 - \sigma_i^2 + 2j\sigma_{ri}, \end{aligned} \tag{2.3}$$

needs to also be considered. Notice that the pseudocovariance now accounts for both the power imbalance, $\sigma_r^2 - \sigma_i^2$ and cross-correlation between the real and imaginary parts, σ_{ri}. Therefore, in order to cater for the generality of complex-valued signals, both the covariance, r, and pseudocovariance, p, have to be taken into account. A special case of signals that have vanishing pseudocovariance (i.e. $p = 0$) are referred to as *second-order circular* or *proper* signals.

Definition 2.1 (Propriety). *A complex random variable, x, is proper if its pseudocovariance, $p = 0$. Conversely, an improper complex random variable has a non-zero pseudocovariance, $p \neq 0$ [43].*

It is also often useful to quantify the level of noncircularity of a complex random variable. This is accomplished using the *circularity quotient*, defined as the ratio of the pseudocovariance, p, to covariance, r, as [45]

$$\rho \overset{\text{def}}{=} \frac{p}{r}, \tag{2.4}$$

while the *circularity coefficient* is given by the magnitude of circularity quotient,

$$|\rho| \stackrel{\text{def}}{=} \frac{|p|}{r}. \tag{2.5}$$

From (2.2) and (2.3), the *circularity coefficient* is bounded by $0 \leq |\rho| \leq 1$ [45].

Definition 2.2 (Circularity). *If x is a circular random variable, then its probability density function (pdf),is rotationally invariant, i.e.,*

$$\mathit{pdf}(xe^{j\phi}) = \mathit{pdf}(x),$$

for any arbitrary rotation with angle ϕ [43].

Generally speaking, proper random variables are not necessarily circular. For example, consider a random variable

$$\tilde{x} = \tilde{x}_r + j\tilde{x}_i,$$

where both $\tilde{x}_i$ and $\tilde{x}_r$ are drawn independently from an identical uniform distribution, $\mathcal{U}(\frac{1}{2}, -\frac{1}{2})$. As illustrated in Figure 2.1, although $\tilde{x}$ is *proper* with $\sigma_r^2 = \sigma_i^2$ and $\sigma_{ri} = 0$, it is clearly noncircular since the $\mathsf{pdf}(\tilde{x})$ is not rotationally invariant.

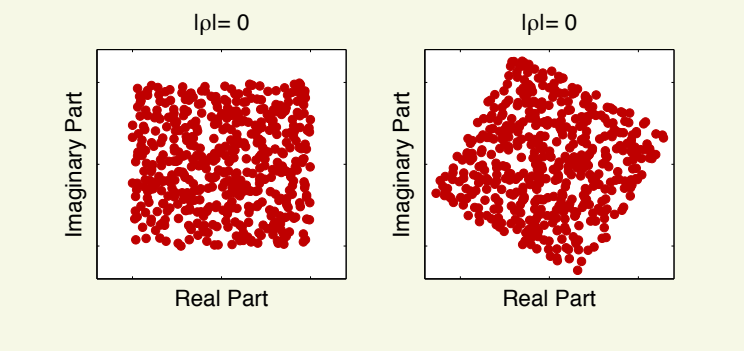

Figure 2.1: Circularity diagrams of zero-mean uniform random variables: (i) Left: $\tilde{x}$ and (ii) Right: $e^{j\frac{\pi}{3}}\tilde{x}$. Notice the random variable $\tilde{x}$ is *proper*, i.e. $\sigma_r^2 = \sigma_i^2$ and $\sigma_{ri} = 0$, but is is not rotationally invariant $\mathsf{pdf}(\tilde{x}) \neq \mathsf{pdf}(\tilde{x}e^{j\phi})$.

For Gaussian random variables, properness implies circularity and therefore for convenience, the terms "circular" and "proper" and shall be used interchangeably. Figure 2.2 shows a Gaussian random variable with different circularity profiles. Observe that noncircular variables exhibit either power imbalances $\mathbb{E}\{x_r^2\} \lesseqgtr \mathbb{E}\{x_i^2\}$, as in the top right and bottom left panels in Figure 2.2 or have correlated components $\mathbb{E}\{x_r x_i\} \neq 0$, as in the bottom right panel.

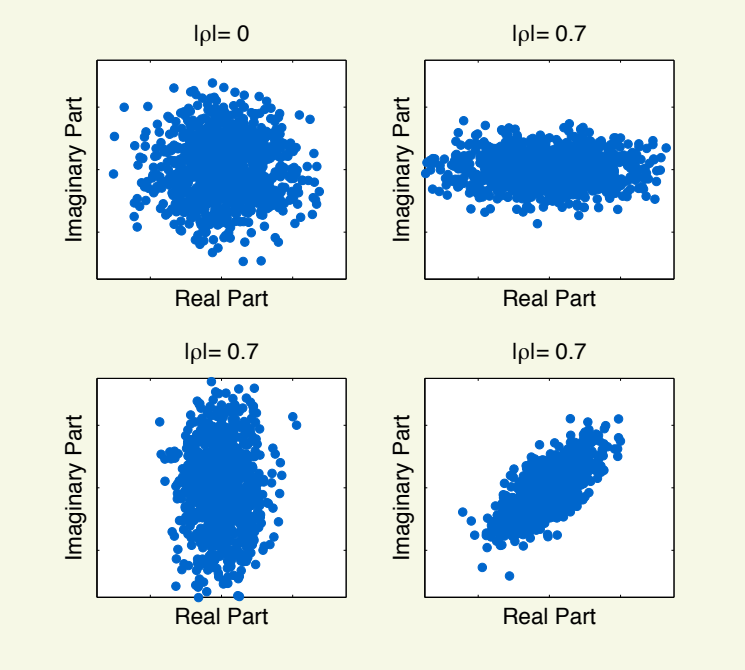

Figure 2.2: Circularity diagrams of zero-mean Gaussian random variables with various non-circularity profiles. Clockwise from top left: (i) proper variable, (ii) improper variable with $\sigma_r^2 > \sigma_i^2$, (iii) improper variable with $\sigma_i^2 > \sigma_r^2$ and (iv) improper variable with correlated real and imaginary components, $\sigma_{ri} \neq 0$.

2.2.1 Extension to Vector Variables

The definitions for the covariance and pseudocovariance from (2.2) and (2.3) can be extended for vector-valued variables, $\boldsymbol{x} \in \mathbb{C}^{M\times 1}$, where $\boldsymbol{x} = \boldsymbol{x}_r + j\boldsymbol{x}_i$ with $\boldsymbol{x}_r = \mathsf{Re}\{\boldsymbol{x}\}$ and $\boldsymbol{x}_i = \mathsf{Im}\{\boldsymbol{x}\}$. The second order statistics is described by the covariance matrix, $\mathbf{R} \in \mathbb{C}^{M\times M}$, and pseudocovariance matrix, $\mathbf{P} \in \mathbb{C}^{M\times M}$, which are defined as

$$\mathbf{R} \overset{\text{def}}{=} \mathbb{E}\left\{\boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}\right\}, \quad \mathbf{P} \overset{\text{def}}{=} \mathbb{E}\left\{\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}\right\}. \tag{2.6}$$

For compactness, we shall introduce an *augmented* data vector, $\bar{\boldsymbol{x}} \in \mathbb{C}^{2M\times 1}$, formed by concatenating $\boldsymbol{x}$ and its conjugate, $\boldsymbol{x}^*$, as[1]

$$\bar{\boldsymbol{x}} \overset{\text{def}}{=} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{x}^* \end{bmatrix}. \tag{2.7}$$

The corresponding *augmented covariance matrix*, $\bar{\mathbf{R}} \in \mathbb{C}^{2M\times 2M}$, is defined as

$$\bar{\mathbf{R}} \overset{\text{def}}{=} \mathbb{E}\left\{\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^{\mathsf{H}}\right\} = \mathbb{E}\left\{\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{x}^* \end{bmatrix}\begin{bmatrix} \boldsymbol{x}^{\mathsf{H}} & \boldsymbol{x}^{\mathsf{T}} \end{bmatrix}\right\} = \begin{bmatrix} \mathbf{R} & \mathbf{P} \\ \mathbf{P}^* & \mathbf{R}^* \end{bmatrix}, \tag{2.8}$$

[1]The top bar, $\bar{(\cdot)}$, is used to denote augmented vectors and matrices.

and comprises both the covariance and pseudocovariances matrices, $\mathbf{R}$ and $\mathbf{P}$, and thus captures the full second-order statistics of the signal.

2.3 Widely Linear Estimation

2.3.1 Widely Linear MMSE

Minimum mean square error (MMSE) estimation aims to find the optimal estimate of a desired signal, $y \in \mathbb{C}$, given a regressor vector, $\boldsymbol{x} \in \mathbb{C}^{M\times 1}$, in the form [46],

$$\hat{y} = \mathbb{E}\left\{y|\boldsymbol{x}\right\}. \tag{2.9}$$

In standard MMSE theory for complex-valued jointly-Gaussian random variables, the solution to (2.9) is given by the *strictly linear* model [46],

$$\hat{y} = \boldsymbol{h}^{\mathsf{H}}\boldsymbol{x}, \tag{2.10}$$

where $\boldsymbol{h} \in \mathbb{C}^{M\times 1}$ is the optimal weight vector and the estimation error is defined as $\varepsilon \overset{\text{def}}{=} y - \hat{y}$. The optimal value for the weight vector $\boldsymbol{h}$ is found by minimising the mean square error (MSE) cost function given by

$$J_{\text{MSE}}(\boldsymbol{h}) \overset{\text{def}}{=} \mathbb{E}\left\{|\varepsilon|^2\right\} = \mathbb{E}\left\{|y - \boldsymbol{h}^{\mathsf{H}}\boldsymbol{x}|^2\right\}. \tag{2.11}$$

Upon expanding (2.11), using the strictly linear model in (2.10), we have

$$J_{\text{MSE}}(\boldsymbol{h}) = \mathbb{E}\left\{|y - \boldsymbol{h}^{\mathsf{H}}\boldsymbol{x}|^2\right\} = \sigma_y^2 - \boldsymbol{h}^{\mathsf{H}}\boldsymbol{r} - \boldsymbol{r}^{\mathsf{H}}\boldsymbol{h} + \boldsymbol{h}^{\mathsf{H}}\mathbf{R}\boldsymbol{h}, \tag{2.12}$$

where $\sigma_y^2 = \mathbb{E}\left\{|y|^2\right\}$ is the signal power, $\mathbf{R} = \mathbb{E}\left\{\boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}\right\}$, is the covariance matrix of the data vector $\boldsymbol{x}$, defined also in (2.6), while $\boldsymbol{r} \overset{\text{def}}{=} \mathbb{E}\left\{\boldsymbol{x}y^*\right\}$ is the cross-covariance vector between the regressor, $\boldsymbol{x}$, and desired signal, y [43]. The minimum of (2.12) with respect to (w.r.t.) $\boldsymbol{h}$ is found via

$$\frac{\partial J_{\text{MSE}}(\boldsymbol{h})}{\partial \boldsymbol{h}^*} = -\boldsymbol{r} + \mathbf{R}\boldsymbol{h} = \mathbf{0}, \tag{2.13}$$

where the optimal weight vector is given by the *strictly linear* Wiener solution

$$\boldsymbol{h} = \mathbf{R}^{-1}\boldsymbol{r}. \tag{2.14}$$

The *strictly linear* Wiener solution in (2.14) can be shown to be restrictive for a general class of complex-valued signals [26]. To understand this limitation, consider expressing the MMSE in (2.9) using the real and imaginary parts of the signal $y = y_r + jy_i$ and $\boldsymbol{x} = \boldsymbol{x}_r + j\boldsymbol{x}_i$ as

$$\hat{y} = \mathbb{E}\left\{y_r|\boldsymbol{x}_r, \boldsymbol{x}_i\right\} + j\mathbb{E}\left\{y_i|\boldsymbol{x}_r, \boldsymbol{x}_i\right\}. \tag{2.15}$$

Since the real and imaginary parts of $\boldsymbol{x}$ can be derived from the data $\boldsymbol{x}$ and its complex conjugate $\boldsymbol{x}^*$, we have

$$\boldsymbol{x}_r = \frac{1}{2}(\boldsymbol{x} + \boldsymbol{x}^*), \qquad \boldsymbol{x}_i = \frac{1}{2j}(\boldsymbol{x} - \boldsymbol{x}^*),$$

thus allowing for the MMSE estimator in (2.15) to be expressed as

$$\begin{aligned} \hat{y} &= \mathbb{E}\{y_r|\boldsymbol{x}, \boldsymbol{x}^*\} + j\mathbb{E}\{y_i|\boldsymbol{x}, \boldsymbol{x}^*\} \\ &= \mathbb{E}\{y|\boldsymbol{x}, \boldsymbol{x}^*\} \\ &= \boldsymbol{h}^{\mathsf{H}}\boldsymbol{x} + \boldsymbol{g}^{\mathsf{H}}\boldsymbol{x}^*. \end{aligned} \tag{2.16a}$$

Notice that the *widely linear* model in (2.16a) is linear in *both* $\boldsymbol{x}$ and $\boldsymbol{x}^*$, giving more degrees of freedom necessary to model the generality of complex-valued signals. The optimal values for the widely linear coefficients $\boldsymbol{h}$ and $\boldsymbol{g}$ are found by first expressing (2.16a) in an augmented form as

$$\hat{y} = \boldsymbol{w}^{\mathsf{H}}\bar{\boldsymbol{x}}, \tag{2.16b}$$

with the augmented vectors $\bar{\boldsymbol{x}} = \begin{bmatrix}\boldsymbol{x}^{\mathsf{T}} & \boldsymbol{x}^{\mathsf{H}}\end{bmatrix}^{\mathsf{T}} \in \mathbb{C}^{2M\times 1}$ and $\boldsymbol{w} = \begin{bmatrix}\boldsymbol{h}^{\mathsf{T}} & \boldsymbol{g}^{\mathsf{T}}\end{bmatrix}^{\mathsf{T}} \in \mathbb{C}^{2M\times 1}$. Following a similar procedure as in (2.11) – (2.14), the MSE now becomes $J_{\text{MSE}}(\boldsymbol{w}) = \mathbb{E}\left\{|y - \boldsymbol{w}^{\mathsf{H}}\bar{\boldsymbol{x}}|\right\}$ and the *widely linear* Wiener solution, $\boldsymbol{w}$, is given by

$$\boldsymbol{w} = \bar{\mathbf{R}}^{-1}\bar{\boldsymbol{r}}, \tag{2.16c}$$

where $\bar{\mathbf{R}}$ is the augmented covariance matrix defined in (2.8) and $\bar{\boldsymbol{r}}$ is the augmented cross-covariance vector is given by

$$\bar{\boldsymbol{r}} \overset{\text{def}}{=} \mathbb{E}\{\bar{\boldsymbol{x}}y^*\} = \mathbb{E}\begin{bmatrix}\boldsymbol{x}y^* \\ \boldsymbol{x}^*y^*\end{bmatrix} = \begin{bmatrix}\boldsymbol{r} \\ \boldsymbol{p}^*\end{bmatrix}, \tag{2.17}$$

with $\boldsymbol{r}$ being identical to the cross-covariance vector in (2.14) while $\boldsymbol{p}$ is referred to as the pseudocross-covariance vector.

Remark 2.1. *Owing to the use of augmented vectors $\bar{\boldsymbol{x}}$ and $\boldsymbol{w}$, the term* widely linear *and* augmented *shall be used interchangeably.*

The individual values for the vectors $\boldsymbol{h}$ and $\boldsymbol{g}$ defined in (2.16a) can be derived by expressing the augmented solution in (2.16c) in terms of its constituent covariance and pseudocovariance matrices and vectors as

$$\begin{bmatrix}\mathbf{R} & \mathbf{P} \\ \mathbf{P}^* & \mathbf{R}\end{bmatrix}\begin{bmatrix}\boldsymbol{h} \\ \boldsymbol{g}\end{bmatrix} = \begin{bmatrix}\boldsymbol{r} \\ \boldsymbol{p}^*\end{bmatrix}. \tag{2.18}$$

Solving the simultaneous equation in (2.18) yields the widely linear weight vectors

$$\begin{aligned} \boldsymbol{h} &= \left[\mathbf{R} - \mathbf{P}\mathbf{R}^{*-1}\mathbf{P}^{*}\right]^{-1}\left[\boldsymbol{r} - \mathbf{P}\mathbf{R}^{*-1}\boldsymbol{p}^{*}\right] \\ \boldsymbol{g} &= \left[\mathbf{R}^{*} - \mathbf{P}^{*}\mathbf{R}^{-1}\mathbf{P}\right]^{-1}\left[\boldsymbol{p}^{*} - \mathbf{P}^{*}\mathbf{R}^{*-1}\boldsymbol{r}\right]. \end{aligned} \tag{2.19}$$

Observe that the standard covariance matrix, $\mathbf{R} = \mathbb{E}\left\{\boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}\right\}$, and cross-covariance vector, $\boldsymbol{r} = \mathbb{E}\left\{\boldsymbol{x}y^{*}\right\}$, in the strictly linear solution in (2.14), do not have the sufficient degrees of freedom to explain the complete second order information in the signals $\boldsymbol{x}$ and y. To capture the second order statistics we also need the pseudocovariance matrix, $\mathbf{P} = \mathbb{E}\left\{\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}\right\}$, and the pseudocross-covariance vector, $\boldsymbol{p} = \mathbb{E}\left\{\boldsymbol{x}y\right\}$. The input and desired signal pairs which exhibit a vanishing pseudocovariance, $\mathbf{P} = \mathbf{0}$ and pseudocross-covariance, $\boldsymbol{p} = \mathbf{0}$, are referred to as *jointly-circular* [26, 47].

2.3.2 Motivating Widely Linear Estimation

An alternative method to motivate the widely linear model for complex signals is by first considering a standard strictly linear relationship between variables $y \in \mathbb{C}$ and $x \in \mathbb{C}$ which is given by

$$y = hx, \tag{2.20}$$

where $h \in \mathbb{C}$, is a linear coefficient. Expressing the strictly linear relationship in (2.20) in terms of its real and imaginary parts gives

$$\begin{aligned} y_r + jy_i &= (h_r + jh_i)(x_r + jx_i) \\ &= h_r x_r - h_i x_i + j(h_r x_i + h_i x_r). \end{aligned} \tag{2.21}$$

Next, expressing the complex-valued relationship in (2.21) as a bivariate vector in $\mathbb{R}^2$ yields

$$\begin{bmatrix} y_r \\ y_i \end{bmatrix} = \begin{bmatrix} h_r & -h_i \\ h_i & h_r \end{bmatrix} \begin{bmatrix} x_r \\ x_i \end{bmatrix}. \tag{2.22}$$

Remark 2.2. *Notice from* (2.22) *that the strictly linear relationship in* (2.20) *restricts the relationship between real parts of the input and output to be identical to the relationship between the imaginary parts of the input and output through the coefficient value* h_r*. A similar argument applies to the cross-relationship between the real and imaginary parts of the input and output through* h_i*.*

If no such restrictions were applied on the estimation procedure, whereby the estimation rule takes the form

$$\begin{bmatrix} y_r \\ y_i \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_r \\ x_i \end{bmatrix}, \tag{2.23}$$

with $a, b, c, d \in \mathbb{R}$ being different coefficients, then the strictly linear estimation procedure in (2.21) is not the appropriate "linear" model for the signal. To circumvent the restrictive nature

of the strictly linear estimation, now consider the widely linear estimation given by

$$\begin{aligned} y &= hx + gx^* \\ y_r + jy_i &= (h_r + jh_i)(x_r + jx_i) + (g_r + jg_i)(x_r - jx_i) \\ &= h_r x_r - h_i x_i + g_r x_r + g_i x_i + j(h_r x_i + h_i x_r + g_i x_r - g_r x_i), \end{aligned} \tag{2.24}$$

which can be again represented in $\mathbb{R}^2$ as

$$\begin{bmatrix} y_r \\ y_i \end{bmatrix} = \begin{bmatrix} h_r + g_r & g_i - h_i \\ h_i + g_i & h_r - g_r \end{bmatrix} \begin{bmatrix} x_r \\ x_i \end{bmatrix}. \tag{2.25}$$

Now, observe that the additional complex coefficient g, gives the necessary degrees of freedom to estimate the signal y. This is equivalent of estimating the complex-valued signal y from data x by treating the complex-valued signal as a bivariate real-valued vector.

Why widely linear?

An inquisitive reader at this point might be tempted to ask why widely linear estimation is needed when working directly in the real-domain can solve this problem of classical strictly linear complex-valued estimation. The answer here is closely related to the motivation of working in the complex domain in the first place.

Like the complex-valued representation, widely linear modelling provides a useful *conceptual framework* to represent several real-world phenonema, especially in the context of power engineering, where quantities like currents and voltages are represented and analysed as complex-valued phasors. For example, as it would be revealed in Chapter 7, the widely linear autoregressive model (WLAR) for a complex-valued voltage, s_k, which is given by

$$s_k = hs_{k-1} + gs_{k-1}^*, \tag{2.26}$$

is a natural way to represent unbalanced Clarke voltages as function of positive and negative sequence phasors. As opposed to working in the real-domain, the widely linear model offers greater "interpetability" whereby the conjugate part of the signal, s_{k-1}^*, represents a voltage phasor rotating at the opposite direction of the signal s_{k-1}. Moreover, the widely linear coefficient, g, corresponds to the *voltage imbalance factor* (VUF), which is standard metric used in power-systems to quantify the level of system imbalance across the three-phase voltages [19].

In addition, based on the WLAR model in (2.26), an adaptive "balancing" three-phase transform which is capable of yielding a voltage sequence that is non-oscillatory even when the underlying three-phase voltages are imbalanced will be developed in Chapter 7. A balancing transform had not been proposed in the past, but by virtue of the enhanced modelling capability of the widely linear model in (2.26), a simple closed form solution was made possible. In a nutshell, both complex-valued and widely linear representations are adopted in this thesis as they allow for a seamless integration between power-system analysis and signal processing algorithms.

2.4 Widely Linear AR modelling

Autoregressive (AR) modelling is a standard tool in time-series analysis, spectral estimation and machine learning applications involving predictions [48]. The intimate relation between AR modelling and sinusoidal frequency estimation motivates this brief background and shall be further explored in Chapter 7. The standard strictly linear AR(M) model of a signal, x_k, in the complex-domain is given by

$$x_k = \sum_{i=1}^{M} h_i^* x_{k-i} + \eta_k, \tag{2.27}$$

where η_k is a (possibly noncircular) white noise sequence with the variance and pseudocovariance, given by

$$\sigma_\eta^2 = \mathbb{E}\left\{|\eta_k|^2\right\}, \quad p_\eta^2 = \mathbb{E}\left\{\eta_k^2\right\}. \tag{2.28}$$

The coefficients $\boldsymbol{h} = [h_1, \ldots, h_M]^{\mathsf{T}}$ can be estimated by first expressing the conjugate of (2.27) as

$$\sum_{i=1}^{M} h_i x_{k-i}^* + \eta_k^* = x_k^*, \tag{2.29}$$

which in vector notation takes the form

$$\boldsymbol{x}_k^{\mathsf{H}} \boldsymbol{h} + \eta_k^* = x_k^*, \tag{2.30}$$

where the input data vector is $\boldsymbol{x}_k = [x_{k-1}, \ldots, x_{k-M}]^{\mathsf{T}}$ and coefficient vector is $\boldsymbol{h} = [h_1, \ldots h_M]^{\mathsf{T}}$. Pre-multiplying (2.30) with $\boldsymbol{x}_k$ and taking the statistical expectation, $\mathbb{E}\{\cdot\}$, gives

$$\mathbb{E}\left\{\boldsymbol{x}_k \boldsymbol{x}_k^{\mathsf{H}}\right\} \boldsymbol{h} + \mathbb{E}\left\{\boldsymbol{x}_k \eta_k^*\right\} = \mathbb{E}\left\{\boldsymbol{x}_k x_k^*\right\}. \tag{2.31}$$

The expression in (2.31) is reminiscent of the MSE solution (2.14), where $\mathbf{R} = \mathbb{E}\left\{\boldsymbol{x}_k \boldsymbol{x}_k^{\mathsf{H}}\right\} \in \mathbb{C}^{M \times M}$ with entries $r_\ell = \mathbb{E}\left\{x_k x_{k-\ell}^*\right\}$. Since the noise, η_k, is white, the cross-correlation values are $\mathbb{E}\left\{x_{k-\ell}\eta_k^*\right\} = \delta_\ell \sigma_\eta^2$, which yields the cross-correlation vectors in (2.31) in the form,

$$\mathbb{E}\left\{\boldsymbol{x}_k \eta_k^*\right\} = \mathbb{E}\left\{\begin{bmatrix} x_{k-1}\eta_k^* \\ \vdots \\ x_{k-M}\eta_k^* \end{bmatrix}\right\} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbb{E}\left\{\boldsymbol{x}_k x_k^*\right\} = \mathbb{E}\left\{\begin{bmatrix} x_{k-1}x_k^* \\ \vdots \\ x_{k-M}x_k^* \end{bmatrix}\right\} = \begin{bmatrix} r_1^* \\ \vdots \\ r_M^* \end{bmatrix}$$

Equation (2.31) can therefore be expressed as

$$\begin{bmatrix} r_0 & r_1 & \cdots & r_{M-1} \\ r_1^* & r_0 & \cdots & r_{M-2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{M-1}^* & r_{M-2}^* & \cdots & r_0 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_M \end{bmatrix} = \begin{bmatrix} r_1^* \\ r_2^* \\ \vdots \\ r_M^* \end{bmatrix}, \tag{2.32}$$

and is referred to as the Yule-Walker equations [48]. Solving the Yule-Walker equations in (2.32), will yield the coefficient vector, $\boldsymbol{h}$.

2.4.1 Widely Linear Autoregressive (WLAR) Modelling

As seen in Section 2.3, the standard strictly linear model is restrictive for general complex-valued signals, and calls for a widely linear variant of the AR model in (2.27). The widely linear AR(M) model which is autoregressive on both x_k and x_k^* can therefore be expressed as [49],

$$x_k = \sum_{i=1}^{M} h_i^* x_{k-i} + \sum_{i=1}^{M} g_i^* x_{k-i}^* + \eta_k, \tag{2.33}$$

where the statistics of the noise term η_k was defined in (2.28). The WLAR(M) model in (2.33) can also be expressed in vector form as

$$\bar{\boldsymbol{x}}_k^{\mathsf{H}} \begin{bmatrix} \boldsymbol{h} \\ \boldsymbol{g} \end{bmatrix} + \eta_k^* = x_k^*, \tag{2.34}$$

where the augmented input data vector $\bar{\boldsymbol{x}}_k = \begin{bmatrix} x_{k-1}, & \dots & , x_{k-M}, & x_{k-1}^*, & \dots & , x_{k-M}^* \end{bmatrix}^{\mathsf{T}}$, $\boldsymbol{h}$ is identical to the strictly linear case and $\boldsymbol{g} = [g_1, \dots, g_M]^{\mathsf{T}}$ is the WLAR coefficients that correspond to the conjugate data terms.

The augmented Yule-Walker equations can be found by pre-multiplying both sides of (2.34) with the augmented data vector, $\bar{\boldsymbol{x}}_k = \begin{bmatrix} \boldsymbol{x}_k^{\mathsf{T}}, & \boldsymbol{x}_k^{\mathsf{H}} \end{bmatrix}^{\mathsf{T}}$, and taking the statistical expectation, to give

$$\begin{bmatrix} \overbrace{\begin{matrix} r_0 & r_1 & \cdots & r_{M-1} \\ r_1^* & r_0 & \cdots & r_{M-2} \\ \vdots & \vdots & \vdots & \ddots \\ r_{M-1}^* & r_{M-2}^* & \cdots & r_0 \end{matrix}}^{\mathbf{R}} & \overbrace{\begin{matrix} p_0 & p_1 & \cdots & p_{M-1} \\ p_1 & p_0 & \cdots & p_{M-2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M-1} & p_{M-2} & \cdots & p_0 \end{matrix}}^{\mathbf{P}} \\ \underbrace{\begin{matrix} p_0^* & p_1^* & \cdots & p_{M-1}^* \\ p_1^* & p_0^* & \cdots & p_{M-2}^* \\ \vdots & \vdots & \ddots & \vdots \\ p_{M-1}^* & p_{M-2}^* & \cdots & p_0^* \end{matrix}}_{\mathbf{P}^*} & \underbrace{\begin{matrix} r_0 & r_1^* & \cdots & r_{M-1}^* \\ r_1 & r_0 & \cdots & r_{M-2}^* \\ \vdots & \vdots & \ddots & \vdots \\ r_{M-1} & r_{M-2} & \cdots & r_0 \end{matrix}}_{\mathbf{R}^*} \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_M \\ g_1 \\ \vdots \\ g_M \end{bmatrix} = \begin{bmatrix} r_1^* \\ \vdots \\ r_M^* \\ p_1^* \\ \vdots \\ p_M^* \end{bmatrix}, \tag{2.35}$$

which uses both the covariance, $\mathbf{R}$, and pseudocovariance, $\mathbf{P}$, of the input data $\boldsymbol{x}_k$. Solving the augmented Yule-Walker equation in (2.35) will then yield the AR coefficient vectors $\boldsymbol{h}$ and $\boldsymbol{g}$.

2.4.2 Semi-Widely Linear Yule-Walker Solution

Consider the special case of the WLAR(M) model in (2.33) where $\boldsymbol{g} = \boldsymbol{0}$. This corresponds to the the strictly linear AR model in (2.27). However, the widely linear Yule-Walker solution

given in (2.35) still holds, as the input data is improper, and the Yule-Walker equations simplify to

$$\left[\begin{array}{cccc} r_0 & r_1 & \cdots & r_{M-1} \\ r_1^* & r_0 & \cdots & r_{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ r_{M-1}^* & r_{M-2}^* & \cdots & r_0 \\ \hline p_0^* & p_1^* & \cdots & p_{M-1}^* \\ p_1^* & p_0^* & \cdots & p_{M-2}^* \\ \vdots & \vdots & \ddots & \vdots \\ p_{M-1}^* & p_{M-2}^* & \cdots & p_0^* \end{array}\right] \left[\begin{array}{c} h_1 \\ h_2 \\ \vdots \\ h_M \end{array}\right] = \left[\begin{array}{c} r_1^* \\ \vdots \\ r_M^* \\ \hline p_1^* \\ \vdots \\ p_M^* \end{array}\right], \tag{2.36}$$

The expression in (2.36) states that even if the AR model is strictly linear, due to the noncircularity of the driving noise, η_k, an augmented solution which uses both the covariance, $\mathbf{R}$, and pseudocovariance, $\mathbf{P}$, leads to more degrees of freedom in estimating $\boldsymbol{h}$.

2.5 Stochastic Gradient Adpative Filters

Both the widely linear Wiener solution in (2.16c) and Yule-Walker equations in (2.35) require the processing of the data statistics in the form of the covariance matrices and cross-covariance vectors. To develop estimators to track signals with time-varying statistics, we require sequential estimators which are capable of estimating the weight vectors $\boldsymbol{h}$ and $\boldsymbol{g}$ by recursively minimising the MSE cost function in (2.11).

For this task, we shall explore the ideas and results from the view of recursive adaptive algorithms which are closely related to stochastic approximations and online learning algorithms. Although the areas of adaptive algorithms and stochastic approximations are treated differently by different communities, they essentially describe the same concepts [50]. This thesis shall therefore adopt the terminology and treatment of this problem from an adaptive algorithms point of view.

Complex Least Mean Square

Consider an unknown system that is to be estimated and for which the measurements $y_k \in \mathbb{C}$, and regressor vector, $\boldsymbol{x}_k \in \mathbb{C}^{M\times 1}$, are related via the strictly linear relationship with true coefficient vector $\boldsymbol{h} \in \mathbb{C}^{M\times 1}$ given by

$$y_k \;=\; \boldsymbol{h}^{\mathsf{H}}\boldsymbol{x}_k + \eta_k, \tag{2.37}$$

where η_k is zero-mean white circular noise with variance $\sigma_\eta^2 = \mathbb{E}\left\{|\eta_k|^2\right\}$ and is independent to the regressor $\boldsymbol{x}_k$.

The well-established complex least mean square (CLMS) algorithm estimates the signal y_k as $\hat{y}_k = \widehat{\boldsymbol{h}}_k^{\mathsf{H}}\boldsymbol{x}_k$, using an estimated weight vector, $\widehat{\boldsymbol{h}}_k$, while the estimation error is defined as $\varepsilon_k = y_k - \hat{y}_k$ [50, 51]. This is accomplished by minimising the instantaneous squared error cost

function (stochastic approximation of the MSE) given by [50, 51]

$$\hat{J}_k(\widehat{\boldsymbol{h}}_k) \stackrel{\text{def}}{=} |\varepsilon_k|^2,$$

via gradient descent. The stochastic gradient technique yields the CLMS in the form [52]

$$\widehat{\boldsymbol{h}}_{k+1} = \widehat{\boldsymbol{h}}_k - \mu_k \frac{\partial \hat{J}_k(\widehat{\boldsymbol{h}}_k)}{\partial \boldsymbol{h}^*}\bigg|_{\boldsymbol{h}=\widehat{\boldsymbol{h}}_k}, \tag{2.38}$$

where the real-valued step-size, $\mu_k > 0$, controls the speed of convergence along the error performance surface.

In the field of stochastic approximation, the step-size, μ_k, is chosen to be a sequence which decays to zero and satisfies the additional conditions of, $\sum_{k=0}^{\infty} \mu_k = \infty$ and $\sum_{k=0}^{\infty} \mu_k^2 < \infty$ [50]. However, for applications that require continuous adaptation (like real-time frequency tracking), choosing a diminishing step-size is problematic since the algorithm might not be able to continuously track any changes in the signal statistics. Therefore, for the applications discussed in this thesis, a constant step-size, $\mu_k = \mu$, will be used to ensure the algorithm does not stop adapting the parameters under non-stationary conditions [97].

The gradient of the instantaneous squared error cost function, $\hat{J}_k(\widehat{\boldsymbol{h}}_k)$, is derived using

$$\frac{\partial \hat{J}_k(\widehat{\boldsymbol{h}}_k)}{\partial \boldsymbol{h}^*}\bigg|_{\boldsymbol{h}=\widehat{\boldsymbol{h}}_k} = \frac{\partial |\varepsilon_k|^2}{\partial \widehat{\boldsymbol{h}}_k^*} = \varepsilon_k^* \frac{\partial \varepsilon_k^*}{\partial \widehat{\boldsymbol{h}}_k^*} + \varepsilon_k \frac{\partial \varepsilon_k}{\partial \widehat{\boldsymbol{h}}_k}^*. \tag{2.39}$$

Since the estimation error is $\varepsilon_k = y_k - \widehat{\boldsymbol{h}}_k^{\mathsf{H}} \boldsymbol{x}_k$, the partial derivatives are given by [42]

$$\frac{\partial \varepsilon_k}{\partial \widehat{\boldsymbol{h}}_k^*} = \frac{\partial \left(y_k - \widehat{\boldsymbol{h}}_k^{\mathsf{H}} \boldsymbol{x}_k \right)}{\partial \widehat{\boldsymbol{h}}_k^*} = -\frac{\partial \widehat{\boldsymbol{h}}_k^{\mathsf{H}} \boldsymbol{x}_k}{\partial \widehat{\boldsymbol{h}}_k^*} = \boldsymbol{x}_k, \quad \& \quad \frac{\partial \varepsilon_k^*}{\partial \boldsymbol{h}_k^*} = \boldsymbol{0}. \tag{2.40}$$

Substituting the gradients obtained in (2.40) and (2.39) into the CLMS step in (2.38) yields the CLMS weight update equation in the form [52]

$$\widehat{\boldsymbol{h}}_{k+1} = \widehat{\boldsymbol{h}}_k + \mu \boldsymbol{x}_k \varepsilon_k^*. \tag{2.41}$$

The term "stochastic gradient" stems from the fact that the gradient of the instantaneous cost function, $\hat{J}_k(\widehat{\boldsymbol{h}}_k)$, is a "stochastic approximation" of the true gradient $\partial J_{\text{MSE}}(\boldsymbol{h})/\partial \boldsymbol{h}^*$. Therefore, unlike the true gradient, the path taken by the stochastic gradient does not always point to the true minimum of the error surface of $J_{\text{MSE}}(\boldsymbol{h})$ but takes a noisier route. Since the gradient update in (2.40) is "stochastic", the weight vector estimates, $\widehat{\boldsymbol{h}}_k$ are also treated as stochastic variables.

Augmented Complex Least Mean Square

In a similar fashion, the augmented complex least mean square (ACLMS) can be derived to estimate widely linear systems given by

$$\begin{aligned} y_k &= \boldsymbol{h}^{\mathsf{H}}\boldsymbol{x}_k + \boldsymbol{g}^{\mathsf{H}}\boldsymbol{x}_k^* + \eta_k \\ &= \boldsymbol{w}^{\mathsf{H}}\bar{\boldsymbol{x}}_k + \eta_k, \end{aligned} \tag{2.42}$$

where once again the true augmented weight vector is $\boldsymbol{w} = [\boldsymbol{h}^{\mathsf{T}},\ \boldsymbol{g}^{\mathsf{T}}]^{\mathsf{T}} \in \mathbb{C}^{2M\times 1}$ and the input vector $\bar{\boldsymbol{x}}_k = [\boldsymbol{x}_k^{\mathsf{T}}, \boldsymbol{x}_k^{\mathsf{H}}]^{\mathsf{T}} \in \mathbb{C}^{2M\times 1}$. The estimation error is $\varepsilon_k = y_k - \widehat{\boldsymbol{w}}_k^{\mathsf{H}}\bar{\boldsymbol{x}}_k$ while the instantaneous squared error cost function is $\hat{J}_k = |\varepsilon_k|^2$. Minimising $\hat{J}_k$ via gradient descent yields the ACLMS in the form

$$\begin{aligned} \widehat{\boldsymbol{w}}_{k+1} &= \widehat{\boldsymbol{w}}_k - \mu \frac{\partial \hat{J}_k}{\partial \boldsymbol{w}^*}\bigg|_{\boldsymbol{w}=\widehat{\boldsymbol{w}}_k} \\ &= \widehat{\boldsymbol{w}}_k + \mu\bar{\boldsymbol{x}}_k\varepsilon_k^*, \end{aligned} \tag{2.43}$$

or equivalently in the individual weight update forms given by [53]

$$\begin{aligned} \widehat{\boldsymbol{h}}_{k+1} &= \widehat{\boldsymbol{h}}_k + \mu\boldsymbol{x}_k\varepsilon_k^* \\ \widehat{\boldsymbol{g}}_{k+1} &= \widehat{\boldsymbol{g}}_k + \mu\boldsymbol{x}_k^*\varepsilon_k^*. \end{aligned} \tag{2.44}$$

2.6 Performance Analysis of Adaptive Algorithms

The stochastic nature of the the CLMS and ACLMS weight updates requires their performance to be analysed in a statistical setting. Specifically, we study the statistical behaviour of the *error* of the weight estimates. As an example, let us consider the ACLMS weight error defined as

$$\tilde{\boldsymbol{w}}_k \overset{\text{def}}{=} \boldsymbol{w} - \widehat{\boldsymbol{w}}_k, \tag{2.45}$$

where $\boldsymbol{w}$ refers to the true system weights from (2.42) and $\widehat{\boldsymbol{w}}_k$ is the ACLMS estimate from (2.43). Since the CLMS can be interpreted as a special case of the ACLMS, (i.e. with fewer parameters to estimate), we shall discuss the statistical performance of the ACLMS without loss of generality.

Definition 2.3 (Mean convergence). *An adaptive algorithm is said to converge in the mean if its weight estimates, $\widehat{\boldsymbol{w}}_k$, approach the true weight vector, $\boldsymbol{w}$, asymptotically, that is,*

$$\lim_{k\to\infty} \mathbb{E}\{\boldsymbol{w} - \widehat{\boldsymbol{w}}_k\} = \lim_{k\to\infty} \mathbb{E}\{\tilde{\boldsymbol{w}}_k\} = \boldsymbol{0}. \tag{2.46}$$

Definition 2.4 (β-Mean square stabilty). *An adaptive algorithm is said to be β-mean square stable if there exists real scalars, $\beta, \gamma > 0$ and $0 < \vartheta < 1$ such that the squared norm of its weight estimation error, $\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_k\|^2\right\}$, satisfies*

$$\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_k\|^2\right\} \leq \vartheta^k \mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_0\|^2\right\}\gamma + \beta, \tag{2.47}$$

for every $k \geq 0$, where $\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_0\|^2\right\}$ refers to the squared norm of the initial estimation error, while its steady-state value is bounded by

$$\limsup_{k\to\infty} \mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_k\|^2\right\} \leq \beta. \tag{2.48}$$

Henceforth, β-mean square stability shall be referred to as "mean square stability".

We shall typically begin with analysing the weight error covariance matrix, $\mathbf{K}_k \overset{\text{def}}{=} \mathbb{E}\left\{\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}}\right\}$ since $\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_k\|^2\right\} = \mathsf{Tr}[\mathbf{K}_k]$. Technically, the weight error covariance matrix is defined as

$$\mathbb{E}\left\{(\tilde{\boldsymbol{w}}_k - \mathbb{E}\{\tilde{\boldsymbol{w}}_k\})(\tilde{\boldsymbol{w}}_k - \mathbb{E}\{\tilde{\boldsymbol{w}}_k\})^{\mathsf{H}}\right\},$$

while the matrix $\mathbf{K}_k = \mathbb{E}\left\{\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}}\right\}$ refers to the second moment. However, for convenience, we shall use both the terms interchangeably.

Steady-state mean square performance. In addition to the mean and mean square performance defined in Definitions 2.3 and 2.4, the steady-state errors of adaptive algorithms are employed to quantify the fluctuations of the filter estimates around their true (optimal) values. The first metric is the mean square deviation (MSD) defined as

$$\text{MSD} \overset{\text{def}}{=} \lim_{k\to\infty} \mathbb{E}\left\{\left\|\tilde{\boldsymbol{w}}_k\right\|^2\right\}. \tag{2.49}$$

Next, the steady-state performance can also be evaluated with respect to the mean square error cost function, whereby

$$\lim_{k\to\infty} \mathbb{E}\left\{\hat{J}_k(\tilde{\boldsymbol{w}}_k)\right\} = \lim_{k\to\infty} \mathbb{E}\left\{|y_k - \hat{y}_k|^2\right\}. \tag{2.50}$$

Upon substituting the true system output, $y_k = \boldsymbol{w}^{\mathsf{H}}\bar{\boldsymbol{x}}_k + \eta_k$ in (2.42) and the ACLMS output estimate, $\hat{y}_k = \widehat{\boldsymbol{w}}_k^{\mathsf{H}}\bar{\boldsymbol{x}}_k$, into the steady-state MSE cost function in (2.50), the MSE can be decomposed into a minimum achievable MSE, $J_{\min}$, and the excess mean square error (EMSE) as

$$\lim_{k\to\infty} \mathbb{E}\left\{\hat{J}_k(\tilde{\boldsymbol{w}}_k)\right\} = \lim_{k\to\infty} \mathbb{E}\left\{|\tilde{\boldsymbol{w}}_k^{\mathsf{H}}\bar{\boldsymbol{x}}_k + \eta_k|^2\right\} = \underbrace{\lim_{k\to\infty} \mathbb{E}\left\{\tilde{\boldsymbol{w}}_k^{\mathsf{H}}\,\bar{\mathbf{R}}\,\tilde{\boldsymbol{w}}_k\right\}}_{\text{EMSE}} + \underbrace{\sigma_\eta^2}_{J_{\min}}. \tag{2.51}$$

From the decomposition in (2.51), and the definition of the MSD in (2.49) the standard metrics used to assess the steady-state performance can be summarised as

$$\text{EMSE} \overset{\text{def}}{=} \lim_{k\to\infty} \mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_k\|^2_{\bar{\mathbf{R}}}\right\} \tag{2.52}$$

$$\text{MSD} = \lim_{k\to\infty} \mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_k\|^2_{\mathbf{I}}\right\}, \tag{2.53}$$

where $\|\tilde{\boldsymbol{w}}_k\|^2_{\bar{\mathbf{R}}} = \tilde{\boldsymbol{w}}_k^{\mathsf{H}}\bar{\mathbf{R}}\tilde{\boldsymbol{w}}_k$ and $\|\tilde{\boldsymbol{w}}_k\|^2_{\mathbf{I}} = \tilde{\boldsymbol{w}}_k^{\mathsf{H}}\tilde{\boldsymbol{w}}_k$. Therefore, the EMSE and MSD can be interpreted as the Euclidean norm of the weight error vector weighted respectively by the data covariance matrix $\bar{\mathbf{R}}$ and the identity matrix, $\mathbf{I}$.

Mean convergence of the ACLMS

In this section, we shall examine the mean behaviour of the ACLMS algorithm by subtracting the true weights, $\boldsymbol{w}$, from both sides of equation (2.43), to give

$$\tilde{\boldsymbol{w}}_{k+1} = \tilde{\boldsymbol{w}}_k - \mu \bar{\boldsymbol{x}}_k \varepsilon_k^*. \tag{2.54}$$

where $\tilde{\boldsymbol{w}}_k = \boldsymbol{w} - \widehat{\boldsymbol{w}}_k$ is the weight error vector. Substituting the true system weights which produced the desired signal y_k in (2.42) into the ACLMS output error, $\varepsilon_k = y_k - \widehat{\boldsymbol{w}}_k^{\mathsf{H}} \bar{\boldsymbol{x}}_k$, gives

$$\varepsilon_k = \boldsymbol{w}_k^{\mathsf{H}} \bar{\boldsymbol{x}}_k + \eta_k - \widehat{\boldsymbol{w}}_k^{\mathsf{H}} \boldsymbol{x}_k = \tilde{\boldsymbol{w}}_k^{\mathsf{H}} \bar{\boldsymbol{x}}_k + \eta_k. \tag{2.55}$$

Substituting the output error from (2.55) into (2.54) yields

$$\tilde{\boldsymbol{w}}_{k+1} = \tilde{\boldsymbol{w}}_k - \mu \bar{\boldsymbol{x}}_k \bar{\boldsymbol{x}}_k^{\mathsf{H}} \tilde{\boldsymbol{w}}_k - \mu \bar{\boldsymbol{x}} \eta_k^*. \tag{2.56}$$

Theorem 2.1 (ACLMS Mean Convergence). *Consider the ACLMS algorithm in (2.43) for the associated true widely linear system in (2.42). Suppose that the following assumptions hold:*

A.1) The augmented input vector, $\bar{\boldsymbol{x}}_k$, has zero-mean and is independent and identically distributed (IID) across time k, with augmented covariance and pseudocovariance matrices respectively given by $\bar{\mathbf{R}} = \mathbb{E}\left\{\bar{\boldsymbol{x}}_k \bar{\boldsymbol{x}}_k^{\mathsf{H}}\right\}$, $\bar{\mathbf{P}} = \mathbb{E}\left\{\bar{\boldsymbol{x}}_k \bar{\boldsymbol{x}}_k^{\mathsf{T}}\right\}$.

A.2) The measurement noise sequence, $\{\eta_k\}_{k\in\mathbb{N}}$, is a zero-mean, IID white noise process which is statistically independent to input vector sequence, $\{\bar{\boldsymbol{x}}_k\}_{k\in\mathbb{N}}$, and its variance is given by, $\sigma_\eta^2 = \mathbb{E}\left\{|\eta_k|^2\right\}\ \forall k$.

Then, the ACLMS asymptotically converges in the mean, that is

$$\lim_{k\to\infty} \mathbb{E}\left\{\tilde{\boldsymbol{w}}_k\right\} = \mathbf{0}, \tag{2.57a}$$

if the step-size, μ, satisfies the following condition

$$0 < \mu < 2/\varrho\left(\bar{\mathbf{R}}\right), \tag{2.57b}$$

where $\varrho(\cdot)$ refers to the spectral radius and $\bar{\mathbf{R}}$ denotes the augmented covariance matrix,

$$\bar{\mathbf{R}} = \mathbb{E}\left\{\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^{\mathsf{H}}\right\} = \mathbb{E}\left\{\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{x}^*\end{bmatrix}\begin{bmatrix}\boldsymbol{x}^{\mathsf{H}} & \boldsymbol{x}^{\mathsf{T}}\end{bmatrix}\right\} = \begin{bmatrix}\mathbf{R} & \mathbf{P}\\ \mathbf{P}^* & \mathbf{R}^*\end{bmatrix}.$$

Proof.
Consider the ACLMS weight error vector recursion in (2.56). Upon taking the statistical expectation $\mathbb{E}\{\cdot\}$ and noticing that under Assumption A.1 in Theorem 2.1, the input vector, $\bar{\boldsymbol{x}}_k$, at time instant k is independent to $\widehat{\boldsymbol{w}}_k$ [54], we arrive at

$$\mathbb{E}\left\{\tilde{\boldsymbol{w}}_{k+1}\right\} = \mathbb{E}\left\{\tilde{\boldsymbol{w}}_k\right\} - \mu\mathbb{E}\left\{\bar{\boldsymbol{x}}_k \bar{\boldsymbol{x}}_k^{\mathsf{H}}\right\}\mathbb{E}\left\{\tilde{\boldsymbol{w}}_k\right\} - \mu\mathbb{E}\left\{\bar{\boldsymbol{x}}_k \eta_k^*\right\}, \tag{2.58}$$

where $\mathbb{E}\left\{\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^{\mathsf{H}}\right\} = \bar{\mathbf{R}}$, and under Assumption A.2 in Theorem 2.1, the measurement noise η_k is zero-mean and independent of the data $\bar{\boldsymbol{x}}_k$, thereby resulting in $\mathbb{E}\left\{\bar{\boldsymbol{x}}_k\eta_k^*\right\} = \mathbb{E}\left\{\bar{\boldsymbol{x}}_k\right\}\mathbb{E}\left\{\eta_k^*\right\} = \mathbf{0}$.

This yields the mean weight error recursion for the ACLMS in the form

$$\mathbb{E}\big\{\tilde{\boldsymbol{w}}_{k+1}\big\} = \left[\mathbf{I} - \mu\bar{\mathbf{R}}\right]\mathbb{E}\big\{\tilde{\boldsymbol{w}}_k\big\}. \tag{2.59}$$

The recursion in (2.59) involves mutual coupling of the individual elements of $\mathbb{E}\big\{\tilde{\boldsymbol{w}}_k\big\}$, due to the non-diagonal augmented covariance matrix, $\bar{\mathbf{R}}$.

To diagonalise (2.59), we use the fact that the augmented covariance matrix, $\bar{\mathbf{R}}$, is Hermitian positive-semidefinite and admits an eigen-decomposition in the form

$$\bar{\mathbf{R}} = \mathbf{V}\mathbf{D}_{\boldsymbol{r}}\mathbf{V}^{\mathsf{H}}, \tag{2.60}$$

where $\mathbf{V}$ is a unitary matrix with $\mathbf{V}^{\mathsf{H}}\mathbf{V} = \mathbf{I}$ and $\mathbf{D}_{\boldsymbol{r}} = \text{diag}\{\lambda_{\max}, \ldots, \lambda_{\min}\}$ is a diagonal matrix with real-valued eigenvalues, λ_i. Left-multiplying both sides of (2.59) with the eigenvector matrix, $\mathbf{V}^{\mathsf{H}}$, and substituting the eigen-decomposition of $\bar{\mathbf{R}}$ from (2.60) into (2.59) gives the "rotated" mean error recursion as

$$\mathbf{V}^{\mathsf{H}}\mathbb{E}\big\{\tilde{\boldsymbol{w}}_{k+1}\big\} = \left[\mathbf{V}^{\mathsf{H}} - \mu\mathbf{V}^{\mathsf{H}}\mathbf{V}\mathbf{D}_{\boldsymbol{r}}\mathbf{V}^{\mathsf{H}}\right]\mathbb{E}\big\{\tilde{\boldsymbol{w}}_k\big\}. \tag{2.61}$$

Upon defining $\boldsymbol{\theta}_k \overset{\text{def}}{=} \mathbf{Q}^{\mathsf{H}}\mathbb{E}\big\{\tilde{\boldsymbol{w}}_k\big\}$, the rotated recursion in (2.61) can be expressed as

$$\boldsymbol{\theta}_{k+1} = \underbrace{\left[\mathbf{I} - \mu\mathbf{D}_{\boldsymbol{r}}\right]}_{\mathbf{B}}\boldsymbol{\theta}_k. \tag{2.62}$$

For the ACLMS recursion in (2.62) to converge in the mean, the diagonal matrix, $\mathbf{B}$, needs to perform a contractive mapping. This is satisfied only if all the diagonal entries of $\mathbf{B}$, given by $(1 - \mu\lambda_i)$, satisfy the condition,

$$|1 - \mu\lambda_i| < 1, \quad i = 1, \ldots, M. \tag{2.63}$$

The most conservative bound, which guarantees convergence of (2.62), is for the eigenvalues in (2.63) to be $\lambda_i = \lambda_{\max} = \varrho(\bar{\mathbf{R}})$. This leads to the mean convergence condition of the ACLMS in the form

$$|1 - \mu\lambda_{\max}| < 1 \tag{2.64}$$

$$\implies 0 < \mu < \frac{2}{\varrho(\bar{\mathbf{R}})}, \tag{2.65}$$

thereby completing the proof. □

2.6.1 Mean Square Analysis of the ACLMS

To aid our discussion on the mean square convergence of the ACLMS as defined in Definition 2.4, we shall first proceed to prove the following lemma.

Lemma 2.2. *For Hermitian positive semi-definite matrices* $\mathbf{B}$ *and* $\bar{\mathbf{K}}_k = \mathbb{E}\left\{\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}}\right\}$*, the following condition holds for all* $k \geq 0$*:*

$$\mathsf{Tr}[\mathbf{B}\bar{\mathbf{K}}_k] \leq \varrho(\mathbf{B})\mathsf{Tr}[\bar{\mathbf{K}}_k], \tag{2.66}$$

where $\varrho(\mathbf{B})$ *denotes the spectral radius of the matrix* $\mathbf{B}$*.*

Proof.
Using the definition of $\bar{\mathbf{K}}_k = \mathbb{E}\left\{\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}}\right\}$, and the linearity of both the trace and expectation operators, we have

$$\mathsf{Tr}[\mathbf{B}\bar{\mathbf{K}}_k] = \mathsf{Tr}[\mathbf{B}\mathbb{E}\left\{\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}}\right\}] = \mathbb{E}\left\{\mathsf{Tr}[\mathbf{B}\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}}]\right\}.$$

Next, using the property that for any two matrices, $\mathbf{M}_1$ and $\mathbf{M}_2$, $\mathsf{Tr}[\mathbf{M}_1\mathbf{M}_2] = \mathsf{Tr}[\mathbf{M}_2\mathbf{M}_1]$ and the eigenvalue decomposition of $\mathbf{B} = \tilde{\mathbf{V}}\mathbf{\Lambda}_{\mathbf{B}}\tilde{\mathbf{V}}^{\mathsf{H}}$, gives

$$\begin{aligned}\mathsf{Tr}[\mathbf{B}\bar{\mathbf{K}}] &= \mathbb{E}\left\{\tilde{\boldsymbol{w}}_k^{\mathsf{H}}\tilde{\mathbf{V}}\mathbf{\Lambda}_{\mathbf{B}}\tilde{\mathbf{V}}^{\mathsf{H}}\tilde{\boldsymbol{w}}_k\right\} \\ &\leq \mathbb{E}\left\{\varrho(\mathbf{B})\|\tilde{\mathbf{V}}^{\mathsf{H}}\tilde{\boldsymbol{w}}_k\|^2\right\} = \varrho(\mathbf{B})\mathsf{Tr}[\bar{\mathbf{K}}_k],\end{aligned}$$

since $\|\tilde{\mathbf{V}}^{\mathsf{H}}\tilde{\boldsymbol{w}}_k\|^2 = \|\tilde{\boldsymbol{w}}_k\|^2$ as $\tilde{\mathbf{V}}$ is a unitary matrix. □

> **Contribution: Statistical analysis framework**
> Although this chapter is presented as a background to the thesis, the analysis framework for the mean square stability of the ACLMS in Theorem 2.3 is novel as it does not require the diagonalisation of the data covariance matrices. Current analyses often impose the requirement of diagonalising the data covariance matrices, which in many cases can prove to be mathematically intractable. To this end, Theorem 2.3 bounds the mean square error of the ACLMS in a simple methodology using the spectral radii of the covariance matrices of the data.

Theorem 2.3 (ACLMS Mean Square Stability). *Consider the ACLMS algorithm in (2.43) for the associated true widely linear system in (2.42). Suppose that in addition to the assumptions in Theorem 2.3, the following assumptions hold:*

A.3) The augmented input vector, $\bar{\boldsymbol{x}}_k$*, is drawn from a multivariate Gaussian distribution.*

Then, the ACLMS is mean square stable, that is, there exits a real-valued scalar $0 < \vartheta < 1$ *such that, the squared norm of the ACLMS weight error vector,* $\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_k\|^2\right\}$*, is bounded by*

$$\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_k\|^2\right\} \leq \vartheta^k\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_0\|^2\right\} + \left(\frac{1-\vartheta^k}{1-\vartheta}\right)\mu^2\sigma_\eta^2\mathsf{Tr}[\bar{\mathbf{R}}], \tag{2.67}$$

for all $k \geq 0$ *and its steady-state value is bounded by*

$$\limsup_{k\to\infty}\ \mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_k\|^2\right\} \leq \frac{\mu^2\sigma_\eta^2\mathsf{Tr}[\bar{\mathbf{R}}]}{1-\vartheta}. \tag{2.68}$$

Proof.
We start from the augmented weight error vector recursion from (2.56) which is given by

$$\tilde{\boldsymbol{w}}_{k+1} = \tilde{\boldsymbol{w}}_k - \mu\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^{\mathsf{H}}\tilde{\boldsymbol{w}}_k - \mu\bar{\boldsymbol{x}}_k\eta_k^*, \tag{2.69}$$

where $\tilde{\boldsymbol{w}}_k = \left[\boldsymbol{h}^{\mathsf{T}} - \widehat{\boldsymbol{h}}_k^{\mathsf{T}},\ \boldsymbol{g}^{\mathsf{T}} - \widehat{\boldsymbol{g}}_k^{\mathsf{T}}\right]^{\mathsf{T}}$, while the augmented weight error covariance matrix is defied as $\bar{\mathbf{K}}_k \overset{\text{def}}{=} \mathbb{E}\left\{\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}}\right\}$. By multiplying both sides of (2.69) by $\tilde{\boldsymbol{w}}_{k+1}^{\mathsf{H}}$, we have

$$\begin{aligned}\tilde{\boldsymbol{w}}_{k+1}\tilde{\boldsymbol{w}}_{k+1}^{\mathsf{H}} = \tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}} - \mu\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}}\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^{\mathsf{H}} - \mu\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^{\mathsf{H}}\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}} \\ + \mu^2\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^{\mathsf{H}}\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}}\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^{\mathsf{H}} + \mu^2\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^{\mathsf{H}}|\eta_k|^2 + \text{cross-terms}.\end{aligned} \tag{2.70}$$

Upon taking the statistical expectation $\mathbb{E}\{\cdot\}$ of (2.70) and considering Assumptions A.1 and A.2 in Theorem 2.1, and the Gaussian data assumption in Theorem 2.3, we arrive at

$$\begin{aligned}\bar{\mathbf{K}}_{k+1} = \bar{\mathbf{K}}_k - \mu\bar{\mathbf{K}}_k\bar{\mathbf{R}} - \mu\bar{\mathbf{R}}\bar{\mathbf{K}}_k + \mu^2\sigma_\eta^2\bar{\mathbf{R}} \\ + \mu^2(\bar{\mathbf{R}}\bar{\mathbf{K}}_k\bar{\mathbf{R}} + \bar{\mathbf{P}}\bar{\mathbf{K}}_k^T\bar{\mathbf{P}}^* + \bar{\mathbf{R}}\mathsf{Tr}[\bar{\mathbf{K}}_k\bar{\mathbf{R}}]),\end{aligned} \tag{2.71}$$

where the Gaussian moment factorizing theorem was employed to decompose the fourth order moments in (2.70) as

$$\mathbb{E}\left\{\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^{\mathsf{H}}\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}}\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^{\mathsf{H}}\right\} = \bar{\mathbf{R}}\bar{\mathbf{K}}_k\bar{\mathbf{R}} + \bar{\mathbf{P}}\bar{\mathbf{K}}_k^T\bar{\mathbf{P}}^* + \bar{\mathbf{R}}\mathsf{Tr}[\bar{\mathbf{K}}_k\bar{\mathbf{R}}],$$

and the augmented covariance matrix is $\bar{\mathbf{R}} = \mathbb{E}\left\{\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^{\mathsf{H}}\right\}$ while the augmented pseudocovariance matrix $\bar{\mathbf{P}}$ is

$$\bar{\mathbf{P}} = \mathbb{E}\left\{\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^{\mathsf{T}}\right\} = \mathbb{E}\left\{\begin{bmatrix}\boldsymbol{x}_k \\ \boldsymbol{x}_k^*\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_k^{\mathsf{T}} & \boldsymbol{x}_k^{\mathsf{H}}\end{bmatrix}\right\} = \begin{bmatrix}\mathbf{P} & \mathbf{R} \\ \mathbf{R}^* & \mathbf{P}^*\end{bmatrix}. \tag{2.72}$$

Applying the trace operator to the (2.71) gives

$$\begin{aligned}\mathsf{Tr}[\bar{\mathbf{K}}_{k+1}] = \ & \mathsf{Tr}[\bar{\mathbf{K}}_k - \mu\bar{\mathbf{K}}_k\bar{\mathbf{R}} - \mu\bar{\mathbf{R}}\bar{\mathbf{K}}_k + \mu^2\sigma_\eta^2\bar{\mathbf{R}}] \\ & + \mathsf{Tr}\left[\mu^2(\bar{\mathbf{R}}\bar{\mathbf{K}}_k\bar{\mathbf{R}} + \bar{\mathbf{P}}\bar{\mathbf{K}}_k^T\bar{\mathbf{P}}^* + \bar{\mathbf{R}}\mathsf{Tr}[\bar{\mathbf{K}}_k\bar{\mathbf{R}}])\right].\end{aligned} \tag{2.73}$$

Upon using the identity $\mathsf{Tr}[\mathbf{M}_1 + \mathbf{M}_2] = \mathsf{Tr}[\mathbf{M}_1] + \mathsf{Tr}[\mathbf{M}_2]$, for any square matrices $\mathbf{M}_1$ and $\mathbf{M}_2$, we have

$$\begin{aligned}\mathsf{Tr}[\bar{\mathbf{K}}_{k+1}] &= \mathsf{Tr}[(\mathbf{I} - 2\mu\bar{\mathbf{R}} + \mu^2\bar{\mathbf{R}}^2 + \mu^2\bar{\mathbf{P}}\bar{\mathbf{P}}^{\mathsf{H}} + \mu^2\mathsf{Tr}[\bar{\mathbf{R}}]\bar{\mathbf{R}})\bar{\mathbf{K}}_k] + \mu^2\sigma_\eta^2\mathsf{Tr}[\bar{\mathbf{R}}] \\ &= \mathsf{Tr}[\mathbf{B}\bar{\mathbf{K}}_k] + \mu^2\sigma_\eta^2\mathsf{Tr}[\bar{\mathbf{R}}],\end{aligned} \tag{2.74}$$

where $\mathbf{B} = \left((\mathbf{I}-\mu\bar{\mathbf{R}})^2 + \mu^2(\bar{\mathbf{P}}\bar{\mathbf{P}}^{\mathsf{H}} + \mathsf{Tr}[\bar{\mathbf{R}}]\bar{\mathbf{R}})\right)$ is a Hermitian positive semi-definite matrix. Using Lemma 2.2, the recursion in (2.74) can be bounded as

$$\mathsf{Tr}[\bar{\mathbf{K}}_{k+1}] \leq \varrho(\mathbf{B})\mathsf{Tr}[\bar{\mathbf{K}}_k] + \mu^2\sigma_\eta^2\mathsf{Tr}[\bar{\mathbf{R}}], \tag{2.75}$$

where $\varrho(\mathbf{B})$ denotes the spectral radius of the matrix $\mathbf{B}$. Since $\mathsf{Tr}[\bar{\mathbf{K}}_{k+1}] = \mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_{k+1}\|^2\right\}$, the

bound in (2.75) can be expressed as

$$\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_{k+1}\|^2\right\} \leq \vartheta \mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_k\|^2\right\} + \mu^2\sigma_\eta^2 \mathsf{Tr}[\bar{\mathbf{R}}],$$

or equivalently as

$$\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_k\|^2\right\} \leq \vartheta \mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_{k-1}\|^2\right\} + \mu^2\sigma_\eta^2 \mathsf{Tr}[\bar{\mathbf{R}}], \tag{2.76}$$

where $\vartheta = \varrho(\mathbf{B})$. For the recursion to converge the following condition needs to be satisfied:

$$\vartheta = \varrho(\mathbf{B}) < 1. \tag{2.77}$$

Using the Weyl inequality for the maximum eigenvalue for sums of Hermitian matrices, $\varrho(\mathbf{B})$ is further bounded by [55]

$$\varrho(\mathbf{B}) \leq \varrho((\mathbf{I} - \mu\bar{\mathbf{R}})^2) + \varrho(\mu^2\bar{\mathbf{P}}\bar{\mathbf{P}}^{\mathsf{H}}) + \varrho(\mu^2\mathsf{Tr}[\bar{\mathbf{R}}]\bar{\mathbf{R}}). \tag{2.78}$$

Therefore, for a conservative bound on the step-size μ, the following condition can be satisfied

$$\varrho((\mathbf{I} - \mu\bar{\mathbf{R}})^2) + \mu^2\varrho(\bar{\mathbf{P}}\bar{\mathbf{P}}^{\mathsf{H}}) + \mu^2\mathsf{Tr}[\bar{\mathbf{R}}]\varrho(\bar{\mathbf{R}}) < 1, \tag{2.79}$$

whereby upon defining $p_{\max}^2 \overset{\text{def}}{=} \varrho(\bar{\mathbf{P}}\bar{\mathbf{P}}^{\mathsf{H}})$, and noticing that $\lambda_{\max} = \varrho(\bar{\mathbf{R}})$ yields

$$\begin{aligned} 1 - 2\mu\lambda_{\min} + \mu^2\lambda_{\max}^2 + \mu^2 p_{\max}^2 + \mu^2\mathsf{Tr}[\bar{\mathbf{R}}]\lambda_{\max} < 1 \\ 2\lambda_{\min} - \mu(\lambda_{\max}^2 + p_{\max}^2 + \mathsf{Tr}[\bar{\mathbf{R}}]\lambda_{\max}) > 0. \end{aligned} \tag{2.80}$$

where $\lambda_{\min}$ refers to the minimum eigenvalue of $\bar{\mathbf{R}}$. The step-size, μ, which guarantees $\vartheta < 1$ and the mean square convergence of $\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_k\|^2\right\}$ is finally given by

$$0 < \mu < \frac{2\lambda_{\min}}{\lambda_{\max}^2 + p_{\max}^2 + \mathsf{Tr}[\bar{\mathbf{R}}]\lambda_{\max}}. \tag{2.81}$$

Applying the inequality in (2.76) successively for each k and using the formula for the sum of a geometric sequence, $1 + \vartheta + \cdots + \vartheta^{k-1} = (1 - \vartheta^k)/(1 - \vartheta)$, we obtain the expression in Theorem 2.3 as

$$\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_k\|^2\right\} \leq \vartheta^k \mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_0\|^2\right\} + \left(\frac{1 - \vartheta^k}{1 - \vartheta}\right)\mu^2\sigma_\eta^2\mathsf{Tr}[\bar{\mathbf{R}}]. \tag{2.82}$$

For the steady-state condition in (2.68), it is sufficient to observe the condition of the bound in (2.82) as $k \to \infty$. □

2.7 Chapter Summary

We have reviewed several important concepts in complex-valued statistical signal processing which are relevant to this thesis. In particular, it has been shown that standard complex-valued

statistical signal processing techniques which are straightforward extensions from the real domain do not take into account the full second order statistics of the signal, namely the covariance and pseudocovariance matrices. This has also been observed in the estimation of complex variables where widely linear estimation and has been shown to be suitable for the generality of complex-valued signals, while standard strictly linear models are only second-order optimal for jointly circular signals. Sequential estimators for the strictly linear and widely linear models, referred to as the CLMS and ACLMS, respectively, have been subsequently introduced. Finally, the statistical analysis for the ACLMS has been presented in Theorem 2.1 and Theorem 2.3.

Noncircularity and widely linear estimation underpins most of the ideas presented in this thesis and shall be further expanded upon in the sequel. For example, the computationally efficient widely linear estimation scheme in Chapter 3 provides a new perspective on the mean square error (MSE) estimation introduced in Section 2.3, while Chapter 4 presents an online algorithm to track the noncircularity of random variables using the second-order statistical measures presented in Section 2.2. In the second part of the thesis, widely linear adaptive filters that are outlined in Section 2.5, will be extended to a distributed setting and analysed in a similar fashion to this chapter. Finally, in Part III, widely linear AR modelling from Section 2.4 is used as a crucial link for the modelling of three-phase voltages in unbalanced power systems. Furthermore, the analysis framework introduced in this chapter in the proofs of Theorem 2.1 and Theorem 2.3 shall serve as a building block for the proofs in the subsequent chapters.

Part I

Adaptive Filters for Widely Linear Estimation

Chapter 3

Low Complexity Complex-Valued Adaptive Filters

> One essential object is to choose that arrangement which shall tend to reduce to a minimum the time necessary for completing the calculation.
>
> *Ada Lovelace*

Chapter Overview

In this chapter, widely linear estimation theory from Chapter 2 is revisited from a perspective of minimising the mean square error (MSE) in estimating the real and imaginary parts of as complex-valued signal. In Chapter 2, we reviewed why complex-valued representations of signals (in particular widely linear modelling) offer enhanced conceptual advantages at the signal modelling and algorithm design stage. However, this conceptual advantage comes at a cost of increasing the computational complexity of the algorithms that are designed for complex-valued data. To observe this, refer to the update equation of the ACLMS in (2.44) which requires twice the number of complex parameters ($\widehat{\boldsymbol{h}}_k$ and $\widehat{\boldsymbol{g}}_k$) to be updated compared the standard strictly linear CLMS in (2.41).

Of course, a sensible method to solve this problem is to work with bivariate real-valued vectors. However, doing so sacrifices the seamless integration between signal modelling and algorithm design that is gained from working with complex-valued signals. In this chapter, we offer a new way to integrate the concepts of complex-valued and real-valued estimation algorithms. Specifically, we show that complex-valued adaptive algorithms can be designed to perform widely linear estimation while also having identical computational requirements (real additions and multiplication) to their real-valued counterparts.

Furthermore, drawing upon ideas from widely linear estimation, we show that the number of real additions and multiplications can be further decreased by a half when the algorithms estimate strictly linear systems. This insight stems from working in the complex domain where the *a priori* knowledge about the signal structure (strictly linear vs. widely linear) can be used design computationally efficient adaptive filters.

3.1 Introduction

Complex-valued linear minimum mean square error (MMSE) estimation is an important statistical technique in communications and signal processing. It has now been accepted that the standard strictly linear model for complex data is not guaranteed to capture the complete

second-order statistical relationship between the input (regressor) and the output (observations), as generic strictly linear extensions of real-valued estimators cater only for data with rotation invariant probability distributions [42, 43, 56].

As shown in Chapter 2, using strictly linear estimators implicitly impose certain constraints on the signals. This constraint is that the linear coefficients relating the real parts of the input and output signals had to be identical to the coefficients relating the imaginary parts of the input and output signals. This constraint could also be explained using the second order statistics of the signals, whereby the strictly linear estimator only utilises the covariance matrix of the input data and cross-covariance vector of the input and output signals, while widely linear estimation exploits the full second order statistics of the data (both covariance and pseudocovariance matrices and vectors).

On the other hand, owing to the use of augmented variables, widely linear algorithms require twice the number of coefficients to update compared to their standard strictly linear counterparts. This adds to both higher computational complexity and a larger excess mean square error (EMSE). To deal with these issues, efficient formulations of widely linear adaptive filters have been proposed – these filters typically make use of the duality between the complex and real domain to cast the computations into the real domain [57, 58]. An efficient implementation of the ACLMS in [59] is one such approach. However, in this way, the input vector no longer resides in the original complex domain and any physical meaning or performance advantage inherent in working in the complex domain may be lost.

To help circumvent these problems and at the same time provide greater insight into complex-valued MMSE estimation, we propose an estimation technique, referred to as the complex dual channel (CDC) estimator, that is capable of performing both widely linear and strictly linear estimation at half the computational complexity (measured in number of real additions and multiplications). Stochastic gradient adaptive filters based on the CDC framework, referred to as the CLMSr, CLMSi (for proper signals) and the dual channel-CLMS (DC-CLMS) (for improper signals), are introduced and their convergence and stability properties are analysed. Simulations on synthetic data verify the analyses.

3.2 Proposed Solution: Complex Dual Channel Estimation

Minimum mean square error (MMSE) estimation aims to find the optimal second order estimate of a desired signal, $y \in \mathbb{C}$, given the regressor vector, $\boldsymbol{x} \in \mathbb{C}^{M\times 1}$. For strictly linear MMSE estimation, the data model is constrained to be strictly linear, $\hat{y} = \boldsymbol{h}^{\mathsf{H}}\boldsymbol{x}$, where $\hat{y}$ is the estimate of the desired signal, $\boldsymbol{h} \in \mathbb{C}^{M\times 1}$ the estimated coefficient vector, and the estimation error is given by $\varepsilon = y - \hat{y}$. The optimal second order mean square error (MSE) fit for the data is obtained by minimising the cost function

$$J_{\text{MSE}}(\boldsymbol{h}) = \mathbb{E}\left\{|\varepsilon|^2\right\} = \mathbb{E}\left\{|y-\hat{y}|^2\right\}. \tag{3.1}$$

Of particular importance to this work is the (often overlooked) observation that the MSE for estimating complex-valued signals combines two separate components: (i) the proportion of MSE in estimating the real part of the signal, and (ii) the proportion of MSE corresponding to

estimating the imaginary part of the signal. This becomes immediately clear when the error, ε, is rewritten in terms of its real component, $\varepsilon_r = \mathsf{Re}\{\varepsilon\}$ and imaginary component, $\varepsilon_i = Image\varepsilon$ as $\varepsilon = \varepsilon_r + j\varepsilon_i$, so that the cost function in (3.1) can be written as the sum of two real-valued cost functions as

$$J_{\text{MSE}}(\boldsymbol{h}) = \mathbb{E}\left\{\varepsilon_r^2\right\} + \mathbb{E}\left\{\varepsilon_i^2\right\} = J_r(\boldsymbol{h}) + J_i(\boldsymbol{h}). \tag{3.2}$$

Remark 3.1. *The linear MMSE estimator aims to minimize two real-valued cost functions, J_r and J_i, using a single complex-valued weight vector $\boldsymbol{h}$. However, one weight vector in general does not have enough degrees of freedom to minimize both $J_r(\boldsymbol{h})$ and $J_i(\boldsymbol{h})$, and the class of signals that admit such an MMSE estimator is very restrictive.*

To cater for a general class of complex-valued signals, a widely linear model model was proposed in [26], and is given by

$$\hat{y}_k = \boldsymbol{w}^{\mathsf{H}}\bar{\boldsymbol{x}} \quad \boldsymbol{w} = \begin{bmatrix}\boldsymbol{h}\\ \boldsymbol{g}\end{bmatrix} \quad \bar{\boldsymbol{x}} = \begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{x}^*\end{bmatrix}. \tag{3.3}$$

This estimator uses an augmented input vector, $\bar{\boldsymbol{x}}$, formed by concatenating the input, $\boldsymbol{x}$, and its conjugate, $\boldsymbol{x}^*$, and the corresponding augmented weight vector, $\boldsymbol{w}$. Based upon the augmented vector $\bar{\boldsymbol{x}}$, the sufficient second order statistics is obtained through the augmented covariance matrix

$$\bar{\mathbf{R}} = \mathbb{E}\left\{\begin{bmatrix}\boldsymbol{x}\boldsymbol{x}^{\mathsf{H}} & \boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}\\ \boldsymbol{x}^*\boldsymbol{x}^{\mathsf{H}} & \boldsymbol{x}^*\boldsymbol{x}^{\mathsf{T}}\end{bmatrix}\right\} = \begin{bmatrix}\mathbf{R} & \mathbf{P}\\ \mathbf{P}^* & \mathbf{R}^*\end{bmatrix}, \tag{3.4}$$

and the augmented cross-covariance vector

$$\bar{\boldsymbol{r}} = \mathbb{E}\begin{bmatrix}\boldsymbol{x}y^*\\ \boldsymbol{x}^*y^*\end{bmatrix} = \begin{bmatrix}\boldsymbol{r}\\ \boldsymbol{p}^*\end{bmatrix}. \tag{3.5}$$

Based on the MSE cost function in (3.2), we propose the use of two *strictly linear* estimators

$$\hat{y}_{cr} = \boldsymbol{w}_{cr}^{\mathsf{H}}\boldsymbol{x}, \quad \text{and} \quad \hat{y}_{ci} = \boldsymbol{w}_{ci}^{\mathsf{H}}\boldsymbol{x}, \tag{3.6}$$

where $\boldsymbol{w}_{cr} \in \mathbb{C}^{M\times 1}$ and $\boldsymbol{w}_{ci} \in \mathbb{C}^{M\times 1}$ and their corresponding estimation errors, $\varepsilon_r = \mathsf{Re}\{y - \hat{y}_{cr}\}$ and $\varepsilon_i = \mathsf{Im}\{y - \hat{y}_{ci}\}$, to minimise the cost functions, $J_r(\boldsymbol{w}_{cr}) = \mathbb{E}\left\{\varepsilon_r^2\right\}$ and $J_i(\boldsymbol{w}_{ci}) = \mathbb{E}\left\{\varepsilon_i^2\right\}$ independently. This makes it possible to have individual Wiener solutions, with the corresponding optimal weights $\boldsymbol{w}_{cr}^{\circ}$ and $\boldsymbol{w}_{ci}^{\circ}$ that minimize the cost functions $J_r(\boldsymbol{w}_{cr})$ and $J_i(\boldsymbol{w}_{ci})$ independently, and therefore the MSE, $J_{\text{MSE}}(\boldsymbol{w}_{cr}, \boldsymbol{w}_{ci})$, in (3.2).

The optimal weights $\boldsymbol{w}_{cr}^{\circ}$ and $\boldsymbol{w}_{ci}^{\circ}$ are next derived using a standard gradient methodology where the MSE in estimating the real part of the signal, $J_r(\boldsymbol{w}_{cr})$, can be expressed as

$$J_r(\boldsymbol{w}_{cr}) = \mathbb{E}\left\{\varepsilon_r^2\right\} = \mathbb{E}\left\{\left(\mathsf{Re}\{y\} - \mathsf{Re}\left\{\boldsymbol{w}_{cr}^{\mathsf{H}}\boldsymbol{x}\right\}\right)^2\right\}. \tag{3.7}$$

Taking the partial derivative of $J_r(\boldsymbol{w}_{cr})$ with respect to $\boldsymbol{w}_{cr}^*$ gives

$$\frac{\partial J_r(\boldsymbol{w}_{cr})}{\partial \boldsymbol{w}_{cr}^*} = \mathbb{E}\left\{2\varepsilon_r \frac{\partial}{\partial \boldsymbol{w}_{cr}^*}\left(\mathsf{Re}\{y\} - \frac{\boldsymbol{w}_{cr}^{\mathsf{H}}\boldsymbol{x} + \boldsymbol{w}_{cr}^{\mathsf{T}}\boldsymbol{x}^*}{2}\right)\right\} = -\mathbb{E}\{\boldsymbol{x}\varepsilon_r\}. \tag{3.8}$$

Next, denoting $\mathsf{Re}\{y\} = y_r$, we have

$$\varepsilon_r = y_r - (1/2)(\boldsymbol{x}^{\mathsf{T}}\boldsymbol{w}_{cr} + \boldsymbol{x}^{\mathsf{H}}\boldsymbol{w}_{ci}),$$

which can be substituted into (3.8) to give

$$\begin{aligned}\mathbb{E}\{\boldsymbol{x}\varepsilon_r\} &= \mathbb{E}\left\{\boldsymbol{x}(y_r - (1/2)(\boldsymbol{x}^{\mathsf{T}}\boldsymbol{w}_{cr}^* + \boldsymbol{x}^{\mathsf{H}}\boldsymbol{w}_{cr})\right\} \\ &= \boldsymbol{p}_{rx} - (1/2)(\mathbf{P}\boldsymbol{w}_{cr}^* + \mathbf{R}\boldsymbol{w}_{cr}),\end{aligned} \tag{3.9}$$

where $\mathbf{P} = \mathbb{E}\left\{\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}\right\}$ is the pseudocovariance matrix of the input, $\mathbf{R} = \mathbb{E}\left\{\boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}\right\}$ is the covariance matrix, $\boldsymbol{p}_{rx} = \mathbb{E}\{\boldsymbol{x}y_r\}$ is the cross-covariance between the real part of the desired signal and the complex input vector $\boldsymbol{x}$ and $\boldsymbol{p}_{ix} = \mathbb{E}\{\boldsymbol{x}y_i\}$ is the cross-covariance between the imaginary part of the desired signal and the input. To obtain the optimal Wiener solution, $\boldsymbol{w}_{cr}^{\circ}$, we set (3.9) to $\mathbf{0}$ to yield

$$\mathbf{P}\boldsymbol{w}^{\circ *}_{cr} + \mathbf{R}\boldsymbol{w}^{\circ}_{cr} = 2\boldsymbol{p}_{rx}, \tag{3.10a}$$

which upon complex conjugation takes the form,

$$\mathbf{P}^*\boldsymbol{w}^{\circ}_{cr} + \mathbf{R}^*\boldsymbol{w}^{\circ *}_{cr} = 2\boldsymbol{p}^*_{rx}. \tag{3.10b}$$

Solving (3.10a) and (3.10b) simultaneously gives the closed form answer,

$$\boldsymbol{w}^{\circ}_{cr} = 2[\mathbf{R} - \mathbf{P}\mathbf{R}^{*-1}\mathbf{P}^*]^{-1}[\boldsymbol{p}_{rx} - \mathbf{P}\mathbf{R}^{*-1}\boldsymbol{p}^*_{rx}]. \tag{3.11}$$

Similarly, the MSE in estimating the imaginary part of the signal, $J_i(\boldsymbol{w}_{ci})$, is given by

$$J_i(\boldsymbol{w}_{ci}) = \mathbb{E}\left\{\varepsilon_i^2\right\} = \mathbb{E}\left\{\left(\mathsf{Im}\{y\} - \mathsf{Im}\left\{\boldsymbol{w}_{ci}^{\mathsf{H}}\boldsymbol{x}\right\}\right)^2\right\}, \tag{3.12}$$

and the corresponding optimal estimator weights are

$$\boldsymbol{w}^{\circ}_{ci} = -2j[\mathbf{R} - \mathbf{P}\mathbf{R}^{*-1}\mathbf{P}^*]^{-1}[\boldsymbol{p}_{ix} - \mathbf{P}\mathbf{R}^{*-1}\boldsymbol{p}^*_{ix}]. \tag{3.13}$$

Remark 3.2. *Observe from* (3.11) *and* (3.13) *that in general, the weights that minimize* $J_r(\boldsymbol{w}_{cr})$ *in* (3.7) *are not those that minimize* $J_i(\boldsymbol{w}_{ci})$ *in* (3.12)*. This justifies the proposed individual minimisation of the contributing terms in the total MSE cost function,* J_{MSE} *in* (3.2)*.*

3.3 Relationship with Strictly Linear and Widely Linear Estimation

The equivalence of the proposed complex dual channel (CDC) model and the widely linear model can be established by factorising the output of the CDC estimator as

$$\hat{y} = \mathsf{Re}\left\{\boldsymbol{w}_{cr}^{\mathsf{H}}\boldsymbol{x}\right\} + j\mathsf{Im}\left\{\boldsymbol{w}_{ci}^{\mathsf{H}}\boldsymbol{x}\right\} \tag{3.14}$$

$$= \left(\frac{\boldsymbol{w}_{cr}+\boldsymbol{w}_{ci}}{2}\right)^{\mathsf{H}}\boldsymbol{x} + \left(\frac{\boldsymbol{w}_{cr}-\boldsymbol{w}_{ci}}{2}\right)^{\mathsf{T}}\boldsymbol{x}^* = \boldsymbol{h}^{\mathsf{H}}\boldsymbol{x} + \boldsymbol{g}^{\mathsf{H}}\boldsymbol{x}^*, \tag{3.15}$$

where $\boldsymbol{h} \overset{\text{def}}{=} \dfrac{\boldsymbol{w}_{cr}+\boldsymbol{w}_{ci}}{2}$ and $\boldsymbol{g}^* \overset{\text{def}}{=} \dfrac{\boldsymbol{w}_{cr}-\boldsymbol{w}_{ci}}{2}$. This form is equivalent to the widely linear model in (3.3), for more details see also [58].

Notice from (3.14) that when $\boldsymbol{w}_{cr} = \boldsymbol{w}_{ci}$, the coefficient vector $\boldsymbol{g} = \boldsymbol{0}$, which indicates that the underlying signal model is strictly linear. We shall now set $\boldsymbol{w}_{cr} = \boldsymbol{w}_{ci}$ in order to find the condition for one set of weights to be able to minimize both J_r, J_i (and therefore J_{MSE}). Upon equating (3.11) and (3.13), we obtain

$$\begin{aligned}[\boldsymbol{p}_{rx} - \mathbf{P}\mathbf{R}^{*-1}\boldsymbol{p}_{rx}^*] &= -j[\boldsymbol{p}_{ix} - \mathbf{P}\mathbf{R}^{*-1}\boldsymbol{p}_{ix}^*] \\ \boldsymbol{p}_{rx} + j\boldsymbol{p}_{ix} &= \mathbf{P}\mathbf{R}^{*-1}(\boldsymbol{p}_{rx}^* + j\boldsymbol{p}_{ix}^*).\end{aligned}$$

Recognising that $\boldsymbol{p}_{rx}^* + j\boldsymbol{p}_{ix}^* = \mathbb{E}\{\boldsymbol{x}^* y\} = \boldsymbol{r}^*$ is the complex valued cross-covariance and $\boldsymbol{p}_{rx} + j\boldsymbol{p}_{ix} = \mathbb{E}\{\boldsymbol{x}y\} = \boldsymbol{p}$ is the pseudocross-covariance, we arrive at the condition,

$$\boldsymbol{p} = \mathbf{P}\mathbf{R}^{*-1}\boldsymbol{r}^*. \tag{3.16}$$

Remark 3.3. *When estimating the real and imaginary parts of a signal with one set of complex weights, optimal performance is only possible when the desired signal y, and the regressor vector, $\boldsymbol{x}$ are jointly-circular, for example, the pseudocovariance, $\mathbf{P} = \mathbf{0}$, and the pseudocross-covariance, $\boldsymbol{p} = \mathbf{0}$.*

Remark 3.4. *Also notice that when $\boldsymbol{w}_{cr} = \boldsymbol{w}_{ci}$, the strictly linear weight $\boldsymbol{h}$ in (3.15) can either be estimated by $\boldsymbol{w}_{cr}$ or $\boldsymbol{w}_{ci}$. In the next section, we will show that this finding can be used to design algorithms that have half the the number of real additions and multiplications compared to the real-valued bivariate LMS.*

3.4 The Design of Adaptive Filters using the CDC Framework

We now provide a new framework to derive complex-valued adaptive filtering algorithms using the CDC estimation model. Our focus is on the stochastic gradient descent type of algorithms, and on benchmarking the CDC framework against the complex least mean square (CLMS) and augmented CLMS (ACLMS), which both minimize the global mean square error cost function $J_{\text{MSE}} = \mathbb{E}\left\{|\varepsilon|^2\right\}$. The CLMS uses a strictly linear model and is given by [52]

$$\begin{aligned}\varepsilon_k &= y_k - \widehat{\boldsymbol{h}}_k^{\mathsf{H}}\boldsymbol{x}_k \\ \widehat{\boldsymbol{h}}_{k+1} &= \widehat{\boldsymbol{h}}_k + \mu\boldsymbol{x}_k\varepsilon_k^*.\end{aligned} \tag{3.17}$$

The ACLMS, proposed in [53], uses the widely linear model that employs both $\boldsymbol{x}_k$ and $\boldsymbol{x}_k^*$, and is given by

$$\begin{aligned} \varepsilon_k &= y_k - \widehat{\boldsymbol{h}}_k^{\mathsf{H}} \boldsymbol{x}_k + \widehat{\boldsymbol{g}}_k^{\mathsf{H}} \boldsymbol{x}_k^* \\ \widehat{\boldsymbol{h}}_{k+1} &= \widehat{\boldsymbol{h}}_k + \mu \boldsymbol{x}_k \varepsilon_k^* \\ \widehat{\boldsymbol{g}}_{k+1} &= \widehat{\boldsymbol{g}}_k + \mu \boldsymbol{x}_k^* \varepsilon_k^*. \end{aligned} \tag{3.18}$$

3.4.1 Proposed Algorithm: Dual Channel CLMS

The dual channel-CLMS (DC-CLMS) is formed by a combining two stochastic gradient descent algorithms: the CLMSr and CLMSi. The CLMSr uses a strictly linear model $\hat{y}_{cr,k} = \boldsymbol{w}_{cr,k}^{\mathsf{H}} \boldsymbol{x}_k$ and aims to find the minimum of the instantaneous cost function $J_{r,k} = (\mathsf{Re}\{y_k - \hat{y}_{cr,k}\})^2$ that corresponds to the error in estimating the real part of the signal. The CLMSr is derived using the stochastic gradient framework [57, 60, 61] as

$$\boldsymbol{w}_{cr,k+1} = \boldsymbol{w}_{cr,k} - \mu \frac{\partial J_{r,k}}{\partial \boldsymbol{w}^*}\bigg|_{\boldsymbol{w}=\boldsymbol{w}_{cr,k}}, \tag{3.19}$$

which yields the following update scheme

$$\boldsymbol{w}_{cr,k+1} = \boldsymbol{w}_{cr,k} + \mu \mathsf{Re}\{\varepsilon_{cr,k}\} \boldsymbol{x}_k, \tag{3.20}$$

where $\varepsilon_{cr,k} = y_k - \hat{y}_{cr,k}$.

Similarly, the CLMSi uses a strictly linear model $\hat{y}_{ci,k} = \boldsymbol{w}_{ci,k}^{\mathsf{H}} \boldsymbol{x}_k$ and aims to find the minimum of the instantaneous cost function $J_{i,k} = (\mathsf{Im}\{y_k - \hat{y}_{ci,k}\})^2$, that corresponds to the error in estimating the imaginary part of the signal. The CLMSi weight update is therefore given by

$$\boldsymbol{w}_{ci,k+1} = \boldsymbol{w}_{ci,k} - j\mu \mathsf{Im}\{\varepsilon_{ci,k}\} \boldsymbol{x}_k \tag{3.21}$$

where $\varepsilon_{ci,k} = y_k - \hat{y}_{ci,k}$.

The DC-CLMS is summarised in Algorithm 1 and as illustrated in Figure 3.1, the DC-CLMS operates in a collaborative fashion by combining the real output of CLMSr with the imaginary output of CLMSi to obtain an output identical to that of the ACLMS.

> **Contribution: Dual Channel Adaptive Filters**
> In addition to theory of dual channel estimation, we propose a low complexity adaptive filter, DC-CLMS in Algorithm 1. The proposed DC-CLMS algorithm is capable of performing widely linear estimation with half the number of computations per iteration. The computational benefits are outlined in Table 3.1. Furthermore, we present a novel framework for statistical analysis of the DC-CLMS in Theorem 3.1 and 3.2.

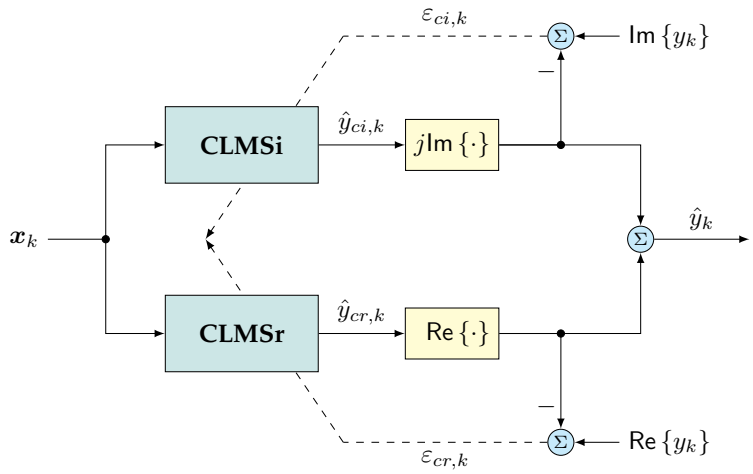

Figure 3.1: Architecture of the DC-CLMS.

Algorithm 1. Dual Channel CLMS (DC-CLMS)

Initialise with: $\boldsymbol{w}_{cr,0} = \mathbf{0}$ and $\boldsymbol{w}_{ci,0} = \mathbf{0}$
At each time instant $k > 0$:

procedure CLMSR
  $\hat{y}_{cr,k} = \boldsymbol{w}_{cr,k}^{\mathsf{H}} \boldsymbol{x}_k$
  $\varepsilon_{cr,k} = y_k - \hat{y}_{cr,k}$
  $\boldsymbol{w}_{cr,k+1} = \boldsymbol{w}_{cr,k} + \mu \mathsf{Re}\{\varepsilon_{cr,k}\} \boldsymbol{x}_k$
end procedure
procedure CLMSI
  $\hat{y}_{ci,k} = \boldsymbol{w}_{ci,k}^{\mathsf{H}} \boldsymbol{x}_k$
  $\varepsilon_{ci,k} = y_k - \hat{y}_{ci,k}$
  $\boldsymbol{w}_{ci,k+1} = \boldsymbol{w}_{ci,k} - j\mu \mathsf{Im}\{\varepsilon_{ci,k}\} \boldsymbol{x}_k$
end procedure
$\hat{y}_k = \mathsf{Re}\{\hat{y}_{cr,k}\} + j\mathsf{Im}\{\hat{y}_{ci,k}\}$

Implementation of the DC-CLMS

Notice from Algorithm 1 that the CLMSr subfilter, only requires the real part of the estimate $\hat{y}_{cr,k}$. Therefore, only the real part of the inner product $\boldsymbol{w}_{cr,k}^{\mathsf{H}}\boldsymbol{x}_k$, needs to be computed. The same reasoning applies for the imaginary part of the CLMSi output. Herein lies the key to the computational effectiveness of the proposed DC-CLMS. Although this scheme effectively describes the bivariate real-valued LMS, our contribution here is to show that the DC-CLMS algorithm can be derived and presented in the complex domain to suit the needs of practitioners who gain physical insight from working in the complex domain. Moreover, our insight in Remark 3.4 shows that we can further reduce the number of operations of the DC-CLMS for strictly linear systems if only the CLMSr or CLMSi is implemented. To the best of the author's knowledge, this finding is novel and has not been presented in the literature of complex-valued adaptive filters in the past.

3.4.2 Comparison with Existing Reduced Complexity Algorithms

For rigour, we now compare the CLMSr/i and DC-CLMS algorithms with the two existing reduced complexity augmented complex least mean square (ACLMS) algorithms:

a. Dual-channel real-valued LMS (DCRLMS) algorithm, a real-valued algorithm that exploits the duality between $\mathbb{R}^2$ and $\mathbb{C}$ [58].

b. Reduced complexity widely linear LMS (RC-WL-LMS), a complex-valued algorithm which employs a complex-valued weight vector and a real-valued input vector (constructed by augmenting the real and imaginary parts of the original complex-valued input vector) [59].

Although, the dual-channel real-valued LMS (DCRLMS) [58] has sufficient degrees of freedom to model complex-valued signals, its parameters reside in $\mathbb{R}$, and any physical insight that can be gained from the signal model is obscured. On the other hand, the RC-WL-LMS [59] does operate in the complex domain but has a limitation since its aim is only to reduce the computational complexity. In contrast to the DCRLMS and RC-WL-LMS, the framework presented in this paper allows for a unified and efficient formulation of both the strictly linear CLMS algorithm in the form of the CLMSr/i sub-filters and the ACLMS algorithm by combining the CLMSr/i sub-filters. Table 3.1 compares the number operations required per iteration for all the complex LMS algorithms considered in this paper.

3.5 Statistical Analysis of the DC-CLMS

Consider the system given by

$$y_k = \mathsf{Re}\left\{\boldsymbol{w}_{cr}^{\circ\,\mathsf{H}}\boldsymbol{x}_k + \eta_k\right\} + j\mathsf{Im}\left\{\boldsymbol{w}_{ci}^{\circ\,\mathsf{H}}\boldsymbol{x}_k + \eta_k\right\}, \tag{3.22}$$

to be the true system which needs to be identified by the CLMSr and CLMSi algorithms. Defining the weight error vectors, $\tilde{\boldsymbol{w}}_{cr,k} \overset{\text{def}}{=} \boldsymbol{w}_{cr}^{\circ} - \boldsymbol{w}_{cr,k}$ and $\tilde{\boldsymbol{w}}_{ci,k} \overset{\text{def}}{=} \boldsymbol{w}_{ci}^{\circ} - \boldsymbol{w}_{ci,k}$ for the CLMSr and

Identification of Strictly Linear Models		
Algorithm	Real Multiplications	Real Additions
CLMS	$8M+2$	$8M$
Proposed: CLMSr	$4M+1$	$4M$
Proposed: CLMSi	$4M+1$	$4M$
Identification of Widely Linear Models		
Algorithm	Real Multiplications	Real Additions
ACLMS	$16M+2$	$16M$
RC-WL-LMS [59]	$8M+2$	$8M$
DCRLMS [58]	$8M+2$	$8M$
Proposed: DC-CLMS	$8M+2$	$8M$

Table 3.1: *Computational requirements per iteration for the considered complex LMS algorithms, where M is the length of the complex-valued input vector $\boldsymbol{x}_k$.*

CLMSi algorithms and subtracting $\boldsymbol{w}_{cr}^{\circ}$ and $\boldsymbol{w}_{ci}^{\circ}$ from the CLMSr/i algorithms in (3.20) – (3.21) respectively gives

$$\text{CLMSr:}\quad \tilde{\boldsymbol{w}}_{cr,k+1} = \tilde{\boldsymbol{w}}_{cr,k} - \mu \mathsf{Re}\left\{\varepsilon_{cr,k}\right\} \boldsymbol{x}_k \tag{3.23a}$$

$$\text{CLMSi:}\quad \tilde{\boldsymbol{w}}_{ci,k+1} = \tilde{\boldsymbol{w}}_{ci,k} + j\mu \mathsf{Im}\left\{\varepsilon_{ci,k}\right\} \boldsymbol{x}_k. \tag{3.23b}$$

Theorem 3.1 (DC-CLMS Mean Convergence). *Consider the CLMSr/i algorithms in (3.20) – (3.21) for the associated true system in (3.22). Suppose that the following assumptions hold:*

A.1) The input vector, $\boldsymbol{x}_k$, has zero-mean and is independent and identically distributed (IID) across time k, with covariance and pseudocovariance matrices respectively given by $\mathbf{R} = \mathbb{E}\left\{\boldsymbol{x}_k \boldsymbol{x}_k^{\mathsf{H}}\right\}$, $\mathbf{P} = \mathbb{E}\left\{\boldsymbol{x}_k \boldsymbol{x}_k^{\mathsf{T}}\right\}$.

A.2) The measurement noise sequence, $\{\eta_k\}_{k\in\mathbb{N}}$, is a zero-mean, IID circular white noise process which is statistically independent to input vector sequence, $\{\boldsymbol{x}_k\}_{k\in\mathbb{N}}$, and its variance given by, $\sigma_\eta^2 = \mathbb{E}\left\{|\eta_k|^2\right\}$.

Then, the DC-CLMS algorithm asymptotically converges in the mean, that is

$$\lim_{k\to\infty} \mathbb{E}\left\{\boldsymbol{w}_{cr}^{\circ} - \boldsymbol{w}_{cr,k}\right\} = \mathbf{0}, \quad \textit{and}, \quad \lim_{k\to\infty} \mathbb{E}\left\{\boldsymbol{w}_{ci}^{\circ} - \boldsymbol{w}_{ci,k}\right\} = \mathbf{0}, \tag{3.24a}$$

if the step-size μ for both the CLMSr and CLMSi algorithms, satisfies the following condition

$$0 < \mu < 4/\varrho\left(\bar{\mathbf{R}}\right), \tag{3.24b}$$

where $\varrho(\cdot)$ refers to the spectral radius of the augmented covariance matrix $\bar{\mathbf{R}}$.

Proof.
For the DC-CLMS to converge in the mean, both of its sub-filters, CLMSr and CLMSi have to

convergence in the mean. Starting from (3.23a) and noticing that $\mathsf{Re}\{\varepsilon_{cr,k}\} = \mathsf{Re}\left\{y_k - \boldsymbol{w}_{cr,k}^{\mathsf{H}}\boldsymbol{x}_k\right\} = \mathsf{Re}\left\{\tilde{\boldsymbol{w}}_{cr,k}^{\mathsf{H}}\boldsymbol{x}_k + \eta_k\right\}$ from the true system in (3.22) gives

$$\text{CLMSr:}\quad \tilde{\boldsymbol{w}}_{cr,k+1} = \tilde{\boldsymbol{w}}_{cr,k} - \mu\boldsymbol{x}_k\mathsf{Re}\left\{\boldsymbol{x}_k^{\mathsf{H}}\tilde{\boldsymbol{w}}_{cr,k} + \eta_k^*\right\} \tag{3.25a}$$

$$\text{CLMSi:}\quad \tilde{\boldsymbol{w}}_{ci,k+1} = \tilde{\boldsymbol{w}}_{ci,k} + j\mu\boldsymbol{x}_k\mathsf{Im}\left\{\boldsymbol{x}_k^{\mathsf{T}}\tilde{\boldsymbol{w}}_{ci,k}^* + \eta_k\right\}, \tag{3.25b}$$

where a similar step was taken for the CLMSi in (3.23b) to arrive at (3.25b). Using the relationship $\mathsf{Re}\{a\} = (1/2)(a + a^*)$ and $\mathsf{Im}\{a\} = (1/2j)(a - a^*)$, we obtain

$$\text{CLMSr:}\quad \tilde{\boldsymbol{w}}_{cr,k+1} = \tilde{\boldsymbol{w}}_{cr,k} - \frac{\mu}{2}\boldsymbol{x}_k\left(\boldsymbol{x}_k^{\mathsf{H}}\tilde{\boldsymbol{w}}_{cr,k} + \boldsymbol{x}_k^{\mathsf{T}}\tilde{\boldsymbol{w}}_{cr,k}^*\right) - \mu\boldsymbol{x}_k\mathsf{Re}\{\eta_k^*\} \tag{3.26}$$

$$\text{CLMSi:}\quad \tilde{\boldsymbol{w}}_{ci,k+1} = \tilde{\boldsymbol{w}}_{ci,k} + j\frac{\mu}{2j}\boldsymbol{x}_k\left(\boldsymbol{x}_k^{\mathsf{T}}\tilde{\boldsymbol{w}}_{ci,k}^* - \boldsymbol{x}_k^{\mathsf{H}}\tilde{\boldsymbol{w}}_{ci,k}\right) + j\mu\boldsymbol{x}_k\mathsf{Im}\{\eta_k\}. \tag{3.27}$$

Upon applying the statistical expectation operator, $\mathbb{E}\{\cdot\}$ and applying Assumptions A.1 and A.2 in Theorem 3.1, we have

$$\text{CLMSr:}\quad \mathbb{E}\{\tilde{\boldsymbol{w}}_{cr,k+1}\} = \left(\mathbf{I} - \frac{\mu}{2}\mathbf{R}\right)\mathbb{E}\{\tilde{\boldsymbol{w}}_{cr,k}\} - \frac{\mu}{2}\mathbf{P}\mathbb{E}\{\tilde{\boldsymbol{w}}_{cr,k}^*\} \tag{3.28a}$$

$$\text{CLMSi:}\quad \mathbb{E}\{\tilde{\boldsymbol{w}}_{ci,k+1}\} = \left(\mathbf{I} - \frac{\mu}{2}\mathbf{R}\right)\mathbb{E}\{\tilde{\boldsymbol{w}}_{ci,k}\} + \frac{\mu}{2}\mathbf{P}\mathbb{E}\{\tilde{\boldsymbol{w}}_{ci,k}^*\}, \tag{3.28b}$$

where $\mathbf{R} = \mathbb{E}\{\boldsymbol{x}_k\boldsymbol{x}_k^{\mathsf{H}}\}$ is the covariance matrix and $\mathbf{P} = \mathbb{E}\{\boldsymbol{x}_k\boldsymbol{x}_k^{\mathsf{T}}\}$ the pseudocovariance matrix of the input data. Recall that the weight error recursion for the CLMS is given by [42]

$$\text{CLMS:}\quad \mathbb{E}\{\tilde{\boldsymbol{h}}_{k+1}\} = (\mathbf{I} - \mu\mathbf{R})\mathbb{E}\{\tilde{\boldsymbol{h}}_k\}. \tag{3.29}$$

Upon concatenating the weight error recursion for the CLMSr in (3.28a) and its conjugate, we have

$$\mathbb{E}\left\{\begin{bmatrix}\tilde{\boldsymbol{w}}_{cr,k+1}\\ \tilde{\boldsymbol{w}}_{cr,k+1}^*\end{bmatrix}\right\} = \left(\begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix} - \frac{\mu}{2}\underbrace{\begin{bmatrix}\mathbf{R} & \mathbf{P}\\ \mathbf{P}^* & \mathbf{R}^*\end{bmatrix}}_{\bar{\mathbf{R}}}\right)\mathbb{E}\left\{\begin{bmatrix}\tilde{\boldsymbol{w}}_{cr,k}\\ \tilde{\boldsymbol{w}}_{cr,k}^*\end{bmatrix}\right\}, \tag{3.30}$$

where $\bar{\mathbf{R}}$ is the augmented covariance matrix defined in (3.4). Observe that the recursion in (3.30) resembles that of the ACLMS introduced in Chapter 2. The CLMSr step-size bound can be derived in a similar fashion with

$$\text{CLMSr: } 0 < \mu < \frac{4}{\varrho(\bar{\mathbf{R}})}, \tag{3.31}$$

where $\varrho(\cdot)$ denotes the spectral radius of a matrix. Since $\bar{\mathbf{R}}$ is positive semi-definite, $\varrho(\bar{\mathbf{R}})$ denotes the maximum eigenvalue of the augmented covariance matrix.

Similarly, for the CLMSi sub-filter in (3.28b), the recursion for the mean error vector and its

conjugate is given by

$$\begin{bmatrix} \mathbb{E}\{\tilde{\boldsymbol{w}}_{ci,k+1}\} \\ \mathbb{E}\left\{\tilde{\boldsymbol{w}}^*_{ci,k+1}\right\} \end{bmatrix} = \left(\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} - \frac{\mu}{2} \underbrace{\begin{bmatrix} \mathbf{R} & -\mathbf{P} \\ -\mathbf{P}^* & \mathbf{R} \end{bmatrix}}_{\bar{\mathbf{R}}'} \right) \begin{bmatrix} \mathbb{E}\{\tilde{\boldsymbol{w}}_{ci,k}\} \\ \mathbb{E}\left\{\tilde{\boldsymbol{w}}^*_{ci,k}\right\} \end{bmatrix}, \tag{3.32}$$

which yields the same stability bound as in (3.31). This completes the proof. □

Remark 3.5. *The spectral radius of* $\bar{\mathbf{R}}'$ *in* (3.32) *is identical to that of the augmented covariance matrix* $\bar{\mathbf{R}} = \mathbb{E}\left\{\bar{\boldsymbol{x}}_k \bar{\boldsymbol{x}}_k^{\mathsf{H}}\right\}$ *in* (3.30). *To prove this, consider an eigenvector-eigenvalue pair,* $(\lambda_i, \boldsymbol{q}_i)$, *of the augmented covariance matrix* $\bar{\mathbf{R}}$ *where the eigenvector* $\boldsymbol{q}_i$ *can be expressed as* $\boldsymbol{q}_i = \left[\boldsymbol{q}_{i,1}^{\mathsf{T}} \; \boldsymbol{q}_{i,2}^{\mathsf{T}}\right]^{\mathsf{T}}$, *so that*

$$\bar{\mathbf{R}}\boldsymbol{q}_i = \begin{bmatrix} \mathbf{R} & \mathbf{P} \\ \mathbf{P}^* & \mathbf{R} \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_{i,1} \\ \boldsymbol{q}_{i,2} \end{bmatrix} = \lambda_i \begin{bmatrix} \boldsymbol{q}_{i,1} \\ \boldsymbol{q}_{i,2} \end{bmatrix} = \lambda_i \boldsymbol{q}_i. \tag{3.33}$$

Now, consider multiplying $\bar{\mathbf{R}}'$ *with the vector* $\boldsymbol{q}_i' \overset{def}{=} \left[\boldsymbol{q}_{i,1}^{\mathsf{T}} \; -\boldsymbol{q}_{i,2}^{\mathsf{T}}\right]^{\mathsf{T}}$ *to give*

$$\bar{\mathbf{R}}'\boldsymbol{q}_i' = \begin{bmatrix} \mathbf{R} & -\mathbf{P} \\ -\mathbf{P}^* & \mathbf{R} \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_{i,1} \\ -\boldsymbol{q}_{i,2} \end{bmatrix} = \lambda_i \begin{bmatrix} \boldsymbol{q}_{i,1} \\ -\boldsymbol{q}_{i,2} \end{bmatrix} = \lambda_i \boldsymbol{q}_i', \tag{3.34}$$

where λ_i *is the eigenvalue of the* $\bar{\mathbf{R}}$. *Performing the same process for all the other eigenvalues* λ_i, *will reveal that the eigenvalues (hence the spectral radius) of* $\bar{\mathbf{R}}$ *are identical to* $\bar{\mathbf{R}}'$.

By comparing the weight error recursions for DC-CLMS in (3.28a)–(3.28b) and the CLMS from (3.29), observe that the evolution of the weight vectors will be identical if the data is proper (i.e. when $\mathbf{P} = \mathbf{0}$). This is supported by the simulation result in Figure 3.2, which shows the average weight trajectories along the error surfaces of the CLMSr, CLMSi and CLMS.

3.5.1 Mean Square Stability

Theorem 3.2 (DC-CLMS Mean Square Stability). *Consider the CLMSr/i algorithms in* (3.20) – (3.21) *for the associated true system in* (3.22). *In addition to the Assumptions A.1 – A.2 in Theorem 3.1, suppose that the following assumptions hold:*

A.3) The input vector, $\boldsymbol{x}_k$, *is drawn from a multivariate Gaussian distribution.*

Then, the CLMSr/i algorithms (hence the DC-CLMS algorithm) are mean-square stable, that is, there exits a real-valued scalar $0 < \vartheta < 1$ *such that*

$$\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_{cr/ci,k}\|^2\right\} \leq \vartheta^k \mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_{cr/ci,0}\|^2\right\} + \frac{1}{4}\left(\frac{1-\vartheta^k}{1-\vartheta}\right)\mu^2\sigma_\eta^2 \mathsf{Tr}[\bar{\mathbf{R}}], \tag{3.35}$$

for all $k \geq 0$ *and its steady-state value is bounded by*

$$\limsup_{k\to\infty} \mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_{cr/ci,k}\|^2\right\} \leq \frac{1}{4}\frac{\mu^2\sigma_\eta^2 \mathsf{Tr}[\bar{\mathbf{R}}]}{1-\vartheta}. \tag{3.36}$$

where the subscript "cr/ci*" indicates that the bounds hold for both the CLMSr and CLMSi algorithms.*

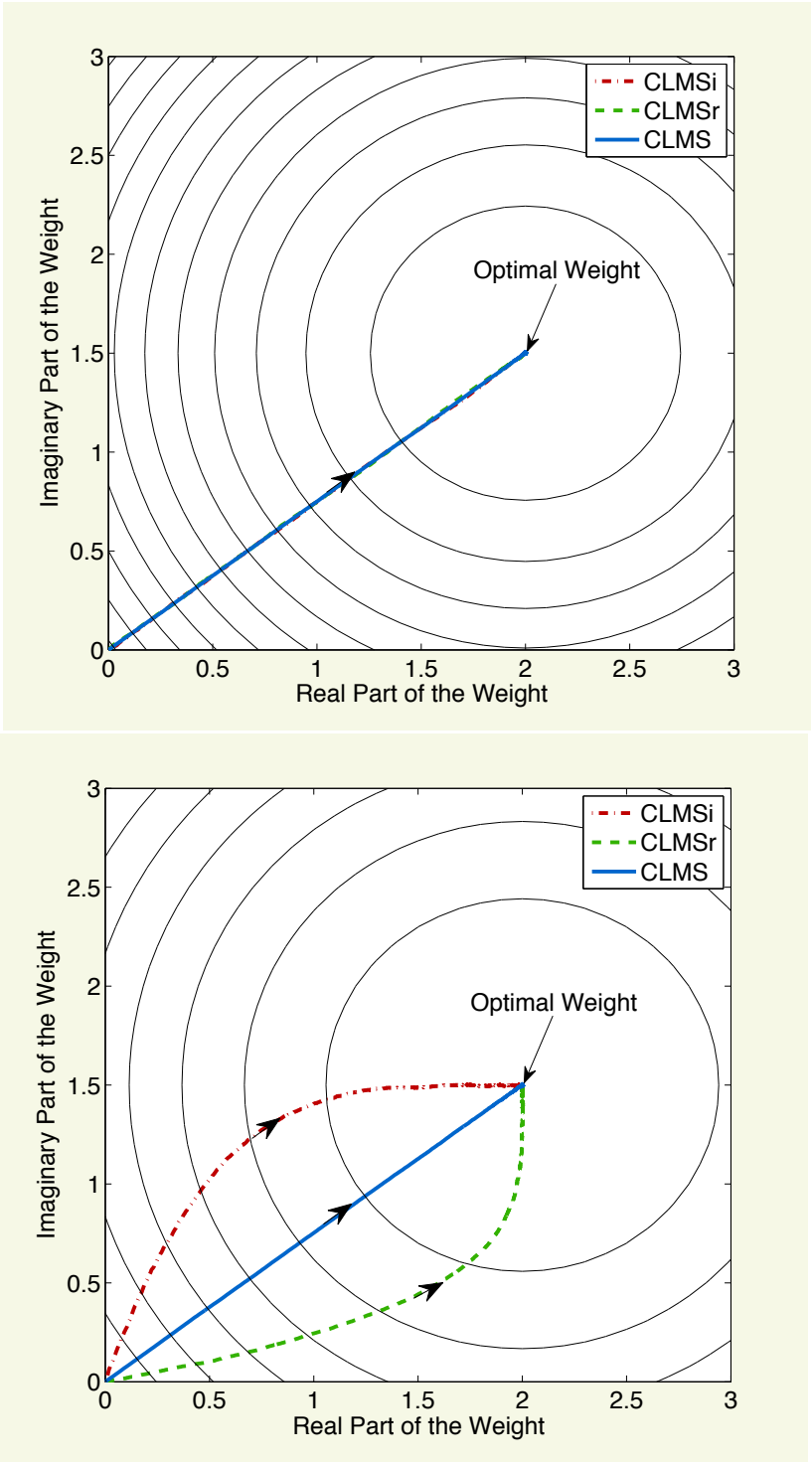

Figure 3.2: Averaged weight trajectories along the error performance contour surface for the estimation of a strictly linear MA(1) system driven by: *Top:* circular and *Bottom:* noncircular white Gaussian noise.

Proof. Recall that the CLMSr weight error recursion in (3.26) is given by

$$\tilde{\boldsymbol{w}}_{cr,k+1} = \tilde{\boldsymbol{w}}_{cr,k} - \frac{\mu}{2}\boldsymbol{x}_k\left(\boldsymbol{x}_k^{\mathsf{H}}\tilde{\boldsymbol{w}}_{cr,k} + \boldsymbol{x}_k^{\mathsf{T}}\tilde{\boldsymbol{w}}_{cr,k}^*\right) - \mu\boldsymbol{x}_k\eta_{r,k}, \tag{3.37}$$

where $\eta_{r,k} = \mathsf{Re}\{\eta_k\}$ and its conjugate is given by

$$\tilde{\boldsymbol{w}}_{cr,k+1}^* = \tilde{\boldsymbol{w}}_{cr,k}^* - \frac{\mu}{2}\boldsymbol{x}_k^*\left(\boldsymbol{x}_k^{\mathsf{T}}\tilde{\boldsymbol{w}}_{cr,k}^* + \boldsymbol{x}_k^{\mathsf{H}}\tilde{\boldsymbol{w}}_{cr,k}\right) - \mu\boldsymbol{x}_k^*\eta_{r,k}, \tag{3.38}$$

whereby upon concatenating (3.37) and (3.38) gives

$$\begin{bmatrix}\tilde{\boldsymbol{w}}_{cr,k+1}\\ \tilde{\boldsymbol{w}}_{cr,k+1}^*\end{bmatrix} = \begin{bmatrix}\tilde{\boldsymbol{w}}_{cr,k}\\ \tilde{\boldsymbol{w}}_{cr,k}^*\end{bmatrix} - \frac{\mu}{2}\begin{bmatrix}\boldsymbol{x}_k\boldsymbol{x}_k^{\mathsf{H}} & \boldsymbol{x}_k\boldsymbol{x}_k^{\mathsf{T}}\\ \boldsymbol{x}^*\boldsymbol{x}_k^{\mathsf{H}} & \boldsymbol{x}_k^*\boldsymbol{x}_k^{\mathsf{T}}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{w}}_{cr,k+1}\\ \tilde{\boldsymbol{w}}_{cr,k+1}^*\end{bmatrix} - \mu\begin{bmatrix}\boldsymbol{x}_k\\ \boldsymbol{x}_k^*\end{bmatrix}\eta_{r,k}. \tag{3.39}$$

Defining the concatenated weight error vector and augmented input vector respectively as

$$\boldsymbol{v}_k \overset{\text{def}}{=} \begin{bmatrix}\tilde{\boldsymbol{w}}_{cr,k}\\ \tilde{\boldsymbol{w}}_{cr,k}^*\end{bmatrix} \quad \text{and} \quad \bar{\boldsymbol{x}}_k \overset{\text{def}}{=} \begin{bmatrix}\boldsymbol{x}_k\\ \boldsymbol{x}_k^*\end{bmatrix}, \tag{3.40}$$

allows (3.39) to be represented compactly as

$$\boldsymbol{v}_{k+1} = \boldsymbol{v}_k - \frac{\mu}{2}\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^{\mathsf{H}}\boldsymbol{v}_k - \mu\bar{\boldsymbol{x}}_k\eta_{r,k}. \tag{3.41}$$

Notice that the recursion in (3.41) is almost identical to the ACLMS weight error vector recursion in Chapter 2. So, to obtain the recursion for the CLMSr weight error covariance matrix, $\bar{\mathbf{K}}_k \overset{\text{def}}{=} \mathbb{E}\{\boldsymbol{v}_k\boldsymbol{v}_k^{\mathsf{H}}\}$, we post-multiply both sides of (3.41) by their Hermitians, and take the statistical expectation, to give,

$$\begin{aligned}\mathbb{E}\{\boldsymbol{v}_{k+1}\boldsymbol{v}_{k+1}^{\mathsf{H}}\} = \mathbb{E}\{&\boldsymbol{v}_k\boldsymbol{v}_k^{\mathsf{H}} - \tfrac{\mu}{2}\boldsymbol{v}_k\boldsymbol{v}_k^{\mathsf{H}}\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^{\mathsf{H}} - \tfrac{\mu}{2}\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^{\mathsf{H}}\boldsymbol{v}_k\boldsymbol{v}_k^{\mathsf{H}}\}\\ &+ \mathbb{E}\left\{\tfrac{\mu^2}{4}\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^{\mathsf{H}}\boldsymbol{v}_k\boldsymbol{v}_k^{\mathsf{H}}\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^{\mathsf{H}} + \mu^2\bar{\boldsymbol{x}}_k\bar{\boldsymbol{x}}_k^{\mathsf{H}}\eta_{r,k}^2\right\}.\end{aligned} \tag{3.42}$$

Utilising Assumptions A.1 – A.3 in Theorem 3.2, (3.42) becomes

$$\begin{aligned}\bar{\mathbf{K}}_{k+1} = \bar{\mathbf{K}}_k - \tfrac{\mu}{2}\bar{\mathbf{K}}_k\bar{\mathbf{R}} - \tfrac{\mu}{2}\bar{\mathbf{R}}\bar{\mathbf{K}}_k + \tfrac{\mu^2}{2}\sigma_\eta^2\bar{\mathbf{R}}\\ + \tfrac{\mu^2}{4}(\bar{\mathbf{R}}\bar{\mathbf{K}}_k\bar{\mathbf{R}} + \bar{\mathbf{P}}\bar{\mathbf{K}}_k^T\bar{\mathbf{P}}^* + \bar{\mathbf{R}}\mathsf{Tr}[\bar{\mathbf{K}}_k\bar{\mathbf{R}}]),\end{aligned} \tag{3.43}$$

where the fact that the noise η_k is IID circular from Assumption A.2 is used in the form $\mathbb{E}\left\{\eta_{r,k}^2\right\} = (1/2)\sigma_\eta^2$. Applying the trace operator to (3.43) gives,

$$\begin{aligned}\mathsf{Tr}[\bar{\mathbf{K}}_{k+1}] = \quad &\mathsf{Tr}[\bar{\mathbf{K}}_k - \tfrac{\mu}{2}\bar{\mathbf{K}}_k\bar{\mathbf{R}} - \tfrac{\mu}{2}\bar{\mathbf{R}}\bar{\mathbf{K}}_k + \tfrac{\mu^2}{2}\sigma_\eta^2\bar{\mathbf{R}}]\\ &+\mathsf{Tr}\left[\tfrac{\mu^2}{4}(\bar{\mathbf{R}}\bar{\mathbf{K}}_k\bar{\mathbf{R}} + \bar{\mathbf{P}}\bar{\mathbf{K}}_k^T\bar{\mathbf{P}}^* + \bar{\mathbf{R}}\mathsf{Tr}[\bar{\mathbf{K}}_k\bar{\mathbf{R}}])\right]\end{aligned}, \tag{3.44}$$

and upon using the identity $\mathsf{Tr}[\mathbf{M}_1 + \mathbf{M}_2] = \mathsf{Tr}[\mathbf{M}_1] + \mathsf{Tr}[\mathbf{M}_2]$, for any square matrices $\mathbf{M}_1$ and

$\mathbf{M}_2$, we have

$$\begin{aligned}\mathsf{Tr}[\bar{\mathbf{K}}_{k+1}] &= \mathsf{Tr}[(\mathbf{I}-\mu\bar{\mathbf{R}}+\tfrac{\mu^2}{4}\bar{\mathbf{R}}^2+\tfrac{\mu^2}{4}\bar{\mathbf{P}}\bar{\mathbf{P}}^{\mathsf{H}}+\tfrac{\mu^2}{4}\mathsf{Tr}[\bar{\mathbf{R}}]\bar{\mathbf{R}})\bar{\mathbf{K}}_k]+\tfrac{\mu^2}{2}\sigma_\eta^2\mathsf{Tr}[\bar{\mathbf{R}}] \\ &= \mathsf{Tr}[\mathbf{B}\bar{\mathbf{K}}_k]+\tfrac{\mu^2}{2}\sigma_\eta^2\mathsf{Tr}[\bar{\mathbf{R}}] \end{aligned} \tag{3.45}$$

where $\mathbf{B} = \big((\mathbf{I}-\frac{\mu}{2}\bar{\mathbf{R}})^2+\frac{\mu^2}{4}(\bar{\mathbf{P}}\bar{\mathbf{P}}^{\mathsf{H}}+\mathsf{Tr}[\bar{\mathbf{R}}]\bar{\mathbf{R}})\big)$ is a Hermitian positive semi-definite matrix. Using Lemma 2.2 in Chapter 2, the recursion in (3.45) can be bounded as

$$\mathsf{Tr}[\bar{\mathbf{K}}_{k+1}] \leq \varrho(\mathbf{B})\mathsf{Tr}[\bar{\mathbf{K}}_k]+\frac{\mu^2}{2}\sigma_\eta^2\mathsf{Tr}[\bar{\mathbf{R}}]. \tag{3.46}$$

For the condition $\vartheta \overset{\text{def}}{=} \varrho(\mathbf{B}) < 1$ to be satisfied, a conservative bound on the step-size μ, can be found with,

$$\varrho((\mathbf{I}-\frac{\mu}{2}\bar{\mathbf{R}})^2)+\frac{\mu^2}{4}\varrho(\bar{\mathbf{P}}\bar{\mathbf{P}}^{\mathsf{H}})+\frac{\mu^2}{4}\mathsf{Tr}[\bar{\mathbf{R}}]\varrho(\bar{\mathbf{R}}) < 1, \tag{3.47}$$

whereby upon defining $p_{\max}^2 \overset{\text{def}}{=} \varrho(\bar{\mathbf{P}}\bar{\mathbf{P}}^{\mathsf{H}})$, and noticing that $\lambda_{\max} = \varrho(\bar{\mathbf{R}})$ yields

$$1-\mu\lambda_{\min}+\frac{\mu^2}{4}\lambda_{\max}^2+\frac{\mu^2}{4}p_{\max}^2+\frac{\mu^2}{4}\mathsf{Tr}[\bar{\mathbf{R}}]\lambda_{\max} < 1,$$

where $\lambda_{\min}$ refers to the minimum eigenvalue of $\bar{\mathbf{R}}$. The step-size, μ, which guarantees $\vartheta < 1$ is therefore given by

$$0 < \mu < \frac{4\lambda_{\min}}{\lambda_{\max}^2+p_{\max}^2+\mathsf{Tr}[\bar{\mathbf{R}}]\lambda_{\max}}. \tag{3.48}$$

Applying the inequality in (3.46) successively for each k and using the formula for the sum of a geometric sequence, $1+\vartheta+\cdots+\vartheta^{k-1} = (1-\vartheta^k)/(1-\vartheta)$, we obtain

$$\mathbb{E}\left\{\|\boldsymbol{v}_k\|^2\right\} \leq \vartheta^k\mathbb{E}\left\{\|\boldsymbol{v}_0\|^2\right\}+\left(\frac{1-\vartheta^k}{1-\vartheta}\right)\frac{\mu^2}{2}\sigma_\eta^2\mathsf{Tr}[\bar{\mathbf{R}}]. \tag{3.49}$$

From the definition of $\boldsymbol{v}_k$ in (3.40), $\mathbb{E}\left\{\|\boldsymbol{v}_k\|^2\right\} = 2\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_{cr,k}\|^2\right\}$. So, dividing (3.49) by two gives

$$\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_{cr,k}\|^2\right\} \leq \vartheta^k\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_{cr,0}\|^2\right\}+\left(\frac{1-\vartheta^k}{1-\vartheta}\right)\frac{\mu^2}{4}\sigma_\eta^2\mathsf{Tr}[\bar{\mathbf{R}}]. \tag{3.50}$$

To complete the proof, we observe (3.50) as $k \to \infty$, which yields

$$\limsup_{k\to\infty}\ \mathbb{E}\left\{\|\|\tilde{\boldsymbol{w}}_{cr,k}\|^2\right\} \leq \frac{1}{1-\vartheta}\frac{\mu^2}{4}\sigma_\eta^2\mathsf{Tr}[\bar{\mathbf{R}}]. \tag{3.51}$$

The proof of the CLMSi mean square stability follows the same arguments as the CLMSr mean square stability proof where the only difference is that the recursions of the CLMSi error vector involves an augmented input vector in the form $\bar{\boldsymbol{x}}_k' = [\boldsymbol{x}_k^{\mathsf{T}}, -\boldsymbol{x}_k^{\mathsf{H}}]^{\mathsf{T}}$. However, as observed in Remark 3.5, the spectral radius of the covariance matrices of $\bar{\boldsymbol{x}}_k'$ are identical to that of $\bar{\boldsymbol{x}}_k$. □

3.6 Simulations

To verify the analyses, all the filters considered were evaluated in the system identification setting with the step-size $\mu = 0.02$. The mean square error (MSE) of the algorithms was calculated at each time instant, k, by averaging the error power from 100 independent trials to give

$$\text{MSE}_k = \frac{1}{100}\sum_{\ell=1}^{100}\left|y_k^{(\ell)} - \hat{y}_k^{(\ell)}\right|^2, \tag{3.52}$$

where $y_k^{(\ell)}$ was the desired signal and $\hat{y}_k^{(\ell)}$ was the estimate given by the algorithms considered at trial (realisation) ℓ. The performances of CLMS, CLMSr and CLMSi were assessed for identifying a strictly linear MA(4) model described by

$$y_k = b_0 x_k + b_1 x_{k-1} + b_2 x_{k-2} + b_3 x_{k-3} + \eta_k, \tag{3.53}$$

for which the coefficients were

$$b_0 = 6 - 6j \qquad b_1 = 0.5 + j \qquad b_2 = -2 + j \qquad b_3 = 2 + 3j, \tag{3.54}$$

and the regressor x_k was a zero-mean unit-variance Gaussian signal which is IID over time k and η_k was a zero-mean IID white Gaussian noise process with variance σ_η^2.

Figure 3.3 shows that CLMSr and CLMSi achieved the same steady state mean square error as the CLMS while requiring only half the operations of the CLMS. This is consistent with Remark 3.4 and implies that the more efficient CLMSr or CLMSi filters can substitute the standard CLMS in strictly linear estimation.

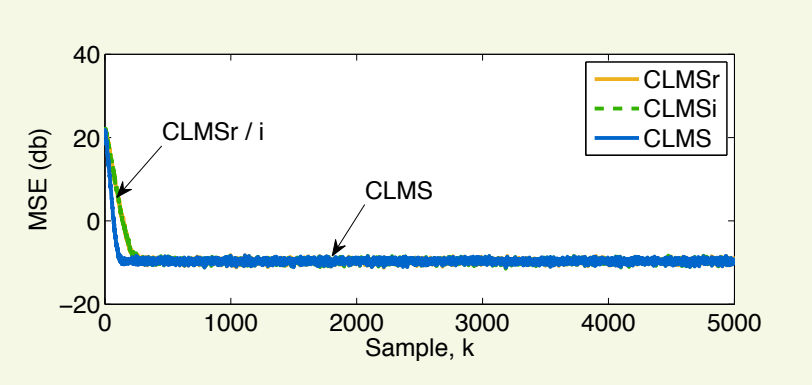

Figure 3.3: Upon convergence, CLMSr and CLMSi have the same steady state mean square error as the CLMS.

Figure 3.4 shows the performances of the DC-CLMS, CLMS and ACLMS when estimating a widely linear system, in terms of the mean square error (MSE). The widely linear MA(4) system

used to generate the signal is given by

$$y_k = b_0 x_k + b_1 x_{k-1} + b_2 x_{k-2} + b_3 x_{k-3} + \bar{b}_0 x_k^* + \bar{b}_1 x_{k-1}^* + \bar{b}_2 x_{k-2}^* + \bar{b}_3 x_{k-3}^* + \eta_k \tag{3.55}$$

where the same system coefficients $\{b_i \in i = 0, \ldots, 3\}$ described in (3.54) and $\bar{b}_0 = 0.2 - 0.2j$, $\bar{b}_1 = 0.1 + 0.1j$, $\bar{b}_2 = 2$, $\bar{b}_3 = 0.4j$, were used. The statistics of the input data, x_k, and noise, η_k, were identical to that of the previous simulation.

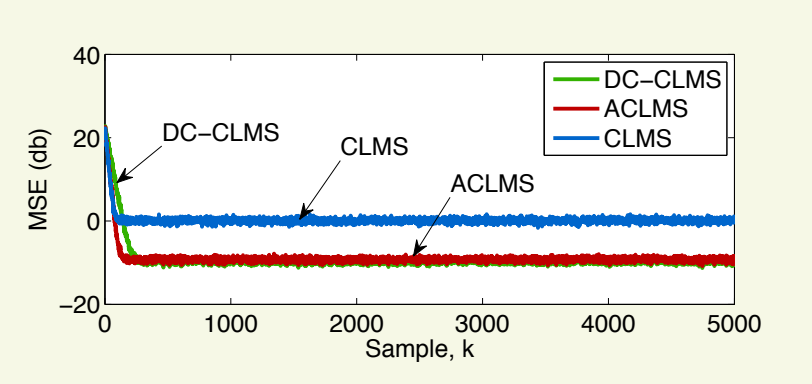

Figure 3.4: Evolution of the mean square error of the DC-CLMS, ACLMS and CLMS. The DC-CLMS and ACLMS have the same mean square error performance when modelling widely linear systems.

Conforming with the analysis, the CLMS exhibited a bias due to the inherent under-modelling (see [63]) while the DC-CLMS and ACLMS were able to model the underlying widely linear system correctly. It is important to re-emphasize that the DC-CLMS was able to achieve the same mean square performance as the ACLMS by using the half the number of operations required by the ACLMS.

The computational complexity of the algorithms considered, represented by the number of multiplications per iteration, is illustrated in Figure 3.5. The number of multiplications of the CLMSr and CLMSi are half of that of the CLMS making the CLMSr or CLMSi more efficient than the CLMS for estimating a strictly linear system. Similarly, the DC-CLMS requires approximately half the number of operations of the ACLMS, and is more efficient than the ACLMS at modelling widely linear systems.

3.7 Chapter Summary

We have introduced an alternative formulation for widely linear estimation and have developed a corresponding adaptive filter referred to as the DC-CLMS. This was accomplished by splitting the MSE cost function in (3.2) into the contributions from estimating the real and imaginary parts of the signal. By optimising individually for those parts, the CDC estimator obtains two separate strictly linear weight vectors, (3.11) and (3.13), which possess the degrees of freedom necessary for widely linear estimation. Furthermore, using the stochastic gradient

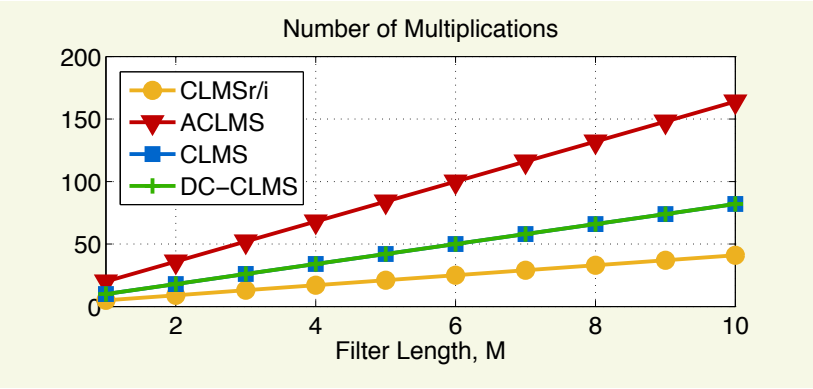

Figure 3.5: Number of multiplications as a function of the filter length.

framework introduced in Chapter 2, an adaptive version of the CDC estimator, referred to as the DC-CLMS, is proposed in Algorithm 1. The DC-CLMS has been shown to be identical to the ACLMS in (3.14), while only requiring approximately half the mathematical operations, as shown in Table 3.1.

In this chapter, we have examined adaptive filters without the context of an application. In the next chapter (Chapter 4), we shall revisit the notion of circularity of complex-random variables and use an adaptive algorithm to estimate the non-circularity in a novel real-time manner. The algorithm used is a scalar version of the CLMS which was introduced in Chapter 2.

Chapter 4

Adaptive Tracking of Complex Circularity

> The most exciting phrase to hear in science, the one that heralds new discoveries, is not "Eureka!" but "That's funny...".
>
> *Isaac Asimov*

Chapter Overview

This chapter re-examines the notion of complex noncircularity, introduced in Chapter 2, to propose a new method to track the degree of noncircularity of a complex-valued signal. This is accomplished by showing that the linear minimum mean square error (LMMSE) estimator of the complex conjugate of a signal from the original signal itself represents its circularity quotient. We exploit this relationship to propose adaptive algorithms for the tracking of circularity and subsequently extend the idea to the tracking of rectilinear signals in wireless communications systems.

4.1 Introduction

Recall from Chapter 2, that the *circularity quotient* of a zero-mean random variable (r.v.) $s_k \in \mathbb{C}$, is given by [45]

$$\rho \overset{\text{def}}{=} \frac{p}{r} = \frac{\mathbb{E}\left\{s_k^2\right\}}{\mathbb{E}\left\{|s_k|^2\right\}}, \tag{4.1}$$

where $r = \mathbb{E}\left\{|s_k|^2\right\}$ and $p = \mathbb{E}\left\{s_k^2\right\}$ are respectively the covariance and pseudocovariance of the signal s_k. The *circularity coefficient* of the signal s_k is given by the magnitude of the circularity quotient i.e. $|\rho| \overset{\text{def}}{=} \frac{|p|}{r} \in \mathbb{R}$. A signal is considered to be rectilinear if its circularity quotient can be expressed as $\rho = e^{j\phi}$, with an arbitrary phase shift ϕ. Therefore, rectilinear signals have circularity coefficients $|\rho| = 1$.

Early methods to test the impropriety of a signal were based on hypothesis tests that use block estimates of the covariance and the pseudocovariance. The authors in [64] developed a hypothesis test for the impropriety of Gaussian signals using a generalized likelihood ratio (GLR) test which was further studied in [65, 66]. This method was extended for signals from a general class of complex elliptically symmetric (CES) distributions in [67]. To deal with measurement errors, a robust circularity coefficient estimator was developed in [68] based on solving M-estimation equations with a novel weighting scheme.

A real time algorithm for tracking the impropriety (see Def. 2.1, Ch. 2) of a signal was introduced in [69] based on a convex combination of a strictly and widely linear estimators. The impropriety of a signal is then tracked by observing the evolution of the adaptive convex mixing parameter which indicates which sub-filter (strictly or widely linear) is a better match to the nature of the data. The limitations of current approaches in identifying the second order circularity are:

(a) Hypothesis tests are limited since they are only able to reveal whether the signal is proper or improper, however, they cannot assess the degree of impropriety;

(b) The value of the convex mixing parameter in [69] depends also on the filter settings and its relationship to the circularity quotient is yet to be established rigorously.

We here provide a solution for real-time circularity tracking by firstly establishing that the circularity quotient is effectively the optimal coefficient of an LMMSE estimator that estimates the complex conjugate of a signal from the original signal itself. The proposed algorithm then utilises an adaptive filter weight to track the circularity quotient of a signal in real time and overcomes the issues mentioned in (a) – (b) above.

4.2 Relationship Between A Complex Variable and Its Conjugate

Deterministic case. Consider the problem of finding a linear mapping that relates a deterministic variable, $\tilde{s} = |\tilde{s}|e^{j\angle\tilde{s}} \in \mathbb{C}$, with magnitude $|\tilde{s}|$ and phase $\angle\tilde{s}$, and its complex conjugate, $\tilde{s}^* = |\tilde{s}|e^{-\angle\tilde{s}}$. This mapping has the form

$$\tilde{s}^* = w^*\tilde{s}. \tag{4.2}$$

The solution for the coefficient w is thus

$$w = \frac{\tilde{s}}{\tilde{s}^*} = \frac{\tilde{s}^2}{|\tilde{s}|^2} = e^{j2\angle\tilde{s}}. \tag{4.3}$$

Physically, the coefficient w in (4.2) rotates the complex variable $\tilde{s}$ by an angle of $-2\angle\tilde{s}$.

Stochastic case. Now, consider the problem of using a zero-mean r.v. $s_k \in \mathbb{C}$ to estimate its complex conjugate, that is

$$\hat{s}_k^* = w^* s_k, \tag{4.4}$$

where $\hat{s}_k^*$ denotes the estimate of the complex conjugate of s_k and w is the coefficient that relates the two variables. Unlike (4.3), every realisation of the r.v. in the data stream has a different phase and we require a stochastic solution.

Our aim is therefore, to find an estimate of w that minimizes the estimation error, $\varepsilon_k = \hat{s}_k^* - s_k^*$, across all realizations of s. To this end, we propose to employ minimum mean square

error (MMSE) estimation whereby the optimal value of w is found by minimising the cost function

$$J_{\text{MSE}}(w) = \mathbb{E}\{|\varepsilon_k|\} = \mathbb{E}\left\{|s_k^* - w^* s_k|^2\right\}. \tag{4.5}$$

The optimal value of w, denoted by w_{opt}, that minimizes the cost function in (4.5) is then given by the Wiener solution

$$w_{\text{opt}} = \frac{\mathbb{E}\left\{s_k^2\right\}}{\mathbb{E}\left\{|s_k|^2\right\}} = \rho. \tag{4.6}$$

Remark 4.1. *From (4.6), the circularity quotient can be interpreted as the LMMSE solution for estimating the complex conjugate of a random variable from the original random variable itself.*

Finding the Wiener solution requires the knowledge of the true statistics of the data (in our case, r and p) which is typically not available. As block based estimators for the Wiener solution are inadequate for non-stationary signals, we next develop an adaptive estimator for this purpose.

4.3 Proposed Algorithm

Interpreting the circularity quotient as the optimal Wiener solution for estimating the complex conjugate of a r.v. from the original r.v. enables us to configure an adaptive filter, in the form illustrated in Figure 4.1, together with an adaptive algorithm like the complex least mean square (CLMS), to track the circularity quotient in real time. The proposed circularity tracking algorithm is

$$\hat{s}_k^* = \hat{\rho}_k^* s_k \tag{4.7a}$$

$$\varepsilon_k = s_k^* - \hat{s}_k^* \tag{4.7b}$$

$$\hat{\rho}_{k+1} = \hat{\rho}_k + \mu \varepsilon_k^* s_k, \tag{4.7c}$$

where the parameter μ in (4.7c) is the step-size which governs the convergence of the algorithm [52], while the degree of second order circularity is represented by the weight $\hat{\rho}_k$.

As CLMS only uses instantaneous estimates of the data statistics, the weights can only asymptotically approach their optimal value, and we need to analyse the contributions of the bias and variance of parameter estimates to the total mean square error.

> **Contribution: Adaptive Circularity Tracker**
> Based on uncovering the relationship between the Wiener solution for estimating a complex-conjugate of a signal from the signal itself in (4.6), a structurally-simple adaptive algorithm is proposed in (4.7a) – (4.7c).

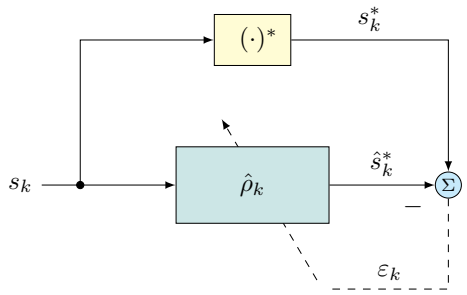

Figure 4.1: Block diagram of the circularity tracker.

4.4 Statistical Analysis

Mean convergence

To study the mean (bias) behaviour of the circularity tracker, consider again the circularity update equation in (4.7c),

$$\hat{\rho}_{k+1} = \hat{\rho}_k + \mu(s_k^2 - \hat{\rho}_k |s_k|^2), \tag{4.8}$$

where we have substituted the error $\varepsilon_k^* = s_k - \hat{\rho}_k s_k^*$.

Theorem 4.1 (Mean Convergence). *Consider the circularity tracker in (4.8). Suppose that the following assumption holds:*

A.1) The sequence, $\{s_k\}_{k\in\mathbb{N}}$, has zero-mean and is independent and identically distributed (IID) across time k, with covariance and pseudocovariance respectively given by $r = \mathbb{E}\{|s_k|^2\}$, $p = \mathbb{E}\{s_k^2\}$.

Then, the circularity quotient estimate from (4.8) asymptotically converges in the mean, that is

$$\lim_{k\to\infty} \mathbb{E}\left\{\frac{p}{r} - \hat{\rho}_k\right\} = 0 \tag{4.9a}$$

if the step-size μ satisfies the following condition

$$0 < \mu < 2/r. \tag{4.9b}$$

Proof.
Defining the circularity quotient error, $\tilde{\rho}_k \overset{\text{def}}{=} \frac{p}{r} - \hat{\rho}_k$, and subtracting $\frac{p}{r}$ from (4.8) gives

$$\tilde{\rho}_{k+1} = \tilde{\rho}_k - \mu\tilde{\rho}_k |s_k|^2 - \mu s_k^2 + \mu\frac{p}{c}|s_k|^2. \tag{4.10}$$

Taking the statistical expectation of (4.10) and using the IID assumption of the signal s_k from

Assumption A.1 in Theorem 4.1 and $r = \mathbb{E}\left\{|s_k|^2\right\}$, $p = \mathbb{E}\left\{s_k^2\right\}$, yields,

$$\mathbb{E}\left\{\tilde{\rho}_{k+1}\right\} = (1-\mu r)\mathbb{E}\left\{\tilde{\rho}_k\right\}. \tag{4.11}$$

The recursion in (4.11) converges in the mean if $|1-\mu r| < 1$, which leads to the bound

$$0 < \mu < \frac{2}{r}, \tag{4.12}$$

where r is the covariance of the input signal s_k. Thus, the algorithm is asymptotically unbiased. □

Remark 4.2. *The mean behaviour of the proposed circularity tracker is not affected by the degree of circularity of the input signal.*

To analyse the mean square behaviour of the circularity tracker, we first define the covariance of the weight error of the proposed circularity tracker as

$$\mathcal{K}_{k+1} \overset{\text{def}}{=} \mathbb{E}\left\{|\tilde{\rho}_{k+1}|^2\right\}. \tag{4.13}$$

Theorem 4.2 (Mean Square Stability). *Consider the circularity tracker in (4.8). In addition to the assumptions in Theorem 4.1, suppose that the following assumption holds:*

A.2) The sequence $\{s_k\}_{k\in\mathbb{N}}$ is drawn from a Gaussian distribution.

Then, the circularity quotient estimate from (4.8) is mean square stable with the steady-state value given by

$$\lim_{k\to\infty} \mathcal{K}_k = \mu r \frac{\left(1-|\rho|^2\right)\left(2-|\rho|^2\right)}{2-\mu r\left(2+|\rho|^2\right)}, \tag{4.14a}$$

if the step-size μ satisfies the following condition

$$0 < \mu < \frac{2}{r(2+\frac{|p|^2}{r^2})} = \frac{2}{r(2+|\rho|^2)}. \tag{4.14b}$$

Proof.
The covariance of the weight error of the circularity tracker is given by

$$\begin{aligned}\mathcal{K}_{k+1} &= \mathbb{E}\left\{|\tilde{\rho}_{k+1}|^2\right\} \\ &= \mathbb{E}\left\{\left|\tilde{\rho}_k - \mu\tilde{\rho}_k|s_k|^2 - \mu s_k^2 + \mu\frac{p}{r}|s_k|^2\right|^2\right\}.\end{aligned} \tag{4.15}$$

Assuming Gaussian data from Assumption A.2 in Theorem 4.2, the fourth order moments in (4.15) can be decomposed into a combination of second order moments as $\mathbb{E}\left\{|s_k|^4\right\} = 2r^2+|p|^2$, $\mathbb{E}\left\{s_k^2|s_k|^2\right\} = 3pr$ and $\mathbb{E}\left\{s_k^{*2}|s_k|^2\right\} = 3p^*r$. After some algebraic manipulations, the weight error covariance becomes

$$\mathcal{K}_{k+1} = \underbrace{\left[(2r^2+|p|^2)\mu^2 - 2r\mu + 1\right]}_{\vartheta}\mathcal{K}_k + \underbrace{\left[2\mu^2 r\left(1-\frac{|p|^2}{r^2}\right)\right]}_{\beta}\mathsf{Re}\left\{p^*\bar{w}_k\right\} + \underbrace{\mu^2 r^2\left[\frac{|p|^4}{r^4} - 3\frac{|p|^2}{r^2} + 2\right]}_{\alpha}, \tag{4.16}$$

where the term $\bar{w}_k \overset{\text{def}}{=} \mathbb{E}\left\{\tilde{\rho}_k\right\}$ refers to the mean of the weight error at time instant k.

The recursion for the weight error covariance, $\mathcal{K}_k$, involves an additional time varying term, $\mathsf{Re}\left\{p^*\bar{w}_k\right\}$, and the evolution of which can be analysed by multiplying the mean weight evolution in (4.11) by p^* to give

$$\mathsf{Re}\left\{p^*\bar{w}_{k+1}\right\} = (1-\mu r)\mathsf{Re}\left\{p^*\bar{w}_k\right\}. \tag{4.17}$$

This allows us to formulate the following vector recursion involving (4.16) and (4.17)

$$\begin{bmatrix}\mathcal{K}_{k+1}\\ \mathsf{Re}\left\{p^*\bar{w}_{k+1}\right\}\end{bmatrix} = \underbrace{\begin{bmatrix}\vartheta & \beta\\ 0 & \gamma\end{bmatrix}}_{\mathbf{B}}\begin{bmatrix}\mathcal{K}_k\\ \mathsf{Re}\left\{p^*\bar{w}_k\right\}\end{bmatrix} + \begin{bmatrix}\alpha\\ 0\end{bmatrix}, \tag{4.18}$$

where $\gamma = (1-\mu r)$. For the mapping in (4.18) to be contractive, all the eigenvalues of $\mathbf{B}$ should lie within the unit circle. Since $\mathbf{B}$ is an upper triangular matrix, its eigenvalues are ϑ and γ. The convergence condition $|\gamma| = |1-\mu r| < 1$ was already addressed in the analysis of the convergence in the mean in (4.12). The second stability condition, $\vartheta < 1$, is satisfied when

$$(2r^2+|p|^2)\mu^2 - 2r\mu < 0, \tag{4.19}$$

which leads to the bound on the learning rate μ

$$0 < \mu < \frac{2}{r(2+\frac{|p|^2}{r^2})} = \frac{2}{r(2+|\rho|^2)}. \tag{4.20}$$

Remark 4.3. *Unlike the condition for convergence in the mean, the mean square stability depends on the degree of circularity of the signal. For proper data, $\rho = 0$, and the condition in (4.20) becomes $\mu \leq 1/r$. Since $|p| \leq r$ [45], for highly non-circular data we have $\mu \leq \frac{2}{3r}$.*

Steady-state mean square performance.

Assuming that the step-size μ satisfies the condition in (4.20), the covariance of the weight error, $\mathcal{K}_k = \mathbb{E}\left\{|\tilde{\rho}_k|^2\right\}$, in (4.16) converges to a steady-state value of

$$\lim_{k\to\infty}\mathcal{K}_k = \frac{\alpha}{1-\vartheta} = \mu r\frac{\left(1-\frac{|p|^2}{r^2}\right)\left(2-\frac{|p|^2}{r^2}\right)}{2-\mu r\left(2+\frac{|p|^2}{r^2}\right)} = \mu r\frac{\left(1-|\rho|^2\right)\left(2-|\rho|^2\right)}{2-\mu r\left(2+|\rho|^2\right)}. \tag{4.21}$$

Note that the steady-state weight error covariance in (4.21) is equivalent to the mean square deviation (MSD) metric introduced in Chapter 2. Thus, the estimate of the circularity quotient has a steady-state error power that is approximately proportional to μ for small step-sizes. The weight error covariance depends on the degree of impropriety of the signal and is lower for signals that are less proper. Moreover, we can see that for small μ

$$\lim_{k\to\infty} \mathcal{K}_k \leq \mu r + \mathcal{O}(\mu^2), \tag{4.22}$$

where equality holds when s_k is proper. □

Remark 4.4. *The weight error covariance, which is identical to the mean square deviation (MSD), depends on the degree of impropriety of the signal and is lower for signals that are less proper. Moreover, the steady state covariance for the circularity estimate is null for rectilinear signals which have a circularity coefficient of* $|\rho| = 1$.

4.5 Simulations

For all the simulations, the CLMS was configured with a filter length $M = 1$, step-size $\mu = 0.01$, input data s_k and desired signal s_k^*. The filter weight was initialized as $\hat{\rho}_0 = 0.5 + 0.5j$.

In the first set of simulations, we demonstrate the circularity tracking ability of the proposed algorithm on a synthetically generated signal that was constructed by concatenating three segments of zero-mean white Gaussian signals, $s_{i,k}$, with different pseudocovariance values given by

$$p_i = \mathbb{E}\left\{z_{i,k}^2\right\}, \qquad i = \{1, 2, 3\} \tag{4.23}$$

These segments had the same covariance, $r = \mathbb{E}\left\{s_{i,k}^2\right\} = 1$, but different pseudocovariances, p_i, and thus different degrees of circularity, $|\rho_i|$ (see Table 4.1).

Sample, k	$1-1000$	$1001-2000$	$2001-3000$
p_i	$0.8j$	$0.6+0.4j$	0

Table 4.1: *Pseudocovariances,* p_i*, of the Gaussian signals.*

Figure 4.2 shows the evolution of the weight estimates within the proposed algorithm. Observe that the algorithm was able to converge to the accurate circularity quotients within 300 samples.

Next, we evaluated our algorithm on complex wind data which was sampled at 50 Hz and measured as a bivariate signal of wind speeds in the East-West and North-South directions, denoted respectively by s_{E} and s_{N}. The complex wind representation is therefore given by $s = s_{\mathrm{E}} + js_{\mathrm{N}}$ [70].

We considered three wind regimes of different dynamics: the "low", "medium" and "high" regimes. Figure 4.3 shows the circularity diagram of empirical distributions of the real and imaginary parts of the wind signal for these regimes, where 98.8% (2.5 standard deviations) of the samples are contained within the ellipses. The circularity diagram suggests that the higher

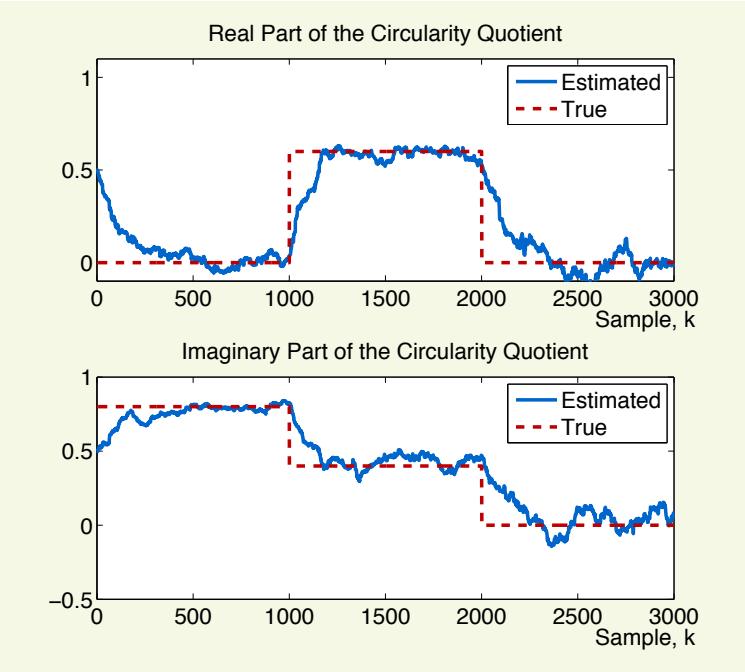

Figure 4.2: The real and imaginary parts of the evolution of the CLMS weights when tracking circularity.

the wind speeds, the higher the degree of non-circularity. This is physically meaningful, as high winds are highly directional, with dominant power in a narrow direction, resulting in a greater degree of non-circularity.

Figure 4.4 shows the estimates of the circularity coefficient, $|\rho|$, under low, medium and high wind regimes. The circularity tracker verifies the observation that the higher the wind speed, the greater its degree of non-circularity.

Steady-State MSD of the Circularity Tracker

Next, we tested the steady-state MSD of the circularity tracker for various levels of non-circularity. To this end, we generated a 1000-sample long white Gaussian noise with unit covariance $r = 1$ and varying degrees of pseudocovariance p (hence circularity coefficients, $|\rho|$) and applied the circularity tracker to estimate the circularity coefficient. The MSD was computed over 5000 independent realisations using the formula

$$\text{MSD} = \frac{1}{5000} \sum_{i=1}^{5000} ||\rho| - |\hat{\rho}_\infty(i)||^2 , \tag{4.24}$$

where $|\rho|$ was the true circularity coefficient and $|\hat{\rho}_\infty(i)|$ the steady-state estimate of the circularity coefficient for simulation i. Figure 4.5 compares the theoretical steady-state misadjustment given in (4.21) to the empirical estimate from the Monte-Carlo simulations given in (4.24). Conforming with the analysis, the steady-state misadjustment of the circularity coefficient estimate decreases with increasing levels of non-circularity and reaches zero for rectilinear signals

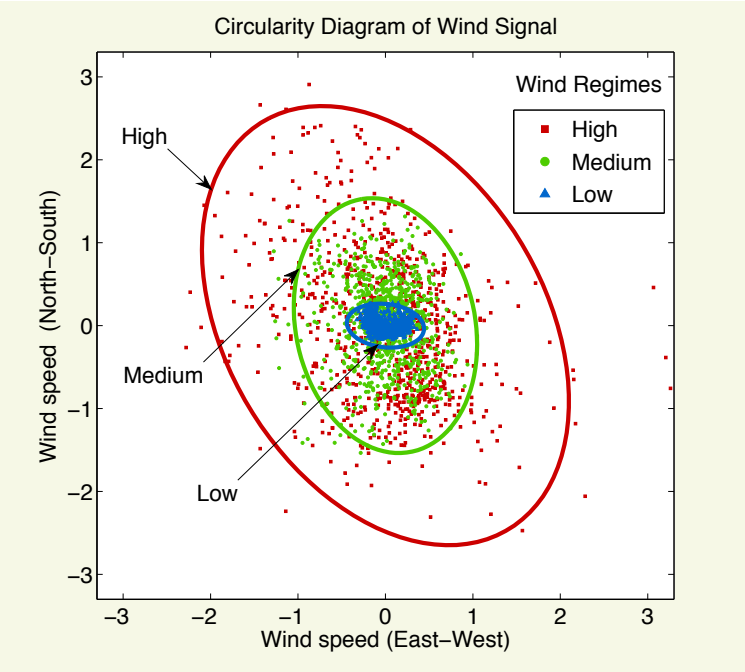

Figure 4.3: The circularity diagram of wind speeds in the low, medium and high dynamic regimes.

($|\rho| = 1$).

4.6 Tracking the Rectilinearity of Communication Signals

Adaptive beamforming is a key tool employed in wireless communication networks as it is used to extract signals of interest (SOI) by rejecting the interference and noise for signals and channels with time varying statistics. Consider a uniform linear array (ULA) of N sensors which receives a signal of interest, m_k, at time instant k according to the following relationship

$$\boldsymbol{x}_k = m_k \boldsymbol{s} + \boldsymbol{n}_k, \tag{4.25}$$

where $\boldsymbol{x}_k \in \mathbb{C}^{N\times 1}$ is the measurement vector with the entries from each ULA sensor, $\boldsymbol{s} \in \mathbb{C}^{N\times 1}$ is the steering or the channel vector of the SOI and $\boldsymbol{n}_k \in \mathbb{C}^{N\times 1}$ is the noise vector.

Traditional complex-valued adaptive beamformers were derived as generic extensions of their real-valued counterparts where the signal m_k is recovered by a strictly linear adaptive filter, $\hat{y}_k = \boldsymbol{h}_k^{\mathsf{H}} \boldsymbol{x}_k$ with output $\hat{y}_k$ which corresponds to the recovered signal and the input $\boldsymbol{x}_k$ is from the antenna array measurements given in (4.25). The beamforming filter weight vector $\boldsymbol{h}_k$ is adaptively estimated by minimising the mean square error (MSE) cost function given by

$$J_{\text{MSE}} = \mathbb{E}\left\{|m_k - \hat{y}_k|^2\right\} = \mathbb{E}\left\{\left|m_k - \boldsymbol{h}_k^{\mathsf{H}} \boldsymbol{x}_k\right|^2\right\}. \tag{4.26}$$

The strictly linear solution to the cost function in (4.26) is only optimal for a restricted class

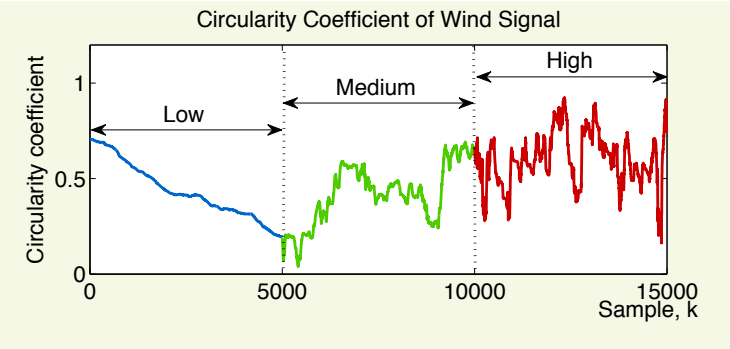

Figure 4.4: The estimate of circularity coefficient for wind signals in the low, medium and high dynamic regimes using the proposed algorithm.

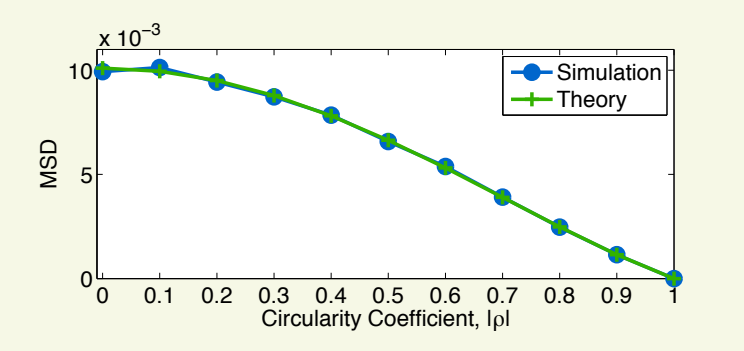

Figure 4.5: Steady-state misadjustment of the circularity tracker for varying levels of signal non-circularity.

of signals which are second order circular and a widely linear beamformer is required for a general class of non-circular signals. The widely linear beamformer is able to recover non-circular SOI using a widely linear model which utilizes both $\boldsymbol{x}_k$ and its conjugate $\boldsymbol{x}_k^*$ to give $\hat{y}_k = \boldsymbol{h}_k^{\mathsf{H}}\boldsymbol{x}_k + \boldsymbol{g}_k^{\mathsf{H}}\boldsymbol{x}_k^*$. The filter weight vectors $\boldsymbol{h}_k$ and $\boldsymbol{g}_k$ are also derived by minimising the MSE cost function given by [71]

$$J_{\text{MSE}} = \mathbb{E}\left\{\left|m_k - \boldsymbol{h}_k^{\mathsf{H}}\boldsymbol{x}_k - \boldsymbol{g}_k^{\mathsf{H}}\boldsymbol{x}_k^*\right|^2\right\}. \tag{4.27}$$

Adaptive (real-time) beamformers are derived by minimising the cost functions given in (4.26) and (4.27) either via the stochastic gradient descent to form the complex least mean square (CLMS) [52] and augmented CLMS (ACLMS) algorithms [42, 53] or by minimising sum of the error squares recursively to form the recursive least squares (RLS) [72] and widely linear RLS (WL–RLS) algorithms [73].

For an $N-$sensor antenna array, widely linear beamformers are able to process more than the conventional limit of $(N-1)$ sources when two or more of the sources are rectilinear [74]. A rectilinear source is a special case of a complex-valued source as it is effectively a single real-valued source with an arbitrary phase shift. However, since the widely linear solution for (4.27) requires twice the number of complex coefficients compared to the strictly linear solution for

(4.26), this results in an increase in computational complexity. Moreover, doubling the number of coefficients for adaptive filters results in a larger steady state excess MSE [30]. Therefore, it is important to determine the rectilinearity of the sources prior to employing either a strictly linear or a widely linear adaptive beamformer[1].

The idea of detecting the rectilinearity of sources after a blind source separation (BSS) step for wireless communication signals was first suggested in [74], where the sample covariance and pseudocovariance were used to estimate the rectilinearity of the signal. Other more sophisticated tools to estimate the non-circularity (hence, rectilinearity) of signals have been suggested in the past, see [64, 65, 75]. However, as most of these algorithms, including the approach in [74], are block-based, they are not suitable for real-time applications or for non-stationary data.

To this end, we show that the rectilinearity detection idea in [74] can be extended for online applications by a combination of an adaptive (online) BSS algorithm [76] and a real-time rectilinearity tracker based on the circularity tracker in proposed in (4.7a)–(4.7c).

> **Contribution: Rectilinear Tracking Algorithm**
> The circularity tracker proposed in this chapter is applied to develop the first online rectilinearity estimation algorithm in communication networks, see Algorithm 2. The motivation of the algorithm was the fact that that the steady-state MSD of the circularity tracker in (4.21) is null for rectilinear signals.

4.6.1 Online Blind Source Separation

Consider representing the signal in (4.25) as a mixture of M sources corrupted by background noise $\boldsymbol{n}_{b,k}$, such that

$$\boldsymbol{x}_k = \sum_{\ell=1}^{M} s_{\ell,k}\boldsymbol{h}_\ell + \boldsymbol{n}_{b,k} = \mathbf{H}\boldsymbol{s}_k + \boldsymbol{n}_{b,k}, \tag{4.28}$$

where $\boldsymbol{s}_k = [s_{1,k}, s_{2,k}, \ldots, s_{M,k}]^{\mathsf{T}}$ is a vector containing all the sources $s_{\ell,k}$, the columns of $\mathbf{H}$ are the channel vectors $\boldsymbol{h}_\ell$ of the sources $\ell = \{1, \ldots, M\}$, and $\boldsymbol{n}_{b,k}$ is assumed to be zero-mean, circular and Gaussian. At each time instant k, the separation task can be expressed as

$$\boldsymbol{y}_k = \mathbf{D}_k\boldsymbol{x}_k, \tag{4.29}$$

where $\boldsymbol{x}_k$ is the mixture of sources given in (4.28).

In this work, we employ the well known class of equivariant adaptive separation via independence (EASI) algorithms for adaptive (real-time) separation of instantaneous mixtures of blind sources [76, 77]. A serial update of the de-mixing matrix $\mathbf{D}_k$ in the EASI algorithm is

[1]Note that we have assumed that the criterion for choosing between strictly linear and widely linear processing is based solely on the number of sources that can be separated.

derived using the relative gradient of the objective function and is given by [76, 78]

$$\mathbf{D}_{k+1} = \mathbf{D}_k + \tilde{\mu}\frac{\mathbf{I} - \mathbf{g}(\boldsymbol{y}_k)\boldsymbol{y}_k^{\mathsf{H}}}{1 + \tilde{\mu}\left|\boldsymbol{y}_k^{\mathsf{H}}\mathbf{g}(\boldsymbol{y}_k)\right|}\mathbf{D}_k, \tag{4.30}$$

where $\tilde{\mu}$ is a positive step-size which governs the trade-off between convergence rate and separation accuracy.

Remark 4.5. *The initial value of the separating matrix, i.e.* $\mathbf{D}_0$*, is formed with IID entries drawn from a zero-mean Gaussian distribution [76].*

The function $\mathbf{g}(\boldsymbol{y}) = [g_1(y_1), g_2(y_2), \ldots, g_N(y_N)]^{\mathsf{T}}$ is a component-wise nonlinear function of the elements in $\boldsymbol{y} = [y_1, y_2, \ldots, y_N]^{\mathsf{T}}$. Choosing the optimal nonlinear function $\mathbf{g}(\cdot)$ is difficult in practice since it depends on the statistical distributions of the sources which are unknown *a priori*. Having said that, a simple choice of a nonlinear function for non-Gaussian sources is found to be $g_i(y_i) = y_i f_i\left(|y_i|^2\right)\ \forall i \in \{1, 2, \ldots, N\}$ where $f_i(\cdot)$ is a suitable real-valued function [79]. The simplest choice of f_i is $f_i\left(|y_i|^2\right) = |y_i|^2$ which gives [80, 81]

$$g_i(y_i) = y_i\,|y_i|^2\,. \tag{4.31}$$

4.6.2 Real-Time Detection of Rectilinear Sources

Consider the problem of estimating the rectilinearity of the sources $s_{\ell,k}$ within the noisy instantaneous mixture $\boldsymbol{x}_k$ given in (4.28) at each time instant k. As a new measurement $\boldsymbol{x}_k$ arrives, the proposed algorithm separates the sources and updates both the de-mixing matrix, $\mathbf{D}_k$, and the circularity quotient, $\hat{\rho}_k$, of the separated sources. If the circularity coefficient exceeds a certain threshold $\beta_{\hat{M}}$, the signal is classified as rectilinear. The proposed algorithm is outlined in Algorithm 2.

Algorithm 2. Real Time Tracker of Rectilinearity

Input data: Array measurements, $\boldsymbol{x}_k$
Initialise with: Circularity tracker, $\hat{\rho}_{y_i,0} = 0$ and demixing matrix, $\mathbf{D}_0$ according to Remark 4.5.
At each time instant $k > 0$:

1: **for** $\hat{M} = \{1, 2, 3, \ldots, N\}$ **do**
2: Separate the $\hat{M}$ blind sources
$$\boldsymbol{y}_k = \mathbf{D}_k\boldsymbol{x}_k$$
3: Update the de-mixing matrix $\mathbf{D}_k$ using (4.30).
$$\mathbf{D}_{k+1} = \mathbf{D}_k + \tilde{\mu}\frac{\mathbf{I}-\mathbf{g}(\boldsymbol{y}_k)\boldsymbol{y}_k^{\mathsf{H}}}{1+\tilde{\mu}\left|\boldsymbol{y}_k^{\mathsf{H}}\mathbf{g}(\boldsymbol{y}_k)\right|}\mathbf{D}_k$$
4: Update the circularity quotient $\rho_{y_i,k}$ for each element of $\boldsymbol{y}_k$ using (4.7a) – (4.7c).
$$\hat{\rho}_{y_i,k+1} = \hat{\rho}_{y_i,k} + \mu(y_{i,k}^2 - \hat{\rho}_{y_i,k}|y_{i,k}|^2)$$
5: If circularity coefficient $|\hat{\rho}_{y_i,k}| > \beta_{\hat{M}}$, then source y_i is rectilinear.
6: **end for**

Remark 4.6. *Although rectilinear sources have a circularity coefficient of unity, the separated sources would be corrupted by noise which may not be rectilinear. Therefore, even in the case of perfect separation, the separated source which is fed into the rectilinear tracker will never be perfectly rectilinear. Consequently, the detection threshold is chosen to be* $\beta_{\hat{M}} < 1$*. The optimal choice of the detection threshold* $\beta_{\hat{M}}$ *is out of the scope of this thesis and in our simulations we used a value that was close to unity, e.g.* $\beta_{\hat{M}} = 0.9$.

4.6.3 Rectilinear Tracking Case Studies

We considered a ULA of 4 omnidirectional sensors, equispaced half a wavelength apart, that received four statistically independent narrow band sources which were either binary phase shift keying (BPSK) or quadrature phase shift keying (QPSK) sources corrupted by zero-mean white Gaussian noise. The directions of arrival of the sources were $\{-45°, 8°, -13°, 30°\}$ and the SNR was chosen to be 10 dB.

The type of narrowband signals received by the antenna array changes so that for the first 500 samples, the sources were chosen to be four independent QPSK signals (circular, non-rectilinear) and for the last 500 samples, three independent BPSK signals (rectilinear) and one independent QPSK signal were chosen. The modified EASI shown in (4.30) was applied to separate sources hidden in the observed signal. The separation matrix $\mathbf{D}_k$ was initialised with random entries and the nonlinear function $\mathbf{g}(\cdot)$ given in (4.31) was used.

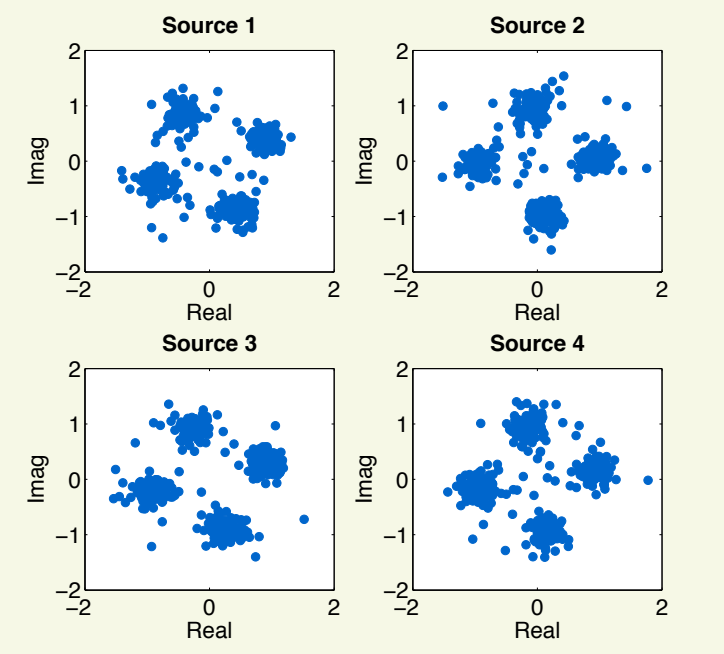

Figure 4.6: Constellations of four separated QPSK sources (circular, non-rectilinear) present in the first 500 samples of the received signal. The step-size used for the BSS process was $\tilde{\mu} = 0.1$.

For the first 100 iterations, the step-size $\tilde{\mu}$ was set to be relatively large ($\tilde{\mu} = 0.1$) for rapid convergence. Then, the value of $\tilde{\mu}$ was reduced to $\tilde{\mu} = 0.03$ for steady-state accuracy. Figure

4.6 and Figure 4.7 show the constellations of separated sources ($\hat{M} = M = 4$) within the first 500 samples and the last 500 samples respectively. The separated sources in Figure 4.6, appear noisier than the sources in Figure 4.7 since a larger step-size $\tilde{\mu}$ was used for the first 100 samples.

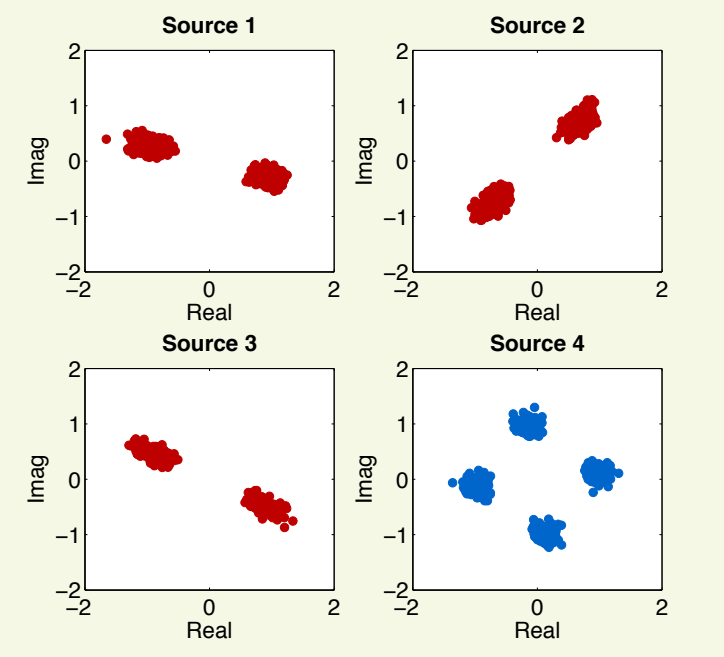

Figure 4.7: Sources 1–3 show the constellations of three separated BPSK sources (rectilinear) while Source 4 shows the constellation of one separated QPSK source (circular, non-rectilinear) present in the second 500 samples of the received signal. The step-size used for the BSS process was $\tilde{\mu} = 0.03$.

For the circularity tracker, the initial estimates were set to zero while the step-size was chosen to be relatively large, $\mu = 0.2$, for a fast convergence rate. The evolution of the circularity coefficient estimates of the sources is shown in Figure 4.8. The estimates of the circularity coefficients of the four separated QPSK sources oscillated between 0 (the true value of the circularity coefficient) and 0.5. This large fluctuation is attributed to the background noise and the relatively large step-size μ which was chosen so that the algorithm converges rapidly. Nevertheless, from the samples 500 onwards, the circularity coefficient estimates of the three separated BPSK sources converged to a value close to unity. As mentioned in Remark 4.6, this is due to the fact that the sources are corrupted by non-rectilinear noise. When the estimated number of sources was not equal to the true number ($\hat{M} = 1, 2, 3$), the circularity coefficients were all less than the threshold $\beta_{\hat{M}} = 0.9$.

4.7 Chapter Summary

A novel algorithm for the estimation of the degree of circularity of a complex-valued signal has been proposed in (4.7a) – (4.7c). This has been achieved by tracking the evolution of the CLMS weight that estimates the complex conjugate of a signal from the original signal itself.

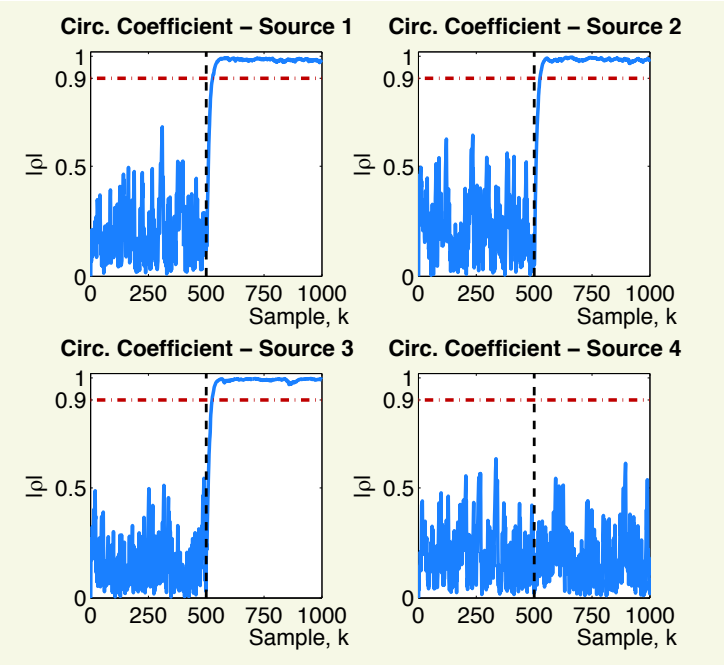

Figure 4.8: Circularity coefficient estimates of the separated sources. Detection threshold was chosen to be $\beta_{\hat{M}} = 0.9$.

The motivation for this algorithm was in the relationship between the circularity quotient and the optimal linear estimator for the complex conjugate of a signal from the original signal in (4.6). The conditions for the stability in the mean and mean square which follow the classic CLMS results have been shown respectively in Theorem 4.1 and Theorem 4.2.

The proposed circularity tracker was also extended to serve as a novel real-time detector of rectilinearity for wireless communication signals as outlined in Algorithm 2. The rectilineartiy detector combines an online BSS algorithm to separate the signal received by an antenna array and the proposed circularity tracker to detect the rectilinearity of the separated sources. The steady-state mean square deviation of the circularity tracker in (4.21) was exploited to tune the algorithm for rectilinear signals so as to achieve rapid convergence. The proposed circularity tracker has been verified on complex-valued wind data, synthetically generated Gaussian data, and synthetically generated BPSK and QPSK sources.

Conclusion to Part I

This chapter concludes the first part of this thesis on adaptive algorithms for widely linear estimation. In Chapter 3, a new methodology for designing widely linear adaptive filters was proposed while this chapter explored a new application of complex-valued adaptive filters in tracking non-circularity. The second part of this thesis (Part II), explores distributed adaptive algorithms which extend the single node adaptive filters to a network-based implementation with multiple agents that act collaboratively in order to jointly estimate a parameter of interest.

Part II

Distributed Adaptive Algorithms

Chapter 5

Diffusion Complex Least Mean Square

> The Internet will disappear. There will be so many IP addresses, so many devices, sensors, things that you are wearing, things that you are interacting with, that you won't even sense it. It will be part of your presence all the time.
>
> *Eric Schmidt*

Chapter Overview

In the next two chapters, the focus shifts to the problem of distributed learning and adaptation over sensor networks. For this chapter, we shall extend the augmented complex least mean square (ACLMS) introduced in Chapter 2, to a distributed setting using the so-called diffusion strategy for distributed adaptive filtering. The main contribution of this chapter, however, is the mean square analysis of the diffusion ACLMS outlined in Section 5.5.3. The proposed mean square analysis is more physically meaningful compared to the current results, while also avoiding the use of the standard small step-size assumption which is routinely used to simplify the analysis.

5.1 Introduction

Sensor networks which found early success from defence-focused applications (e.g. collaborative target tracking) are now experiencing widespread adoption in other areas like environmental monitoring, industrial sensing and traffic control [82]. This growth is mainly fuelled by the emergence of low-cost sensors with communications and computational capabilities. The task of developing fast and robust signal processing algorithms for these sensor networks has therefore been a subject of great interest in the signal processing, control and machine learning communities [83].

Specifically we consider *ad hoc* sensor networks where agents (nodes) do not communicate to a fusion centre but are only able to communicate with their neighbours via single-hop communications [84, 85]. A special case of this communication protocol is known as "gossiping" where at a particular time instant, a node only communicates to one other node [83]. We shall next address the problem formally, using sensor network and graph theory conventions.

5.2 Background

Consider a general network that is represented by an undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, with node set $\mathcal{N} = \{1, 2, \ldots, N\}$, where N is the number of nodes in the network and edge set $\mathcal{E}$ denotes the connections between the nodes in the network. The neighbourhood of a node i, denoted by $\mathcal{N}_i$, is defined as all the nodes connected to node i including itself, that is $\mathcal{N}_i = \{j \,|\, (i,j) \in \mathcal{E}\}$ [23]. The communication links are also assumed to be symmetric, that is, node i both transmits and receives messages from its neighbours.

The cardinality of set $N_i \overset{\text{def}}{=} |\mathcal{N}_i|$ is defined as the number of connections node i has with its neighbours including itself, as illustrated in Figure 5.1. The term $N_{\max} \overset{\text{def}}{=} \max_i \{N_i\}$, denotes the maximum number of connections in the network. The connectivity information is represented using a graph adjacency matrix, $\mathbf{N} \in \mathbb{R}^{N \times N}$, whereby, the $[\mathbf{N}]_{im} = 1$ if node i and m have a communication link and $[\mathbf{N}]_{im} = 0$, otherwise.

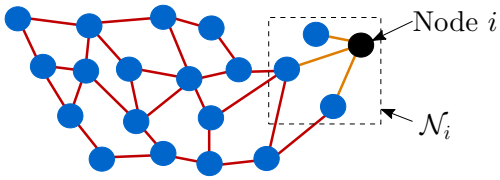

Figure 5.1: A distributed network with $N = 20$ nodes.

The task of the network is to estimate a global parameter vector, $\boldsymbol{w}$, (assumed to be identical throughout the network) observed from noisy local measurements, $(y_{i,k}, \boldsymbol{x}_{i,k})$, whereby the subscript i denotes the node and k the time instant. This is accomplished by minimising the global mean square error (MSE) cost function [84, 85], to yield

$$\boldsymbol{w}_{\text{cen}}^{\circ} = \underset{\boldsymbol{w}}{\text{argmin}} \, \frac{1}{N} \sum_{\ell=1}^{N} \mathbb{E}\left\{|y_{\ell,k} - \boldsymbol{w}^{\mathsf{H}} \boldsymbol{x}_{\ell,k}|^2\right\}. \tag{5.1}$$

In the centralised estimation solution for the model in (5.1), observations collected by each node are transmitted to a fusion centre where a desired algorithm performs the estimation task. Specifically, the optimal solution to (5.1) is given by the centralised Wiener solution [86],

$$\boldsymbol{w}_{\text{cen}}^{\circ} = \left(\frac{1}{N} \sum_{\ell=1}^{N} \mathbf{R}_{\ell}\right)^{-1} \left(\frac{1}{N} \sum_{\ell=1}^{N} \boldsymbol{r}_{\ell}\right), \tag{5.2}$$

where $\mathbf{R}_{\ell} = \mathbb{E}\left\{\boldsymbol{x}_{\ell,k} \boldsymbol{x}_{\ell,k}^{\mathsf{H}}\right\}$, is the data covariance matrix at node ℓ and $\boldsymbol{r}_{\ell} = \mathbb{E}\left\{\boldsymbol{x}_{\ell,k} y_{\ell,k}^{*}\right\}$ is the cross-covariance vector. Alternatively, a recursive stochastic gradient solution to the LMMSE problem is the centralised complex-least mean square (C-CLMS) given by [84]

$$\widehat{\boldsymbol{w}}_{\text{cen},k+1} = \widehat{\boldsymbol{w}}_{\text{cen},k} + \frac{\mu}{N} \sum_{\ell=1}^{N} \boldsymbol{x}_{\ell,k} (y_{\ell,k}^{*} - \boldsymbol{x}_{\ell,k}^{\mathsf{H}} \widehat{\boldsymbol{w}}_{\text{cen},k}), \tag{5.3}$$

where μ is the step-size. The C-CLMS solution statistically converges to the optimal Wiener

solution in (5.2) but allows for an online implementation which is suitable for non-stationary real-world data.

Although the centralised solutions in (5.2)–(5.3) would be globally optimal, excessive communication constraints are placed on the nodes. The aim is to somehow conduct the estimation task in a distributed manner, where each node performs the estimation based on the measurements available to itself and its neighbours, and propagate the estimated weights throughout the network to reach a solution close to the centralised solution.

Notice that both the centralised Wiener solution on (5.2) and the CLMS solution (5.3), require the average network data statistics, e.g. $\frac{1}{N}\sum_{\ell=1}\mathbf{R}_\ell$. The problem of computing a global average quantity in a network in which nodes communicate only with their neighbours, has been long studied in the distributed averaging and consensus filtering communities [87]. Therefore, we shall review some basic ideas on distributed averaging which forms the basis of many distributed estimation algorithms.

5.3 Distributed Averaging

Now, consider the problem of finding an average quantity in a sensor network where each node takes a single measurement of the field (say, the temperature) and aims to compute the average temperature of the field in the form

$$\bar{y} = \frac{1}{N}\sum_{\ell=1}^{N} y_\ell, \tag{5.4}$$

where y_ℓ is a measurement made at node ℓ. The computation of the mean value in (5.4) clearly requires the data collected in the entire network. Now, suppose that the nodes are only able to communicate their values to other neighbouring nodes, whereby no node has access to all the measurements in the network. To compute the global average, each node obtains a local weighted average $\psi_{i,t}$, given by [88]

$$\psi_{i,t} = \sum_{\ell\in\mathcal{N}_i} a_{\ell i}\psi_{\ell,t-1}, \tag{5.5}$$

where t is the iteration number and the intermediate estimates are initialised with the field measurement $\psi_{i,0} = y_i$, while $a_{\ell i}$ is the weight node i assigns to its neighbours. For an unbiased estimation the weighting coefficients sum to unity, $\sum_\ell a_{\ell i} = 1$ [88, 89]. Furthermore, the nodes share their intermediate local averages, $\psi_{i,t}$ with their neighbours and the averaging step in (5.5) is repeated multiple times. The local averages in (5.5) at each node can be shown to converge the global average in (5.4) asymptotically [89].

To prove the result, the local averaging steps throughout the network can be collectively represented by first defining the intermediate local estimates as $\boldsymbol{\psi}_t \overset{\text{def}}{=} [\psi_{1,t}, \psi_{2,t}, \ldots, \psi_{N,t}]^{\mathsf{T}}$.

Next, the weighting coefficients, $a_{\ell i}$, can be combined into a matrix, $\mathbf{A} \in \mathbb{R}^{N\times N}$ as [90]

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1i} & \cdots & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2i} & \cdots & \cdots & a_{2N} \\ \vdots & \vdots & \vdots & a_{3i} & \cdots & \vdots & \vdots \\ a_{i1} & a_{i2} & a_{i3} & a_{ii} & a_{i5} & \cdots & a_{iN} \\ \vdots & \vdots & \cdots & a_{5i} & \ddots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \cdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{Ni} & \cdots & \cdots & a_{NN} \end{bmatrix} \tag{5.6}$$

Weights used by node i to scale information from its neighbours.

Weights assigned to the estimates from node i by its neighbours.

where the weights that node i assigns to nodes $\ell \in \mathcal{N}_i$ are in the i-th column of $\mathbf{A}$. Observe that the matrix $\mathbf{A}$ is left-stochastic since all its columns sum up to unity, i.e.

$$\mathbf{1}_N^{\mathsf{T}}\mathbf{A} = \mathbf{1}_N^{\mathsf{T}}, \quad \mathbf{A}^{\mathsf{T}}\mathbf{1}_N = \mathbf{1}_N, \tag{5.7}$$

which implies that the principal eigenvalue-eigenvector pair of the matrix $\mathbf{A}^{\mathsf{T}}$ is $(1, \frac{1}{\sqrt{N}}\mathbf{1}_N)$. The local averaging step in the network from (5.5) can therefore take the form

$$\boldsymbol{\psi}_{t+1} = \mathbf{A}^{\mathsf{T}}\boldsymbol{\psi}_t, \tag{5.8}$$

with $\boldsymbol{\psi}_0 = \boldsymbol{y}$, where $\boldsymbol{y} = [y_1, y_2, \ldots, y_N]^{\mathsf{T}}$ is the collection of fields measurements. It can be shown that when the local averaging and sharing is repeated many times, the network measurements converge to

$$\lim_{t\to\infty} \boldsymbol{\psi}_{t+1} = \mathbf{A}^{\mathsf{T}}\boldsymbol{\psi}_t = \lim_{t\to\infty} (\mathbf{A}^t)^{\mathsf{T}}\boldsymbol{\psi}_0 = \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^{\mathsf{T}}\boldsymbol{y}, \tag{5.9}$$

whereby, the asymptotic solution, $\frac{1}{N}\mathbf{1}_N^{\mathsf{T}}\boldsymbol{y}$, is the global average solution in (5.4).

5.3.1 Decoupling Adaptation and Diffusion

Consider the network of N nodes, each running the centralised CLMS algorithm in (5.3) in the following manner

$$\textit{Adapt}: \quad \boldsymbol{\psi}_{i,k+1} = \widehat{\boldsymbol{w}}_{i,k} + \mu\boldsymbol{x}_{i,k}(y^*_{i,k} - \boldsymbol{x}^{\mathsf{H}}_{i,k}\widehat{\boldsymbol{w}}_{i,k}) \tag{5.10}$$

$$\textit{Combine}: \quad \widehat{\boldsymbol{w}}_{i,k+1} = \frac{1}{N}\sum_{\ell=1}^{N}\boldsymbol{\psi}_{\ell,k+1}, \tag{5.11}$$

where each node only updates its weight estimate $\boldsymbol{\psi}_{i,k+1}$ using the local data in step (5.10), but by virtue of access to the intermediate estimates $\boldsymbol{\psi}_{i,k+1}$ from all the other nodes in (5.11), it is able to exploit all the information in the network. In fact, if all the nodes are initialised with the

same value, $\boldsymbol{w}_{i,0} = \boldsymbol{\theta}, \forall i$, the decoupled adapt-and-combine steps implemented at each node in (5.10)–(5.11) yields the identical weight estimates to the centralised solution in (5.3) at each iteration.

Although the updates in (5.10)–(5.11) appear to only use local measurements, they require each node to have access to intermediate estimates from all the other nodes, thereby effectively replicating a centralised solution in (5.3). The combination step in (5.11), which shall also be referred to as the diffusion step, is a global averaging procedure which naturally admits a distributed solution. Therefore, the decoupled centralised diffusion step can be configured for *ad hoc* sensor networks by replacing the global average step in (5.11) by local averaging similar to (5.5), which results in the diffusion CLMS algorithm given by [86, 91, 92]

$$\textit{Adapt}: \quad \boldsymbol{\psi}_{i,k+1} = \boldsymbol{w}_{i,k} + \mu \boldsymbol{x}_{i,k}(y^*_{i,k} - \boldsymbol{x}^{\mathsf{H}}_{i,k}\boldsymbol{w}_{i,k}) \tag{5.12}$$

$$\textit{Combine}: \quad \boldsymbol{w}_{i,k+1} = \sum_{\ell \in \mathcal{N}_i} a_{\ell i}\boldsymbol{\psi}_{\ell,k+1}. \tag{5.13}$$

In the next section, we shall formally derive the diffusion augmented complex least mean square algorithm (D-ACLMS) which is capable of estimating data generated from widely linear models.

5.4 Diffusion Augmented Complex Least Mean Square

Suppose that a node i receives streaming measurement data, $\{y_{i,k},\ \bar{\boldsymbol{x}}_{i,k}\}$, at each time instant k, where the measurement $y_{i,k} \in \mathbb{C}$ and input vector (regressor) $\boldsymbol{x}_{i,k} \in \mathbb{C}^{M\times 1}$, are related via a widely linear model [93],

$$y_{i,k} = \boldsymbol{h}^{\mathsf{H}}\boldsymbol{x}_{i,k} + \boldsymbol{g}^{\mathsf{H}}\boldsymbol{x}^*_{i,k} + \eta_{i,k}, \tag{5.14}$$

corrupted with zero-mean white noise $\eta_{i,k}$ with variance σ^2_η. Note that the widely linear model in (5.14) is a generalisation of the standard strictly linear model for which $\boldsymbol{g} = \boldsymbol{0}$. For the compactness of the analysis, we shall represent both the strictly linear and widely linear models using

$$y_{i,k} = \boldsymbol{w}^{\mathsf{H}}\bar{\boldsymbol{x}}_{i,k} + \eta_{i,k}, \tag{5.15}$$

where $\boldsymbol{w} \in \mathbf{C}^{\overline{M}\times 1}$ and $\bar{\boldsymbol{x}}_{i,k} \in \mathbf{C}^{\overline{M}\times 1}$ with $\overline{M} = M$ for strictly linear models and $\overline{M} = 2M$ for widely linear models, see Table 5.1 for details.

Model	$\boldsymbol{w}$	$\bar{\boldsymbol{x}}_{i,k}$	**Dimension**, $\overline{M}$
Strictly Linear	$\boldsymbol{h}$	$\boldsymbol{x}_{i,k}$	$M \times 1$
Widely Linear	$\begin{bmatrix}\boldsymbol{h}\\ \boldsymbol{g}\end{bmatrix}$	$\begin{bmatrix}\boldsymbol{x}_{i,k}\\ \boldsymbol{x}^*_{i,k}\end{bmatrix}$	$2M \times 1$

Table 5.1: *Input and optimal weight vectors for the strictly linear and widely linear models.*

Adaptation Step

The local mean square error (MSE) between the desired signal and the estimate is given by

$$J_{\text{MSE}}^{(i,k)}(\boldsymbol{w}) \stackrel{\text{def}}{=} \mathbb{E}\left\{|y_{i,k} - \boldsymbol{w}^{\mathsf{H}}\bar{\boldsymbol{x}}_{i,k}|^2\right\}. \tag{5.16}$$

At each node, the diffusion augmented complex least mean square (D-ACLMS) algorithm performs a stochastic gradient descent minimisation procedure on two cost functions in succession. Firstly, all the nodes minimise the stochastic approximation of the MSE cost function via

$$\boldsymbol{\psi}_{i,k} = \widehat{\boldsymbol{w}}_{i,k-1} - \mu \frac{\partial \hat{J}_{\text{MSE}}^{(i,k)}(\boldsymbol{w})}{\partial \boldsymbol{w}^*}\bigg|_{\boldsymbol{w}=\widehat{\boldsymbol{w}}_{i,k-1}}, \tag{5.17}$$

where $\hat{J}_{\text{MSE}}^{(i,k)} = |y_{i,k} - \boldsymbol{w}^{\mathsf{H}}\bar{\boldsymbol{x}}_{i,k}|^2$ is an instantaneous estimate of the MSE cost function in (5.16) using local measurements $y_{i,k}$ and input vectors $\bar{\boldsymbol{x}}_{i,k}$ and the step-size is μ. Finding the gradient in from (5.17) yields the ACLMS algorithm[1] [53, 93]

$$\boldsymbol{\psi}_{i,k} = \widehat{\boldsymbol{w}}_{i,k-1} + \mu\bar{\boldsymbol{x}}_{i,k}\left[y_{i,k}^* - \bar{\boldsymbol{x}}_{i,k}^{\mathsf{H}}\widehat{\boldsymbol{w}}_{i,k-1}\right]. \tag{5.18}$$

The term $\boldsymbol{\psi}_{i,k}$ is used instead of $\boldsymbol{w}_{i,k}$ to indicate that the estimate in the adaptation step in (5.18) is an intermediate step and will be succeeded by another operation.

Diffusion Step

The nodes then transmit their local (intermediate) estimates, $\boldsymbol{\psi}_{i,k}$, to neighbouring nodes so that each local node performs the second step of the adaptation process, whereby every node minimises the difference between its own weight estimate in (5.18) and the weights in its neighbourhood $\mathcal{N}_i$, by minimising the cost function

$$J_{\text{AV}}^{(i,k)}(\boldsymbol{\psi}) \stackrel{\text{def}}{=} \sum_{\ell\in\mathcal{N}_i} b_{\ell i}\|\boldsymbol{\psi}_{\ell,k} - \boldsymbol{\psi}\|^2, \tag{5.19}$$

which is the weighted norm (using weights $b_{\ell i}$) of the difference between the estimates in the neighbourhood. The minimisation of this cost function is carried out using the gradient descent method where

$$\begin{aligned}\widehat{\boldsymbol{w}}_{i,k} &= \boldsymbol{\psi}_{i,k} - \mu \frac{\partial J_{\text{AV}}^{(i,k)}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}^*}\bigg|_{\boldsymbol{\psi}=\boldsymbol{\psi}_{i,k}} \\ &= \boldsymbol{\psi}_{i,k} + \sum_{\ell=1}^{N} \mu b_{\ell i}(\boldsymbol{\psi}_{\ell,k} - \boldsymbol{\psi}_{i,k})\end{aligned} \tag{5.20}$$

Rearranging equation in (5.20) gives

$$\widehat{\boldsymbol{w}}_{i,k} = \left(1 - \sum_{\ell\neq i}\mu b_{\ell i}\right)\boldsymbol{\psi}_{i,k} + \sum_{\ell\neq i}\mu b_{\ell i}\boldsymbol{\psi}_{\ell,k} \tag{5.21}$$

[1] See Chapter 2 for more details on the derivation of the ACLMS.

Algorithm 3. Diffusion Augmented Complex Least Mean Square (D-ACLMS)

Initialise each node with: $\widehat{\boldsymbol{w}}_{i,0} = \mathbf{0}$
Training data: $(y_{i,k}, \bar{\boldsymbol{x}}_{i,k})$
At each time instant $k > 0$:

1: **for** Nodes $i = \{1, 2, \ldots, N\}$ **do**
2: $\quad \boldsymbol{\psi}_{i,k|k} = \widehat{\boldsymbol{w}}_{i,k|k-1} + \mu \bar{\boldsymbol{x}}_{i,k} \left[y^*_{i,k} - \bar{\boldsymbol{x}}^{\mathsf{H}}_{i,k} \widehat{\boldsymbol{w}}_{i,k-1} \right]$
3: $\quad$ Diffuse the states from the network:
$\quad\quad \widehat{\boldsymbol{w}}_{i,k|k} = \sum_{\ell \in \mathcal{N}_i} a_{\ell i} \boldsymbol{\psi}_{\ell,k|k}$
4: **end for**

Upon defining $a_{\ell i} \overset{\text{def}}{=} \mu b_{\ell i}$, we can see that $a_{ii} = 1 - \sum_{\ell \neq i} a_{\ell i}$, which yields

$$\widehat{\boldsymbol{w}}_{i,k} = \sum_{\ell=1}^{N} a_{\ell i} \boldsymbol{\psi}_{\ell,k}. \tag{5.22}$$

The weighting coefficients (sometimes referred to as the "trust coefficients") are zero, $a_{\ell i} = 0$, if the node i and ℓ are not connected to each other. For sufficiently small step-sizes, the combiners are also positive, $a_{\ell i} \geq 0$, and are chosen so that they sum up to one [90], that is

$$\sum_{\ell=1}^{N} a_{\ell i} = 1 \implies a_{ii} = 1 - \sum_{\ell \neq i} a_{\ell i}. \tag{5.23}$$

The diffusion ACLMS algorithm implemented at each node is then given by [33, 93]

$$\textit{Adapt:} \quad \boldsymbol{\psi}_{i,k} = \widehat{\boldsymbol{w}}_{i,k-1} + \mu \bar{\boldsymbol{x}}_{i,k} \left[y^*_{i,k} - \bar{\boldsymbol{x}}^{\mathsf{H}}_{i,k} \widehat{\boldsymbol{w}}_{i,k-1} \right] \tag{5.24a}$$

$$\textit{Combine:} \quad \widehat{\boldsymbol{w}}_{i,k} = \sum_{\ell=1}^{N} a_{\ell i} \boldsymbol{\psi}_{\ell,k} \tag{5.24b}$$

The configuration in (5.24a)–(5.24b) is referred to as the *Adapt-then-Combine* configuration while in the *Combine-then-Adapt* scheme, the steps (5.24a)–(5.24b) are reversed. The ATC configuration was shown to have a slight performance advantage in the steady state and is therefore chosen for the rest of this chapter.

Notes on the Combination Weights

Although the determination of the optimal weights for an arbitrary network of nodes is challenging without accurate knowledge of the data statistics for every node [94], it is possible to set combination rules based on information like the cardinality of the neighbourhood. Particularly, the combination weights $a_{\ell i}$ used by the diffusion step in (5.24b) can obey a number of rules, including the Metropolis [91], Laplacian [95] or the nearest neighbour [96] rules, however, finding the set of optimal weights remains an open issue though progress has been made in some particular cases [90, 94, 97]. Table 5.2 shows the most commonly employed combination coefficients [90].

Rules	Weighting Coefficients
Averaging	$a_{\ell i} = \dfrac{1}{N_i},$
Laplacian	$a_{\ell i} = \begin{cases} \dfrac{1}{N_{\max}}, & \ell \neq i \\ 1 - \dfrac{(1-N_i)}{N_{\max}}, & \ell = i \end{cases}$
Maximum Degree	$a_{\ell i} = \begin{cases} \dfrac{1}{N}, & \ell \neq i \\ 1 - \dfrac{(1-N_i)}{N}, & \ell = i \end{cases}$
Relative Degree	$a_{\ell i} = \dfrac{N_\ell}{\sum_{m \in \mathcal{N}_i} N_m},$
Metropolis	$a_{\ell i} = \begin{cases} \dfrac{1}{\max\{N_i, N_\ell\}}, & \ell \neq i \\ 1 - \sum\limits_{\ell \neq i} a_{\ell i}, & \ell = i \end{cases}$

Table 5.2: *Weighting coefficients for distributed averaging.*

5.5 Mean and Mean Square Analysis

Although quantifying the performance of diffusion adaptive networks is not straightforward, as the interactions between multiple connected adaptive filters add to the complexity of the analysis, considerable advances have been made in the area [86, 91]. To make the analysis mathematically tractable, current approaches assume that the step-sizes are small enough so that the second-order terms in the analysis can be neglected [86, 98, 99, 100, 101, 102]. However, this assumption somewhat compromises the steady-state analysis since important characteristics of the performance can be influenced by the second-order terms.

To this end, we propose a new method to bound the mean square performance of an adaptive network while incorporating the second-order terms in the analyses. The proposed bound is based on what we refer to as the "similarity assumption" which states that, upon convergence, the steady-state filter weights in the network are almost identical. The similarity conjecture is physically meaningful as it exploits the inherent behaviour of the diffusion scheme in which the individual filters are adapting towards the same optimal weight, while diffusing their intermediate estimates with neighbouring nodes. For a fully connected network, the bound we propose is exact.

The proposed approach is applied to the analysis of the diffusion complex least mean square (D-CLMS) [91] and its widely linear counterpart, the diffusion augmented CLMS (D-ACLMS) [33, 93]. The closed form expressions for the mean square deviation (MSD) and excess mean square error (EMSE) allow us to quantify the steady-state performance of both the D-CLMS and D-ACLMS as a function of the input data noncircularity; this is not possible with the current methods as the noncircularity of the input data is a second-order effect which is neglected if the small step-size theory is used.

5.5.1 Network Behaviour

To aid our analysis, we shall represent the behaviour of all the nodes in the network using a single collective recursion. The D-ACLMS filter variables in the network can be collected using

$$\begin{aligned}
\boldsymbol{w}_k &= \mathsf{col}\{\boldsymbol{w}_{1,k}, \ldots, \boldsymbol{w}_{N,k}\}, & \boldsymbol{w}_k &\in \mathbb{C}^{\overline{M}N\times 1} \\
\boldsymbol{\psi}_k &= \mathsf{col}\{\boldsymbol{\psi}_{1,k}, \ldots, \boldsymbol{\psi}_{N,k}\}, & \boldsymbol{\psi}_k &\in \mathbb{C}^{\overline{M}N\times 1} \\
\boldsymbol{y}_k &= \mathsf{col}\{y_{1,k}, \ldots, y_{N,k}\}, & \boldsymbol{y}_k &\in \mathbb{C}^{N\times 1} \\
\mathbf{X}_k &= \begin{bmatrix} \bar{\boldsymbol{x}}_{1,k} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \bar{\boldsymbol{x}}_{2,k} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \bar{\boldsymbol{x}}_{N,k} \end{bmatrix}, & \mathbf{X}_k &\in \mathbb{C}^{\overline{M}N\times N} \\
\boldsymbol{\mathcal{A}} &= \mathbf{A}\otimes\mathbf{I}_{\overline{M}}, & \boldsymbol{\mathcal{A}} &\in \mathbb{R}^{\overline{M}N\times\overline{M}N},
\end{aligned} \tag{5.25}$$

where $\mathsf{col}\{\cdot\}$ operator creates a column vector of its arguments and $\otimes$ is the Kronecker product operator. This set-up applies to both the D-CLMS with $\overline{M} = M$ and D-ACLMS with $\overline{M} = 2M$, see Table 5.1.

The D-ACLMS algorithm from (5.24a) and (5.24b) for the network can therefore be expressed as

$$\textit{Adapt:}\ \boldsymbol{\psi}_{k+1} = \boldsymbol{w}_k + \mu\mathbf{X}_k\left(\boldsymbol{y}_k^* - \mathbf{X}_k^{\mathsf{H}}\boldsymbol{w}_k\right) \tag{5.26a}$$

$$\textit{Combine:}\ \boldsymbol{w}_{k+1} = \boldsymbol{\mathcal{A}}^{\mathsf{T}}\boldsymbol{\psi}_{k+1}. \tag{5.26b}$$

Before the mean and mean square converge analyses are carried out, we shall define the optimal weight vector for the network as $\boldsymbol{w}_{\text{net}} \overset{\text{def}}{=} \mathbf{1}_N \otimes \boldsymbol{w}$, since all the nodes in the network are estimating the same optimal weights $\boldsymbol{w} \in \mathbb{C}^{\overline{M}}$.

Subtracting $\boldsymbol{w}_{\text{net}}$ from both sides of (5.26a)–(5.26b), using the desired signal model in (5.15) and recognising that $\boldsymbol{\mathcal{A}}^{\mathsf{T}}\boldsymbol{w}_{\text{net}} = \boldsymbol{w}_{\text{net}}$ enables the network weight error recursion to be formulated as

$$\tilde{\boldsymbol{w}}_{k+1} = \boldsymbol{\mathcal{A}}^{\mathsf{T}}\left(\tilde{\boldsymbol{w}}_k - \mu\widehat{\boldsymbol{\mathcal{R}}}_k\tilde{\boldsymbol{w}}_k - \mu\boldsymbol{n}_k\right), \tag{5.27}$$

where $\widehat{\boldsymbol{\mathcal{R}}}_k \overset{\text{def}}{=} \mathbf{X}_k\mathbf{X}_k^{\mathsf{H}}$ and $\boldsymbol{n}_k \overset{\text{def}}{=} \mathbf{X}_k\boldsymbol{\eta}_k^*$, while $\boldsymbol{\eta}_k = \mathsf{col}\{\eta_{1,k}, \ldots, \eta_{N,k}\}$ is the vector containing measurement noise from all the nodes.

5.5.2 Mean Convergence of the D-ACLMS

Both the mean and mean square analysis of the D-ACLMS will rely upon the network weight error vector in (5.27). Specifically, the mean convergence of the D-ACLMS as addressed in Theorem 5.1 states that if all the individual ACLMS filters in the network convergence in the mean, the network also converges.

Theorem 5.1 (D-ACLMS Mean Convergence). *Consider the D-ACLMS algorithm in (5.24a) – (5.24b) for the associated true system in (5.15). Suppose that the following assumptions hold:*

A.1) The input vector, $\bar{\boldsymbol{x}}_{i,k}$, has zero-mean and is independent and identically distributed (IID) across time k, and nodes i, with covariance and pseudocovariance matrices respectively given by $\bar{\mathbf{R}} = \mathbb{E}\left\{\bar{\boldsymbol{x}}_{i,k}\bar{\boldsymbol{x}}_{i,k}^{\mathsf{H}}\right\}$, $\bar{\mathbf{P}} = \mathbb{E}\left\{\bar{\boldsymbol{x}}_{i,k}\bar{\boldsymbol{x}}_{i,k}^{\mathsf{T}}\right\}$.

A.2) The measurement noise sequence, $\{\eta_{i,k}\}_{i,k\in\mathbb{N}}$ is a zero-mean, white noise process which is IID across time k, and nodes i, and is statistically independent to input vector sequence, $\{\boldsymbol{x}_{i,k}\}_{i,k\in\mathbb{N}}$. Its variance given by, $\sigma_\eta^2 = \mathbb{E}\left\{|\eta_{i,k}|^2\right\}$ $\forall i, k$.

Then, the D-ACLMS algorithm converges in the mean if all the individual ACLMS algorithms in the network convergence in the mean, that is

$$\lim_{k\to\infty} \mathbb{E}\left\{\tilde{\boldsymbol{w}}_k\right\} = \mathbf{0} \quad \textit{if} \quad \lim_{k\to\infty} \mathbb{E}\left\{\tilde{\boldsymbol{w}}_{i,k}\right\} = \mathbf{0} \;\; \forall\, i. \tag{5.28}$$

Proof.
Applying the statistical expectation operator to the weight error recursion in (5.27) and applying Assumptions A.1 – A.2 in Theorem 5.1 yields

$$\mathbb{E}\left\{\tilde{\boldsymbol{w}}_{k+1}\right\} = \boldsymbol{\mathcal{A}}^{\mathsf{T}} \underbrace{\left(\mathbf{I} - \mu\,\boldsymbol{\mathcal{R}}\right)}_{\boldsymbol{\mathcal{B}}} \mathbb{E}\left\{\tilde{\boldsymbol{w}}_k\right\}, \tag{5.29}$$

where the matrix $\boldsymbol{\mathcal{R}} = \mathbb{E}\left\{\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}}\right\}$ is a block-diagonal matrix given by

$$\boldsymbol{\mathcal{R}} = \begin{bmatrix} \bar{\mathbf{R}}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{R}}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \bar{\mathbf{R}}_N \end{bmatrix}. \tag{5.30}$$

with input data covariance (or augmented covariance) matrices at each node, $\bar{\mathbf{R}}_i = \mathbb{E}\left\{\bar{\boldsymbol{x}}_{i,k}\bar{\boldsymbol{x}}_{i,k}^{\mathsf{H}}\right\} = \mathbf{R}$, according to Assumption A.1 in Theorem 5.1. Therefore, the weight error recursion can be written as

$$\mathbb{E}\left\{\tilde{\boldsymbol{w}}_{k+1}\right\} = \boldsymbol{\mathcal{A}}^{\mathsf{T}}\boldsymbol{\mathcal{B}}\,\mathbb{E}\left\{\tilde{\boldsymbol{w}}_k\right\}. \tag{5.31}$$

For this recursion to be a contraction, the spectral radius of the matrix $\boldsymbol{\mathcal{A}}^{\mathsf{T}}\boldsymbol{\mathcal{B}}$ has to be less than unity. The spectral radius of a matrix, denoted by $\varrho(\cdot)$, is bounded by any induced matrix norm as [103]

$$\varrho(\boldsymbol{\mathcal{A}}^{\mathsf{T}}\boldsymbol{\mathcal{B}}) \leq \|\boldsymbol{\mathcal{A}}^{\mathsf{T}}\boldsymbol{\mathcal{B}}\| \leq \|\boldsymbol{\mathcal{A}}^{\mathsf{T}}\|\|\boldsymbol{\mathcal{B}}\|. \tag{5.32}$$

Since the matrices and vectors in the analysis are of block-form, we shall introduce the so-called "block-maximum norm". Let $\boldsymbol{w} = \mathsf{col}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots \boldsymbol{w}_N\}$ denote a block vector with N column

vectors, $\boldsymbol{w}_i$, each with dimension M. The block maximum norm of $\boldsymbol{w}$ is [90, App. D]

$$\|\boldsymbol{w}\|_{b,\infty} \stackrel{\text{def}}{=} \max_{1\leq i\leq N} \|\boldsymbol{w}_i\|_2,$$

where $\|\cdot\|_2$ denotes the Euclidean norm. The corresponding induced block maximum norm for a matrix, $\mathbf{M}$, is defined as

$$\|\mathbf{M}\|_{b,\infty} \stackrel{\text{def}}{=} \max_{\boldsymbol{w}\neq 0} \frac{\|\mathbf{M}\boldsymbol{w}\|_{b,\infty}}{\|\boldsymbol{w}\|_{b,\infty}}.$$

The block maximum norm of the stochastic matrix in (5.31) is given by $\|\boldsymbol{\mathcal{A}}\|_{b,\infty} = 1$. Therefore, selecting the block-maximum norm for the inequality in (5.32) gives

$$\varrho(\boldsymbol{\mathcal{A}}^{\mathsf{T}}\boldsymbol{\mathcal{B}}) \leq \|\boldsymbol{\mathcal{A}}^{\mathsf{T}}\|_{b,\infty}\|\boldsymbol{\mathcal{B}}\|_{b,\infty} = \|\boldsymbol{\mathcal{B}}\|_{b,\infty}. \tag{5.33}$$

Also, as $\boldsymbol{\mathcal{B}}$ is a Hermitian block-diagonal matrix, its block maximum norm is equal to its spectral radius $\|\boldsymbol{\mathcal{B}}\|_{b,\infty} = \varrho(\mathbf{B})$ which implies that [90, App. D]

$$\varrho(\boldsymbol{\mathcal{A}}^{\mathsf{T}}\boldsymbol{\mathcal{B}}) \leq \varrho(\boldsymbol{\mathcal{B}}) \tag{5.34}$$

for any left stochastic matrix $\boldsymbol{\mathcal{A}}$ [90]. For the mean convergence of the individual ACLMS algorithms, please refer to Chapter 2. □

Remark 5.1. *The diffusion strategy enhances the stability range of the network by reducing the spectral radius of* $\boldsymbol{\mathcal{A}}^{\mathsf{T}}\boldsymbol{\mathcal{B}}$.

> **Contribution: Mean Square Analysis of the D-ACLMS**
> The contribution of this chapter is the mean square analysis of the D-CLMS and D-ACLMS in Theorem 5.2. The proposed analysis avoids the small step-size assumption used in the current theory. This is accomplished by proposing and using the similarity assumption in (5.36). This results in the structurally elegant steady-state MSD in (5.52) which is able reveal the effects of the second-order terms like noncircularity of the input data on the network MSD.

5.5.3 Mean Square Analysis of the D-ACLMS

The mean square behaviour of the network is analysed using the weight error covariance matrix

$$\boldsymbol{\mathcal{K}}_k \stackrel{\text{def}}{=} \mathbb{E}\left\{\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}}\right\}, \tag{5.35}$$

where $\tilde{\boldsymbol{w}}_k = \boldsymbol{w}_{\text{net}} - \boldsymbol{w}_k$ is the network weight error vector defined in (5.27). Notice that the weight error covariance matrix is a block matrix in which the ij-th block is given by $[\,\boldsymbol{\mathcal{K}}_k\,]_{\boldsymbol{ij}} \stackrel{\text{def}}{=} \mathbf{K}_{ij,k} = \mathbb{E}\left\{\tilde{\boldsymbol{w}}_{i,k}\tilde{\boldsymbol{w}}_{j,k}^{\mathsf{H}}\right\}$.

Theorem 5.2 (D-ACLMS Steady-state Mean Square Deviation). *Consider the D-ACLMS algorithm in (5.24a) –(5.24b) for the associated true system in (5.15). In addition to the Assumptions A.1 – A.2 in Theorem 5.1, suppose that the following assumptions hold:*

A.3) The elements of the input vectors, $\bar{\boldsymbol{x}}_{i,k}$, are drawn from a IID Gaussian distribution, so that the covariance and pseudocovariance matrices of $\bar{\boldsymbol{x}}_{i,k}$, are given by $\bar{\mathbf{R}} = r\mathbf{I}$, $\bar{\mathbf{P}} = p\mathbf{I}$.

A.4) The D-ACLMS converges to a steady-state, with the weight vectors at each node given by

$$\lim_{k\to\infty} \tilde{\boldsymbol{w}}_k = \mathbf{1}_N \otimes \tilde{\boldsymbol{w}}_{ss}, \quad (a.s.) \tag{5.36}$$

where $\tilde{\boldsymbol{w}}_{ss} \in \mathbf{C}^{\overline{M}\times 1}$.

Then, there exists real scalars, $0 < \alpha \leq 1$ and

$$\vartheta \stackrel{def}{=} \left(1 - 2\mu r + \mu^2 r^2 + \alpha\mu^2(|p|^2 + r^2\overline{M})\right), \tag{5.37}$$

with $0 < \vartheta < 1$ when the step-sizes μ of the individual filters in the node are chosen to satisfy

$$0 < \mu < \frac{2r}{r^2 + \alpha(|p|^2 + r^2\overline{M})}, \tag{5.38}$$

so that the average steady-state mean square deviation of the D-ACLMS is bounded by

$$\lim_{k\to\infty} \frac{1}{N}\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_k\|^2\right\} = \frac{\alpha\mu^2\sigma_\eta^2 \mathsf{Tr}[\bar{\mathbf{R}}]}{1-\vartheta}. \tag{5.39}$$

Remark 5.2. *The contribution of Theorem 5.2 is the removal a restrictive assumption employed in current analysis of distributed adaptive filters which is that the step-size μ is small, so that terms which contain the variable μ^2 can be ignored [90]. We however, introduce Assumption A.4 to keep the analysis mathematically tractable while considering the second order, i.e. μ^2, terms. Assumption A.4 is realistic as the "combine step" (5.26a) reduces the difference between filter weight vectors via averaging and the "adapt step" in (5.26b) forces the filter weights to evolve towards the true value, $\boldsymbol{w}$, which is identical throughout the network.*

Proof.
The weight error covariance matrix, $\boldsymbol{\mathcal{K}}_k$, is obtained by first post-multiplying both sides of (5.27) by their Hermitian transpose and taking the statistical expectation to give

$$\begin{aligned}\mathbb{E}\left\{\tilde{\boldsymbol{w}}_{k+1}\tilde{\boldsymbol{w}}_{k+1}^{\mathsf{H}}\right\} = \mathbb{E}\Big\{\boldsymbol{\mathcal{A}}^{\mathsf{T}}\Big(&\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}} - \mu\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}}\widehat{\boldsymbol{\mathcal{R}}}_k - \mu\widehat{\boldsymbol{\mathcal{R}}}_k\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}} \\ &+ \mu^2\,\widehat{\boldsymbol{\mathcal{R}}}_k\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}}\widehat{\boldsymbol{\mathcal{R}}}_k + \mu^2\boldsymbol{n}_k\boldsymbol{n}_k^{\mathsf{H}}\Big)\boldsymbol{\mathcal{A}}\Big\}.\end{aligned} \tag{5.40}$$

First, notice that the network input data statistics are given by both the covariance and pseudocovariance matrices, respectively denoted by $\boldsymbol{\mathcal{R}} = \mathbb{E}\left\{\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}}\right\}$ and $\boldsymbol{\mathcal{P}} = \mathbb{E}\left\{\mathbf{X}_k\mathbf{X}_k^{\mathsf{T}}\right\}$. Next,

the ij-th block of the block-matrix $\mathbb{E}\left\{\widehat{\boldsymbol{\mathcal{R}}}_k\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}}\widehat{\boldsymbol{\mathcal{R}}}_k\right\}$ in (5.40) is given by

$$\mathbb{E}\left\{\widehat{\boldsymbol{\mathcal{R}}}_k\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}}\widehat{\boldsymbol{\mathcal{R}}}_k\right\}_{\boldsymbol{ij}} = \mathbb{E}\left\{\bar{\boldsymbol{x}}_{i,k}\bar{\boldsymbol{x}}_{i,k}^{\mathsf{H}}\tilde{\boldsymbol{w}}_{i,k}\tilde{\boldsymbol{w}}_{j,k}^{\mathsf{H}}\bar{\boldsymbol{x}}_{j,k}\bar{\boldsymbol{x}}_{j,k}^{\mathsf{H}}\right\}. \tag{5.41}$$

Using the Gaussian data assumption in Assumption A.3 in Theorem 5.2 and the Gaussian moment factorising theorem, the expression in (5.41) becomes

$$\mathbb{E}\left\{\widehat{\boldsymbol{\mathcal{R}}}_k\tilde{\boldsymbol{w}}_k\tilde{\boldsymbol{w}}_k^{\mathsf{H}}\widehat{\boldsymbol{\mathcal{R}}}_k\right\}_{\boldsymbol{ij}} = \begin{cases} \bar{\mathbf{R}}_i\mathbf{K}_{i,k}\bar{\mathbf{R}}_i + \bar{\mathbf{P}}_i\mathbf{K}_{i,k}^{\mathsf{T}}\bar{\mathbf{P}}_i^{*} + \bar{\mathbf{R}}_i\mathsf{Tr}[\bar{\mathbf{R}}_i\mathbf{K}_{i,k}], & \boldsymbol{i}=\boldsymbol{j}, \\ \bar{\mathbf{R}}_i\mathbf{K}_{ij,k}\bar{\mathbf{R}}_j, & \boldsymbol{i}\neq\boldsymbol{j}, \end{cases}$$

where $\mathbf{R}_i = \mathbb{E}\left\{\bar{\boldsymbol{x}}_{i,k}\bar{\boldsymbol{x}}_{i,k}^{\mathsf{H}}\right\}$ and $\mathbf{P}_i = \mathbb{E}\left\{\bar{\boldsymbol{x}}_{i,k}\bar{\boldsymbol{x}}_{i,k}^{\mathsf{T}}\right\}$ are respectively the covariance and pseudo-covariance matrices of the input vector at node i. This results in,

$$\begin{aligned}\bar{\boldsymbol{\mathcal{K}}}_k &\overset{\text{def}}{=} \mathsf{bdiag}\left\{\mathbf{K}_{1,k},\ldots,\mathbf{K}_{N,k}\right\} \\ \boldsymbol{\mathcal{T}}_k &\overset{\text{def}}{=} \mathsf{bdiag}\left\{\mathsf{Tr}[\bar{\mathbf{R}}_1\mathbf{K}_{1,k}]\mathbf{I}_{\overline{M}},\ldots,\mathsf{Tr}[\bar{\mathbf{R}}_N\mathbf{K}_{N,k}]\mathbf{I}_{\overline{M}}\right\},\end{aligned}$$

and where the operator $\mathsf{bdiag}\{\cdot\}$ creates a block diagonal matrix. Then, upon applying employing the Assumptions A.1 – A.2 from Theorem 5.1 and Assumption 5.2, we have

$$\begin{aligned}\boldsymbol{\mathcal{K}}_{k+1} = \boldsymbol{\mathcal{A}}^{\mathsf{T}}\Big(&\boldsymbol{\mathcal{K}}_k - \mu\boldsymbol{\mathcal{K}}_k\boldsymbol{\mathcal{R}} - \mu\boldsymbol{\mathcal{R}}\boldsymbol{\mathcal{K}}_k + \mu^2\boldsymbol{\mathcal{R}}\boldsymbol{\mathcal{K}}_k\boldsymbol{\mathcal{R}} \\ &+ \mu^2\boldsymbol{\mathcal{P}}\bar{\boldsymbol{\mathcal{K}}}_k^{\mathsf{T}}\boldsymbol{\mathcal{P}} + \mu^2\boldsymbol{\mathcal{T}}_k\boldsymbol{\mathcal{R}} + \mu^2\sigma_\eta^2\boldsymbol{\mathcal{R}}\Big)\boldsymbol{\mathcal{A}},\end{aligned} \tag{5.42}$$

where apart from the weight error covariance matrix, $\boldsymbol{\mathcal{K}}_k$, and the combination matrix, $\boldsymbol{\mathcal{A}}$, all the other matrices are block diagonal. A consequence of Assumption A.4 in Theorem (5.2) which will be exploited in the analysis is

$$\lim_{k\to\infty}\boldsymbol{\mathcal{K}}_k = \mathbf{1}_N\mathbf{1}_N^{\mathsf{T}}\otimes\mathbf{K}_{ss}, \quad \mathbf{K}_{ss}\in\mathbf{C}^{\overline{M}\times\overline{M}}. \tag{5.43}$$

Remark 5.3. *For a fully connected network (equivalent to a centralised system), the approximations in* (5.36) *and* (5.43) *are exact not only at the steady-state but at every iteration.*

For a fully connected network, the combination matrix becomes $\mathbf{A} = \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^{\mathsf{T}}$, [89] therefore from the combine step in (5.26b)

$$\tilde{\boldsymbol{w}}_k = \boldsymbol{\mathcal{A}}^{\mathsf{T}}\tilde{\boldsymbol{\psi}}_k = \mathbf{1}_N\otimes\tilde{\boldsymbol{\psi}}_{\mathrm{av},k},$$

where $\tilde{\boldsymbol{\psi}}_{\mathrm{av},k} = \frac{1}{N}\sum_{\ell=1}^{N}\tilde{\boldsymbol{\psi}}_{\ell,k}$. This implies that $\boldsymbol{\mathcal{K}}_k = \mathbf{1}_N\mathbf{1}_N^{\mathsf{T}}\otimes\mathbb{E}\left\{\tilde{\boldsymbol{\psi}}_{\mathrm{av},k}\tilde{\boldsymbol{\psi}}_{\mathrm{av},k}^{\mathsf{H}}\right\}$, thereby completing the proof.

Remark 5.4. *For all other topologies, the similarity assumption in* (5.43) *acts as an upper bound for the mean square performance. This is because equation* (5.43) *can be interpreted as replacing the weight error cross-covariance matrices between the different nodes,* $\mathbf{K}_{ij,k}$*, with the weight error auto-covariance matrix,* $\mathbf{K}_{ii,k}$*, which acts as an upper-bound to* $\mathbf{K}_{ij,k}$*.*

Single Filter Representation of the D-ACLMS

Finally, incorporating Assumption A.1 which states that the covariance and pseudocovariance matrices are identical throughout the network, i.e. $\bar{\mathbf{R}}_i = \mathbf{R}$ and $\bar{\mathbf{P}}_i = \mathbf{P}$, $\forall i$, together with the similarity assumption in (5.43) enables the terms in the network weight error covariance recursion in (5.42) to be reduced to $\boldsymbol{\mathcal{R}} = \mathbf{I}_N \otimes \bar{\mathbf{R}}$, $\boldsymbol{\mathcal{P}} = \mathbf{I}_N \otimes \bar{\mathbf{P}}$, $\bar{\boldsymbol{\mathcal{K}}}_k = \mathbf{I}_N \otimes \mathbf{K}_k$, $\boldsymbol{\mathcal{T}}_k = \mathsf{Tr}[\bar{\mathbf{R}}\mathbf{K}_k]\mathbf{I}_{N\overline{M}}$.

Taking the average of the matrices in the block diagonal entries of (5.42) as

$$\mathbf{K}_k \overset{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} [\boldsymbol{\mathcal{K}}_k]_{\boldsymbol{ii}}\,, \tag{5.44}$$

now gives the evolution of the average weight error covariance matrix in the network in the form

$$\begin{aligned}\mathbf{K}_{k+1} = \mathbf{K}_k - \mu\mathbf{K}_k\bar{\mathbf{R}} - \mu\bar{\mathbf{R}}\mathbf{K}_k + \mu^2\bar{\mathbf{R}}\mathbf{K}_k\bar{\mathbf{R}} \\ + \alpha\mu^2\left(\bar{\mathbf{P}}\mathbf{K}_k^{\mathsf{T}}\bar{\mathbf{P}}^* + \mathsf{Tr}[\bar{\mathbf{R}}\mathbf{K}_k]\bar{\mathbf{R}} + \sigma_\eta^2\bar{\mathbf{R}}\right),\end{aligned} \tag{5.45}$$

where $\alpha = \mathsf{Tr}[\mathbf{A}^{\mathsf{T}}\mathbf{A}]/N$ is the factor that captures the effect of the diffusion strategy.

Remark 5.5. *The factor $\alpha \leq 1$ acts as a contractive term, which reduces the mean square error along the iterations.*

Proof: Since $a_{\ell i} \geq 0$, and $\sum_\ell a_{\ell i} = 1$, it can be shown that

$$\begin{aligned}\alpha = \frac{1}{N}\mathsf{Tr}[\mathbf{A}^{\mathsf{T}}\mathbf{A}] = \frac{1}{N}\sum_i\sum_\ell a_{\ell i}^2 \\ \leq \frac{1}{N}\sum_i\Big(\sum_\ell a_{\ell i}\Big)^2 = 1.\end{aligned} \tag{5.46}$$

It is important to note that employing the similarity assumption and uniform data statistics transforms the network mean square performance recursion in (5.42) into a representation in (5.45), which resembles a classical single filter weight error covariance recursion [104, 105, 106] as reviewed in Chapter 2. Although the diffusion network has been analysed from a single-node perspective in the past [107, 108], the proposed similarity assumption simplifies the analysis by giving a means to deal with cross-nodal weight error covariance terms as explained in Remark 5.4.

Now, we proceed by analysing the D-ACLMS similarly to the individual ACLMS in Chapter 2. Applying the trace operator, we have

$$\begin{aligned}\mathsf{Tr}[\mathbf{K}_{k+1}] = \mathsf{Tr}[\mathbf{K}_k - \mu\mathbf{K}_k\bar{\mathbf{R}} - \mu\bar{\mathbf{R}}\mathbf{K}_k + \mu^2\bar{\mathbf{R}}\mathbf{K}_k\bar{\mathbf{R}}] \\ + \alpha\mu^2\mathsf{Tr}[\big(\bar{\mathbf{P}}\mathbf{K}_k^{\mathsf{T}}\bar{\mathbf{P}}^* + \mathsf{Tr}[\bar{\mathbf{R}}\mathbf{K}_k]\bar{\mathbf{R}} + \sigma_\eta^2\mathsf{Tr}[\bar{\mathbf{R}}]\big)].\end{aligned} \tag{5.47}$$

Next using the property $\mathsf{Tr}[\mathbf{M}_1 + \mathbf{M}_2] = \mathsf{Tr}[\mathbf{M}_1] + \mathsf{Tr}[\mathbf{M}_2]$ and $\mathsf{Tr}[\mathbf{M}_1'\mathbf{M}_2'] = \mathsf{Tr}[\mathbf{M}_2'\mathbf{M}_1']$ gives

$$\mathsf{Tr}[\mathbf{K}_{k+1}] = \mathsf{Tr}[\big(\mathbf{I} - 2\mu\bar{\mathbf{R}} + \mu^2\bar{\mathbf{R}}^2 + \alpha\mu^2(\mathbf{P}\mathbf{P}^{\mathsf{H}} + \mathsf{Tr}[\bar{\mathbf{R}}]\bar{\mathbf{R}})\big)\,\mathbf{K}_k] + \alpha\mu^2\sigma_\eta^2\mathsf{Tr}[\bar{\mathbf{R}}]. \tag{5.48}$$

Using Assumption A.3 in Theorem 5.2, which states that the augmented covariance and pseudocovariance matrices are diagonal, we have

$$\mathsf{Tr}[\mathbf{K}_{k+1}] = \mathsf{Tr}[\left(1 - 2\mu r + \mu^2 r^2 + \alpha\mu^2(|p|^2 + r^2\overline{M})\right)\mathbf{K}_k] + \alpha\mu^2\sigma_\eta^2\mathsf{Tr}[\bar{\mathbf{R}}]. \tag{5.49}$$

Defining $\vartheta = \left(1 - 2\mu r + \mu^2 r^2 + \alpha\mu^2(|p|^2 + r^2\overline{M})\right)$ gives

$$\mathsf{Tr}[\mathbf{K}_{k+1}] = \vartheta\mathsf{Tr}[\mathbf{K}_k] + \alpha\mu^2\sigma_\eta^2\mathsf{Tr}[\bar{\mathbf{R}}]. \tag{5.50}$$

To satisfy the condition $\vartheta < 1$, the step-size μ has to be chosen such that

$$0 < \mu < \frac{2r}{r^2 + \alpha(|p|^2 + r^2\overline{M})}.$$

Applying the equality in (5.50) successively and letting $k \to \infty$ gives

$$\lim_{k\to\infty} \mathsf{Tr}[\mathbf{K}_k] = \frac{\alpha\mu^2}{1-\vartheta}\sigma_\eta^2\mathsf{Tr}[\bar{\mathbf{R}}]. \tag{5.51}$$

Also, from (5.44), we finally have

$$\lim_{k\to\infty} \frac{1}{N}\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_k\|^2\right\} = \frac{\alpha\mu^2\sigma_\eta^2\mathsf{Tr}[\bar{\mathbf{R}}]}{1-\vartheta}. \tag{5.52}$$

□

Remark 5.6. *The closed form expressions in (5.52) have much simpler forms compared to the existing results. As stated in Remark 5.4, the so obtained MSD and EMSE expressions are exact for a fully connected network and represent an upper bound on the performance for other topologies.*

Remark 5.7. *The steady-state mean square deviation and bounds for the step-sizes in Theorem 5.2 are related to the results of Theorem 2.3 in Chapter 2. However, we would like to point out a several differences:*

- *Firstly, by virtue of the diagonal covariance and pseudocovariance matrices in Assumption A.3, the expression for ϑ in (5.37) involves the scalars r and p which are also the eigenvalues of the matrices $\bar{\mathbf{R}}$ and $\bar{\mathbf{P}}$. This also enables the expression for the MSD in (5.39) to be stated as an equality rather than a bound.*
- *Secondly, the term α which appears in both the step-size bound in (5.38) and MSD in (5.39), represents the network topology and the combination scheme used in the D-ACLMS.*

5.6 Simulations

In the following simulations, the proposed MSD expression in (5.52), denoted by "(New)" was benchmarked against the current mean square analysis [90] denoted by "(Old)" and Monte-Carlo simulations for MSD values for a system identification task in a network of 20 nodes. The system was an FIR channel with weight vector $\boldsymbol{h} = [-0.2,\ 0.8]^\mathsf{T}$. Unless stated otherwise,

the input $\boldsymbol{x}_k$ was a zero-mean Gaussian process with covariance matrix $\mathbf{R} = \mathbf{I}_M$ and pseudo-covariance matrix $\mathbf{P} = \mathbf{0}$ and the step-size was $\mu = 0.1$. The Metropolis rule was chosen as the combination rule according to Table 5.2.

The simulations were repeated with other well-known combination rules like the uniform, Laplacian, relative degree and maximum degree rules [101], and the relationship between the MSD and the level of connectivity, step-sizes and input circularity were identical. We therefore have chosen to present the results with the Metropolis rule.

Case Study #1: Varying levels of connectivity.

In the first set of simulations, the MSD was evaluated at different levels of connectivity within the network. For each level of connectivity (i.e. number of connections in the network), the performance was averaged over an ensemble of 5000 randomly generated network topologies. A single realisation of a 20-node network with 30 connections is shown in Figure 5.1. Figure 5.2 shows that, as discussed in Remark 5.4, the theoretical MSD proposed in (5.52) was closest to the actual MSDs obtained from the Monte-Carlo simulation in a fully connected network because it mimics the centralised implementation where the filter weights in all the nodes are identical. For networks with only a few connections, the theoretical MSD from (5.52) acts as an upper bound for the algorithm performance.

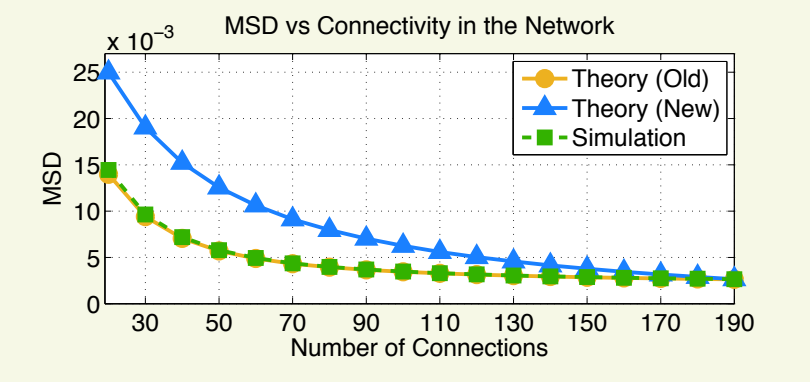

Figure 5.2: MSD values for in a 20-node network with varying levels of connectivity.

Case Study #2: Varying the step-sizes.

Figure 5.3 shows the MSD values for a fully connected 20-node network with different step-sizes. For small step-sizes, as expected, the existing expression for the MSD matches the simulated values. However, for larger step-sizes, the proposed MSD better models the simulated values.

Case Study #3: Effects of noncircularity.

In the third set of simulations (as shown in Figure 5.4), the input signal was noncircular with varying levels of noncircularity modelled as $\mathbf{P} = \rho\mathbf{I}$. Current mean square analyses [90] use a

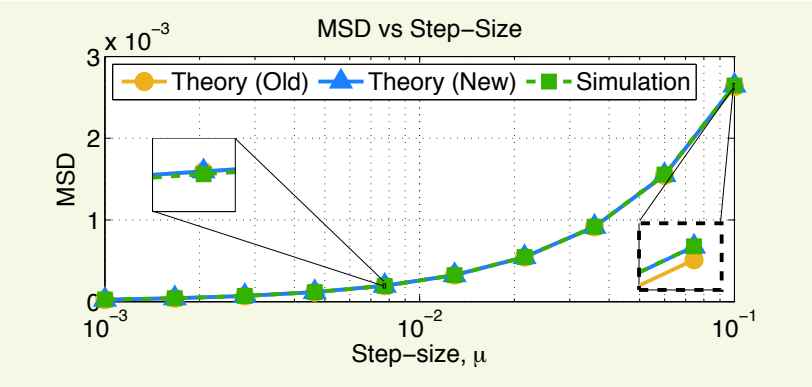

Figure 5.3: MSD levels for different step-sizes.

small step-size assumption to neglect the second-order terms associated with μ^2, and in doing so they do not capture the impact of noncircularity of the input signal on the performance. The proposed expression in (5.52), however, produced the MSD performance that almost identically matches the values obtained from the Monte-Carlo simulations. As shown in [35], the MSD is directly proportional to the noncircularity of the input.

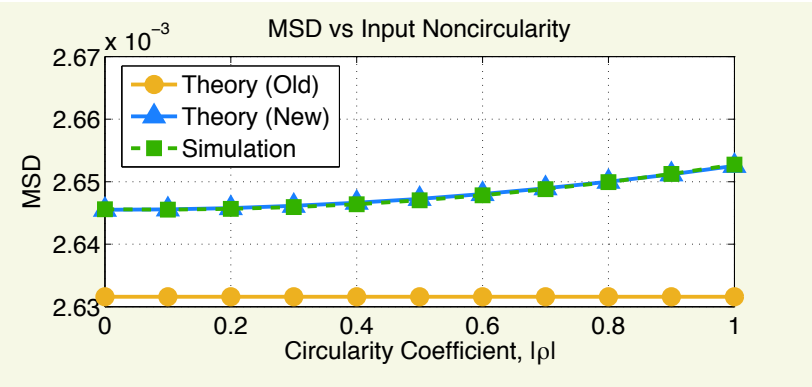

Figure 5.4: MSD values for different levels of input noncircularity.

5.7 Chapter Summary

We have introduced the strictly linear and widely linear D-CLMS and D-ACLMS algorithms which generalize their single-node counterparts to collaborative estimation tasks, as outlined in Algorithm 3. For rigour, a novel, compact and physically meaningful mean square analysis for both the D-CLMS and D-ACLMS algorithms has been proposed in Theorem 5.2. This has been achieved based on the so-called "similarity assumption" in (5.36) which allows for the incorporation of the second-order terms, while maintaining the mathematical tractability of the analysis. This has resulted in the structurally elegant steady-state MSD in (5.52). The proposed method has then been applied to study the effect of the noncircularity of the input signal

on the steady-state mean square performance of the D-CLMS and D-ACLMS. Simulations on synthetic data have verified the proposed analysis.

In Chapter 6, we shall extend the applicability of diffusion adaptive filters by generalising the widely linear model in (5.14) to cater for non-linear models. This will be accomplished using a complex-valued extended Kalman filtering scheme which, for the first time, can also cater for noncircular noise and widely linear models state-space models. The application of the proposed diffusion nonlinear estimators shall be explored in Chapter 8.

Chapter 6

Diffusion Complex Extended Kalman Filter

> Intelligence is deeply and inevitably coupled with interaction.
>
> *Gerhard Weiss*

Chapter Overview

The diffusion-ACLMS algorithm, introduced in Chapter 5, was shown to be suitable for widely linear models. Although the stochastic-gradient type algorithms like the ACLMS are structurally simple and robust, more general learning algorithms that are able to account for a wider range of models are needed in many practical applications.

To this end, we introduce a distributed extended Kalman filtering algorithm which is capable of estimating nonlinear state space models. The proposed distributed extended Kalman filtering is an extension of the diffusion strategy-based Kalman filter in [109], and caters for widely linear state and observation models, noncircular data, and spatially correlated measurement noise. In fact, the diffusion augmented complex Kalman filter (D-ACEKF) could also be considered as a generalisation of the class of diffusion Kalman filters proposed in [96, 109].

6.1 Introduction

Early work in the field of distributed Kalman filtering focused on decentralising the Kalman filtering operations using individual agents which communicate to a fusion centre [110]. This method will be referred to as the centralised Kalman filter, since a central fusion centre has access to all the information in the network. Since communicating to a single fusion centre makes the network vulnerable to a single point of failure, more decentralised solutions were proposed. Fully distributed Kalman filters then began to emerge where each node was required to share all its information with every other node in the network [111, 112], i.e. effectively replicating the operation of the centralised Kalman filter at each node in the network [111].

More general distributed Kalman filters were proposed in the consensus estimation framework, where the constraint that nodes communicate with every other node in the network was relaxed [113, 114]. To compensate for the fact that nodes only access the measurements in their neighbourhood, consensus Kalman filters included a consensus step for their state estimates whereby the individual nodes exchange and average their intermediate state estimates with their neighbours several times before the next measurement is obtained [115]. Therefore the consensus filters operated at two time-scales, a longer time-scale for the measurement updates and a shorter time-scale for the consensus update. The consensus protocol is therefore unsuitable for problems where measurements are taken in a similar time-scale as the communication

protocol.

The diffusion Kalman filter, proposed in [96], is based on a wider class of diffusion adaptive algorithms [90, 97], and enabled both the measurement update and information fusion throughout the network to be applied in a single time-scale. Furthermore, it was shown that diffusion strategies enable information to diffuse more thoroughly in the network compared to consensus strategies [116]. A fundamental feature in the diffusion strategy is that only the state estimates, together with observation variables, are shared in the network.

However, traditional diffusion strategies only account for strictly linear models with circular noise processes. To cater for widely linear models, the diffusion augmented Kalman filter (D-ACKF) was proposed in [117]. In this chapter, we extend the D-ACKF for non-linear models using the theory of extended Kalman filtering and refer to the proposed algorithm as the diffusion augmented complex extended Kalman filter (D-ACEKF). Mean and mean square analyses are included for rigour.

6.2 Single-node Complex Kalman Filters

Consider a general network represented by an undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, with node set $\mathcal{N} = \{1, 2, \ldots, N\}$, where N is the number of nodes in the network and edge set $\mathcal{E}$ designates the connections between the nodes in the network. At each time instant, k, the node i is tasked to estimate a parameter (or state) vector $\boldsymbol{w}_k \in \mathbb{C}^{M\times 1}$ which is assumed to be identical throughout the network but observed locally through measurements $\boldsymbol{y}_{i,k} \in \mathbb{C}^{L\times 1}$. The measurements and state are coupled via a state-space model given by

$$\begin{aligned} \boldsymbol{w}_k &= \mathbf{F}_k \boldsymbol{w}_{k-1} + \boldsymbol{q}_k \\ \boldsymbol{y}_{i,k} &= \mathbf{H}_{i,k} \boldsymbol{w}_k + \boldsymbol{\eta}_{i,k}, \end{aligned} \tag{6.1}$$

where $\mathbf{F}_k \in \mathbb{C}^{M\times M}$ is the state transition matrix and $\mathbf{H}_{i,k} \in \mathbb{C}^{L\times M}$ is the observation matrix. In standard complex-valued Kalman filtering literature, the state noise $\boldsymbol{q}_k \in \mathbb{C}^{M\times 1}$ and observation noise $\boldsymbol{\eta}_{i,k} \in \mathbb{C}^{L\times 1}$ are temporally uncorrelated and spatially independent zero-mean white Gaussian noise processes with a joint covariance and pseudocovariance matrices defined as [96]

$$\mathbf{C}_k \overset{\text{def}}{=} \mathbb{E}\left\{\boldsymbol{q}_k \boldsymbol{q}_k^{\mathsf{H}}\right\},\ \tilde{\mathbf{C}}_k \overset{\text{def}}{=} \mathbb{E}\left\{\boldsymbol{q}_k \boldsymbol{q}_k^{\mathsf{T}}\right\},\ \boldsymbol{\Sigma}_{i,k} \overset{\text{def}}{=} \mathbb{E}\left\{\boldsymbol{\eta}_{i,k} \boldsymbol{\eta}_{i,k}^{\mathsf{H}}\right\},\ \tilde{\boldsymbol{\Sigma}}_{i,k} \overset{\text{def}}{=} \mathbb{E}\left\{\boldsymbol{\eta}_{i,k} \boldsymbol{\eta}_{i,k}^{\mathsf{T}}\right\}, \tag{6.2a}$$

$$\mathbb{E}\left\{\begin{bmatrix}\boldsymbol{q}_k \\ \boldsymbol{\eta}_{i,k}\end{bmatrix}\begin{bmatrix}\boldsymbol{q}_n^{\mathsf{H}} & \boldsymbol{\eta}_{\ell,n}^{\mathsf{H}}\end{bmatrix}\right\} = \begin{bmatrix}\mathbf{C}_k & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{i,k}\delta_{i,\ell}\end{bmatrix}\delta_{k,n} \tag{6.2b}$$

$$\mathbb{E}\left\{\begin{bmatrix}\boldsymbol{q}_k \\ \boldsymbol{\eta}_{i,k}\end{bmatrix}\begin{bmatrix}\boldsymbol{q}_n^{\mathsf{T}} & \boldsymbol{\eta}_{\ell,n}^{\mathsf{T}}\end{bmatrix}\right\} = \begin{bmatrix}\tilde{\mathbf{C}}_k & \mathbf{0} \\ \mathbf{0} & \tilde{\boldsymbol{\Sigma}}_{i,k}\delta_{i,\ell}\end{bmatrix}\delta_{k,n}, \tag{6.2c}$$

where $\delta_{k,n}$ and $\delta_{i,\ell}$ are the Kronecker delta functions which satisfy

$$\delta_{a,b} = \begin{cases} 1, & \text{if } a = b, \\ 0, & \text{otherwise.} \end{cases}$$

If no collaboration is allowed, each node is able to perform its own estimation scheme with the Kalman filter[1] given by [118]

$$\textit{Predict}: \begin{cases} \widehat{\boldsymbol{w}}_{i,k|k-1} &= \mathbf{F}_k \widehat{\boldsymbol{w}}_{i,k-1|k-1} \\ \mathbf{M}_{i,k|k-1} &= \mathbf{F}_k \mathbf{M}_{i,k-1|k-1} \mathbf{F}_k^{\mathsf{H}} + \mathbf{C}_{i,k} \end{cases} \tag{6.3a}$$

$$\textit{Update}: \begin{cases} \mathbf{M}_{i,k|k}^{-1} &= \mathbf{M}_{i,k|k-1}^{-1} + \mathbf{H}_{i,k}^{\mathsf{H}} \boldsymbol{\Sigma}_{i,k}^{-1} \mathbf{H}_{i,k} \\ \widehat{\boldsymbol{w}}_{i,k|k} &= \widehat{\boldsymbol{w}}_{i,k|k-1} + \mathbf{M}_{i,k|k} \mathbf{H}_{i,k}^{\mathsf{H}} \boldsymbol{\Sigma}_{i,k}^{-1} \left[\boldsymbol{y}_{i,k} - \mathbf{H}_{i,k} \widehat{\boldsymbol{w}}_{i,k|k-1} \right], \end{cases} \tag{6.3b}$$

where the state estimate $\widehat{\boldsymbol{w}}_{i,k|k}$ and state error covariance matrix $\mathbf{M}_{i,k|k}$ at each node are initialised as

$$\begin{aligned} \widehat{\boldsymbol{w}}_{i,0|0} &= \mathbb{E}\{\boldsymbol{w}_0\} \\ \mathbf{M}_{i,0|0} &= \mathbb{E}\left\{ (\boldsymbol{w}_0 - \mathbb{E}\{\boldsymbol{w}_0\})(\boldsymbol{w}_0 - \mathbb{E}\{\boldsymbol{w}_0\})^{\mathsf{H}} \right\}. \end{aligned} \tag{6.4}$$

Remark 6.1. *The initialisation procedure in (6.4) is technically based on some a priori information about the mean and covariance of the initial state, $\boldsymbol{w}_0$. However, linear Kalman filters will converge (in the mean square sense) even when initialised arbitrarily [119]. Therefore it is common practice to set $\widehat{\boldsymbol{w}}_{i,0|0} = \mathbf{0}$ and $\mathbf{M}_{i,0|0} = c\mathbf{I}$, where c is a real positive constant, when no prior information is available [120].*

6.2.1 Extension to Nonlinear Models

The linear state-space model in (6.1) can be extended to a nonlinear state-space formulation with the state and measurement evolutions given by

$$\begin{aligned} \boldsymbol{w}_k &= \boldsymbol{f}_k(\boldsymbol{w}_{k-1}) + \boldsymbol{q}_k \\ \boldsymbol{y}_{i,k} &= \boldsymbol{\varphi}_{i,k}(\boldsymbol{w}_k) + \boldsymbol{\eta}_{i,k}, \end{aligned} \tag{6.5}$$

where the nonlinear functions $\boldsymbol{f}_k(\cdot)$ is the state-transition function while $\boldsymbol{\varphi}_{i,k}(\cdot)$ is a general observation function. The statistics of the process noise, $\boldsymbol{q}_k$, and measurement noise, $\boldsymbol{\eta}_{i,k}$, remain as (6.2b) – (6.2c).

The Linearised State Space Equations

The original Kalman filter was derived for linear Gaussian systems and to cater for nonlinear models in (6.5), the complex extended Kalman filter (CEKF) linearises the nonlinear state

[1] A novel derivation of the Kalman in (6.3a) – (6.3b) is included in the Appendix A.

and observation functions by their first order Taylor series expansions (TSE) about the state estimates $\widehat{\boldsymbol{w}}_{i,k|k}$ and $\widehat{\boldsymbol{w}}_{i,k|k-1}$ for each node i, so that [121]

$$\begin{aligned} \boldsymbol{w}_k &\approx \mathbf{F}_{i,k}\boldsymbol{w}_{k-1} + \tilde{\mathbf{F}}_{i,k}\boldsymbol{w}_{k-1}^* + \boldsymbol{u}_{i,k} + \boldsymbol{q}_k \\ \boldsymbol{y}_{i,k} &\approx \mathbf{H}_{i,k}\boldsymbol{w}_k + \tilde{\mathbf{H}}_{i,k}\boldsymbol{w}_k^* + \boldsymbol{\epsilon}_{i,k} + \boldsymbol{\eta}_{i,k} \end{aligned}, \tag{6.6}$$

where the Jacobians of functions $\boldsymbol{f}(\cdot)$ and $\boldsymbol{\varphi}_i(\cdot)$ are defined as

$$\begin{aligned} \mathbf{F}_{i,k} &= \frac{\partial \boldsymbol{f}_k(\boldsymbol{w})}{\partial \boldsymbol{w}^{\mathsf{T}}}\Big|_{\boldsymbol{w}=\widehat{\boldsymbol{w}}_{i,k-1|k-1}} & \tilde{\mathbf{F}}_{i,k} &= \frac{\partial \boldsymbol{f}_k(\boldsymbol{w})}{\partial \boldsymbol{w}^{\mathsf{H}}}\Big|_{\boldsymbol{w}^*=\widehat{\boldsymbol{w}}^*_{i,k-1|k-1}} \\ \mathbf{H}_{i,k} &= \frac{\partial \boldsymbol{\varphi}_{i,k}(\boldsymbol{w})}{\partial \boldsymbol{w}^{\mathsf{T}}}\Big|_{\boldsymbol{w}=\widehat{\boldsymbol{w}}_{i,k|k-1}} & \tilde{\mathbf{H}}_{i,k} &= \frac{\partial \boldsymbol{\varphi}_{i,k}(\boldsymbol{w})}{\partial \boldsymbol{w}^{\mathsf{H}}}\Big|_{\boldsymbol{w}^*=\widehat{\boldsymbol{w}}^*_{i,k|k-1}}. \end{aligned} \tag{6.7}$$

while the variables $\boldsymbol{u}_{i,k}$ and $\boldsymbol{\epsilon}_{i,k}$ in (6.6) represent the Taylor series expansion errors and can be treated as deterministic inputs to the state equations given by

$$\begin{aligned} \boldsymbol{u}_{i,k} &= \boldsymbol{f}(\widehat{\boldsymbol{w}}_{i,k-1|k-1}) - \mathbf{F}_{i,k-1}\widehat{\boldsymbol{w}}_{i,k-1|k-1} - \tilde{\mathbf{F}}_{i,k-1}\widehat{\boldsymbol{w}}^*_{i,k-1|k-1} \\ \boldsymbol{\epsilon}_{i,k} &= \boldsymbol{\varphi}_{i,k}(\widehat{\boldsymbol{w}}_{i,k|k-1}) - \mathbf{H}_{i,k}\widehat{\boldsymbol{w}}_{i,k|k-1} - \tilde{\mathbf{H}}_{i,k}\widehat{\boldsymbol{w}}^*_{i,k|k-1}. \end{aligned}$$

Notice that in (6.6), the linearised equations for a general state-space admits a widely linear form by involving both the state $\boldsymbol{w}_k$ and its conjugate[1] $\boldsymbol{w}_k^*$. For this reason, it shall be useful to always consider the augmented state-space to account for the widely linear nature of the linearisation procedure in (6.6). This is accomplished using an augmented version of the state, $\bar{\boldsymbol{w}}_k = \begin{bmatrix}\boldsymbol{w}_k^{\mathsf{T}} & \boldsymbol{w}_k^{\mathsf{H}}\end{bmatrix}^{\mathsf{T}}$, and observations vectors, $\bar{\boldsymbol{y}}_{i,k} = \begin{bmatrix}\boldsymbol{y}_{i,k}^{\mathsf{T}} & \boldsymbol{y}_{i,k}^{\mathsf{H}}\end{bmatrix}^{\mathsf{T}}$, so that the linearised state space in (6.6) admits an augmented form

$$\begin{aligned} \bar{\boldsymbol{w}}_k &= \bar{\mathbf{F}}_{i,k}\bar{\boldsymbol{w}}_{k-1} + \bar{\boldsymbol{u}}_{i,k-1} + \bar{\boldsymbol{q}}_k \\ \bar{\boldsymbol{y}}_{i,k} &= \bar{\mathbf{H}}_{i,k}\bar{\boldsymbol{w}}_k + \bar{\boldsymbol{\epsilon}}_{i,k} + \bar{\boldsymbol{\eta}}_{i,k}, \end{aligned} \tag{6.8}$$

where the augmented state transition matrix, $\bar{\mathbf{F}}_{i,k}$, observation matrix, $\bar{\mathbf{H}}_{i,k}$, and the Taylor series errors (deterministic inputs), $\bar{\boldsymbol{u}}_{i,k}$, $\bar{\boldsymbol{\epsilon}}_{i,k}$, are given by

$$\bar{\mathbf{F}}_{i,k} = \begin{bmatrix}\mathbf{F}_{i,k} & \tilde{\mathbf{F}}_{i,k} \\ \tilde{\mathbf{F}}^*_{i,k} & \mathbf{F}^*_{i,k}\end{bmatrix}, \quad \bar{\mathbf{H}}_{i,k} = \begin{bmatrix}\mathbf{H}_{i,k} & \tilde{\mathbf{H}}_{i,k} \\ \tilde{\mathbf{H}}^*_{i,k} & \mathbf{H}^*_{i,k}\end{bmatrix}, \quad \bar{\boldsymbol{u}}_{i,k} = \begin{bmatrix}\boldsymbol{u}_{i,k} \\ \boldsymbol{u}^*_{i,k}\end{bmatrix}, \quad \bar{\boldsymbol{\epsilon}}_{i,k} = \begin{bmatrix}\boldsymbol{\epsilon}_{i,k} \\ \boldsymbol{\epsilon}^*_{i,k}\end{bmatrix}.$$

6.2.2 Single-node ACEKF

The complex Kalman filter methodology in (6.3a) – (6.3b) can now be applied to the linearised state-space equation in (6.8), to give the augmented complex extended Kalman filter (ACEKF)

[1] Note that the widely linear model degenerates into a strictly linear one when the state vector is real (i.e. $\boldsymbol{w}_k^* = \boldsymbol{w}_k$) or if the observation and state transition functions are analytic (i.e. $\partial\boldsymbol{\varphi}_i(\cdot)/\partial\boldsymbol{w}_k^{\mathsf{H}} = \mathbf{0}$, or $\partial\boldsymbol{f}(\cdot)/\partial\boldsymbol{w}_k^{\mathsf{H}} = \mathbf{0}$).

algorithm as [121]

$$\textit{Predict}: \begin{cases} \bar{\mathbf{F}}_{i,k} &= \text{Jacobian}\left(\bar{\boldsymbol{f}}_k,\ \widehat{\boldsymbol{w}}_{i,k-1|k-1}\right) \\ \widehat{\boldsymbol{w}}_{i,k|k-1} &= \bar{\boldsymbol{f}}_k(\widehat{\boldsymbol{w}}_{i,k-1|k-1}) \\ \mathbf{M}_{i,k|k-1} &= \bar{\mathbf{F}}_{i,k}\mathbf{M}_{i,k-1|k-1}\bar{\mathbf{F}}_{i,k}^{\mathsf{H}} + \bar{\mathbf{C}}_k \end{cases} \tag{6.9a}$$

$$\textit{Update}: \begin{cases} \bar{\mathbf{H}}_{i,k} &= \text{Jacobian}\left(\bar{\boldsymbol{\varphi}}_{i,k},\ \widehat{\boldsymbol{w}}_{i,k|k-1}\right) \\ \mathbf{M}_{i,k|k}^{-1} &= \mathbf{M}_{i,k|k-1}^{-1} + \bar{\mathbf{H}}_{i,k}^{\mathsf{H}}\bar{\boldsymbol{\Sigma}}_{i,k}^{-1}\bar{\mathbf{H}}_{i,k} \\ \widehat{\boldsymbol{w}}_{i,k|k} &= \widehat{\boldsymbol{w}}_{i,k|k-1} + \mathbf{M}_{i,k|k}\bar{\mathbf{H}}_{i,k}^{\mathsf{H}}\bar{\boldsymbol{\Sigma}}_{i,k}^{-1}\left[\bar{\boldsymbol{y}}_{i,k} - \bar{\boldsymbol{\varphi}}_{i,k}(\widehat{\boldsymbol{w}}_{i,k|k-1})\right] \end{cases}, \tag{6.9b}$$

where the augmented nonlinear functions are $\bar{\boldsymbol{f}}_k = \begin{bmatrix} \boldsymbol{f}_k^{\mathsf{T}} & \boldsymbol{f}_k^{\mathsf{H}} \end{bmatrix}^{\mathsf{T}}$ and $\bar{\boldsymbol{\varphi}}_{i,k} = \begin{bmatrix} \boldsymbol{\varphi}_{i,k}^{\mathsf{T}} & \boldsymbol{\varphi}_{i,k}^{\mathsf{H}} \end{bmatrix}^{\mathsf{T}}$ and the Jacobian$(\cdot)$ operator designates the computation of the Jacobian matrices based on (6.7). The ACEKF is initialised with the augmented state vector and state error covariance matrix as

$$\begin{aligned} \widehat{\boldsymbol{w}}_{i,0|0} &= \mathbb{E}\left\{\bar{\boldsymbol{w}}_0\right\} \\ \mathbf{M}_{i,0|0} &= \mathbb{E}\left\{(\bar{\boldsymbol{w}}_0 - \mathbb{E}\left\{\bar{\boldsymbol{w}}_0\right\})(\bar{\boldsymbol{w}}_0 - \mathbb{E}\left\{\bar{\boldsymbol{w}}_0\right\})^{\mathsf{H}}\right\}. \end{aligned}$$

6.3 Collaboration Schemes

6.3.1 Centralised ACEKF

The most straightforward distributed multi-sensor Kalman filter is the centralised scheme where all nodes transmit their measurements to a fusion centre which assumes the following centralised augmented state-space formulation

$$\bar{\boldsymbol{w}}_k = \bar{\boldsymbol{f}}_k(\bar{\boldsymbol{w}}_{k-1}) + \bar{\boldsymbol{q}}_k \tag{6.10a}$$

$$\bar{\boldsymbol{y}}_{\text{cen},k} = \bar{\boldsymbol{\varphi}}_{\text{cen},k}(\bar{\boldsymbol{w}}_k) + \bar{\boldsymbol{\eta}}_{\text{cen},k}, \tag{6.10b}$$

where the network observation vector, $\bar{\boldsymbol{y}}_{\text{cen},k} \in \mathbb{C}^{NL\times 1}$, in (6.10b) represents all the observations in the network, with the collective measurement and measurement noise vectors given by

$$\bar{\boldsymbol{y}}_{\text{cen},k} = \begin{bmatrix} \bar{\boldsymbol{y}}_{1,k} \\ \bar{\boldsymbol{y}}_{2,k} \\ \vdots \\ \bar{\boldsymbol{y}}_{N,k} \end{bmatrix}, \quad \bar{\boldsymbol{\varphi}}_{\text{cen},k}(\cdot) = \begin{bmatrix} \bar{\boldsymbol{\varphi}}_{1,k}(\cdot) \\ \bar{\boldsymbol{\varphi}}_{2,k}(\cdot) \\ \vdots \\ \bar{\boldsymbol{\varphi}}_{N,k}(\cdot) \end{bmatrix}, \quad \bar{\boldsymbol{\eta}}_{\text{cen},k} = \begin{bmatrix} \bar{\boldsymbol{\eta}}_{1,k} \\ \bar{\boldsymbol{\eta}}_{2,k} \\ \vdots \\ \bar{\boldsymbol{\eta}}_{N,k} \end{bmatrix}, \tag{6.11}$$

while the augmented covariance matrices of the measurement and state noise are respectively $\bar{\boldsymbol{\Sigma}}_{\text{cen},k} \overset{\text{def}}{=} \mathbb{E}\left\{\boldsymbol{\eta}_{\text{cen},k}\boldsymbol{\eta}_{\text{cen},k}^{\mathsf{H}}\right\}$ and $\bar{\mathbf{C}}_k \overset{\text{def}}{=} \mathbb{E}\left\{\bar{\boldsymbol{q}}_k\bar{\boldsymbol{q}}_k^{\mathsf{H}}\right\}$. Following the augmented state-spaced formulation which lead to the ACEKF in (6.9a) – (6.9b), the centralised ACEKF can be derived as

$$\textit{Predict}: \begin{cases} \bar{\mathbf{F}}_k &= \text{Jacobian}\left(\bar{\boldsymbol{f}}_k,\ \widehat{\boldsymbol{w}}_{\text{cen},k-1|k-1}\right) \\ \widehat{\boldsymbol{w}}_{\text{cen},k|k-1} &= \bar{\boldsymbol{f}}_k(\widehat{\boldsymbol{w}}_{\text{cen},k-1|k-1}) \\ \mathbf{M}_{k|k-1} &= \bar{\mathbf{F}}_k \mathbf{M}_{k-1|k-1} \bar{\mathbf{F}}_k^{\mathsf{H}} + \bar{\mathbf{C}}_k \end{cases} \tag{6.12a}$$

$$\textit{Update}: \begin{cases} \bar{\mathbf{H}}_{\text{cen},k} &= \text{Jacobian}\left(\bar{\boldsymbol{\varphi}}_{\text{cen},k},\ \widehat{\boldsymbol{w}}_{\text{cen},k|k-1}\right) \\ \mathbf{M}_{k|k}^{-1} &= \mathbf{M}_{k|k-1}^{-1} + \bar{\mathbf{H}}_{\text{cen},k}^{\mathsf{H}} \bar{\boldsymbol{\Sigma}}_{\text{cen},k}^{-1} \bar{\mathbf{H}}_{\text{cen},k} \\ \widehat{\boldsymbol{w}}_{\text{cen},k|k} &= \widehat{\boldsymbol{w}}_{\text{cen},k|k-1} + \mathbf{M}_{k|k} \bar{\mathbf{H}}_{\text{cen},k}^{\mathsf{H}} \bar{\boldsymbol{\Sigma}}_{\text{cen},k}^{-1} \left[\bar{\boldsymbol{y}}_{\text{cen},k} - \bar{\boldsymbol{\varphi}}_{\text{cen},k}(\widehat{\boldsymbol{w}}_{\text{cen},k|k-1})\right] \end{cases} . \tag{6.12b}$$

Although the centralised ACEKF in (6.12a) – (6.12b), is the optimal state-estimation algorithm given access to all the measurements in the network, it requires excessive communications to a single fusion center. This causes bottlenecks and imposes the risk of the distributed estimation task to a single point of failure. To avoid these issues, a distributed formulation of (6.12a) – (6.12b) is accomplished by restricting nodes in the network to only communicate within the neighbourhood.

6.3.2 Local ACEKF

To remove the dependence on a fusion center, the so-called "Local" ACEKF replicates the centralised ACEKF at each node using only the observations from its neighbourhood. Specifically, the node i observes $\bar{\boldsymbol{y}}_{i_{\text{col}},k} \in \mathbb{C}^{LN_i \times 1}$, where $N_i = |\mathcal{N}_i|$, denotes the number of nodes in neighbourhood of node i, such that the local state equations are

$$\bar{\boldsymbol{w}}_k = \bar{\boldsymbol{f}}_k(\boldsymbol{w}_{k-1}) + \bar{\boldsymbol{q}}_k \tag{6.13a}$$

$$\bar{\boldsymbol{y}}_{i_{\text{col}},k} = \bar{\boldsymbol{\varphi}}_{i_{\text{col}},k}(\boldsymbol{w}_k) + \bar{\boldsymbol{\eta}}_{i_{\text{col}},k}, \tag{6.13b}$$

with the collection of observation variables from the neighbourhood of the node i

$$\bar{\boldsymbol{y}}_{i_{\text{col}},k} = \begin{bmatrix} \boldsymbol{y}_{i_1,k} \\ \boldsymbol{y}_{i_2,k} \\ \vdots \\ \boldsymbol{y}_{i_{N_i},k} \end{bmatrix}, \quad \bar{\boldsymbol{\varphi}}_{i_{\text{col}},k}(\cdot) = \begin{bmatrix} \bar{\boldsymbol{\varphi}}_{i_1,k}(\cdot) \\ \bar{\boldsymbol{\varphi}}_{i_2,k}(\cdot) \\ \vdots \\ \bar{\boldsymbol{\varphi}}_{i_{N_i},k}(\cdot) \end{bmatrix}, \quad \bar{\boldsymbol{\eta}}_{i_{\text{col}},k} = \begin{bmatrix} \bar{\boldsymbol{\eta}}_{i_1,k} \\ \bar{\boldsymbol{\eta}}_{i_2,k} \\ \vdots \\ \bar{\boldsymbol{\eta}}_{i_{N_i},k} \end{bmatrix}. \tag{6.14}$$

The augmented covariance matrix of the collective observation noise vector is then given by

$$\bar{\boldsymbol{\Sigma}}_{i_{\text{col}},k} = \mathbb{E}\left\{\bar{\boldsymbol{\eta}}_{i_{\text{col}},k} \bar{\boldsymbol{\eta}}_{i_{\text{col}},k}^{\mathsf{H}}\right\} = \begin{bmatrix} \bar{\boldsymbol{\Sigma}}_{i_1} & \bar{\boldsymbol{\Sigma}}_{i_1 i_2} & \cdots & \boldsymbol{\Sigma}_{i_1 i_{|\mathcal{N}_i|}} \\ \bar{\boldsymbol{\Sigma}}_{i_2 i_1} & \bar{\boldsymbol{\Sigma}}_{i_2} & \cdots & \bar{\boldsymbol{\Sigma}}_{i_2 i_{N_i}} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\boldsymbol{\Sigma}}_{i_{N_i} i_1} & \bar{\boldsymbol{\Sigma}}_{i_{N_i} i_2} & \cdots & \boldsymbol{\Sigma}_{i_{N_i}} \end{bmatrix}. \tag{6.15}$$

The local ACEKF therefore takes the form

$$\textit{Predict}: \begin{cases} \bar{\mathbf{F}}_{i,k} &= \text{Jacobian}\left(\bar{\boldsymbol{f}}_k,\ \widehat{\boldsymbol{w}}_{i,k-1|k-1}\right) \\ \widehat{\boldsymbol{w}}_{i,k|k-1} &= \bar{\boldsymbol{f}}_k(\widehat{\boldsymbol{w}}_{i,k-1|k-1}) \\ \mathbf{M}_{i,k|k-1} &= \bar{\mathbf{F}}_{i,k}\mathbf{M}_{i,k-1|k-1}\bar{\mathbf{F}}_{i,k}^{\mathsf{H}} + \bar{\mathbf{C}}_k \end{cases} \tag{6.16a}$$

$$\textit{Update}: \begin{cases} \bar{\mathbf{H}}_{i_{\text{col}},k} &= \text{Jacobian}\left(\bar{\boldsymbol{\varphi}}_{i_{\text{col}},k},\ \widehat{\boldsymbol{w}}_{i,k|k-1}\right) \\ \mathbf{M}_{i,k|k}^{-1} &= \mathbf{M}_{i,k|k-1}^{-1} + \bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\boldsymbol{\Sigma}}_{i_{\text{col}},k}^{-1}\bar{\mathbf{H}}_{i_{\text{col}},k} \\ \widehat{\boldsymbol{w}}_{i,k|k} &= \widehat{\boldsymbol{w}}_{i,k|k-1} + \mathbf{M}_{i,k|k}\bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\boldsymbol{\Sigma}}_{i_{\text{col}},k}^{-1}\left[\bar{\boldsymbol{y}}_{i_{\text{col}},k} - \bar{\boldsymbol{\varphi}}_{i_{\text{col}},k}(\widehat{\boldsymbol{w}}_{i,k|k-1})\right] \end{cases}. \tag{6.16b}$$

One of the most widely used assumptions in the distributed Kalman filtering literature is that the observation noise vectors $\bar{\boldsymbol{\eta}}_{i,k}$ are spatially independent, resulting in a block-diagonal local observation noise covariance matrix in (6.15) in the form

$$\bar{\boldsymbol{\Sigma}}_{i_{\text{col}},k} = \begin{bmatrix} \bar{\boldsymbol{\Sigma}}_{i_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \bar{\boldsymbol{\Sigma}}_{i_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \bar{\boldsymbol{\Sigma}}_{i_{N_i}} \end{bmatrix}. \tag{6.16c}$$

This enables the local ACEKF update equations in (6.16b) to be expressed as sums of the individual measurement matrices as

$$\textit{Update}: \begin{cases} \bar{\mathbf{H}}_{\ell,k} &= \text{Jacobian}\left(\bar{\boldsymbol{\varphi}}_{\ell,k},\ \widehat{\boldsymbol{w}}_{i,k|k-1}\right), \quad \ell = i_1, \ldots, i_{N_i} \\ \mathbf{M}_{i,k|k}^{-1} &= \mathbf{M}_{i,k|k-1}^{-1} + \sum_{\ell\in\mathcal{N}_i}\bar{\mathbf{H}}_{\ell,k}^{\mathsf{H}}\bar{\boldsymbol{\Sigma}}_{\ell,k}^{-1}\bar{\mathbf{H}}_{\ell,k} \\ \widehat{\boldsymbol{w}}_{i,k|k} &= \widehat{\boldsymbol{w}}_{i,k|k-1} + \mathbf{M}_{i,k|k}\sum_{\ell\in\mathcal{N}_i}\bar{\mathbf{H}}_{\ell,k}^{\mathsf{H}}\bar{\boldsymbol{\Sigma}}_{\ell,k}^{-1}\left[\bar{\boldsymbol{y}}_{\ell,k} - \bar{\boldsymbol{\varphi}}_{\ell,k}(\widehat{\boldsymbol{w}}_{i,k|k-1})\right] \end{cases}. \tag{6.16d}$$

6.4 Proposed: Diffusion Augmented Complex Extended Kalman Filter (D-ACEKF)

In Chapter 5, the diffusion scheme was introduced as a method of diffusing the ACLMS weight (state) estimates $\widehat{\boldsymbol{w}}_{i,k}$ throughout the network. The diffusion step was introduced as a weighted averaging scheme and was derived using the the stochastic gradient technique to minimise the cost function originally defined in (5.19) as

$$J_{\text{AV}}^{(i,k)}(\boldsymbol{\psi}) = \sum_{\ell\in\mathcal{N}_i} b_{\ell i}\|\boldsymbol{\psi}_{\ell,k} - \boldsymbol{\psi}\|^2,$$

which minimises the difference between the intermediate state estimates, $\boldsymbol{\psi}_{i,k}$, within the network, thereby promoting consensus between the nodes. The diffusion strategy can also be

derived using the Kalman filtering formulation by observing the relationship between the centralised state estimate, $\widehat{\boldsymbol{w}}_{\text{cen},k|k}$ in (6.12b) and the local EKF state estimates $\widehat{\boldsymbol{w}}_{i,k|k}$ in (6.9b) which are related through their respective weight error covariance matrices, $\mathbf{M}_{i,k|k}$, as

$$\mathbf{M}_{k|k}^{-1}\widehat{\boldsymbol{w}}_{\text{cen},k|k} = \sum_{\ell=1}^{N} \mathbf{M}_{\ell,k|k}^{-1}\widehat{\boldsymbol{w}}_{\ell,k|k}. \tag{6.17}$$

Pre-multiplying both sides of (6.17) with the centralised state error covariance matrix $\mathbf{M}_{k|k}$ yields

$$\widehat{\boldsymbol{w}}_{\text{cen},k|k} = \sum_{\ell=1}^{N} \mathbf{M}_{k|k}\mathbf{M}_{\ell,k|k}^{-1}\widehat{\boldsymbol{w}}_{\ell,k|k}, \tag{6.18}$$

which implies that the centralised solution $\widehat{\boldsymbol{w}}_{\text{cen},k|k}$ in (6.12b) can be obtained by a weighted averaging operation of the individual state estimates $\widehat{\boldsymbol{w}}_{\ell,k|k}$ with weighting matrices $\mathbf{M}_{k|k}\mathbf{M}_{\ell,k|k}^{-1}$. Since the centralised state error covariance matrix, $\mathbf{M}_{k|k}$, is related to its single-node counterparts $\mathbf{M}_{\ell,k|k}$ as

$$\mathbf{M}_{k|k}^{-1} = \sum_{\ell=1}^{N} \mathbf{M}_{\ell,k|k}^{-1}, \tag{6.19}$$

the weighting matrices satisfy the power-invariance condition, such that

$$\sum_{\ell=1}^{N} \mathbf{M}_{k|k}\mathbf{M}_{\ell,k|k}^{-1} = \mathbf{I}.$$

To obtain the effect of the fusion rule in (6.18) without requiring the communication (sharing) of the state error covariance matrices $\mathbf{M}_{\ell,k|k}$, we shall replace the weighting matrices $\mathbf{M}_{k|k}\mathbf{M}_{\ell,k|k}^{-1}$ with combination coefficients $a_{\ell,\text{cen}}$ such that the fusion rule in (6.18) is approximately given by

$$\widehat{\boldsymbol{w}}_{\text{cen},k|k} \approx \sum_{\ell=1}^{N} a_{\ell,\text{cen}}\widehat{\boldsymbol{w}}_{\ell,k|k}, \tag{6.20}$$

where the weighting coefficients $a_{\ell,\text{cen}}$ are only restricted to satisfy the power-invariance condition $\sum_{\ell} a_{\ell,\text{cen}} = 1$. The step in (6.20) resembles the diffusion step introduced in Chapter 5, and can now be incorporated with the local ACEKF in (6.16b) – (6.16a), to give the Diffusion

ACEKF in the form

$$\bar{\mathbf{F}}_{i,k} = \text{Jacobian}\left(\bar{\boldsymbol{f}}_k,\ \widehat{\boldsymbol{w}}_{i,k-1|k-1}\right) \tag{6.21a}$$
$$\widehat{\boldsymbol{w}}_{i,k|k-1} = \bar{\boldsymbol{f}}_k(\widehat{\boldsymbol{w}}_{i,k-1|k-1}) \tag{6.21b}$$
$$\mathbf{M}_{i,k|k-1} = \bar{\mathbf{F}}_{i,k}\mathbf{M}_{i,k-1|k-1}\bar{\mathbf{F}}_{i,k}^{\mathsf{H}} + \bar{\mathbf{C}}_k \tag{6.21c}$$
$$\bar{\mathbf{H}}_{i_{\text{col}},k} = \text{Jacobian}\left(\bar{\boldsymbol{\varphi}}_{i_{\text{col}},k},\ \widehat{\boldsymbol{w}}_{i,k|k-1}\right) \tag{6.21d}$$
$$\mathbf{M}_{i,k|k}^{-1} = \mathbf{M}_{i,k|k-1}^{-1} + \bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\boldsymbol{\Sigma}}_{i_{\text{col}},k}^{-1}\bar{\mathbf{H}}_{i_{\text{col}},k} \tag{6.21e}$$
$$\boldsymbol{\psi}_{i,k|k} = \widehat{\boldsymbol{w}}_{i,k|k-1} + \mathbf{M}_{i,k|k}\bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\boldsymbol{\Sigma}}_{i_{\text{col}},k}^{-1}\left[\bar{\boldsymbol{y}}_{i_{\text{col}},k} - \bar{\boldsymbol{\varphi}}_{i_{\text{col}},k}(\widehat{\boldsymbol{w}}_{i,k|k-1})\right] \tag{6.21f}$$
$$\widehat{\boldsymbol{w}}_{i,k|k} = \sum_{\ell\in\mathcal{N}_i} a_{\ell i}\boldsymbol{\psi}_{\ell,k|k}. \tag{6.21g}$$

Contribution: D-ACEKF
The distinguishing feature of the proposed class of distributed Kalman filters in (6.21a) – (6.21g) is that we generalise the diffusion strategy in [96] by equipping the model of with state and noise models that do not impose any restrictions on: i) the correlation properties of the cross-nodal observation noises, ii) the signal and noise circularity at different nodes, iii) the widely linear nature of the underpinning system. This also allows distributed Kalman filtering algorithms proposed in [96, 113, 122] to be used in wider application scenarios.

For more information on how the widely linear state-space model is able to incorporate *a priori* information on the circularity of the measurement and state noise process, please refer to [118]. An alternative version of the D-ACEKF is given by the so-called "one-step" Kalman filter whereby the standard EKF update equations in (6.21a) – (6.21f) are modified to

$$\bar{\mathbf{F}}_{i,k} = \text{Jacobian}\left(\bar{\boldsymbol{f}}_k,\ \widehat{\boldsymbol{w}}_{i,k-1}\right) \tag{6.22a}$$
$$\bar{\mathbf{H}}_{i_{\text{col}},k} = \text{Jacobian}\left(\bar{\boldsymbol{\varphi}}_{i_{\text{col}},k},\ \widehat{\boldsymbol{w}}_{i,k-1}\right) \tag{6.22b}$$
$$\mathbf{G}_{i,k} = \mathbf{M}_{i,k-1}\bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\left(\bar{\boldsymbol{\Sigma}}_{i_{\text{col}},k} + \bar{\mathbf{H}}_{i_{\text{col}},k}\mathbf{M}_{i,k-1}\bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\right)^{-1} \tag{6.22c}$$
$$\boldsymbol{\psi}_{i,k} = \boldsymbol{f}_k(\widehat{\boldsymbol{w}}_{i,k-1}) + \bar{\mathbf{F}}_{i,k}\mathbf{G}_{i,k}\left[\bar{\boldsymbol{y}}_{i_{\text{col}},k} - \bar{\boldsymbol{\varphi}}_{i_{\text{col}},k}(\widehat{\boldsymbol{w}}_{i,k-1})\right] \tag{6.22d}$$
$$\mathbf{M}_{i,k} = \bar{\mathbf{F}}_{i,k}\left(\mathbf{M}_{i,k-1} - \mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k}\mathbf{M}_{i,k-1}\right)\bar{\mathbf{F}}_{i,k}^{\mathsf{H}} + \bar{\mathbf{C}}_k \tag{6.22e}$$
$$\widehat{\boldsymbol{w}}_{i,k} = \sum_{\ell\in\mathcal{N}_i} a_{\ell i}\boldsymbol{\psi}_{\ell,k}. \tag{6.22f}$$

The one-step D-ACEKF filter in (6.22a) – (6.22f) has the same convergence property as the D-ACEKF in (6.21a) – (6.21g), but has the benefit of having less cumbersome notation in the convergence analysis [120].

Remark 6.2. *Unlike the linear Kalman filter, the state vector estimate for the EKF has to be initialised to a value which is "close" to the true state* $\bar{\boldsymbol{w}}_0$. *This can be accomplished with* a priori *knowledge about the system. For example, in frequency estimation tasks, the state (frequency) can be initialised to the nominal frequency of the grid of* $\widehat{\boldsymbol{w}}_{i,0} = \boldsymbol{w}_0 = 50$ *Hz, since the grid frequency is known to stay within a small interval of 49 Hz and 51 Hz. The small initial state estimation error shall be investigated in the*

Algorithm 4. Diffusion Augmented Complex Extended Kalman Filter (D-ACEKF)

Initialise each node with: $\widehat{\boldsymbol{w}}_{i,0|0} = \boldsymbol{w}_0$ and $\mathbf{M}_{i,0|0} = c\mathbf{I}$.
At each time instant $k > 0$:

1: **for** Nodes $i = \{1, 2, \ldots, N\}$ **do**
2: $\quad \bar{\mathbf{F}}_{i,k} = \mathsf{Jacobian}\left(\bar{\boldsymbol{f}}_k,\ \widehat{\boldsymbol{w}}_{i,k-1|k-1}\right)$
3: $\quad \widehat{\boldsymbol{w}}_{i,k|k-1} = \bar{\boldsymbol{f}}_k(\widehat{\boldsymbol{w}}_{i,k-1|k-1})$
4: $\quad \mathbf{M}_{i,k|k-1} = \bar{\mathbf{F}}_{i,k}\mathbf{M}_{i,k-1|k-1}\bar{\mathbf{F}}_{i,k}^{\mathsf{H}} + \bar{\mathbf{C}}_k$
5: $\quad \bar{\mathbf{H}}_{i_{\text{col}},k} = \mathsf{Jacobian}\left(\bar{\boldsymbol{\varphi}}_{i_{\text{col}},k},\ \widehat{\boldsymbol{w}}_{i,k|k-1}\right)$
6: $\quad \mathbf{M}_{i,k|k}^{-1} = \mathbf{M}_{i,k|k-1}^{-1} + \bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\boldsymbol{\Sigma}}_{i_{\text{col}},k}^{-1}\bar{\mathbf{H}}_{i_{\text{col}},k}$
7: $\quad \boldsymbol{\psi}_{i,k|k} = \widehat{\boldsymbol{w}}_{i,k|k-1} + \mathbf{M}_{i,k|k}\bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\boldsymbol{\Sigma}}_{i_{\text{col}},k}^{-1}\left[\bar{\boldsymbol{y}}_{i_{\text{col}},k} - \bar{\boldsymbol{\varphi}}_{i_{\text{col}},k}(\widehat{\boldsymbol{w}}_{i,k|k-1})\right]$
8: $\quad$ Diffuse the states from the network:
$\quad\quad \widehat{\boldsymbol{w}}_{i,k|k} = \sum_{\ell\in\mathcal{N}_i} a_{\ell i}\boldsymbol{\psi}_{\ell,k|k}$
9: **end for**

proof of Theorem 6.1. The state error covariance matrix can be initialised as $\mathbf{M}_{i,0|0} = c\mathbf{I}$*, where* c *is positive constant [120].*

6.5 Performance Analysis

First consider subtracting the true state $\bar{\boldsymbol{w}}_k$ from both sides of the diffusion step in (6.22f),

$$\tilde{\boldsymbol{w}}_{i,k} = \sum_{\ell\in\mathcal{N}_i} a_{\ell i}\tilde{\boldsymbol{\psi}}_{i,k}, \tag{6.23}$$

where the state error vectors are $\tilde{\boldsymbol{w}}_{i,k} \overset{\text{def}}{=} \bar{\boldsymbol{w}}_k - \widehat{\boldsymbol{w}}_{i,k}$ and $\tilde{\boldsymbol{\psi}}_{i,k} \overset{\text{def}}{=} \bar{\boldsymbol{w}}_k - \boldsymbol{\psi}_{i,k}$. Taking the squared norms of both sides of (6.23) and applying the statistical expectation yields

$$\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_{i,k}\|^2\right\} = \sum_{\ell} a_{\ell i}^2\mathbb{E}\left\{\|\tilde{\boldsymbol{\psi}}_{i,k}\|^2\right\} + \sum_{m}\sum_{\ell\neq m} a_{\ell i}a_{mi}\mathbb{E}\left\{\tilde{\boldsymbol{\psi}}_{m,k}^{\mathsf{H}}\tilde{\boldsymbol{\psi}}_{\ell,k}\right\}. \tag{6.24}$$

Denoting $\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_{\max,k}\|^2\right\} \overset{\text{def}}{=} \max\left(\mathbb{E}\left\{\|\tilde{\boldsymbol{\psi}}_{1,k}\|^2\right\}, \ldots, \mathbb{E}\left\{\|\tilde{\boldsymbol{\psi}}_{N,k}\|^2\right\}\right)$, yields the inequality

$$\begin{aligned}\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_{i,k}\|^2\right\} &\leq \left(\sum_{\ell} a_{\ell i}^2 + \sum_{m}\sum_{\ell\neq m} a_{\ell i}a_{mi}\right)\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_{\max,k}\|^2\right\} \\ &= \left(\sum_{\ell} a_{\ell i}\right)^2\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_{\max,k}\|^2\right\} = \mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_{\max,k}\|^2\right\},\end{aligned} \tag{6.25}$$

since $\sum_{\ell} a_{\ell i} = 1$.

Remark 6.3. *The inequality in (6.25) states that under the diffusion scheme, the state estimation error at each node is bounded by the worst case state estimation error in the network. Therefore, it shall be sufficient to prove the mean square converge of each node in the D-ACEKF to claim the mean square convergence of the network.*

Now, consider the state-space model in (6.13a) – (6.13b) with time-invariant state-space functions $\bar{\boldsymbol{f}}$ and $\bar{\boldsymbol{\varphi}}_{i_{\text{col}}}$, such that

$$\bar{\boldsymbol{w}}_k = \bar{\boldsymbol{f}}(\bar{\boldsymbol{w}}_{k-1}) + \bar{\boldsymbol{q}}_k \tag{6.26a}$$

$$\bar{\boldsymbol{y}}_{i_{\text{col}},k} = \bar{\boldsymbol{\varphi}}_{i_{\text{col}}}(\bar{\boldsymbol{w}}_{k-1}) + \bar{\boldsymbol{\eta}}_{i_{\text{col}},k}, \tag{6.26b}$$

where the time index of the state, $\bar{\boldsymbol{w}}$, in the observation equation (6.26b) is modified to suit "one-step" D-ACEKF. Recall the D-ACEKF state estimate update step in (6.22d),

$$\widehat{\boldsymbol{w}}_{i,k} = \bar{\boldsymbol{f}}(\widehat{\boldsymbol{w}}_{i,k-1}) + \bar{\mathbf{F}}_{i,k}\mathbf{G}_{i,k}\Big[\bar{\boldsymbol{y}}_{i_{\text{col}},k} - \bar{\boldsymbol{\varphi}}_{i_{\text{col}}}(\widehat{\boldsymbol{w}}_{i,k-1})\Big]. \tag{6.27}$$

In light of Remark 6.3, we have omitted the diffusion step since it is sufficient to bound the mean square error behaviour of each individual node. Next, we express the nonlinear estimation errors in terms of the Taylor series expansion errors as

$$\bar{\boldsymbol{f}}(\bar{\boldsymbol{w}}_{k-1}) - \bar{\boldsymbol{f}}(\widehat{\boldsymbol{w}}_{i,k-1}) = \bar{\mathbf{F}}_{i,k}(\tilde{\boldsymbol{w}}_{i,k-1}) + \boldsymbol{\xi}(\tilde{\boldsymbol{w}}_{i,k-1}) \tag{6.28}$$

$$\bar{\boldsymbol{\varphi}}_{i_{\text{col}}}(\bar{\boldsymbol{w}}_{k-1}) - \bar{\boldsymbol{\varphi}}_{i_{\text{col}},k}(\widehat{\boldsymbol{w}}_{i,k-1}) = \bar{\mathbf{H}}_{i_{\text{col}}}(\tilde{\boldsymbol{w}}_{i,k-1}) + \boldsymbol{\chi}(\tilde{\boldsymbol{w}}_{i,k-1}), \tag{6.29}$$

where $\boldsymbol{\xi}(\tilde{\boldsymbol{w}}_{i,k-1})$ and $\boldsymbol{\chi}(\tilde{\boldsymbol{w}}_{i,k-1})$ refer respectively to the linearization errors in the state evolution and observation functions. Subtracting the true state vector $\bar{\boldsymbol{w}}_k$ from the one-step Kalman filter in (6.27) gives

$$\bar{\boldsymbol{w}}_k - \widehat{\boldsymbol{w}}_{i,k} = \bar{\boldsymbol{f}}(\bar{\boldsymbol{w}}_{k-1}) - \bar{\boldsymbol{f}}(\widehat{\boldsymbol{w}}_{i,k-1}) - \bar{\mathbf{F}}_{i,k}\mathbf{G}_{i,k}\Big[\bar{\boldsymbol{y}}_{i_{\text{col}},k} - \bar{\boldsymbol{\varphi}}_{i_{\text{col}}}(\widehat{\boldsymbol{w}}_{i,k-1})\Big] + \bar{\boldsymbol{q}}_k. \tag{6.30}$$

Using the relationships in (6.28) and (6.29) gives,

$$\tilde{\boldsymbol{w}}_{i,k} = \bar{\mathbf{F}}_{i,k}\tilde{\boldsymbol{w}}_{i,k-1} + \boldsymbol{\xi}(\tilde{\boldsymbol{w}}_{i,k-1}) - \bar{\mathbf{F}}_{i,k}\mathbf{G}_{i,k}\Big[\bar{\mathbf{H}}_{i_{\text{col}},k}\tilde{\boldsymbol{w}}_{i,k-1} + \boldsymbol{\chi}(\tilde{\boldsymbol{w}}_{i,k-1}) + \bar{\boldsymbol{\eta}}_{i_{\text{col}},k}\Big] + \bar{\boldsymbol{q}}_k \tag{6.31}$$

which can be simplified into

$$\tilde{\boldsymbol{w}}_{i,k} = \mathbf{B}_{i,k}\tilde{\boldsymbol{w}}_{i,k-1} + \boldsymbol{r}_{i,k-1} + \boldsymbol{n}_{i,k} \tag{6.32}$$

with

$$\mathbf{B}_{i,k} = \Big[\bar{\mathbf{F}}_{i,k} - \bar{\mathbf{F}}_{i,k}\mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k}\Big] \tag{6.33}$$

$$\boldsymbol{r}_{i,k-1} = \boldsymbol{\xi}(\tilde{\boldsymbol{w}}_{i,k-1}) - \bar{\mathbf{F}}_{i,k}\mathbf{G}_{i,k}\boldsymbol{\chi}(\tilde{\boldsymbol{w}}_{i,k-1}) \tag{6.34}$$

$$\boldsymbol{n}_{i,k} = \bar{\boldsymbol{q}}_k - \bar{\mathbf{F}}_{i,k}\mathbf{G}_{i,k}\bar{\boldsymbol{\eta}}_{i_{\text{col}},k}. \tag{6.35}$$

The recursion for the state estimation error in (6.32) is different to that of the linear Kalman filter due to the presence of the nonlinear error term, $\boldsymbol{r}_{i,k-1}$. Next, we shall present a theorem on the mean square stability of the D-ACEKF which is based on the stochastic stability theorem of the extended Kalman filter which was first proposed in [123].

Theorem 6.1 (D-ACEKF Mean Square Stability). *Consider the nonlinear state space model in (6.26a) –(6.26b) and D-ACEKF filter in (6.22a) – (6.22f). Suppose the following assumptions hold:*

A.1) Each element of the functions $\bar{\boldsymbol{f}}$ and $\bar{\boldsymbol{\varphi}}_{i_{col}}$ is smooth.

A.2) There exists real scalars $\bar{f}, \bar{h}, \underline{m}, \bar{m}, \underline{s}, \underline{c}, \delta > 0$ such the the following bounds are fulfilled almost surely (a.s.) for every node i and time instant, $k \geq 0$:

$$\|\bar{\mathbf{F}}_{i,k}\| \leq \bar{f}, \qquad \|\bar{\mathbf{H}}_{i,k}\| \leq \bar{h} \quad (a.s.) \tag{6.36a}$$

$$\underline{m}\mathbf{I} \preceq \mathbf{M}_{i,k} \preceq \bar{m}\mathbf{I} \quad (a.s.) \tag{6.36b}$$

$$\underline{c}\mathbf{I} \preceq \bar{\mathbf{C}}_k \preceq \delta\mathbf{I}, \qquad \underline{s}\mathbf{I} \preceq \boldsymbol{\Sigma}_{i_{col},k} \preceq \delta\mathbf{I}. \tag{6.36c}$$

A.3) $\bar{\mathbf{F}}_{i,k}$ is invertible for every i and $k \geq 0$.

A.4) There exists real scalars, $\epsilon_\xi, \epsilon_\chi, \kappa_\xi, \kappa_\chi > 0$ such that the nonlinear functions are bounded by

$$\|\boldsymbol{\xi}(\tilde{\boldsymbol{w}}_{i,k})\| \leq \kappa_\xi \|\tilde{\boldsymbol{w}}_{i,k}\|^2 \quad \textit{with } \|\tilde{\boldsymbol{w}}_{i,k}\| \leq \epsilon_\xi \quad (a.s.) \tag{6.36d}$$

$$\|\boldsymbol{\chi}(\tilde{\boldsymbol{w}}_{i,k})\| \leq \kappa_\chi \|\tilde{\boldsymbol{w}}_{i,k}\|^2 \quad \textit{with } \|\tilde{\boldsymbol{w}}_{i,k}\| \leq \epsilon_\chi \quad (a.s.) \tag{6.36e}$$

for every $\tilde{\boldsymbol{w}}_k \in \mathbb{C}^{M\times 1}$.

Then, the D-ACEKF at each node is stable in the mean square when the initial error, $\|\tilde{\boldsymbol{w}}_0\| \leq \epsilon$ with $\epsilon > 0$ and $0 < \vartheta < 1$, that is

$$\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_{i,k}\|^2\right\} \leq \frac{\bar{m}}{\underline{m}}(1-\vartheta)^k \mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_{i,0}\|^2\right\} + \frac{\kappa_{noise}\delta}{\underline{m}\vartheta}, \tag{6.36f}$$

where $\kappa_{noise} = \dfrac{M}{\underline{m}} + \dfrac{\bar{f}^2\bar{h}^2\bar{m}^2 L}{\underline{m}\underline{s}^2}$ and $\vartheta = \dfrac{1}{2} - \dfrac{1}{2}\left[1 + \dfrac{\underline{c}}{\bar{m}(\bar{f} + \bar{f}\bar{m}\bar{h}^2/\underline{c})^2}\right]^{-1}$ with M and L denoting the dimensions of the state and observation[1] vectors respectively and the initial estimation error bound, $\epsilon = \min\left(\epsilon_\xi, \epsilon_\chi, \dfrac{\vartheta}{2\bar{m}\kappa_{nonl}}\right)$ where κ_{nonl} is defined in Lemma 6.4 with $\kappa_{nonl} = \mathcal{O}(\kappa^2)$, $\kappa = \max(\kappa_\xi, \kappa_\chi)$.

A few remarks about the various assumptions in Theorem 6.1.

1. The boundedness of the Jacobian matrices in (6.36a) follow from the smoothness assumption of the functions $\bar{\boldsymbol{f}}$ and $\bar{\boldsymbol{\varphi}}_{i_{\text{col}}}$.
2. The bounds on the state error covariance matrix $\mathbf{M}_{i,k}$ in (6.36b) imply the observability and controllability of the linearised system, that is, the observability of the pair $(\mathbf{F}_{i,k}, \mathbf{H}_{i,k})$ and controllability of the pair $(\mathbf{F}_{i,k}, \mathbf{C}_k^{-\frac{1}{2}})$ [124, 125].
3. The bounds on the noise covariance matrices $\bar{\mathbf{C}}_k$ and $\boldsymbol{\Sigma}_{i_{\text{col}},k}$ in (6.36c) imply that the matrices are positive-definite and yet the level of noise in the system is not "too large" [123].
4. The nonlinear error bounds in (6.36d) – (6.36e) are related to the smoothness of the state transition and observation functions and imply that either the functions are "weakly" nonlinear or, the initial state estimate is "close" to the true state [126].

[1] The dimensions of the original observation vector as defined in (6.13b) is LN_i but is shortened to L here for notational compactness.

5. The proof also relies upon the initial state estimation error bound, ϵ which is inversely proportional to square of the constants $\kappa = \max(\kappa_\xi, \kappa_\chi)$. In other words, the more "non-linear" the function is, the larger the values of κ, hence the initial estimate, $\widehat{\boldsymbol{w}}_0$ should be closer to the true value, $\boldsymbol{w}_0$, for the D-ACEKF be mean square stable.

We shall first consider several Lemmas which will be useful in the proofs of Theorem 6.1. Note that the matrix bounds in this proof that involve stochastic matrices are assumed to hold almost surely.

Lemma 6.2. *Assume that there exists a stochastic process, $V_k(\tilde{\boldsymbol{w}}_k)$, and real numbers $\underline{v}, \bar{v}, \beta \geq 0$ and $0 \leq \vartheta < 1$, which satisfies*

$$\underline{v}\|\tilde{\boldsymbol{w}}_{i,k}\|^2 \leq V_k(\tilde{\boldsymbol{w}}_{i,k}) \leq \bar{v}\|\tilde{\boldsymbol{w}}_{i,k}\|^2 \ \ (a.s.) \quad \textit{and} \tag{6.37}$$

$$\mathbb{E}\left\{V_k(\tilde{\boldsymbol{w}}_{i,k})|\tilde{\boldsymbol{w}}_{i,k-1}\right\} - V_{k-1}(\tilde{\boldsymbol{w}}_{i,k-1}) \leq \beta - \vartheta V_{k-1}(\tilde{\boldsymbol{w}}_{i,k-1}) \ \ (a.s.), \tag{6.38}$$

for every solution of (6.32). Then, the stochastic process $\tilde{\boldsymbol{w}}_k$ is exponentially bounded in the mean square, with

$$\mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_{i,k}\|^2\right\} \leq \frac{\bar{v}}{\underline{v}}(1-\vartheta)^k \mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_{i,0}\|^2\right\} + \frac{\beta}{\underline{v}\vartheta}. \tag{6.39}$$

Proof. We refer the reader to [123, Lem. 2.1][119, Lem. 3.8] for the proof of this Lemma. □

Lemma 6.3. *Under the assumptions of Theorem 6.1, there exists $0 < \vartheta < 1$, such that*

$$\mathbf{B}_{i,k}^{\mathsf{H}} \mathbf{M}_{i,k} \mathbf{B}_{i,k} \preceq (1-\vartheta)\mathbf{M}_{i,k-1} \quad (a.s.)$$

for $k \geq 0$ and $\vartheta = 1 - \left[1 + \dfrac{\underline{c}}{\bar{m}(\bar{f} + \bar{f}\bar{m}\bar{h}^2/\underline{c})^2}\right]^{-1}$.

Proof.
The proof of this Lemma follows the proof from [123, Lem 3.1]. We begin by bounding several matrices in the algorithm. First, since $\bar{\mathbf{H}}_{i_{\text{col}},k}\mathbf{M}_{i,k-1}\bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}} \succeq 0$, the following bound holds,

$$\left\|\left(\bar{\boldsymbol{\Sigma}}_{i_{\text{col}},k} + \bar{\mathbf{H}}_{i_{\text{col}},k}\mathbf{M}_{i,k-1}\bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\right)^{-1}\right\| \leq \left\|\bar{\boldsymbol{\Sigma}}_{i_{\text{col}},k}^{-1}\right\| \leq 1/\underline{s}. \tag{6.40}$$

Furthermore, the norm of the Kalman gain matrix, $\mathbf{G}_{i,k} = \mathbf{M}_{i,k-1}\bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\left(\bar{\boldsymbol{\Sigma}}_{i_{\text{col}},k} + \bar{\mathbf{H}}_{i_{\text{col}},k}\mathbf{M}_{i,k-1}\bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\right)^{-1}$ can be bounded by

$$\left\|\mathbf{G}_{i,k}\right\| \leq \left\|\mathbf{M}_{i,k-1}\right\|\left\|\bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\right\|\left\|\bar{\boldsymbol{\Sigma}}_{i_{\text{col}},k}^{-1}\right\| \leq \bar{m}\bar{h}/\underline{s}. \tag{6.41}$$

Next, the norm of the matrix $\mathbf{B}_{i,k} = \bar{\mathbf{F}}_{i,k} - \bar{\mathbf{F}}_{i,k}\mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k}$ is therefore bounded by

$$\left\|\mathbf{B}_{i,k}\right\| \leq \left\|\bar{\mathbf{F}}_{i,k}\right\| + \left\|\bar{\mathbf{F}}_{i,k}\right\|\left\|\mathbf{G}_{i,k}\right\|\left\|\bar{\mathbf{H}}_{i_{\text{col}},k}\right\| \leq \bar{f} + \bar{f}\bar{m}\bar{h}^2/\underline{s}. \tag{6.42}$$

The proof of Lemma 6.3 is derived by re-expressing the state error covariance matrix update in (6.22e) as

$$\begin{aligned}\mathbf{M}_{i,k} &= \bar{\mathbf{F}}_{i,k}\left(\mathbf{M}_{i,k-1} - \mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k}\mathbf{M}_{i,k-1}\right)\bar{\mathbf{F}}_{i,k}^{\mathsf{H}} + \bar{\mathbf{C}}_k \\ &= \mathbf{B}_{i,k}\mathbf{M}_{i,k-1}\mathbf{B}_{i,k}^{\mathsf{H}} + \bar{\mathbf{F}}_{i,k}\mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k}\mathbf{M}_{i,k-1}\mathbf{B}_{i,k}^{\mathsf{H}} + \bar{\mathbf{C}}_k \qquad (6.43)\\ &\succeq \mathbf{B}_{i,k}\mathbf{M}_{i,k-1}\mathbf{B}_{i,k}^{\mathsf{H}} + \bar{\mathbf{C}}_k, \qquad (6.44)\end{aligned}$$

where, $\bar{\mathbf{F}}_{i,k}\mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k}\mathbf{M}_{i,k-1}\mathbf{B}_{i,k}^{\mathsf{H}} \succeq 0$. This can be shown by expressing

$$\bar{\mathbf{F}}_{i,k}\mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k}\mathbf{M}_{i,k-1}\mathbf{B}_{i,k}^{\mathsf{H}} = \bar{\mathbf{F}}_{i,k}\mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k}\mathbf{M}_{i,k-1}(\mathbf{I} - \mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k})^{\mathsf{H}}\bar{\mathbf{F}}_{i,k}^{\mathsf{H}}. \qquad (6.45)$$

We proceed to show that expression,

$$\mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k}\mathbf{M}_{i,k-1}(\mathbf{I} - \mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k})^{\mathsf{H}} \succeq 0$$

by first using the matrix inversion lemma (Woodbury identity [46]) to state that

$$\left[\mathbf{M}_{i,k-1}^{-1} + \bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\mathbf{\Sigma}}_{i_{\text{col}},k}^{-1}\bar{\mathbf{H}}_{i_{\text{col}},k}\right]^{-1} = \mathbf{M}_{i,k-1} - \mathbf{M}_{i,k-1}\bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\left(\bar{\mathbf{\Sigma}}_{i_{\text{col}},k} + \bar{\mathbf{H}}_{i_{\text{col}},k}\mathbf{M}_{i,k-1}\bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\right)^{-1}\bar{\mathbf{H}}_{i_{\text{col}},k}\mathbf{M}_{i,k-1} \qquad (6.46)$$

Noting that $\mathbf{G}_{i,k} = \mathbf{M}_{i,k-1}\bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\left(\bar{\mathbf{\Sigma}}_{i_{\text{col}},k} + \bar{\mathbf{H}}_{i_{\text{col}},k}\mathbf{M}_{i,k-1}\bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\right)^{-1}$, allows (6.46) to be expressed as

$$\begin{aligned}\left[\mathbf{M}_{i,k-1}^{-1} + \bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\mathbf{\Sigma}}_{i_{\text{col}},k}^{-1}\bar{\mathbf{H}}_{i_{\text{col}},k}\right]^{-1} &= \mathbf{M}_{i,k-1} - \mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k}\mathbf{M}_{i,k-1} \\ &= \left(\mathbf{I} - \mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k}\right)\mathbf{M}_{i,k-1}. \qquad (6.47)\end{aligned}$$

Since $\mathbf{M}_{i,k-1}^{\mathsf{H}} = \mathbf{M}_{i,k-1}$, taking the Hermitian of (6.47) gives

$$\mathbf{M}_{i,k-1}\left(\mathbf{I} - \mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k}\right)^{\mathsf{H}} = \left[\mathbf{M}_{i,k-1}^{-1} + \bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\mathbf{\Sigma}}_{i_{\text{col}},k}^{-1}\bar{\mathbf{H}}_{i_{\text{col}},k}\right]^{-\mathsf{H}}. \qquad (6.48)$$

Next, the term $\mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k}$, can be re-expressed by using (6.47) whereby

$$\begin{aligned}\left(\mathbf{I} - \mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k}\right)\mathbf{M}_{i,k-1} &= \left[\mathbf{M}_{i,k-1}^{-1} + \bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\mathbf{\Sigma}}_{i_{\text{col}},k}^{-1}\bar{\mathbf{H}}_{i_{\text{col}},k}\right]^{-1} \\ \left(\mathbf{I} - \mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k}\right) &= \left[\mathbf{M}_{i,k-1}^{-1} + \bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\mathbf{\Sigma}}_{i_{\text{col}},k}^{-1}\bar{\mathbf{H}}_{i_{\text{col}},k}\right]^{-1}\mathbf{M}_{i,k-1}^{-1} \\ \mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k} &= \mathbf{I} - \left[\mathbf{M}_{i,k-1}^{-1} + \bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\mathbf{\Sigma}}_{i_{\text{col}},k}^{-1}\bar{\mathbf{H}}_{i_{\text{col}},k}\right]^{-1}\mathbf{M}_{i,k-1}^{-1} \\ &= \left[\mathbf{M}_{i,k-1}^{-1} + \bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\mathbf{\Sigma}}_{i_{\text{col}},k}^{-1}\bar{\mathbf{H}}_{i_{\text{col}},k}\right]^{-1}\left[\mathbf{M}_{i,k-1}^{-1} + \bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\mathbf{\Sigma}}_{i_{\text{col}},k}^{-1}\bar{\mathbf{H}}_{i_{\text{col}},k} - \mathbf{M}_{i,k-1}^{-1}\right] \\ &= \left[\mathbf{M}_{i,k-1}^{-1} + \bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\mathbf{\Sigma}}_{i_{\text{col}},k}^{-1}\bar{\mathbf{H}}_{i_{\text{col}},k}\right]^{-1}\bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\mathbf{\Sigma}}_{i_{\text{col}},k}^{-1}\bar{\mathbf{H}}_{i_{\text{col}},k}. \qquad (6.49)\end{aligned}$$

Combining the results from (6.47) and (6.49) yields

$$\mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k}\mathbf{M}_{i,k-1}(\mathbf{I}-\mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k})^{\mathsf{H}} = \tag{6.50}$$
$$\left(\mathbf{M}_{i,k-1}^{-1}+\mathbf{M}_{i,k-1}\bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\mathbf{\Sigma}}_{i_{\text{col}},k}^{-1}\bar{\mathbf{H}}_{i_{\text{col}},k}\right)^{-1}\bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\mathbf{\Sigma}}_{i_{\text{col}},k}^{-1}\bar{\mathbf{H}}_{i_{\text{col}},k}\left(\mathbf{M}_{i,k-1}^{-1}+\bar{\mathbf{H}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\mathbf{\Sigma}}_{i_{\text{col}},k}^{-1}\bar{\mathbf{H}}_{i_{\text{col}},k}\right)^{-\mathsf{H}}\succeq 0$$

because $\bar{\mathbf{\Sigma}}_{i_{\text{col}},k}^{-1}\succeq 0$. From (6.50), observe that $(\mathbf{I}-\mathbf{G}_{i,k}\bar{\mathbf{H}}_{i_{\text{col}},k})\succeq 0$ and is invertible. From the invertibility of $\mathbf{F}_{i,k}$, the matrix $\mathbf{B}_{i,k}$ is also invertible. Therefore, the inequality in (6.44) becomes

$$\mathbf{M}_{i,k}\succeq\mathbf{B}_{i,k}(\mathbf{M}_{i,k-1}+\mathbf{B}_{i,k}^{-1}\bar{\mathbf{C}}_k\mathbf{B}_{i,k}^{-\mathsf{H}})\mathbf{B}_{i,k}^{\mathsf{H}} \tag{6.51}$$

Using the assumptions (6.36a) – (6.36c), we have $\|\mathbf{B}_{i,k}\|\leq\bar{f}+\bar{f}\bar{m}\bar{h}^2/\underline{s}$ and therefore

$$\|\mathbf{B}_{i,k}^{-1}\bar{\mathbf{C}}_k\mathbf{B}_{i,k}^{-\mathsf{H}}\|\geq\underline{c}/(\bar{f}+\bar{f}\bar{m}\bar{h}^2/\underline{s})^2. \tag{6.52}$$

So, we can express, (6.51) as

$$\mathbf{M}_{i,k}\succeq\mathbf{B}_{i,k}\left(\mathbf{M}_{i,k-1}+\frac{\underline{c}}{(\bar{f}+\bar{f}\bar{m}\bar{h}^2/\underline{s})^2}\mathbf{I}\right)\mathbf{B}_{i,k}^{\mathsf{H}}. \tag{6.53}$$

Taking the inverse of both sides of (6.53), we have

$$\mathbf{M}_{i,k}^{-1}\preceq\mathbf{B}_{i,k}^{-\mathsf{H}}\left(\mathbf{M}_{i,k-1}+\frac{\underline{c}}{(\bar{f}+\bar{f}\bar{m}\bar{h}^2/\underline{s})^2}\mathbf{I}\right)^{-1}\mathbf{B}_{i,k}^{-1}. \tag{6.54}$$

Next, by pre- and post multiplying (6.54) with $\mathbf{B}_{i,k}^{\mathsf{H}}$ and $\mathbf{B}_{i,k}$, yields

$$\mathbf{B}_{i,k}^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}\mathbf{B}_{i,k}\preceq\left(\mathbf{M}_{i,k-1}+\frac{\underline{c}}{(\bar{f}+\bar{f}\bar{m}\bar{h}^2/\underline{s})^2}\mathbf{I}\right)^{-1} \tag{6.55}$$
$$=\left(\mathbf{I}+\frac{\underline{c}}{(\bar{f}+\bar{f}\bar{m}\bar{h}^2/\underline{s})^2}\mathbf{M}_{i,k-1}^{-1}\right)^{-1}\mathbf{M}_{i,k-1}^{-1} \tag{6.56}$$
$$\preceq\left(1+\frac{\underline{c}}{\bar{m}(\bar{f}+\bar{f}\bar{m}\bar{h}^2/\underline{s})^2}\right)^{-1}\mathbf{M}_{i,k-1}^{-1}. \tag{6.57}$$

Setting $1-\vartheta=\left(1+\frac{\underline{c}}{\bar{m}(\bar{f}+\bar{f}\bar{m}\bar{h}^2/\underline{s})^2}\right)^{-1}$, completes the proof. □

Lemma 6.4. *Under the conditions of Theorem 6.1, there exists positive scalars ϵ', $\kappa_{nonl}>0$ such that*

$$\text{Re}\left\{\boldsymbol{r}_{i,k-1}^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}(2\mathbf{B}_{i,k}\tilde{\boldsymbol{w}}_{i,k-1}+\boldsymbol{r}_{i,k-1})\right\}\leq\kappa_{nonl}\|\tilde{\boldsymbol{w}}_{i,k-1}\|^3\quad\textit{(a.s.)}$$

holds for $\|\boldsymbol{w}_k\|\leq\varepsilon'$, where $\kappa_{nonl}\overset{def}{=}\frac{\kappa'}{\underline{m}}\left(2\left(\bar{f}+\frac{\bar{f}\bar{m}\bar{h}^2}{\underline{s}}\right)+\kappa'\epsilon'\right)$.

Proof. This proof was modified from [123, Lem. 3.2] for complex variables.
We first begin by bounding the norm of the nonlinear function, $\boldsymbol{r}_{i,k-1}=\boldsymbol{\xi}(\tilde{\boldsymbol{w}}_{i,k-1})-\bar{\mathbf{F}}_{i,k}\mathbf{G}_{i,k}\boldsymbol{\chi}(\tilde{\boldsymbol{w}}_{i,k-1})$,

whereby,

$$\|\boldsymbol{r}_{i,k-1}\| \leq \|\boldsymbol{\xi}(\tilde{\boldsymbol{w}}_{i,k-1})\| + \frac{\bar{f}\bar{h}\bar{m}}{\underline{s}}\|\boldsymbol{\chi}(\tilde{\boldsymbol{w}}_{i,k-1})\|. \tag{6.58}$$

Using the bounds for the nonlinear errors in (6.36d) – (6.36e), we have,

$$\|\boldsymbol{r}_{i,k-1}\| \leq \kappa_\xi\|\tilde{\boldsymbol{w}}_{i,k}\|^2 + \frac{\bar{f}\bar{h}\bar{m}}{\underline{s}}\kappa_\chi\|\tilde{\boldsymbol{w}}_{i,k}\|^2 \tag{6.59}$$

for $\|\tilde{\boldsymbol{w}}_{i,k}\| \leq \epsilon'$, where $\epsilon' = \min(\epsilon_\xi, \epsilon_\chi)$. Therefore,

$$\|\boldsymbol{r}_{i,k-1}\| \leq \kappa'\|\tilde{\boldsymbol{w}}_{i,k}\|^2 \tag{6.60}$$

with $\kappa' = \left(\kappa_\xi + \frac{\bar{f}\bar{h}\bar{m}}{\underline{s}}\kappa_\chi\right)$. Next, by using the inequality $\mathsf{Re}\{a\} \leq |\mathsf{Re}\{a\}| \leq |a|$, we have

$$\begin{aligned}
\mathsf{Re}\left\{\boldsymbol{r}_{i,k-1}^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}(2\mathbf{B}_{i,k}\tilde{\boldsymbol{w}}_{i,k-1} + \boldsymbol{r}_{i,k-1})\right\} &\leq \left|\boldsymbol{r}_{i,k-1}^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}(2\mathbf{B}_{i,k}\tilde{\boldsymbol{w}}_{i,k-1} + \boldsymbol{r}_{i,k-1})\right| \\
&\leq \|\boldsymbol{r}_{i,k-1}\|\|\mathbf{M}_{i,k}^{-1}\|(2\|\mathbf{B}_{i,k}\|\|\tilde{\boldsymbol{w}}_{i,k-1}\| + \|\boldsymbol{r}_{i,k-1}\|) \\
&\leq \kappa'\|\tilde{\boldsymbol{w}}_{i,k-1}\|^2\frac{1}{\underline{m}}\left(2\left(\bar{f} + \frac{\bar{f}\bar{m}\bar{h}^2}{\underline{s}}\right)\|\tilde{\boldsymbol{w}}_{i,k-1}\| + \kappa'\epsilon'\|\tilde{\boldsymbol{w}}_{i,k-1}\|\right) \\
&= \frac{\kappa'}{\underline{m}}\left(2\left(\bar{f} + \frac{\bar{f}\bar{m}\bar{h}^2}{\underline{s}}\right) + \kappa'\epsilon'\right)\|\tilde{\boldsymbol{w}}_{i,k-1}\|^3
\end{aligned}$$

Setting, $\kappa_{\text{nonl}} = \frac{\kappa'}{\underline{m}}\left(2\left(\bar{f} + \frac{\bar{f}\bar{m}\bar{h}^2}{\underline{s}}\right) + \kappa'\epsilon'\right)$, we have,

$$\mathsf{Re}\left\{\boldsymbol{r}_{i,k-1}^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}(2\mathbf{B}_{i,k}\tilde{\boldsymbol{w}}_{i,k-1} + \boldsymbol{r}_{i,k-1})\right\} \leq \kappa_{\text{nonl}}\|\tilde{\boldsymbol{w}}_{i,k-1}\|^3 \quad \text{(a.s.)}$$

for $\|\tilde{\boldsymbol{w}}_{i,k-1}\| \leq \epsilon'$. □

Lemma 6.5. *Under the assumptions of Theorem 6.1, there exists a positive real scalar $\kappa_{noise} > 0$ independent of δ, such that*

$$\mathbb{E}\left\{\boldsymbol{n}_{i,k}^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}\boldsymbol{n}_{i,k}\middle|\tilde{\boldsymbol{w}}_{i,k-1}\right\} \leq \kappa_{noise}\delta \tag{6.61}$$

holds for all $k \geq 0$, where $\kappa_{noise} = \dfrac{M}{\underline{m}} + \dfrac{\bar{f}^2\bar{h}^2\bar{m}^2L}{\underline{m}\underline{s}^2}$.

Proof. This proof involves a minor improvement from [123, Lem. 3.3].
Recall that the noise term is $\boldsymbol{n}_{i,k} = \bar{\boldsymbol{q}}_k - \bar{\mathbf{F}}_{i,k}\mathbf{G}_{i,k}\bar{\boldsymbol{\eta}}_{i_{\text{col}},k}$. Therefore,

$$\boldsymbol{n}_{i,k}^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}\boldsymbol{n}_{i,k} = \bar{\boldsymbol{q}}_k^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}\bar{\boldsymbol{q}}_k + \bar{\boldsymbol{\eta}}_{i_{\text{col}},k}^{\mathsf{H}}\mathbf{G}_{i,k}^{\mathsf{H}}\bar{\mathbf{F}}_{i,k}^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}\bar{\mathbf{F}}_{i,k}\mathbf{G}_{i,k}\bar{\boldsymbol{\eta}}_{i_{\text{col}},k} \tag{6.62}$$

$$- \bar{\boldsymbol{q}}_k^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}\bar{\mathbf{F}}_{i,k}\mathbf{G}_{i,k}\bar{\boldsymbol{\eta}}_{i_{\text{col}},k} - \bar{\boldsymbol{\eta}}_{i_{\text{col}},k}^{\mathsf{H}}\mathbf{G}_{i,k}^{\mathsf{H}}\bar{\mathbf{F}}_{i,k}^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}\bar{\boldsymbol{q}}_k. \tag{6.63}$$

Taking the conditional expectation $\mathbb{E}\{\cdot|\boldsymbol{w}_{i,k-1}\}$ of (6.62) – (6.63) gives

$$\mathbb{E}\left\{\boldsymbol{n}_{i,k}^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}\boldsymbol{n}_{i,k}\big|\boldsymbol{w}_{i,k-1}\right\} = \mathbb{E}\left\{\bar{\boldsymbol{q}}_k^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}\bar{\boldsymbol{q}}_k + \bar{\boldsymbol{\eta}}_{i_{\text{col}},k}^{\mathsf{H}}\mathbf{G}_{i,k}^{\mathsf{H}}\bar{\mathbf{F}}_{i,k}^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}\bar{\mathbf{F}}_{i,k}\mathbf{G}_{i,k}\bar{\boldsymbol{\eta}}_{i_{\text{col}},k}\big|\tilde{\boldsymbol{w}}_{i,k-1}\right\}, \quad (6.64)$$

where the cross-terms in (6.63) disappear. This is because from (6.22a) – (6.22e), the algorithm variables $\mathbf{G}_{i,k}$, $\mathbf{F}_{i,k}$ and $\mathbf{M}_{i,k}$ are computed as functions of $\boldsymbol{w}_{i,k-1}$, hence also are functions of $\tilde{\boldsymbol{w}}_{i,k-1}$. Therefore, taking conditional expectation of (6.62) with respect to $\tilde{\boldsymbol{w}}_{i,k-1}$ allows us to treat these variables as constants. Also, both the system noise, $\bar{\boldsymbol{q}}_k$, and measurement noise $\bar{\boldsymbol{\eta}}_{i_{\text{col}},k}$ are statistically independent to the each other and the algorithm variables prior to time k, as shown in (6.2b) – (6.2c). From the various matrix bounds in Theorem 6.1, we now have

$$\mathbb{E}\left\{\boldsymbol{n}_{i,k}^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}\boldsymbol{n}_{i,k}\big|\tilde{\boldsymbol{w}}_{i,k-1}\right\} \leq \mathbb{E}\left\{\frac{1}{\underline{m}}\bar{\boldsymbol{q}}_k^{\mathsf{H}}\bar{\boldsymbol{q}}_k + \frac{(\bar{m}\bar{h}/\underline{s})^2\bar{f}^2}{\underline{m}}\bar{\boldsymbol{\eta}}_{i_{\text{col}},k}^{\mathsf{H}}\bar{\boldsymbol{\eta}}_{i_{\text{col}},k}\big|\tilde{\boldsymbol{w}}_{i,k-1}\right\} \quad (6.65)$$

$$= \frac{1}{\underline{m}}\mathbb{E}\left\{\mathsf{Tr}[\bar{\boldsymbol{q}}_k\bar{\boldsymbol{q}}_k^{\mathsf{H}}]\right\} + \frac{(\bar{m}\bar{h}/\underline{s})^2\bar{f}^2}{\underline{m}}\mathbb{E}\left\{\mathsf{Tr}[\bar{\boldsymbol{\eta}}_{i_{\text{col}},k}\bar{\boldsymbol{\eta}}_{i_{\text{col}},k}^{\mathsf{H}}]\right\}. \quad (6.66)$$

Using the definitions of the noise error covariance matrices and their respecives bounds in (6.36c), we have

$$\mathbb{E}\left\{\boldsymbol{n}_{i,k}^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}\boldsymbol{n}_{i,k}\big|\tilde{\boldsymbol{w}}_{i,k-1}\right\} \leq \left(\frac{1}{\underline{m}}M + \frac{(\bar{m}\bar{h}/\underline{s})^2\bar{f}^2}{\underline{m}}L\right)\delta, \quad (6.67)$$

and setting $\kappa_{\text{noise}} = \dfrac{M}{\underline{m}} + \dfrac{\bar{m}^2\bar{h}^2\bar{f}^2}{\underline{m}\underline{s}^2}L$, completes the proof. □

Proof of the Mean Square Stability of D-ACEKF (Theorem 6.1)

Proof. This proof is a combination of the proofs from [119, Th. 5.4] [123, Th. 3.1].
Consider defining the stochastic process $V_k(\tilde{\boldsymbol{w}}_{i,k})$ as

$$V_k(\tilde{\boldsymbol{w}}_{i,k}) \overset{\text{def}}{=} \tilde{\boldsymbol{w}}_{i,k}^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}\tilde{\boldsymbol{w}}_{i,k}, \quad (6.68)$$

which from assumptions on $\mathbf{M}_{i,k}$ in (6.36b) immediately satisfies the first condition of Lemma 6.2, that is,

$$\frac{1}{\bar{m}}\|\tilde{\boldsymbol{w}}_{i,k}\|^2 \leq V_k(\tilde{\boldsymbol{w}}_{i,k}) \leq \frac{1}{\underline{m}}\|\tilde{\boldsymbol{w}}_{i,k}\|^2. \quad (6.69)$$

Next, using the state error vector in (6.32), we have

$$\begin{aligned} V_k(\tilde{\boldsymbol{w}}_{i,k}) &= (\mathbf{B}_{i,k}\tilde{\boldsymbol{w}}_{i,k-1} + \boldsymbol{r}_{i,k-1} + \boldsymbol{n}_{i,k})^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}(\mathbf{B}_{i,k}\tilde{\boldsymbol{w}}_{i,k-1} + \boldsymbol{r}_{i,k-1} + \boldsymbol{n}_{i,k}) \\ &= \tilde{\boldsymbol{w}}_{i,k-1}^{\mathsf{H}}\mathbf{B}_{i,k-1}^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}\mathbf{B}_{i,k}\tilde{\boldsymbol{w}}_{i,k-1} + \mathsf{Re}\left\{\boldsymbol{r}_{i,k-1}^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}(2\mathbf{B}_{i,k}\tilde{\boldsymbol{w}}_{i,k-1} + \boldsymbol{r}_{i,k-1})\right\} \\ &\quad + 2\mathsf{Re}\left\{\boldsymbol{n}_{i,k}^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}(\mathbf{B}_{i,k}\tilde{\boldsymbol{w}}_{i,k-1} + \boldsymbol{r}_{i,k-1})\right\} + \boldsymbol{n}_{i,k-1}^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}\boldsymbol{n}_{i,k}. \end{aligned} \quad (6.70)$$

Following that, taking the conditional expectation, $\mathbb{E}\{V_k(\tilde{\boldsymbol{w}}_{i,k})|\tilde{\boldsymbol{w}}_{i,k-1}\}$, and noticing that

$$\mathbb{E}\left\{2\mathsf{Re}\left\{\boldsymbol{n}_{i,k}^{\mathsf{H}}\mathbf{M}_{i,k}^{-1}(\mathbf{B}_{i,k}\tilde{\boldsymbol{w}}_{i,k-1} + \boldsymbol{r}_{i,k-1})\right\}\big|\tilde{\boldsymbol{w}}_{i,k-1}\right\} = \mathbf{0},$$

since the noise process $\boldsymbol{n}_{i,k}$ is zero-mean and independent to the algorithm variables and data prior to k, and using Lemma 6.3 and Lemma 6.4, the recursion in (6.70), can be bounded by

$$\mathbb{E}\left\{V_k(\tilde{\boldsymbol{w}}_{i,k})|\tilde{\boldsymbol{w}}_{i,k-1}\right\} \leq (1-\vartheta)V_{k-1}(\tilde{\boldsymbol{w}}_{i,k-1}) + \kappa_{\text{nonl}}\|\tilde{\boldsymbol{w}}_{i,k-1}\|^3 + \kappa_{\text{noise}}\delta. \tag{6.71}$$

Restating (6.71) gives

$$\mathbb{E}\left\{V_k(\tilde{\boldsymbol{w}}_{i,k})|\tilde{\boldsymbol{w}}_{i,k-1}\right\} - V_{k-1}(\tilde{\boldsymbol{w}}_{i,k-1}) \leq -\vartheta V_{k-1}(\tilde{\boldsymbol{w}}_{i,k-1}) + \kappa_{\text{nonl}}\|\tilde{\boldsymbol{w}}_{i,k-1}\|^3 + \kappa_{\text{noise}}\delta, \tag{6.72}$$

for $\|\tilde{\boldsymbol{w}}_{i,k-1}\| \leq \epsilon'$. Defining $\epsilon = \min\left(\epsilon', \dfrac{\vartheta}{2\bar{m}\kappa_{\text{nonl}}}\right)$, we have the condition

$$\kappa_{\text{nonl}}\|\tilde{\boldsymbol{w}}_{i,k-1}\|\|\tilde{\boldsymbol{w}}_{i,k-1}\|^2 \leq \frac{\vartheta}{2\bar{m}}\|\tilde{\boldsymbol{w}}_{i,k-1}\|^2 \leq \frac{\vartheta}{2}V_{k-1}(\tilde{\boldsymbol{w}}_{i,k-1}), \tag{6.73}$$

which further bounds (6.72) as

$$\mathbb{E}\left\{V_k(\tilde{\boldsymbol{w}}_{i,k})|\tilde{\boldsymbol{w}}_{i,k-1}\right\} - V_{k-1}(\tilde{\boldsymbol{w}}_{i,k-1}) \leq -\frac{\vartheta}{2}V_{k-1}(\tilde{\boldsymbol{w}}_{i,k-1}) + \kappa_{\text{noise}}\delta, \tag{6.74}$$

for $\|\tilde{\boldsymbol{w}}_{i,k-1}\| \leq \epsilon$. By setting $\vartheta' = \vartheta/2 < 1$, the conditions of (6.38) in Lemma 6.2 are satisfied. Therefore, the weight error $\tilde{\boldsymbol{w}}_{i,k}$ is bounded in the mean square. By choosing $\delta = \dfrac{\vartheta\tilde{\epsilon}^2}{2\bar{m}\kappa_{\text{noise}}}$ with some $\tilde{\epsilon} < \epsilon$, for $\|\tilde{\boldsymbol{w}}_{i,k}\| > \tilde{\epsilon}$, reduces the inequality in (6.74) to

$$\begin{aligned}\mathbb{E}\left\{V_k(\tilde{\boldsymbol{w}}_{i,k})|\tilde{\boldsymbol{w}}_{i,k-1}\right\} - V_{k-1}(\tilde{\boldsymbol{w}}_{i,k-1}) \leq & -\frac{\vartheta}{2}V_{k-1}(\tilde{\boldsymbol{w}}_{i,k-1}) + \frac{\vartheta}{2\bar{m}}\|\tilde{\boldsymbol{w}}_{i,k}\|^2 \\ \leq & -\frac{\vartheta}{2}V_{k-1}(\tilde{\boldsymbol{w}}_{i,k-1}) + \frac{\vartheta}{2}V_{k-1}(\tilde{\boldsymbol{w}}_{i,k-1}) = 0,\end{aligned} \tag{6.75}$$

so as to ensure that the estimation error is bounded for all k. □

6.6 Chapter Summary

In this chapter, we have reviewed the concepts underpinning Kalman filtering for widely linear state spaces and noncircular noise processes. In addition, various distributed Kalman filtering strategies were explored for collaborative estimation tasks. The key contribution of this chapter, is the D-ACEKF algorithm in (6.21a) – (6.21g) which is developed through widely linear state-space modelling so as to cater for widely linear system models and noncircular noise processes. Secondly, through Theorem 6.1, we showed that the D-ACEKF is mean square stable under certain assumptions on the smoothness of the nonlinear functions and boundedness of the noise variances and initial estimation errors.

Conclusion to Part II

In Part II, standalone adaptive algorithms (some of which were introduced in Part I) have been extended to a multi-agent setting. In Part III of the thesis, the contribution of casting the frequency estimation problem as a distributed adaptive filtering problem shall be presented. Par-

ticularly, Chapter 7 introduces three-phase transforms which are routinely used in frequency estimation tasks from a signal processing perspective. The discussion in Chapter 7, will help the reader to understand the challenges modern and future grids face in frequency estimation during voltage imbalances. Following that, practical benefits of a distributed approach to frequency estimation shall be demonstrated in Chapter 8.

Part III

Distributed Frequency Estimation in the Smart Grid

Chapter 7

Modern View of Three-Phase Transforms

> The existence of mysterious relations between all these different domains is the most striking and delightful feature of mathematics.
>
> *Vladamir Arnold*

Chapter Overview

In the following two chapters, the focus shall shift back to the electricity grid. In this chapter, three-phase transformations which are commonly used in signal processing applications for power networks are introduced. More importantly, the contribution of this chapter is that we provide a signal processing perspective of three-phase transforms and elucidate their links with discrete Fourier transform (DFT), dimensionality reduction and frequency demodulation techniques.

Specifically, the Clarke transform is shown to represent a principal component analysis (PCA) based dimensionality reduction algorithm for a balanced three-phase voltage signal, while the symmetrical transform is shown to represent the discrete Fourier transform of unbalanced phasors and that the Park transform is viewed as frequency (de)modulation (FM) scheme. We then employ recent results in complex-valued statistics to show that the complex-domain allows for a more compact and physically meaningful representations for the Clark and Park transforms. Finally, based on these novel interpretations and recent developments in widely linear modelling, an adaptive three-phase balancing transformation is proposed for real-time dynamic power systems.

7.1 Introduction

Three-phase transforms are the backbone of power system analysis as they allow for dimensionality reduction and the simplification of analysis of unbalanced power networks. Most of the underlying theory of three-phase transforms was developed through the diagonalisation of circulant impedance matrices so as to enable the analysis of a general unbalanced three-phase network through separate balanced networks. This common principle is also the reason for an intimate relationship between these transforms. For example, the symmetrical transform [127], Clarke transform [128, 129] and their closely related variants, like the Kimbark, Boyajian, and Concordia transforms are related through performing elementary column and row operations (e.g. scaling) on the original Clarke transformation matrix [130]. In [131], the Clarke and Park transforms [132] were interpreted as time-domain counterparts of the symmetrical transform.

Owing to their generality, three-phase transformations have found applications in signal processing, control and machine learning tasks, such as frequency estimation and fault detection [24, 133, 134]. However, sufficient justification and physical interpretation from a signal processing perspective is still lacking. Motivated by the statement of A. Boyajian, an early contributor to the theory of three-phase transforms [135],

> *"A transformation which replaces a given system of circuits by another, presumably more convenient, equivalent system of circuits, should be subjected to two tests: (a) physical interpretation and (b) mathematical consistency,"*

we next offer a modern signal processing perspective to both (a) physical interpretation and (b) mathematical consistency of the three-phase transforms.

We start with the symmetrical transform and show that it can be developed from a spatial discrete Fourier transform (DFT) perspective, to decompose a general unbalanced phasor vector into three-separate "spatial" sinusoids. Next, the Clark transform is derived as a general second-order optimal dimensionality reduction algorithm for balanced three-phase signals. This derivation establishes a direct link between the principal components analysis (PCA) and the Clark transform.

We then employ a complex-valued representation to show that although the Clarke transform is not the optimal dimensionality reduction scheme for a set of unbalanced phase voltages, it yields a meaningful relationship with respect to the symmetrical transform [131]. Furthermore, the chosen complex-valued representation allows for a straightforward link between the Clarke and Park transforms, with the latter shown to be a frequency demodulation scheme.

Finally, these ideas together with recently introduced widely linear autoregressive modelling of the $\alpha\beta$ voltages [22] and so-called "balancing" transformation [136] are employed to introduce a new class of adaptive three-phase transformations suitable for real-time application in dynamic power networks. The proposed adaptive three-phase transforms is referred to as the "adaptive Clarke-Park transform" owing to its resemblance to the outputs of the original Clarke and Park transforms for balanced systems. Simulations on synthetically generated signals benchmark the performance of the proposed adaptive Clark-Park transform against their original static counterparts.

7.2 Background on Three-Phase Transforms

7.2.1 Clarke Transform

Consider a sampled three-phase voltage measurement vector at the discrete time instant k, given by

$$\boldsymbol{s}_k = \begin{bmatrix} v_{a,k} \\ v_{b,k} \\ v_{c,k} \end{bmatrix} = \begin{bmatrix} V_a \cos(\omega k + \phi_a) \\ V_b \cos(\omega k + \phi_b - \frac{2\pi}{3}) \\ V_c \cos(\omega k + \phi_c + \frac{2\pi}{3}) \end{bmatrix}, \tag{7.1}$$

where V_a, V_b, V_c are the amplitudes of the phase voltages $v_{a,k}$, $v_{b,k}$, $v_{c,k}$, $\omega = 2\pi f T$ is the fundamental angular frequency, with f the fundamental power system frequency and T the sampling

interval. The phase values for each phase are denoted by ϕ_a, ϕ_b, ϕ_c.

The Clarke or $\alpha\beta$ transform changes the basis of the three-phase signal $\boldsymbol{s}_k$ in (7.1) to yield the $v_{0,k}, v_{\alpha,k}, v_{\beta,k}$ voltages

$$\begin{bmatrix} v_{0,k} \\ v_{\alpha,k} \\ v_{\beta,k} \end{bmatrix} = \underbrace{\sqrt{\frac{2}{3}} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix}}_{\text{Clarke matrix}} \begin{bmatrix} v_{a,k} \\ v_{b,k} \\ v_{c,k} \end{bmatrix}, \tag{7.2}$$

where $v_{\alpha,k}$ and $v_{\beta,k}$ are the rotating components referred to as the α and β sequences. The term $v_{0,k}$ is known as the zero-sequence, as it is null when the three-phase signal $\boldsymbol{s}_k$ is balanced (equal voltage magnitudes, $V_a = V_b = V_c$, and the $\frac{2\pi}{3}$ phase angle separation, $\phi_a = \phi_b = \phi_c$, across the phase voltages). Therefore, typically only $v_{\alpha,k}$ and $v_{\beta,k}$ are used in practice and can be obtained using the last two rows of the Clarke matrix in (7.2), such that the three-phase voltage in (7.1) is projected onto a 2D subspace as

$$\begin{bmatrix} v_{\alpha,k} \\ v_{\beta,k} \end{bmatrix} = \sqrt{\frac{2}{3}} \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} v_{a,k} \\ v_{b,k} \\ v_{c,k} \end{bmatrix}, \tag{7.3}$$

Figure 7.1 shows the geometric interpretation of the Clarke transform for a balanced system. Observe the orthogonal nature of the $v_{\alpha,k}$ and $v_{\beta,k}$ components, which allows their convenient combination into a complex-valued voltage, $s_k = v_{\alpha,k} + jv_{\beta,k}$.

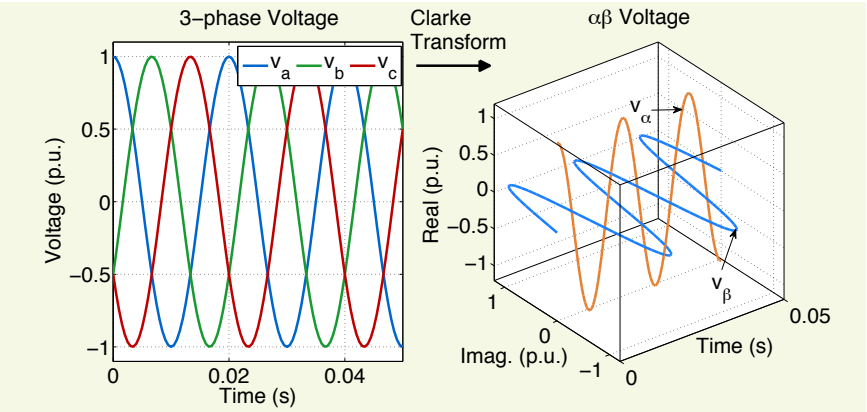

Figure 7.1: Geometric representation of the Clarke transform with voltages normalised to per unit (p.u.) voltages.

7.2.2 Park Transform

The Park transform (also known as the dq transform) projects the three-phase signal $\boldsymbol{s}_k$ on to an orthogonal, time-varying frame, which by virtue of rotating at the fundamental frequency $\omega_\circ$, yields stationary outputs. In other words, the Park voltages $v_{d,k}, v_{q,k}$ are obtained from the

$\alpha\beta$ voltages in (7.3) using a time-varying transformation given by [132]

$$\begin{bmatrix} v_{d,k} \\ v_{q,k} \end{bmatrix} = \underbrace{\begin{bmatrix} \cos(\omega_\circ k) & \sin(\omega_\circ k) \\ -\sin(\omega_\circ k) & \cos(\omega_\circ k) \end{bmatrix}}_{\text{Park matrix}} \begin{bmatrix} v_{\alpha,k} \\ v_{\beta,k} \end{bmatrix}, \tag{7.4}$$

where the orthogonal direct and quadrature components, $v_{d,k}$ and $v_{q,k}$, can be combined into a complex variable $v_k = v_{d,k} + jv_{q,k}$.

7.2.3 Symmetrical Transform

The symmetrical transform operates in the "phasor" domain of the three-phase signal, $\boldsymbol{s}_k$, that is, the rotation becomes inherent to the representation in (7.1). The phasors within $\boldsymbol{s}_k$ are given by the voltage magnitudes and phase angles of each phase voltage. In this way, $\bar{V}_a \stackrel{\text{def}}{=} V_a e^{j\phi_a}$, $\bar{V}_a \stackrel{\text{def}}{=} V_b e^{j(\phi_b - \frac{2\pi}{3})}$, $\bar{V}_c \stackrel{\text{def}}{=} V_c e^{j(\phi_c + \frac{2\pi}{3})}$, which can be comprised into the phasor vector

$$\boldsymbol{v} \stackrel{\text{def}}{=} \begin{bmatrix} \bar{V}_a & \bar{V}_b & \bar{V}_c \end{bmatrix}^{\mathsf{T}}. \tag{7.5}$$

The symmetrical transform converts a general unbalanced phasor in (7.5) into three separate balanced components, referred to as the zero–, positive– and negative–sequence phasors, denoted respectively by $\bar{V}_0$, $\bar{V}_+$, $\bar{V}_-$. Unlike the Clarke and Park transforms, the symmetrical transform is an inherent complex-valued transformation on the phasor vector $\boldsymbol{v}$ and is given by

$$\begin{bmatrix} \bar{V}_0 \\ \bar{V}_+ \\ \bar{V}_- \end{bmatrix} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 1 \\ 1 & a & a^2 \\ 1 & a^2 & a \end{bmatrix} \begin{bmatrix} \bar{V}_a \\ \bar{V}_b \\ \bar{V}_c \end{bmatrix}, \tag{7.6}$$

where $a = e^{-j\frac{2\pi}{3}}$.

> **Contribution: New Derivations of Canonical Three-Phase Transforms**
> Rest of this chapter shall motivate and derive the Symmetrical, Clarke and Park transformations from a novel signal analysis perspective. These novel derivations are in contrast to the traditional circuit theory methods and serve to aid algorithm designers in designing algorithms for dynamic grids.

7.3 Symmetrical Transform as a Spatial DFT

A three-phase time-domain signal, $\boldsymbol{s}_k$, in (7.1) is typically considered as a collection of three separate univariate signals. However, $\boldsymbol{s}_k$ can also be treated as three samples of a monocomponent signal rotating at a *spatial frequency*. In this sense, for a balanced system, the phasor vector

$\boldsymbol{v}$ in (7.14), is given by

$$\boldsymbol{v} = \begin{bmatrix} 1 & e^{j\Omega} & e^{j2\Omega} \end{bmatrix}^{\mathsf{T}}.$$

Therefore $\boldsymbol{v}$ can be treated as a single sinusoid rotating at a spatial frequency of $\Omega = -\frac{2\pi}{3}$, although every component of $\boldsymbol{v}$ comes from a separate phase voltage $v_{a,k}, v_{b,k}$, and $v_{c,k}$. However, under unbalanced conditions, the phasor vector $\boldsymbol{v}$, does not represent a single complex-valued spatial sinusoid since it contains the individual phasors with unequal amplitudes and a non-uniform phase separation, as defined in (7.5).

To resolve this issue, we consider decomposing $\boldsymbol{v}$ into monocomponent spatial sinusoids using the framework of a discrete Fourier transform (DFT) operating in the spatial domain. The DFT of the phasor vector, $\boldsymbol{v} = [v_0, v_1, v_2]^{\mathsf{T}} \in \mathbb{C}^{3\times 1}$, is given by

$$X[k] = \frac{1}{\sqrt{3}} \sum_{n=0}^{2} v_n e^{-j\frac{2\pi}{3}nk}, \quad k = 0, 1, 2,$$

which can expressed in a matrix form as

$$\begin{bmatrix} X[0] \\ X[1] \\ X[2] \end{bmatrix} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 1 \\ 1 & a & a^2 \\ 1 & a^2 & a \end{bmatrix} \begin{bmatrix} \bar{V}_a \\ \bar{V}_b \\ \bar{V}_c \end{bmatrix}, \tag{7.7}$$

where $a = e^{-j\frac{2\pi}{3}}$. The three-point DFT in (7.7) therefore decomposes the phasor vector $\boldsymbol{v}$ into a stationary component $X[0]$ and two components rotating at spatial frequencies $\frac{2\pi}{3}$, and $-\frac{2\pi}{3}$.

Remark 7.1. *The spatial DFT in (7.7) is identical to the symmetrical transform in (7.6). More specifically, upon comparing the DFT in (7.7) and the symmetrical transform in (7.6), the stationary DFT component, $X[0]$, corresponds to the zero-sequence phasor, $\bar{V}_0$, while the fundamental DFT components, $X[1]$ and $X[2]$, are respectively the positive– and negative–sequence phasors.*

7.4 Clarke Transform as a Principal Components Analyser

7.4.1 Principal Components Analysis (PCA)

In many signal processing, control and machine learning operations based on multi-channel data, it is often useful to reduce the dimensionality of a signal while maintaining the useful information. This reduces the computational complexity of any algorithm while preserving the physical meaning in the data. Consider a general data vector, $\boldsymbol{x}_k \in \mathbb{R}^{M\times 1}$, for which the covariance matrix is defined as

$$\text{cov}(\boldsymbol{x}_k) \overset{\text{def}}{=} \mathbf{R}_{\boldsymbol{x}} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} \boldsymbol{x}_k \boldsymbol{x}_k^{\mathsf{T}}. \tag{7.8}$$

The symmetric covariance matrix $\mathbf{R}_{\boldsymbol{x}}$ admits the following eigenvalue decomposition

$$\mathbf{Q}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{x}}\mathbf{Q} = \mathbf{\Lambda}, \tag{7.9}$$

where the diagonal eigenvalue matrix, $\mathbf{\Lambda} = \mathsf{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_M\}$, indicates the power of each component of $\boldsymbol{x}_k$ while the matrix of eigenvectors, $\mathbf{Q}_r = [\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_M]$, designates the principal directions in the data.

Suppose the signal $\boldsymbol{x}_k$ is to be transformed into a vector, $\boldsymbol{u}_k \in \mathbb{R}^{M\times 1}$, with the same dimensionality as the original signal $\boldsymbol{x}_k$, using a linear transformation matrix $\mathbf{W}$, to give

$$\boldsymbol{u}_k = \mathbf{W}\boldsymbol{x}_k, \quad \text{where} \quad \mathsf{cov}(\boldsymbol{u}_k) = \mathbf{I}. \tag{7.10}$$

The principal component analysis (PCA), also known as the Karhunen-Loeve transform, states that the above transformation matrix, $\mathbf{W}$, can be obtained from the eigenvector and eigenvalue matrices in (7.9) as $\mathbf{W} = \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{Q}^{\mathsf{T}}$ [137].

Proof:

$$\begin{aligned}
\mathsf{cov}(\boldsymbol{u}_k) &= \lim_{N\to\infty}\frac{1}{N}\sum_{i=0}^{N-1}\boldsymbol{u}_k\boldsymbol{u}_k^{\mathsf{T}} \\
&= \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{Q}^{\mathsf{T}}\left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=0}^{N-1}\boldsymbol{x}_k\boldsymbol{x}_k^{\mathsf{T}}\right)\mathbf{Q}\mathbf{\Lambda}^{-\frac{1}{2}} \\
&= \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{Q}^{\mathsf{T}}\mathbf{R}_{\boldsymbol{x}}\mathbf{Q}\mathbf{\Lambda}^{-\frac{1}{2}} = \mathbf{I}.
\end{aligned}$$

Since the transformation in (7.10) does not change the dimensionality of $\boldsymbol{x}_k$, a dimensionality reduction scheme based on the PCA in (7.10) can be applied to obtain a transformed data vector $\boldsymbol{u}_{r,k} \in \mathbb{R}^{r\times 1}$ of dimension $r < M$ as

$$\boldsymbol{u}_{r,k} = \mathbf{W}_r\boldsymbol{x}_k = \mathbf{\Lambda}_{1:r}^{-\frac{1}{2}}\mathbf{Q}_{1:r}^{\mathsf{T}}\boldsymbol{x}_k, \tag{7.11}$$

where $\mathbf{\Lambda}_{1:r} = \mathsf{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_r\}$ and $\mathbf{Q}_{1:r} = [\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_r]$, while r stands for the r largest eigenvalues in $\mathbf{\Lambda}$. The PCA-based dimensionality reduction scheme in (7.11) selects the directions in which the data expresses maximal variance, that is, the directions of the principal eigenvectors of the data matrix.

Since the three-phase signal, $\boldsymbol{s}_k$, in (7.1) is essentially a multichannel signal, it admits the PCA-based dimensionality reduction in (7.11). Employing the identity $\cos(x) = (e^{jx} + e^{-jx})/2$, allows for a phasor representation of the three-phase voltage in (7.1) in the form

$$\boldsymbol{s}_k = \frac{1}{2}\begin{bmatrix}\bar{V}_a\\ \bar{V}_b\\ \bar{V}_c\end{bmatrix}e^{j\omega k} + \frac{1}{2}\begin{bmatrix}\bar{V}_a^*\\ \bar{V}_b^*\\ \bar{V}_c^*\end{bmatrix}e^{-j\omega k}, \tag{7.12}$$

where the time-independent phasors are given by $\bar{V}_a = V_a e^{j\phi_a}$, $\bar{V}_b = V_b e^{j\left(\phi_b - \frac{2\pi}{3}\right)}$ and $\bar{V}_c = V_c e^{j\left(\phi_c + \frac{2\pi}{3}\right)}$. Without loss in generality, we shall normalise the phasors relative to $\bar{V}_a$. To this end, defining $\delta_i \overset{\text{def}}{=} \bar{V}_i/\bar{V}_a$, and $\bar{V}_a = 1$, gives

$$\boldsymbol{s}_k = \frac{1}{2}\left(\boldsymbol{v}e^{j\omega k} + \boldsymbol{v}^* e^{-j\omega k}\right), \tag{7.13}$$

where the phasor vector $\boldsymbol{v}$, defined originally in (7.5), now becomes

$$\boldsymbol{v} = \begin{bmatrix} 1 & \delta_b & \delta_c \end{bmatrix}^{\mathsf{T}}. \tag{7.14}$$

The PCA based dimensionality reduction requires the computation of the covariance matrix, $\mathsf{cov}(\boldsymbol{s}_k) = \mathbf{R}_{\boldsymbol{s}}$, in the form

$$\mathbf{R}_{\boldsymbol{s}} = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \boldsymbol{s}_k \boldsymbol{s}_k^{\mathsf{H}}. \tag{7.15}$$

From the definition of the three-phase signal in (7.13), we have

$$\boldsymbol{s}_k \boldsymbol{s}_k^{\mathsf{H}} = \frac{1}{4}\left(\boldsymbol{v}\boldsymbol{v}^{\mathsf{H}} + \boldsymbol{v}^*\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} e^{2j\omega k} + \boldsymbol{v}^*\boldsymbol{v}^{\mathsf{H}} e^{-2j\omega k}\right). \tag{7.16}$$

Substituting the outer-product in (7.16) into the definition of the covariance matrix in (7.15) and considering that for, $0 < \omega < \pi$, the following holds [138, p. 56]

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} e^{\pm 2j\omega k} = 0, \tag{7.17}$$

gives the following covariance matrix for general 3-phase voltages

$$\mathbf{R}_{\boldsymbol{s}} = \frac{1}{4}\left(\boldsymbol{v}\boldsymbol{v}^{\mathsf{H}} + \boldsymbol{v}^*\boldsymbol{v}^{\mathsf{T}}\right) = \frac{1}{2}\mathsf{Re}\left\{\boldsymbol{v}\boldsymbol{v}^{\mathsf{H}}\right\} = \frac{1}{4}\left(\boldsymbol{v}_r\boldsymbol{v}_r^{\mathsf{T}} + \boldsymbol{v}_i\boldsymbol{v}_i^{\mathsf{T}}\right), \tag{7.18}$$

where $\boldsymbol{v}_r = \mathsf{Re}\{\boldsymbol{v}\}$ and $\boldsymbol{v}_i = \mathsf{Im}\{\boldsymbol{v}\}$ denote the real and imaginary parts of the phasor vector $\boldsymbol{v}$ defined in (7.14).

Remark 7.2. *Observe from (7.18) that the covariance matrix of the trivariate signal $\boldsymbol{s}_k$ is rank-deficient (rank 2) as it represents as a sum of two rank-one outer products $\boldsymbol{v}_r\boldsymbol{v}_r^{\mathsf{T}}$ and $\boldsymbol{v}_i\boldsymbol{v}_i^{\mathsf{T}}$. This implies that the use of all three data channels (system phases) is redundant. Therefore, without loss in information, the three-phase signal in (7.12) can be projected onto a subspace of lower dimensionality (two).*

7.4.2 Balanced Systems

For a balanced system, the imbalance values δ_b and δ_i correspond to the $\frac{2\pi}{3}$ phase separations, so that $\delta_b = e^{-j\frac{2\pi}{3}}$ and $\delta_c = e^{j\frac{2\pi}{3}}$, so that the phasor vector, $\boldsymbol{v}$, becomes

$$\boldsymbol{v} = \begin{bmatrix} 1 & e^{-j\frac{2\pi}{3}} & e^{j\frac{2\pi}{3}} \end{bmatrix}^{\mathsf{T}}. \tag{7.19}$$

From (7.18), this yields the covariance matrix in the form

$$\mathbf{R}_{\boldsymbol{s}} = \frac{1}{2}\mathsf{Re}\left\{\boldsymbol{v}\boldsymbol{v}^{\mathsf{H}}\right\} = \frac{1}{4}\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}. \tag{7.20}$$

As seen in (7.11), the PCA-based dimensionality reduction matrix, $\mathbf{W}_r$, is obtained from the eigen-decomposition of the sample covariance matrix $\mathbf{R}_{\boldsymbol{s}}$, that is

$$\mathbf{R}_{\boldsymbol{s}} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{\mathsf{T}}. \tag{7.21}$$

By inspection of $\mathbf{R}_{\boldsymbol{s}}$ in (7.20), from the first eigenvector-eigenvalue pair, $(\boldsymbol{q}_1, \lambda_1)$, we have

$$\mathbf{R}_{\boldsymbol{s}}\boldsymbol{q}_1 = \mathbf{0} \implies \boldsymbol{q}_1 = \frac{1}{\sqrt{3}}\mathbf{1},\ \lambda_1 = 0. \tag{7.22}$$

To find the remaining eigenvector-eigenvalue pairs, consider the expression for the covariance matrix in (7.18) and the value of the phasor vector $\boldsymbol{v}$ in (7.19) and notice that $\boldsymbol{v}_r = \mathsf{Re}\{\boldsymbol{v}\} = [1, -\frac{1}{2}, -\frac{1}{2}]^{\mathsf{T}}$ and $\boldsymbol{v}_i = \mathsf{Im}\{\boldsymbol{v}\} = [0, -\frac{\sqrt{3}}{2}, \frac{\sqrt{3}}{2}]^{\mathsf{T}}$ are orthogonal, i.e. $\boldsymbol{v}_r^{\mathsf{T}}\boldsymbol{v}_i = 0$.

Therefore, the remaining two eigenvectors of $\mathbf{R}_{\boldsymbol{s}}$ are $\boldsymbol{q}_2 = \boldsymbol{v}_r/\|\boldsymbol{v}_r\|$ and $\boldsymbol{q}_3 = \boldsymbol{v}_i/\|\boldsymbol{v}_i\|$ with the corresponding eigenvalues $(1/4)\|\boldsymbol{v}_r\|$ and $(1/4)\|\boldsymbol{v}_i\|$. In summary, the eigenvector matrix, $\mathbf{Q}^{\mathsf{T}}$, and the diagonal matrix of eigenvalues, $\boldsymbol{\Lambda}$, in (7.21) are given by

$$\mathbf{Q}^{\mathsf{T}} = \sqrt{\frac{2}{3}}\begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \quad \boldsymbol{\Lambda} = \frac{1}{4}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1.5 & 0 \\ 0 & 0 & 1.5 \end{bmatrix}. \tag{7.23}$$

By inspection of the diagonal elements of $\boldsymbol{\Lambda}$ in (7.23), observe only two non-zero eigenvalues which reinforces Remark 7.2 which states that the matrix $\mathbf{R}_{\boldsymbol{s}}$ is of Rank 2.

Remark 7.3. *The eigenvector matrix $\mathbf{Q}^{\mathsf{T}}$ in (7.23) is identical to the Clarke transformation matrix defined in (7.2), whereby each row of $\mathbf{Q}^{\mathsf{T}}$ corresponds to the eigenvector associated with an eigenvalue on the diagonal of $\boldsymbol{\Lambda}$.*

Since all of the variance in three-phase power system data can be explained by the two eigenvectors associated with the non-zero eigenvalues (principle axes), the Clarke transform can be interpreted as a projection of the $\mathbb{R}^3$ data onto a 2D subspace spanned by the orthogonal eigenvectors $[1, -\frac{1}{2}, -\frac{1}{2}]^{\mathsf{T}}$ and $[0, \frac{\sqrt{3}}{2}, -\frac{\sqrt{3}}{2}]^{T}$, as illustrated in Figure 7.2. Specifically, based on the definition of the PCA transformation matrix in (7.11), the optimal dimensionality reduction scheme takes the form

$$\begin{aligned} \mathbf{W}_2 &= \boldsymbol{\Lambda}_{1:2}^{-\frac{1}{2}}\mathbf{Q}_{1:2}^{\mathsf{T}} \\ &= \frac{4}{3}\begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix}. \end{aligned}$$

which has the same form as the Clarke matrix in (7.3) but with a different normalisation con-

stant. Observe that, this does not affect the operation of the Clarke transform since the scaling effects equally $v_{\alpha,k}$ and $v_{\beta,k}$.

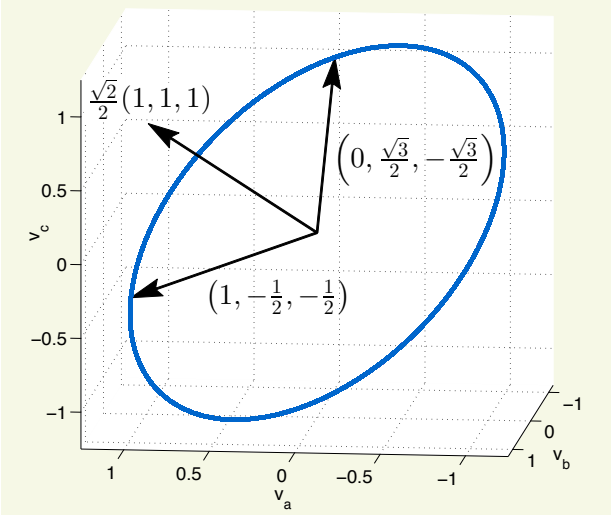

Figure 7.2: A balanced system with three principal axes. All the information is contained within a 2-dimensional subspace spanned by the eigenvectors $\left(1,\ -\frac{1}{2},\ -\frac{1}{2}\right)$ and $\left(0,\ \frac{\sqrt{3}}{2},\ -\frac{\sqrt{3}}{2}\right)$.

7.4.3 Generalised Balancing Transform for Unbalanced Power Systems

For unbalanced systems, the imbalance ratios, $\delta_b = |\delta_b|e^{j\phi_b}$, and $\delta_c = |\delta_c|e^{j\phi_c}$, depend on the type of imbalance. This yields a covariance matrix, $\mathbf{R}_{\boldsymbol{s}}^{u}$ that is different from that for the balanced case in (7.18), which has the form

$$\begin{aligned}
\mathbf{R}_{\boldsymbol{s}}^{u} &= \frac{1}{2}\mathsf{Re}\left\{\boldsymbol{v}\boldsymbol{v}^{\mathsf{H}}\right\} \\
&= \frac{1}{2}\mathsf{Re}\left\{\begin{bmatrix} 1 & \delta_b^* & \delta_c^* \\ \delta_b & |\delta_b|^2 & \delta_b {\delta_c}^* \\ \delta_c & \delta_b^*\delta_c & |\delta_c|^2 \end{bmatrix}\right\} \\
&= \frac{1}{2}\begin{bmatrix} 1 & |\delta_b|\cos(\phi_b) & |\delta_c|\cos(\phi_c) \\ |\delta_b|\cos(\phi_b) & |\delta_b|^2 & |\delta_b||\delta_c|\cos(\Delta_{bc}) \\ |\delta_c|\cos(\phi_c) & |\delta_b||\delta_c|\cos(\Delta_{bc}) & |\delta_c|^2 \end{bmatrix}.
\end{aligned} \tag{7.24}$$

where $\Delta_{bc} = \phi_b - \phi_c$.

Notice that the eigenvector and eigenvalue matrices are different when the system is unbalanced, and therefore different projection matrices are required [24]. Consider a special case of power imbalance characterised by symmetrical faults where $\delta_c = \delta_b^*$. The typical voltage sags (Types A – D) [139], fall into this category, see the phasor diagrams for the Type C and Type D

sags in Figure 7.3. The covariance matrix in (7.24) for symmetrical unbalances then becomes

$$\mathbf{R}_{\boldsymbol{s}}^{u} = \frac{1}{2}\begin{bmatrix} 1 & |\delta_b|\cos(\phi_b) & |\delta_b|\cos(\phi_b) \\ |\delta_b|\cos(\phi_b) & |\delta_b|^2 & |\delta_b|^2\cos(2\phi_b) \\ |\delta_b|\cos(\phi_b) & |\delta_b|^2\cos(2\phi_b) & |\delta_b|^2 \end{bmatrix}.$$

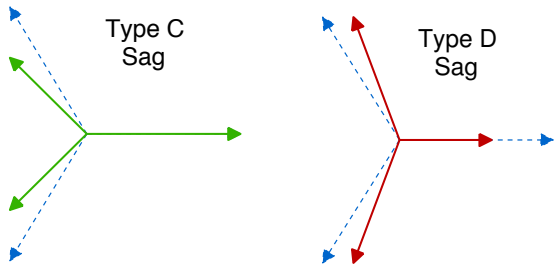

Figure 7.3: Phasor diagrams of voltage sags, where the dashed blue arrows show a set of balanced three-phase voltage phasors. Notice the symmetrical nature of the sags.

Similar to the balanced case in (7.23), the real and imaginary components of the unbalanced phasor vectors $\boldsymbol{v}_r$ and $\boldsymbol{v}_i$ are orthogonal, so that the eigenvectors of the system are again given by $\boldsymbol{q}_2 = \boldsymbol{v}_r/\|\boldsymbol{v}_r\|$ and $\boldsymbol{q}_3 = \boldsymbol{v}_i/\|\boldsymbol{v}_i\|$. Consequently, the eigen-decomposition of $\mathbf{R}_{\boldsymbol{s}}^{u}$ becomes

$$\mathbf{Q}^{\mathsf{T}} = \frac{1}{\sqrt{2m^2+1}}\begin{bmatrix} \sqrt{2}m & -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ 1 & m & m \\ 0 & \sqrt{m^2+\frac{1}{2}} & -\sqrt{m^2+\frac{1}{2}} \end{bmatrix},$$

$$\mathbf{\Lambda} = \frac{1}{2}\begin{bmatrix} 0 & 0 & 0 \\ 0 & m^2+0.5 & 0 \\ 0 & 0 & |\delta_b|^2 - m^2 \end{bmatrix}. \tag{7.25}$$

where $m = |\delta_b|\cos(\phi_b)$. For a balanced case, $m = -\frac{1}{2}$, and the above system simplifies into the eigendecomposition for the balanced case in (7.23). This is illustrated in Figure 7.4, where regardless of the imbalance level in the system, the three-phase voltages lie in a 2-dimensional subspace of $\mathbb{R}^3$. However, as the type and level of unbalance change, the PCA based dimensionality reduction scheme derived in (7.25) will not be identical to the Clarke transform in (7.3). This conforms with the well-known phenomenon that the application of the Clarke transform to unbalanced system voltages will yield different forms of $\alpha\beta$ voltages compared to those in the balanced case [24].

7.4.4 Clarke and Park Transforms in the Complex Domain

To interpret the output of the Clarke transform for a general (unbalanced) three-phase system, it is useful to cast our discussion to the complex domain, where the $\alpha\beta$ voltage in (7.3) is

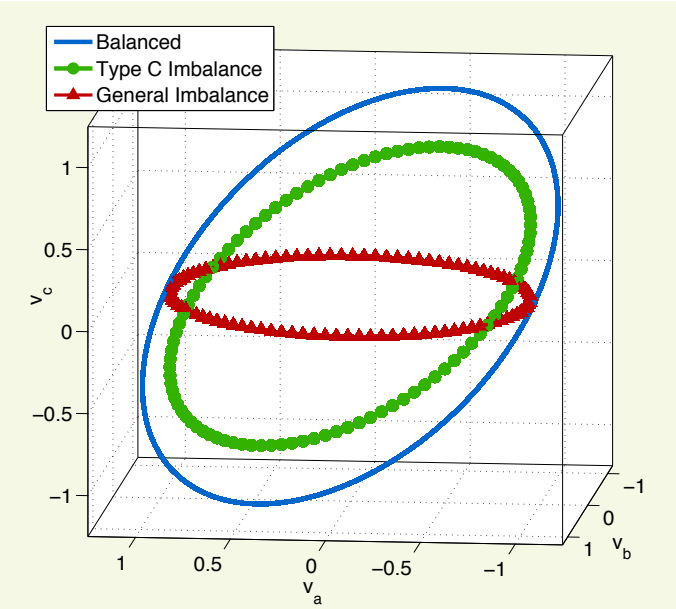

Figure 7.4: Three-phase voltages under balanced conditions, a symmetrically imbalanced condition (Type C Sag), and a general asymmetrical imbalance.

represented using a complex variable

$$s_k \stackrel{\text{def}}{=} v_{\alpha,k} + j v_{\beta,k}. \tag{7.26}$$

Using the complex-valued $\alpha\beta$ voltage in (7.26), the Clarke transform defined in (7.3) can also be represented in the complex domain as

$$s_k = \boldsymbol{c}^{\mathsf{H}} \boldsymbol{s}_k, \quad \boldsymbol{c} \stackrel{\text{def}}{=} \sqrt{\frac{2}{3}} \begin{bmatrix} 1 & e^{-j\frac{2\pi}{3}} & e^{j\frac{2\pi}{3}} \end{bmatrix}^{\mathsf{T}}, \tag{7.27}$$

where $\boldsymbol{c}$ is the complex Clarke transformation vector.

Replacing the phasors of $\boldsymbol{s}_k$ from (7.12) into the Clarke transform in (7.27) allows for the complex $\alpha\beta$ voltage, s_k, to be expressed in terms of the positive– and negative– sequence voltages as

$$s_k = \frac{1}{\sqrt{2}} \left(\bar{V}_+ e^{j\omega k} + \bar{V}_-^* e^{-j\omega k} \right), \tag{7.28}$$

where the scalar phasors $\bar{V}_+$ and $\bar{V}_-^*$ correspond to the positive and negative sequence voltages defined in (7.6) as [131]

$$\begin{aligned} \bar{V}_+ &= \frac{1}{\sqrt{3}} \left[V_a e^{j\phi_a} + V_b e^{j\phi_b} + V_c e^{j\phi_c} \right] \\ \bar{V}_-^* &= \frac{1}{\sqrt{3}} \left[V_a e^{-j\phi_a} + V_b e^{-j\left(\phi_b + \frac{2\pi}{3}\right)} + V_c e^{-j\left(\phi_c - \frac{2\pi}{3}\right)} \right]. \end{aligned} \tag{7.29}$$

Notice that for a balanced system with equal voltage magnitudes, $V_a = V_b = V_c$ and equal phase separations, $\phi_a = \phi_b = \phi_c$, the negative sequence $\bar{V}_- = 0$, which yields the Clark trans-

form output

$$s_k = \frac{1}{\sqrt{2}} \bar{V}_+ e^{j\omega k}. \tag{7.30}$$

In a similar fashion, the complex-valued output of the Park transform introduced in (7.4) is given by

$$v_k \overset{\text{def}}{=} v_{d,k} + j v_{q,k}. \tag{7.31}$$

Therefore, the complex-valued Park transformation can also be compactly represented as

$$v_k = e^{-j\omega_\circ k} \boldsymbol{c}^{\mathsf{H}} \boldsymbol{s}_k = e^{-j\omega_\circ k} s_k. \tag{7.32}$$

where $s_k = v_{\alpha,k} + j v_{\beta,k}$ is the Clarke voltage.

Remark 7.4. *From (7.32), the Park transform can be interpreted as a frequency demodulation (FM) scheme [140] of the $\alpha\beta$ voltage, where the demodulating frequency is chosen as the nominal system frequency $\omega_\circ$.*

It is well known that demodulating a signal with frequency $\omega_\circ$ extract the message which is modulated with the same frequency. Therefore, for a balanced system operating at the fundamental frequency $\omega_\circ$, the Park transform yields the stationary positive sequence phasor

$$v_k = \frac{1}{\sqrt{2}} \bar{V}_+. \tag{7.33}$$

However, for a general unbalanced three-phase system, the complex-valued dq voltage is given by

$$v_k = \frac{1}{\sqrt{2}} \left(\bar{V}_+ e^{j(\omega - \omega_\circ)k} + \bar{V}_-^* e^{-j(\omega + \omega_\circ)k} \right). \tag{7.34}$$

Notice that the system is operating at an off-nominal frequency ω while the Park transform uses the nominal frequency, $\omega_\circ$. On the other hand, if an unbalanced system is operating at the nominal system frequency, $\omega = \omega_\circ$, the dq voltage in (7.34) becomes

$$v_k = \frac{1}{\sqrt{2}} \left(\bar{V}_+ + \bar{V}_-^* e^{-j2\omega_\circ k} \right), \tag{7.35}$$

which resembles a typical output of a FM demodulator.

7.5 Adaptive Clarke/Park Transform

In real-world power systems, the three-phase voltages are rarely perfectly balanced and the system frequency is never exactly the fundamental frequency [19]. Therefore, the Clarke and Park transforms do not yield the ideal intended outputs. To this end, we require adaptive transformations, capable of tracking: i) the voltage imbalances within an "adaptive" Clarke transform, and ii) the system frequency for an "adaptive" Park transform.

We now develop the adaptive Clarke and Park transforms by first introducing the widely linear autoregressive (WLAR) model of the $\alpha\beta$ voltage in (7.28). The WLAR model for the Clarke transform is given by [24]

$$s_k = h^* s_{k-1} + g^* s_{k-1}^*, \tag{7.36}$$

where the WLAR coefficients h and g contain the information of the system frequency, ω, and the level of imbalance in the system defined through the voltage unbalance factor (VUF) given by [19],

$$\kappa \overset{\text{def}}{=} \bar{V}_-/\bar{V}_+. \tag{7.37}$$

Comparing the output of the WLAR model in (7.36) with the actual $\alpha\beta$ voltage in (7.28) gives the relationships [22, 24]

$$e^{j\omega} = h^* + g^*\kappa \quad \text{and} \quad e^{-j\omega} = h^* + \frac{g^*}{\kappa^*}. \tag{7.38}$$

Solving the simultaneous equation in (7.38) gives the expressions for the system frequency and VUF in the form

$$e^{j\omega} = \mathsf{Re}\{h\} + j\sqrt{\mathsf{Im}^2\{h\} - |g|^2}, \tag{7.39}$$

$$\kappa = \frac{\bar{V}_-}{\bar{V}_+} = \frac{j}{g^*}\left(\mathsf{Im}\{h\} + \sqrt{\mathsf{Im}^2\{h\} - |g|^2}\right). \tag{7.40}$$

Contribution: Adaptive Three-Phase Transform
Unlike the static transforms, we next propose an adaptive filter based three-phase transformation in Algorithm 5. The proposed adaptive transform is capable of yielding non-oscillatory outputs even when the system is unbalanced, as shown in Table 7.1. This algorithm is therefore envisioned to be a central component of future control units in the dynamic smart grid.

7.5.1 Balancing Transform

It was shown in [136], that the VUF, κ, defined in (7.37) and (7.40) can be used to eliminate the negative sequence phasor, $\bar{V}_-$, from the $\alpha\beta$ voltage s_k, as

$$\begin{aligned} m_k &\overset{\text{def}}{=} \sqrt{2}\left(s_k - \kappa^* s_k^*\right) \qquad (7.41)\\ &= \bar{V}_+ e^{j\omega k} + \bar{V}_-^* e^{-j\omega k} - \frac{\bar{V}_-^*}{\bar{V}_+^*}\left(\bar{V}_+^* e^{-j\omega k} + \bar{V}_- e^{j\omega k}\right)\\ &= \bar{V}_+ e^{j\omega k} - \frac{|\bar{V}_-|^2}{\bar{V}_+^*} e^{j\omega k}\\ &= \bar{V}_+\left(1 - |\kappa|^2\right) e^{j\omega k}. \qquad (7.42) \end{aligned}$$

Remark 7.5. *The voltage m_k in (7.42) can be regarded as the output of an adaptive Clarke transform*

Algorithm 5. Adaptive Clarke/Park Transform

Input: Three-phase voltages, $\boldsymbol{s}_k$
At each time instant $k > 0$:

1: Obtain Clarke transform

$$s_k = \sqrt{2}\, \boldsymbol{c}^{\mathsf{H}} \boldsymbol{s}_k$$

2: Update ACLMS weights

$$\begin{aligned} \varepsilon_k &= s_k - (h_{k-1}^* s_{k-1} + g_{k-1}^* s_{k-1}^*) \\ h_k &= h_{k-1} + \mu s_{k-1} \varepsilon_k^* \\ g_k &= g_{k-1} + \mu s_{k-1}^* \varepsilon_k^* \\ \kappa_k &= \tfrac{j}{g_k^*} \left(\mathsf{Im}\{h_k\} + \sqrt{\mathsf{Im}^2\{h_k\} - |g_k|^2} \right) \\ e^{j\omega} &= h_k^* + \tfrac{g_k^*}{\kappa_k^*} \end{aligned}$$

3: Adaptive Clarke transform

$$m_k = s_k - \kappa_k^* s_k^*$$

4: Adaptive Park transform

$$\tilde{m}_k = e^{-j\omega k} m_k$$

as it closely resembles the $\alpha\beta$ voltage for a balanced system in (7.30).

The method in [136], requires the evaluation of the second-order statistics of the $\alpha\beta$ voltage s_k. In this work, we show that the WLAR coefficients h and g in (7.38) are sufficient to obtain the adaptive Clarke voltage in (7.41), while also enabling an adaptive Park transform. Specifically, using the estimated time-varying values of $e^{j\omega_k}$ in (7.39) and κ_k in (7.40), the adaptive Clarke and Park transforms are defined as

$$m_k = \sqrt{2}(s_k - \kappa_k^* s_k^*) \tag{7.43a}$$

$$\tilde{m}_k = e^{-j\omega_k k} m_k, \tag{7.43b}$$

where m_k is the adaptive $\alpha\beta$ (Clark) voltage while $\tilde{m}_k$ is the adaptive dq (Park) voltage.

Adaptive Clarke and Park transforms can therefore be implemented using (7.43a)–(7.43b), with a suitable adaptive algorithm (e.g. least mean square, Kalman filter) employed to track the VUF, κ_k, and system frequency ω_k. For simplicity, we present the adaptive Clarke/Park transform in Algorithm 5 which utilises the augmented complex least mean square (ACLMS) to estimate the WLAR coefficients h and g.

Remark 7.6. *The proposed adaptive Clarke and Park transforms are now able to yield non-oscillatory outputs, regardless of the system frequency or level of imbalance, as shown in Table 7.1. This would enable standard techniques designed for nominal conditions*[1] *to be applied to a general set of unbalanced voltages, resulting in a bias-free operation.*

Furthermore, the ease of implementation of the adaptive Clarke-Park transform in Algorithm 5 offers electronic inverter manufacturers a practical method to configure their equipment for unbalanced voltages.

[1]Nominal conditions refer to a balanced three-phase system operating at the system frequency of 50 Hz.

	System Condition	
Transforms	**Balanced**	**Unbalanced**
Clarke [128]	$\bar{V}_+ e^{j\omega k}$	$\bar{V}_+ e^{j\omega k} + \bar{V}_- e^{-j\omega k}$
Balancing [136]	$\bar{V}_+ e^{j\omega k}$	$(1 - \lvert\kappa\rvert^2)\bar{V}_+ e^{j\omega k}$
Park [132]	$\bar{V}_+$	$\bar{V}_+ e^{j(\omega-\omega_\circ)k} + \bar{V}_-^* e^{-j(\omega+\omega_\circ)k}$
This work	$\bar{V}_+$	$(1 - \lvert\kappa\rvert^2)\bar{V}_+$

Table 7.1: *Output of the various three-phase transformations.*

Type	a	b_i	b_q	c_i	c_q
Balanced	1	1	1	1	1
D	R	R	1	R	1

Table 7.2: *Characterisation of different types of voltage imbalances. For the case study in this section,* $R = 0.5$.

7.6 Simulations

The performance of the ACLMS based adaptive Clark/Park transform in Algorithm 5 was tested on a three-phase signal under nominal (balanced voltages and frequency, $f = 50$ Hz) and off-nominal conditions (unbalanced voltages with $f \neq 50$ Hz), over a period of four seconds. For these case studies, we have chosen to illustrate the adaptive Park transform as it simultaneously reveals the performance of both the adaptive Clark and Park transforms. The step-size for the ACLMS was chosen[1] to be $\mu = 0.05$ and the $\alpha\beta$ voltage, s_k, was generated by performing the Clarke transform on a three-phase signal for which the phasors are given by

$$\bar{V}_a = a,\ \bar{V}_b = \frac{1}{2}(-b_i - j\sqrt{3}b_q),\ \bar{V}_c = \frac{1}{2}(-c_i - j\sqrt{3}c_q),$$

The parameters a, b_i, b_q, c_i and c_q were chosen according to Table 7.2, depending on the type of voltage imbalance [25]. The fundamental frequency was 50 Hz and the sampling period $T = 1/500$ s.

In the first case study, the three-phase signal was operating under balanced conditions for the first two seconds and when it was subjected to a Type D sag, with the phasors given in Table 7.2. Figure 7.5 shows the direct and quadrature Park voltages, $v_{d,k}$ and $v_{q,k}$, from both the original Park transform defined in (7.4) and the adaptive Park transform in Algorithm 5. Notice that when the system in unbalanced, the Park voltage oscillates according to (7.35), while the adaptive Park transform was able to obtain a stationary phasor as indicated in Table 7.1.

Figure 7.6 illustrates the second case study where the phase voltages were chosen to be operating under a balanced condition for the whole 4 s period, but the system frequency changed from the nominal frequency of 50 Hz to 53 Hz at 2 s. As shown in (7.34), the original Park transform yielded an oscillating output when the system frequency was at an off-nominal value, while the adaptive Park transform was able to converge to a stationary (non-oscillatory) pha-

[1] The step-size was chosen arbitrarily as the claims in these simulations only serve to demonstrate that the proposed adaptive transform is able to yield non-oscillatory outputs regardless of the operating condition of the three-phase voltages.

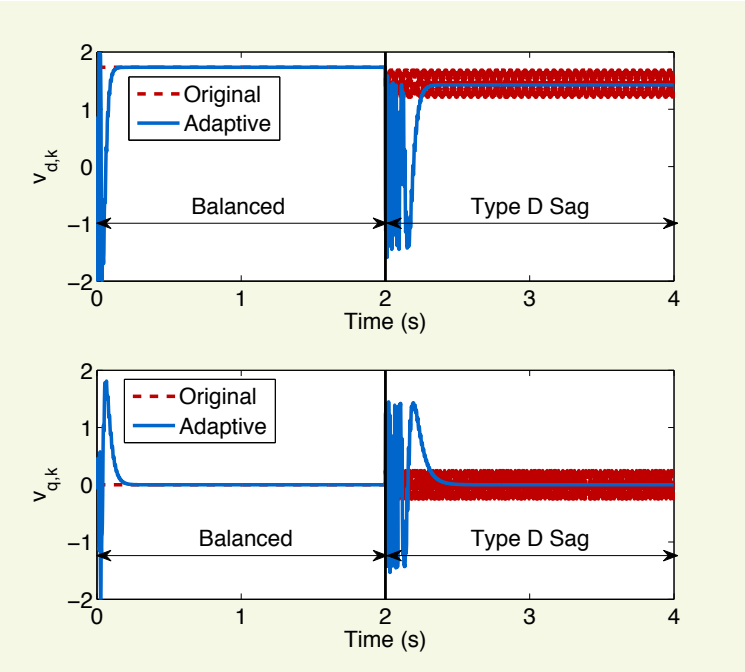

Figure 7.5: Performance of the fixed Park transform in (7.4) (Original) and adaptive Park Transform in Algorithm 5 (Adaptive) under voltage imbalances.

sor after the frequency change point.

7.7 Chapter Summary

We have introduced a modern perspective on the symmetrical, Clarke and Park transforms. The symmetrical transform has been shown to be a spatial DFT applied to an unbalanced set of phasors in (7.7) while the Clarke transform has been derived as a PCA-based optimal dimensionality reduction scheme for a balanced three-phase system in (7.23). Moreover, as shown in (7.32), the Park transform has been interpreted as a frequency demodulation scheme in the complex-domain.

Table 7.3 summarises the signal processing interpretations of the most widely used three-phase transforms.

Transforms	**Interpretations**
Symmetrical [127]	Spatial DFT
Clarke [128]	PCA
Park [132]	FM demodulation

Table 7.3: *Signal processing interpretations of three-phase transformations.*

This has allowed us to show that the Clarke and Park transforms do not yield the optimal dimensionality reduction for unbalanced voltages, a theoretical justification of a frequently experienced phenomenon in practice. To solve this issue for dynamic power systems, we have

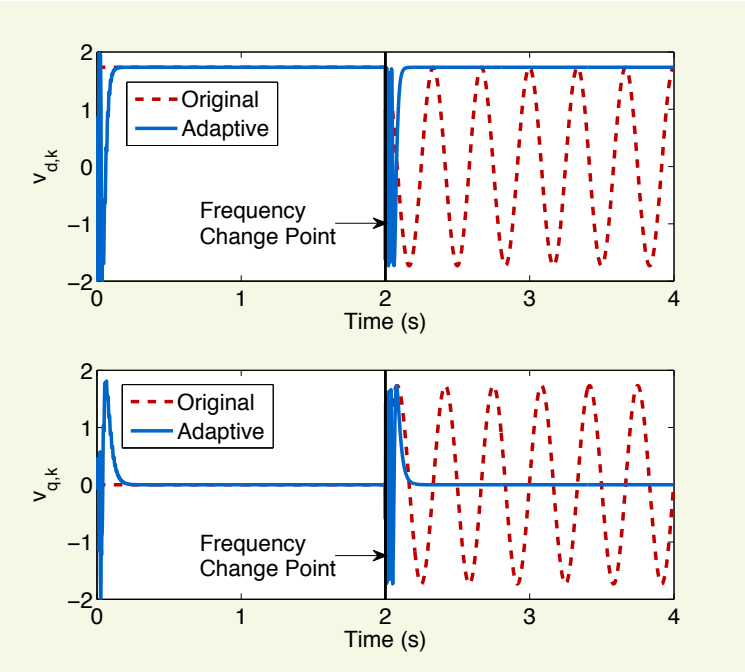

Figure 7.6: Performance of the fixed Park transform in (7.4) (Original) and adaptive Park Transform in Algorithm 5 (Adaptive) under frequency deviations.

proposed a time-varying, adaptive transformation (see Algorithm 5) capable of transforming an unbalanced set of voltages operating at off-nominal frequencies to values that resemble the ideal outputs of the original Clarke and Park transformations of balanced systems. This has opened new possibilities in online dimensionality reduction algorithms and a stabilising platform for accurate and robust signal processing applications which are critical to future low-inertia grids.

In Chapter 8, the transform-domain voltages (particularly the Clarke voltage in 7.28) will be used for the frequency estimation tasks. In particular, the distributed adaptive algorithms developed in Part II will be applied to estimate the frequency, ω. Moreover, we develop strategies for distributed frequency estimation when voltages across different nodes have different levels of imbalance, that is, with different values of positive and negative sequence phasors, V_+ and V_-.

Chapter 8

Distributed Frequency Estimation Examples

> Errors using inadequate data are much less than those using no data at all.
>
> *Charles Babbage*

Chapter Overview

Thus far we have studied adaptive algorithms that employ general linear or non-linear models but have not specified the exact model structure for the frequency tracking problem. In this chapter, we shall investigate specific models for frequency estimation based on the Clarke voltage which was introduced in Chapter 7. It should be noted that several frequency estimation models outlined in this chapter were extended from their original real-valued representations in order to cater for complex-valued sinusoids. Next, the diffusion based extended Kalman filters proposed in Chapter 6 are applied to the problem of frequency estimation in a networked grid. Simulations over synthetic and real-world data will aid the discussion in this chapter.

8.1 Introduction

The task of estimating the instantaneous frequency of a signal has been of practical and theoretical interest for decades and has attracted a plethora of solutions from various fields [141, 142]. This section provides a brief overview of the different methods used to track the instantaneous frequency, with a particular emphasis on adaptive estimation methods.

Recall the complex-valued Clarke voltage embedded in white noise, η_k, is given by

$$y_k = s_k + \eta_k, \qquad s_k = Ae^{j\omega k} + Be^{-j\omega k}, \tag{8.1}$$

where $\omega = 2\pi fT$, is the unknown frequency of interest and T the sampling period and η_k is a complex-valued white Gaussian noise with variance $\mathbb{E}\left\{|\eta_k|^2\right\} = \sigma_\eta^2$. The complex-valued phasors, $A = |A|e^{j\phi_A}$ and $B = |B|e^{j\phi_B}$, are also unknown.

The signal model in (8.1) is a generalisation of single-frequency sinusoids and degenerates into the well-studied cisoid, $s_k = Ae^{j\omega k}$, for $B = 0$, while for $B = A^*$, the signal in (8.1) simplifies into a real valued sinusoid, $s_k = 2|A|\cos(\omega k + \phi_A)$.

The processes of designing parameter estimation algorithms can be separated into the following three categories:

Step 1: Choose a model, for the signal of interest, $\mathcal{M}_{\boldsymbol{w}}$, parametrised by a set of parameters $\boldsymbol{w}$.

Figure 8.1: A general estimation problem.

Step 2: Determine the objective function, $\mathcal{J}_{\boldsymbol{w}}$, to assess the modelling accuracy.

Step 3: Estimate model parameters, $\boldsymbol{w}$, to achieve the desired modelling accuracy, as determined by the objective function in Step 2.

Since the discussion on the mean square error (MSE) cost function and corresponding minimisation algorithms, e.g. stochastic gradient-type methods (CLMS, ACLMS), or state-space methods (CEKF, ACEKF) were focus of the previous chapters, we shall now review several adaptive models used for the frequency estimation problem and extend them to cater for unbalanced Clarke-voltages.

8.2 Linear Prediction of Sinusoids

The basis of many frequency estimation algorithms is the linear prediction of sinusoids. This theory goes back to the work of Prony who exploited the fact that uniformly sampled sums of exponentials can be modelled, hence predicted, using their previous samples [143]. To illustrate this concept, consider the sinusoidal model in (8.1) for $B = 0$, given by[1]

$$\begin{aligned} s_{k+1} &= A_k e^{j\omega(k+1)} \\ &= e^{j\omega} A_k e^{j\omega k} \\ \implies s_{k+1} &= h s_k, \end{aligned} \tag{8.2}$$

where $h = e^{j\omega}$. The model in (8.2), shall be referred to as the *strictly linear autoregressive model* of order one (SLAR-1), and is reminiscent of the complex-valued AR models which were introduced in Chapter 2. The AR modelling of a sinusoid forms the basis of many frequency estimation algorithms as it implies that a sinusoid can be *linearly* predicted from its past samples and that the prediction coefficients are functions of the signal frequencies. For the canonical model in (8.2), the frequency of the signal, ω, and the prediction coefficient is related through

$$\omega = \arg[h] = \tan^{-1}\left[\frac{\mathsf{Im}\{h\}}{\mathsf{Re}\{h\}}\right]. \tag{8.3}$$

However, the SLAR-1 model in (8.2) is inaccurate when the power system is operating under unbalanced conditions, such as when the voltage amplitudes V_a, V_b and V_c are no longer equal or phases are not equally separated, $\phi_a \neq \phi_b \neq \phi_c$.

To overcome the under-modelling issue with the SLAR-1 model for unbalanced systems,

[1] In the context of $\alpha\beta$ voltages, the condition $B = 0$ represents a balanced operating condition where the three phase voltages have the same amplitudes and equal phase separations.

consider again the general case of the signal model in (8.1)

$$s_k = Ae^{j\omega k} + Be^{-j\omega k},$$

At the time-instant, k, adding the previous sample, s_{k-1}, to the subsequent sample, s_{k+1}, yields

$$\begin{array}{rcl} s_{k+1} & = & e^{j\omega}Ae^{j\omega k} + e^{-j\omega}Be^{-j\omega k} \\ + \quad s_{k-1} & = & e^{-j\omega}Ae^{j\omega k} + e^{j\omega}Be^{-j\omega k} \\ \hline s_{k+1} + s_{k-1} & = & \left(e^{j\omega} + e^{-j\omega}\right)Ae^{j\omega k} + \left(e^{j\omega} + e^{-j\omega}\right)Be^{-j\omega k} \\ & = & 2\cos(\omega)\left(Ae^{j\omega k} + Be^{-j\omega k}\right) \\ & = & 2\cos(\omega)s_k, \end{array}$$

so that,

$$s_{k+1} = as_k - s_{k-1}. \tag{8.4}$$

where $a = 2\cos(\omega)$. The model in (8.4) is referred to as the *strictly linear* AR model of order two (SLAR-2), as two previous samples s_k and s_{k-1} are needed to predict the next sample, s_{k+1} [144, 145, 146]. However, the SLAR-2 model exhibits high bias due to modelling inaccuracies in the presence of noise [147]. The frequency is related to the prediction coefficient a as

$$\omega = \cos^{-1}\left[a/2\right]. \tag{8.5}$$

An alternative method to model the general sinusoid in (8.1) is the widely linear model, introduced in Chapter 2. Since the $\alpha\beta$ voltage in (8.1) can be interpreted as the sum of two phasors, one rotating clockwise (positive sequence) and the other rotating counter clockwise (negative sequence) at the same frequency, it is only natural and intuitive to consider the previous value s_k *and* its conjugate s_k^* (where the conjugate represents the phasor rotating in the opposite direction) within the autoregressive model. The widely linear model for the complex-valued $\alpha\beta$ voltage is therefore given by

$$\begin{aligned} \hat{s}_{k+1} &= h^* s_k + g^* s_k^* \\ &= h^*\left(A_k e^{j\omega k} + B_k e^{-j\omega k}\right) + g^*\left(A_k^* e^{-j\omega k} + B_k^* e^{j\omega k}\right) \\ &= \left(h^* A_k + g^* B_k^*\right)e^{j\omega k} + \left(h^* B_k + g^* A_k^*\right)e^{-j\omega k}. \end{aligned} \tag{8.6}$$

Comparing (8.6) to the signal s_{k+1} in (8.1) which is given by

$$s_{k+1} = \left(e^{j\omega}A\right)e^{j\omega k} + \left(e^{-j\omega}B\right)e^{-j\omega k}, \tag{8.7}$$

yields

$$e^{j\omega}A = h^* A + g^* B^*, \quad \text{and} \quad e^{-j\omega}B = h^* B + g^* A^*. \tag{8.8}$$

Solving the simultaneous equation in (8.8) gives

$$e^{j\omega} = \mathsf{Re}\{h\} + j\sqrt{\mathsf{Im}^2\{h\} - |g|^2}.$$

Therefore, the widely linear estimate of the system frequency assumes the form [22, 24, 148]

$$\omega = \tan^{-1}\left[\frac{\sqrt{\mathsf{Im}^2\{h\} - |g|^2}}{\mathsf{Re}\{h\}}\right]. \tag{8.9}$$

The signal model in (8.6) and its corresponding frequency estimate in (8.9) shall be referred to as the widely linear AR-I model (WLAR-I).

Remark 8.1. *From (8.9), notice that when the system is balanced ($B = 0$), the coefficient $g = 0$, and the widely linear frequency estimate is identical to its strictly linear counterpart in (8.3). While the strictly linear AR(2) model in (8.4) is identical for both balanced and unbalanced voltages, the widely linear model provides an intuitive advantage as the coefficient g represents the* negative sequence *which characterises the imbalance of the system voltage.*

8.3 Selecting an Estimator

8.3.1 Linear Adaptive Filters

The first use of adaptive filters to track the frequency of a sinusoid can be traced back to the work of Griffiths who used the LMS to estimate a time-varying AR model [149]. Griffiths did not use the explicit relationship between the AR coefficients of the signal to the actual frequencies, as shown by Prony, but performed this implicitly by tracking the peaks of the power spectrum constructed using the AR model. This development was closely related to adaptive line enhancer (ALE) which uses a prediction configuration to "enhance" a sinusoid. Although the focus of the ALE is its output and not the filter coefficients, the Griffiths' frequency tracker is identical to the Widrow's ALE if the filter order is set to the number of sinusoids present in the signal [150]. An adaptive linear predictor can be configured using the CLMS or ACLMS algorithm introduced in Chapter 2, as

$$\begin{aligned} \varepsilon_k &= y_k - \boldsymbol{w}_k^{\mathsf{H}}\boldsymbol{x}_k \\ \boldsymbol{w}_{k+1} &= \boldsymbol{w}_k + \mu\varepsilon_k^*\boldsymbol{x}_k, \end{aligned} \tag{8.10}$$

where the input $\boldsymbol{x}_k$ and weight vectors $\boldsymbol{w}_k$ take different forms depending on the model used, for more details see Table 8.1.

It is important to note that the ALE in (8.10) is a finite impulse response (FIR) feed-forward structure, and can be extended to infinite impulse response (IIR) feedback structures to increase the frequency resolution and tracking capability at the expense of possible stability and robustness issues [151, 152, 153]. We shall however limit our discussion to feed-forward structures and move on to more robust nonlinear state-space estimators.

Linear Prediction Models		
Model	Input and Weight Vectors	Frequency Estimate
LP-SLAR-1 [154]	$\boldsymbol{x}_k = y_{k-1}$ $\boldsymbol{w}_k = h_k$	$\hat{\omega}_k = \tan^{-1}\left[\frac{\mathsf{Im}\{h_k\}}{\mathsf{Re}\{h_k\}}\right]$
LP-WLAR-I [22]	$\boldsymbol{x}_k = [y_{k-1}, y_{k-1}^*]^{\mathsf{T}}$ $\boldsymbol{w}_k = [h_k, g_k]^{\mathsf{T}}$	$\tilde{a}_k = \sqrt{\mathsf{Im}^2\{h_k\} - \lvert g_k\rvert^2}$ $\hat{\omega}_k = \tan^{-1}\left[\frac{\tilde{a}_k}{\mathsf{Re}\{h_k\}}\right]$
LP-SLAR-2 [147]	$\boldsymbol{x}_k = [y_{k-1}, y_{k-2}]^{\mathsf{T}}$ $\boldsymbol{w}_k = [a_k, -1]^{\mathsf{T}}$	$\hat{\omega}_k = \cos^{-1}\left[\frac{a_k}{2}\right]$

Table 8.1: *Summary of linear prediction models. The prefix "LP"– indicates the models use linear prediction scheme which requires the current observation, y_k and its past observations y_{k-1} and y_{k-2}.*

Contribution: Unification of Nonlinear State Space Frequency Estimators
In Section 8.3.2, we unify the various non-linear state space estimators proposed in the power system frequency tracking literature. Furthermore, we also extend the estimators which were originally proposed for real-valued variables to cater for the complex Clarke voltage. The summary of the estimators is presented in Table 8.2 while their performances are benchmarked in Section 8.4.

8.3.2 Non-Linear State Space Estimators

Nonlinear state-space models have been investigated for frequency tracking tasks since the nonlinear Kalman filters were first introduced [155]. This is due to the fact that nonlinear models provide greater modelling flexibility and may have near optimal performance under noisy scenarios. For generality, consider a nonlinear state space evolution given by

$$\begin{aligned} \boldsymbol{w}_k &= \boldsymbol{f}_{k-1}(\boldsymbol{w}_{k-1}) + \boldsymbol{q}_k, \\ y_k &= \varphi_k(\boldsymbol{w}_k) + \eta_k, \end{aligned} \tag{8.11}$$

where $y_k \in \mathbb{C}$ is the observation (or measurement) of the state vector $\boldsymbol{w}_k \in \mathbb{C}^{M\times 1}$ through a known nonlinear function $\varphi_k(\cdot)$. The state vector $\boldsymbol{w}_k$ is also time-varying with a known state transition function $\boldsymbol{f}_k(\cdot)$. The zero-mean white Gaussian observation noise η_k is independent of the white Gaussian state noise vector $\boldsymbol{q}_k$. The nonlinear function $\varphi_k(\cdot)$ caters for a wide-range of models considered in frequency estimation problems and the most commonly employed models shall be reviewed in this section. The state $\boldsymbol{w}_k$ can be estimated/tracked using an complex (or augmented complex) extend Kalman filter (ACEKF) introduced in Chapter 6.

Nonlinear AR models

We shall first extend the linear and widely linear AR models in Table 8.1 to their nonlinear versions. Firstly, let us consider the SLAR-1 model in (8.2), whereby the signal s_k and frequency

estimate $e^{j\omega}$ are represented as elements in the state vector $\boldsymbol{w}_k = [e^{j\omega}, s_k]^{\mathsf{T}}$. This admits the state transition model given by [156]

$$\begin{aligned} \boldsymbol{w}_k = \begin{bmatrix} e^{j\omega} \\ s_k \end{bmatrix} &= \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & e^{j\omega} \end{bmatrix}}_{\boldsymbol{f}_{k-1}(\boldsymbol{w}_{k-1})} \begin{bmatrix} e^{j\omega} \\ s_{k-1} \end{bmatrix} + \boldsymbol{q}_k \\ y_k &= \underbrace{\begin{bmatrix} 0 & 1 \end{bmatrix}}_{\varphi_k(\boldsymbol{w}_k)} \begin{bmatrix} e^{j\omega} \\ s_k \end{bmatrix} + \eta_k, \end{aligned} \tag{8.12}$$

where $\boldsymbol{f}_{k-1}(\boldsymbol{w}_{k-1})$ is nonlinear due to the presence of the state variable $e^{j\omega}$ within the state transition matrix, while the observation function, $\varphi_k(\boldsymbol{w}_{k-1}) = \begin{bmatrix} 0 & 1 \end{bmatrix}$, is linear and time-invariant.

The state space formulation in (8.12) was first proposed in [156] and was applied to track the power system frequency in [157]. In [156], a complex extended Kalman filter employing the state-space model in (8.12) was shown to be stable and robust due to the weak non-linearity of the state transition function $\boldsymbol{f}_k(\cdot)$. The nonlinear AR-1 model in (8.12) shall be referred to as the SLAR-1 model due to the underlying strictly linear AR model of that is employed.

In a similar fashion, the widely linear model in (8.6) can be configured within a nonlinear state space method whereby the state vector contains the WLAR-I coefficients, $\boldsymbol{w}_k = [h_k^*, g_k^*, s_k]^{\mathsf{T}}$, and the nonlinear state transition model is given by [148]

$$\begin{aligned} \boldsymbol{w}_k = \begin{bmatrix} h_k^* \\ g_k^* \\ s_k \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ s_{k-1} & s_{k-1}^* & 0 \end{bmatrix} \begin{bmatrix} h_{k-1}^* \\ g_{k-1}^* \\ s_{k-1} \end{bmatrix} + \boldsymbol{q}_k \\ y_k &= \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} h_k^* \\ g_k^* \\ s_k \end{bmatrix} + \eta_k, \end{aligned} \tag{8.13}$$

In a similar fashion, the strictly linear AR-2 model (SLAR-2) in (8.4) shall be represented in the form [158]

$$\begin{aligned} \boldsymbol{w}_k = \begin{bmatrix} 2\cos(\omega) \\ s_k \\ s_{k-1} \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2\cos(\omega) & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2\cos(\omega) \\ s_{k-1} \\ s_{k-2} \end{bmatrix} + \boldsymbol{q}_k \\ y_k &= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2\cos(\omega) \\ s_k \\ s_{k-1} \end{bmatrix} + \eta_k. \end{aligned} \tag{8.14}$$

A more physically meaningful nonlinear model which takes inspiration from the WLAR-I and SLAR-1 models is the following. Consider the observed signal, y_k, which can be rep-

resented as a sum of two phasors, ν_k^+ and ν_k^-, (positive and negative sequence) rotating in opposite directions, given by

$$y_k = e^{j\omega} \underbrace{\left(A_{k-1}e^{j\omega(k-1)}\right)}_{\stackrel{\text{def}}{=}\nu_{k-1}^+} + e^{-j\omega} \underbrace{\left(B_k e^{-j\omega(k-1)}\right)}_{\stackrel{\text{def}}{=}\nu_{k-1}^-} + \eta_k.$$

Both phasors, ν_k^+ and ν_k^-, independently obey a strictly linear model where their phases are incremented by ω at the every time instant. Therefore, the process admits a state space model in the form [159]

$$\begin{aligned} \boldsymbol{w}_k = \begin{bmatrix} e^{j\omega} \\ \nu_k^+ \\ \nu_k^- \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{j\omega} & 0 \\ 0 & 0 & e^{-j\omega} \end{bmatrix} \begin{bmatrix} e^{j\omega} \\ \nu_{k-1}^+ \\ \nu_{k-1}^- \end{bmatrix} + \boldsymbol{q}_k \\ y_k &= \begin{bmatrix} 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} e^{j\omega} \\ \nu_k^+ \\ \nu_k^- \end{bmatrix} + \eta_k, \end{aligned} \tag{8.15}$$

and was proposed independently by [159, 160] and [161]. We shall refer to the model in (8.15) as the widely linear AR model-II (WLAR-II), due to its resemblance to the intuition of the widely linear model in (8.13). It is important to observe that the nonlinear state space models in (8.12) – (8.15), which in essense employ the linear prediction (or AR) idea, do not use the signal y_k and its past samples y_{k-1} to predict the frequency. Instead only the signal y_k is used in the observation function while the past sample y_{k-1} is also estimated as a nuisance state.

Nonlinear Sinusoidal Model

The nonlinear model in (8.11) is not restrictive to only the models that arise from the linear prediction property of sinusoids. For example, a straightforward nonlinear state space model for the sinusoidal signal in (8.1) is given by

$$\boldsymbol{w}_k = \begin{bmatrix} A_k \\ B_k \\ \omega_k \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\boldsymbol{f}_{k-1}(\boldsymbol{w}_{k-1})} \begin{bmatrix} A_{k-1} \\ B_{k-1} \\ \omega_{k-1} \end{bmatrix} + \boldsymbol{q}_k \tag{8.16}$$

$$y_k = \underbrace{A_k e^{j\omega_k k} + B_k e^{-j\omega_k k}}_{\varphi_k(\boldsymbol{w}_k)} + \eta_k, \tag{8.17}$$

where the frequency estimate is obtained directly from the state vector $\boldsymbol{w}_k$. The function $\boldsymbol{f}_{k-1}(\cdot) = \mathbf{I}$ in the above model is linear, while observation function, $\varphi_k(\cdot)$, is nonlinear. The state-space model in (8.16) will be referred to as the nonlinear sinusoidal model I (NLS-I). Variants of NLS-I in (8.16) have been employed to estimate the parameters for real-valued sinusoids using the standard extended Kalman filter [162, 163]. Equivalently, recursive Gauss-Newton algorithms have also been used to estimate the NLS-I model parameters [164, 165]. It is worth

mentioning that the recursive Gauss-Newton and EKF are functionally equivalent [166], which is analogous to the duality of the recursive least squares and linear Kalman filter for linear state-space models [167]. The NLS-I model in (8.16), contains an observation function that is time-dependent. To simplify this, the frequency and the phase of the signal can be decoupled via the state transition matrix as [142]

$$\begin{aligned} \boldsymbol{w}_k = \begin{bmatrix} |A_k| \\ |B_k| \\ \phi_k \\ \omega_k \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} |A_{k-1}| \\ |B_{k-1}| \\ \phi_{k-1} \\ \omega_{k-1} \end{bmatrix} + \boldsymbol{q}_k \\ y_k &= |A_k| e^{j\phi_k} + |B_k| e^{-j\phi_k} + \eta_k, \end{aligned} \tag{8.18}$$

so that the observation function $\varphi_k(\boldsymbol{w}_k)$ can be made independent of the time index k. The model in (8.18) will be referred to as the nonlinear sinusoidal model-II (NLS-II). An extended Kalman filter applied to the real-valued version of (8.18) is shown to be the equivalent to phase locked-loop (PLL) estimator [155].

A summary of nonlinear state space models used in tracking the frequency of sinusoid is outlined in Table 8.2.

8.4 Single Node Case Studies

We next assess the models shown in Table 8.2 with the ACEKF algorithm for frequency tracking task under balanced and unbalanced system conditions. Figure 8.2 shows the case where the frequency of the $\alpha\beta$ voltage was initialised at 50.5 Hz but experienced a step-change to 49.5 Hz at 0.15 s. Notice that almost all the algorithms were able to react to the change in frequency except the algorithm that employed the NLS-I model.

In the second case study, an unbalanced $\alpha\beta$ voltage with the same frequency conditions was tested on the different models. The SLAR-1 model was unable to track the correct voltage due to fact that the strictly linear model being inadequate for unbalanced systems.

The steady state behaviour of the algorithms were next benchmarked at various noise levels, as illustrated in Figure 8.4. Notice that the worst performing algorithm is the SLAR-2 model in (8.14). This is a well-known result which states that the SLAR-2 model has poor performance for noisy data [147]. The model which exhibits the best performance was the NLS-I model on the other hand in fact has a very poor tracking ability. We therefore conclude that the widely linear, WLAR-I and WLAR-II models and the nonlinear NLS-II model offer the best trade-off between tracking ability, robustness against noise and steady-state performance. We shall therefore investigate these three models in the case studies on distributed frequency estimation.

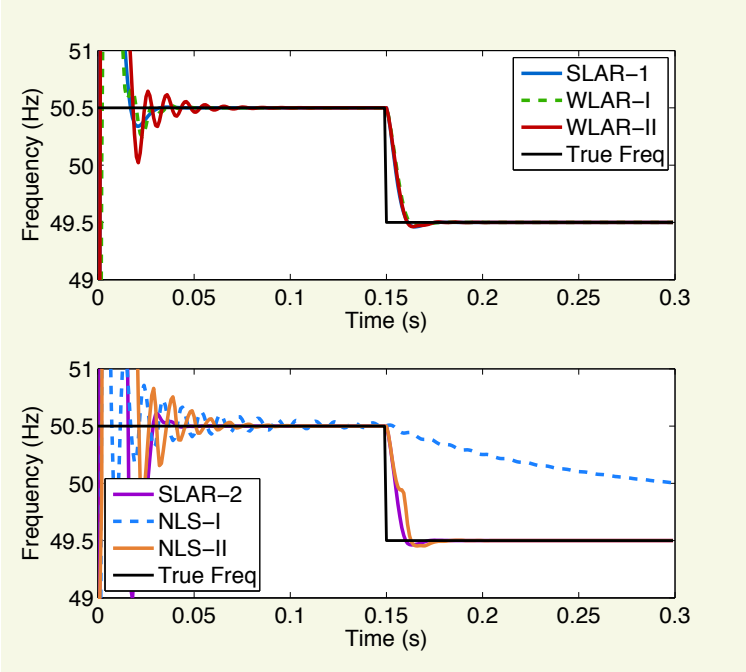

Figure 8.2: Frequency tracking performance for a noise-free signal under a balanced condition.

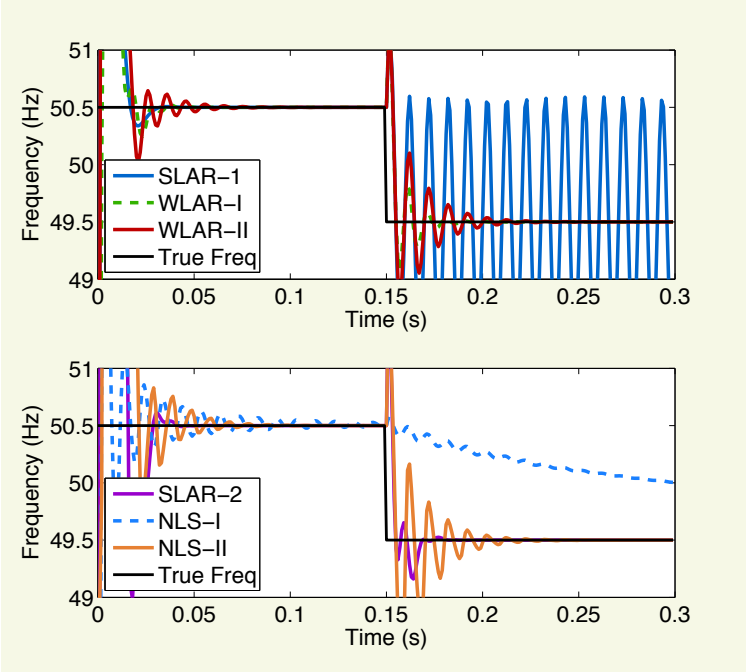

Figure 8.3: Frequency tracking performance under unbalanced conditions for a noise-free signal.

Nonlinear State Space Models		
Model	State Space Models	Frequency Estimate
SLAR-1 [156]	$\begin{bmatrix} h_k \\ s_k \end{bmatrix} = \begin{bmatrix} h_{k-1} \\ h_{k-1}s_{k-1} \end{bmatrix} + \boldsymbol{q}_k$ $y_k = s_k + \eta_k$	$\hat{\omega}_k = \tan^{-1}\left[\frac{\mathsf{Im}\{h_k\}}{\mathsf{Re}\{h_k\}}\right]$
WLAR-I [148]	$\begin{bmatrix} h_k^* \\ g_k^* \\ s_k \end{bmatrix} = \begin{bmatrix} h_{k-1}^* \\ g_{k-1}^* \\ h_{k-1}^* s_{k-1} + g_{k-1}^* s_{k-1}^* \end{bmatrix} + \boldsymbol{q}_k$ $y_k = s_k + \eta_k$	$\tilde{a}_k = \sqrt{\mathsf{Im}^2\{h_k\} - \lvert g_k\rvert^2}$ $\hat{\omega}_k = \tan^{-1}\left[\frac{\tilde{a}_k}{\mathsf{Re}\{h_k\}}\right]$
WLAR-II [159]	$\begin{bmatrix} h_k \\ \nu_k^+ \\ \nu_k^- \end{bmatrix} = \begin{bmatrix} h_{k-1} \\ h_{k-1}\nu_{k-1}^+ \\ h_{k-1}^*\nu_{k-1}^- \end{bmatrix} + \boldsymbol{q}_k$ $y_k = \nu_k^+ + \nu_k^- + \eta_k$	$\hat{\omega}_k = \tan^{-1}\left[\frac{\mathsf{Im}\{h_k\}}{\mathsf{Re}\{h_k\}}\right]$
SLAR-II [158]	$\begin{bmatrix} a_k \\ s_k \\ s_{k-1} \end{bmatrix} = \begin{bmatrix} a_{k-1} \\ a_{k-1}s_{k-1} - s_{k-2} \\ s_{k-1} \end{bmatrix} + \boldsymbol{q}_k$ $y_k = s_k + \eta_k$	$\hat{\omega}_k = \cos^{-1}\left[\frac{a_k}{2}\right]$
NLS-I [162]	$\begin{bmatrix} A_k \\ B_k \\ \omega_k \end{bmatrix} = \begin{bmatrix} A_{k-1} \\ B_{k-1} \\ \omega_{k-1} \end{bmatrix} + \boldsymbol{q}_k$ $y_k = A_k e^{j\omega_k k} + B_k e^{-j\omega_k k} + \eta_k$	$\hat{\omega}_k = \omega_k$
NLS-II [142]	$\begin{bmatrix} \lvert A_k\rvert \\ \lvert B_k\rvert \\ \phi_k \\ \omega_k \end{bmatrix} = \begin{bmatrix} \lvert A_{k-1}\rvert \\ \lvert B_{k-1}\rvert \\ \phi_{k-1} + \omega_{k-1} \\ \omega_{k-1} \end{bmatrix} + \boldsymbol{q}_k$ $y_k = \lvert A_k\rvert e^{j\phi_k} + \lvert B_k\rvert e^{-j\phi_k} + \eta_k$	$\hat{\omega}_k = \omega_k$

Table 8.2: *Summary of the non-linear state space models.*

8.5 Distributed Frequency Estimation Examples

Contribution: Distributed Frequency Estimation
In this section, we apply distributed adaptive filters to exploit spatially diverse voltage measurements to produce rapid frequency estimates. In particular, the performance advantage of the proposed distributed framework for frequency estimation is examined through case studies in model of a real-world distributed power system.

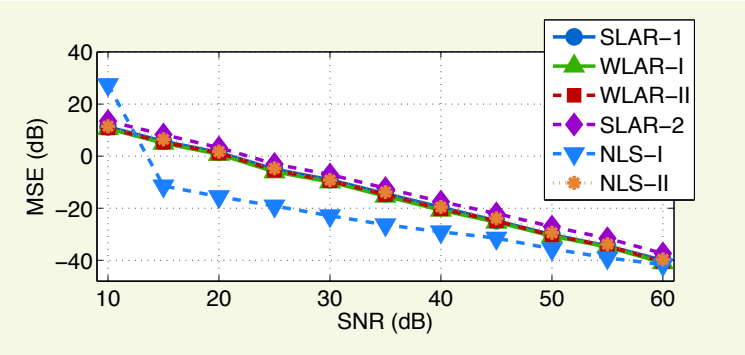

Figure 8.4: Steady State frequency estimation performance.

8.5.1 Experiment (Synthetic Data) Set-Up

The simulations were based on a network of six buses (nodes) where each substation had access to three-phase voltage measurements via transformers with metering capabilities. The number of connections in the network was chosen to be 9 (each node is connected to less than two other nodes) as it reflected the topology of substations in a distribution network. The power system under consideration had a nominal frequency of 50 Hz, and was sampled at a rate of 1 kHz while the signal to noise ratio (SNR) was determined by the metering accuracy class of the potential transformer. The BS EN 61869-1:2009 standard for the metering accuracy of potential transformers considers six separate classes for metering requirements, which translates to an SNR range of 30 dB to 60 dB [168]. To illuminate the robustness of our proposed augmented diffusion Kalman filters, we chose an SNR level of 35 dB in all our simulations, unless stated otherwise.

Each node in the network has access to noisy measurements of a common complex-valued $\alpha\beta$ voltage, that is

$$v_{i,k} = v_{\alpha,k} + jv_{\beta,k} + \eta_{i,k} \quad \eta_i \sim \mathcal{N}(0, \sigma_\eta^2) \tag{8.19}$$

where the measurement noise, $\eta_{i,k}$, is modelled as complex-valued circular white Gaussian noise with the same variance, σ_η^2, at each node. The complex-valued $\alpha\beta$ voltage was generated by applying the Clarke transformation on a three-phase voltage signal

$$\begin{bmatrix} v_{\alpha,k} \\ v_{\beta,k} \end{bmatrix} = \sqrt{\frac{2}{3}} \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \text{Re} \left\{ \begin{bmatrix} \bar{V}_a \\ \bar{V}_b \\ \bar{V}_c \end{bmatrix} e^{j\omega kT} \right\} \tag{8.20}$$

where $\bar{V}_a$, $\bar{V}_b$ and $\bar{V}_c$ are phasor representations of the three-phase voltage. Under a balanced operating condition, $\bar{V}_a = 1$, $\bar{V}_b = -\frac{1}{2} - j\frac{\sqrt{3}}{2}$ and $\bar{V}_c = -\frac{1}{2} + j\frac{\sqrt{3}}{2}$. An imbalance in the system causes voltage sags which can be represented by a change in phasors as shown in Table 8.3 [169].

Voltage Sag	$\bar{V}_a$	$\bar{V}_b$	$\bar{V}_c$
Type A	γ	$-\frac{\gamma}{2} - j\frac{\sqrt{3}\gamma}{2}$	$-\frac{\gamma}{2} + j\frac{\sqrt{3}\gamma}{2}$
Type B	γ	$-\frac{1}{2} - j\frac{\sqrt{3}}{2}$	$-\frac{1}{2} + j\frac{\sqrt{3}}{2}$
Type C	1	$-\frac{1}{2} - j\frac{\sqrt{3}\gamma}{2}$	$-\frac{1}{2} + j\frac{\sqrt{3}\gamma}{2}$
Type D	γ	$-\frac{\gamma}{2} - j\frac{\sqrt{3}}{2}$	$-\frac{\gamma}{2} + j\frac{\sqrt{3}}{2}$

Table 8.3: *Voltage sags and their phasor representations.*

Case Study #1: Noise suppression

Firstly, the performance of the proposed D-ACEKF was examined under noisy voltage inputs. The system voltages were operating under a balanced condition (see Table 8.3) and contaminated with circular white Gaussian noise with an SNR of 35 dB. Figure 8.5 shows the overlay of the frequency estimates from all the six nodes in the network for both the distributed (D-ACEKF) and non-collaborative (ACEKF) algorithms. The distributed algorithm, D-ACEKF, had a better noise attenuation performance compared to its uncooperative counterpart, ACEKF, owing to the diffusion process.

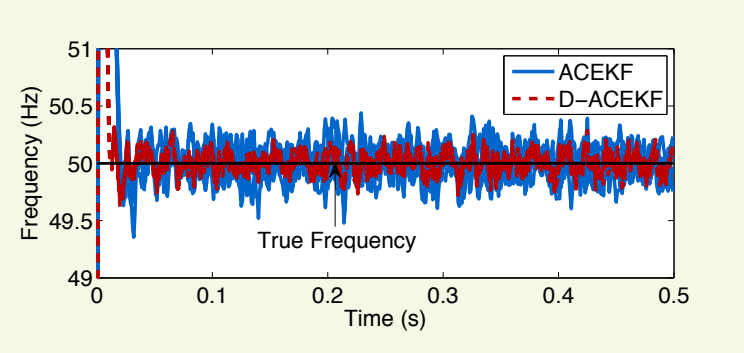

Figure 8.5: The increase in frequency estimation accuracy of the distributed algorithm (D-ACEKF) over the non-cooperative algorithms (ACEKF) when the phase voltages of three nodes in the network are contaminated with random Gaussian noise.

Next, Figure 8.6 illustrates frequency estimation in the presence of random spike noise, which models the switching noise from the inverters which interface the renewable sources to the grid. The distributed algorithm, D-ACEKF, outperformed its uncooperative counterpart, ACEKF, owing to the fact that the spike noise is spread over the network at different times and averaging spiky estimates with cleaner estimates attenuates the level of spike noise contamination.

Case Study #2: Voltage sags

In the second case study, the performances of the algorithms were evaluated for an initially balanced system which became unbalanced after undergoing a Type D voltage sag starting at 0.1 s, followed by a balanced condition starting at 0.3 s. Figure 8.7 shows that, conforming with the analysis, the widely linear algorithm, D-ACEKF, was able to converge to the correct system frequency for both balanced and unbalanced operating conditions, while the strictly

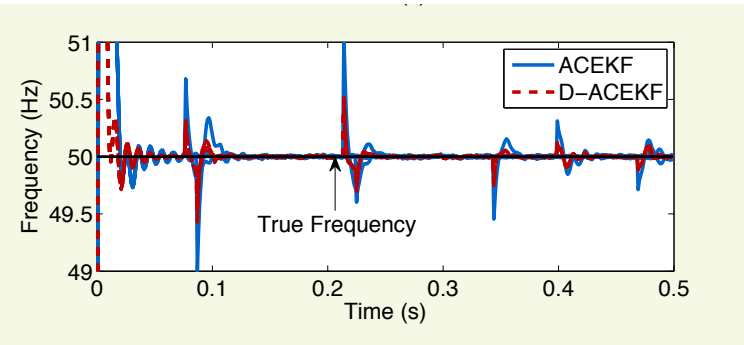

Figure 8.6: Frequency estimation performance of single node (ACEKF) and distributed (D-ACEKF) algorithms when the phase voltages of three nodes in the network are contaminated with random spike noise.

linear algorithm, D-CEKF, was unable to accurately estimate the frequency during the voltage sag due to under-modelling of the system (not accounting for its widely linear nature). As expected, the widely linear and strictly linear algorithms had similar performances under balanced conditions, as illustrated in the time interval 0-0.1 s and 0.3-0.5 s.

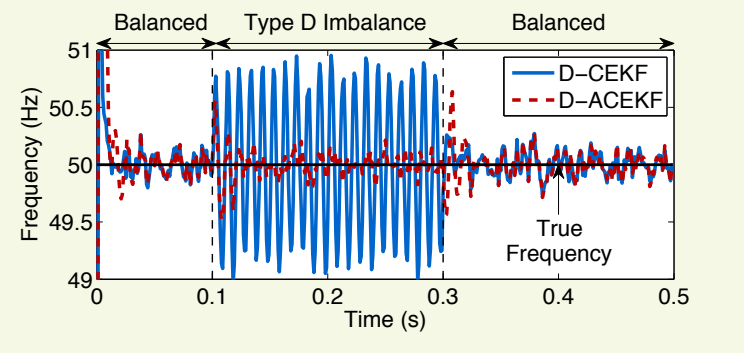

Figure 8.7: Frequency estimation performance of the distributed algorithms (D-CEKF and D-ACEKF) for a system at 35 dB SNR. The system is balanced up to 0.1s, it then undergoes a Type D voltage imbalance followed by a balanced condition at 0.3s.

Furthermore, it is important to note that although the standard diffusion scheme is formulated to share all the states in the network, in the case of frequency tracking in the electricity grid, only the frequency is common throughout the network while the imbalance levels, amplitudes and phase angles are not necessarily the same. Therefore, diffusing other states besides the frequency results in biased estimates.

Figure 8.8 shows the profile of the voltages at different nodes in the network. Each substation underwent different faults, which is reflected in the different relative amplitudes and phase shifts of the $\alpha\beta$ voltages. In addition, Substations 1 and 2 underwent total line failures from 0.3 s to 0.5 s and 0.1 s to 0.3 s respectively.

Figure 8.9 shows distributed implementations of two widely linear state-space models described in the Table 8.2, namely the WLAR-I and WLAR-II models. Notice that both models are capable of estimating the the frequency under balanced and unbalanced conditions. However,

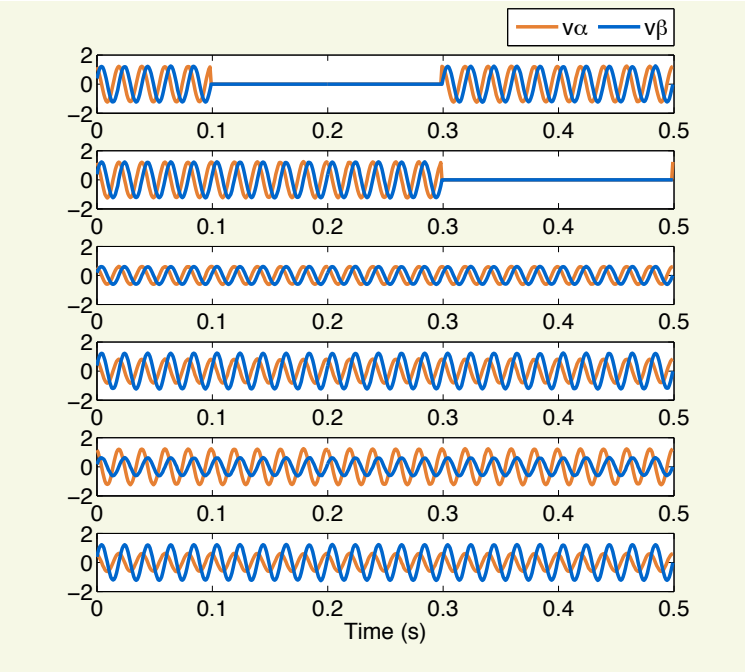

Figure 8.8: Each substation (node) has a different $\alpha\beta$ voltages, including cases where the voltage drops to zero (line cut).

within the WLAR-I model, the frequency is embedded within the coefficients, h and g, which also contain information about the level of imbalance in the system. Diffusing the widely linear coefficients, will therefore lead to biased estimates, since the level of imbalance accross the network is not necessarily the same, as illustrated in Figure 8.8.

However, the WLAR-II model contains the frequency as an isolated state (this is also true for the NLS-II model), which can be diffused across the network. Indeed, Figure 8.9 shows that the diffusion based WLAR-II algorithm was able to estimate the frequency of the network even with different levels of imbalance at each node, while the diffusion based WLAR-I model produces biased estimates.

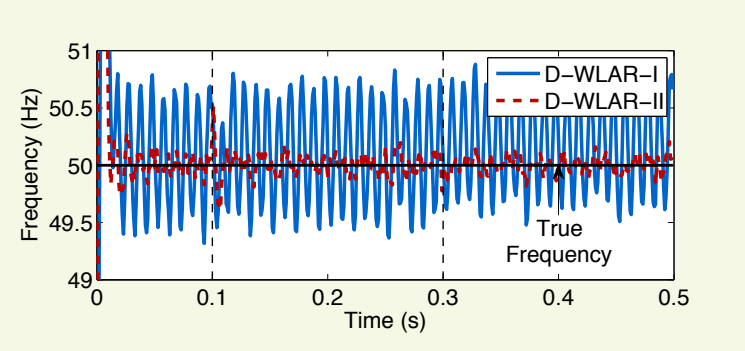

Figure 8.9: The D-ACEKF is able to estimate the frequency in a distributed setting in the presence of different types of faults at each node, see Fig. 8.8 for voltage profiles.

Case Study #3: Frequency variations

Figure 8.10 illustrates the performance the D-ACEKF when a power network was contaminated with white noise at 35 dB and 60 dB SNR and underwent a gradual drop and increase in frequency from 0.1 s to 0.3 s. This is a typical scenario when generation does not match the load and system inertia keeps the frequency from changing too quickly. From 0.3 s to 0.5 s the system undergoes a step-change followed by linear ramp in frequency. The D-ACEKF was able to track the frequency in both cases, illustrating that its suitability for both the current electricity grid and future smart grids.

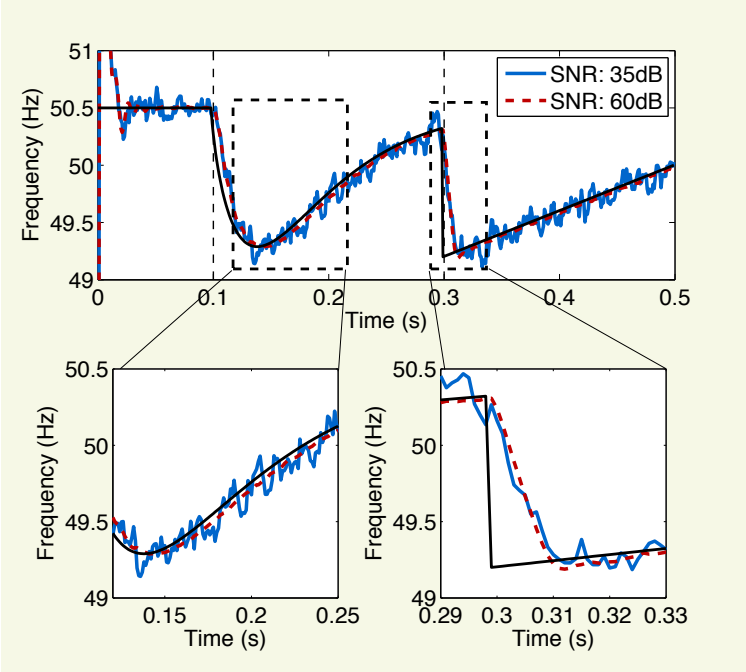

Figure 8.10: Frequency tracking performance of D-ACEKF at 35 dB and 60 dB SNR, which experiences a gradual change in frequency from 0.1 s to 0.3 s and a step change in system frequency to 49.2 Hz at 0.3 s and a linear ramp from 0.3 s to 0.5 s. The solid black line shows the true instantaneous frequency of the voltage. The D-ACEKF is able to track both slow and rapid changes in frequency.

Case Study #4: Steady-state mean square error

Figure 8.11 illustrates the mean square error (MSE) for the proposed distributed frequency estimators. The steady state frequency estimate at a node i for the trial m is denoted by $\hat{f}_{i,ss}[m]$. The mean square error (MSE) of the frequency estimators were calculated over 200 independent trials, as

$$\text{MSE} = \frac{1}{200 \cdot 6} \sum_{m=1}^{200} \sum_{i=1}^{6} \left(\hat{f}_{i,ss}[m] - f_0 \right)^2 \tag{8.21}$$

where $f_0 = 50$ Hz is the fundamental frequency.

The algorithms were evaluated at different SNR levels for an unbalanced system undergoing a Type D voltage sag. Observe that the distributed estimation algorithm outperformed its non-cooperative counterpart, while the only consistent distributed estimator was the proposed D-ACEKF.

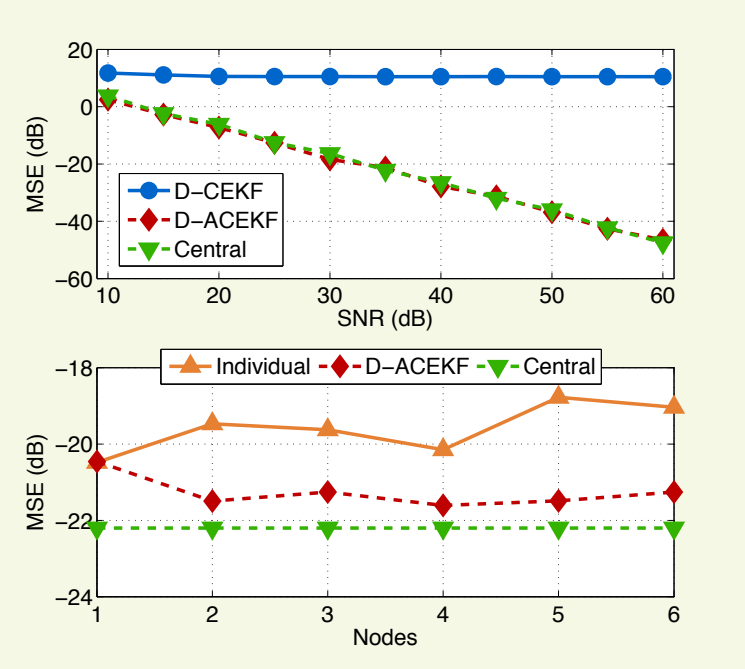

Figure 8.11: *Top panel*: The average MSE of a frequency estimate by D-ACEKF is lower than that of the D-CEKF under a Type D sag. *Bottom panel*: MSE for each node in the network with and without cooperation at SNR = 35 dB shows that the diffusion strategy reduces the steady state error for all the nodes.

Case Study #5: Real world data

We next assessed the performance of the proposed algorithms in a real world case study using three-phase voltage measurements from two adjacent sub-stations in Malaysia during a brief line-to-earth fault. This caused voltage sags, similar to those in Case Study #2. The three-phase measurements were sampled at 5 kHz and the voltage values were normalized. The left panel in Figure 8.12 shows the normalized $\alpha\beta$ voltages at one of the sub-stations. The fault that occurs in phase a, $v_{a,k}$, around 0.1 s is reflected in the voltage dip in $v_{\alpha,k}$. Figure 8.12 shows the frequency estimate from the D-ACEKF which conforms with the analysis and the scenario in Figure 8.7, where the collaborative widely linear D-ACEKF was able to track the real world frequency of a power network under both balanced and unbalanced conditions.

8.6 Chapter Summary

We reviewed the concepts of adaptive frequency estimation and applied the D-ACEKF proposed in Chapter 6 for the task of distributed tracking of the power system frequency for both

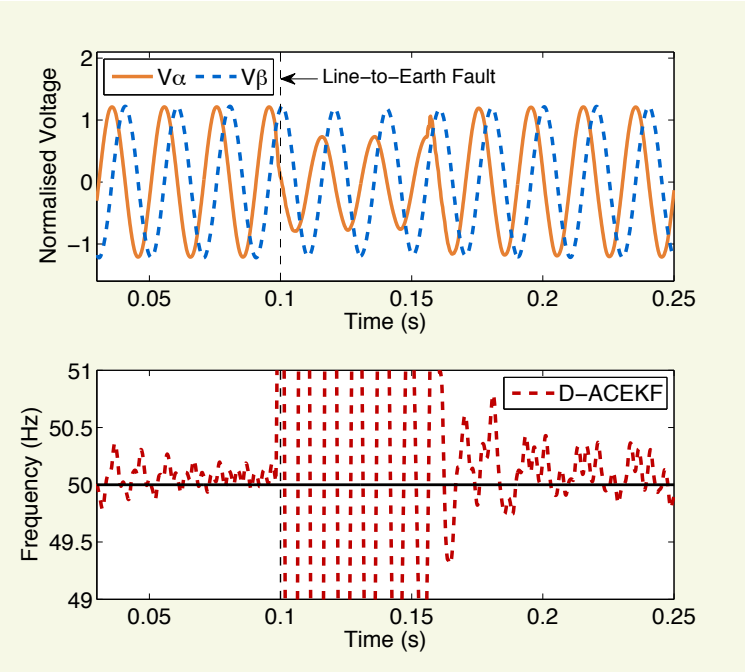

Figure 8.12: *Top*: The $\alpha\beta$ voltages at Sub-station 1 before and during the fault event. *Bottom*: Frequency estimation using the proposed algorithm.

synthetic and real data. This has been accomplished by first developing frequency estimation models for the output of the Clarke voltage in (8.1). For the single node case, amongst the various frequency estimation models outlined in Table 8.2, the WLAR-II and NLS-II models given respectively in (8.15) and (8.18) were the most suited to be effectively implemented in a distributed fashion with the D-ACEKF (Algorithm 4).

Following this insight, the D-ACEKF was tested under a variety of case studies which are summarised in Table 8.4. The origin of their performance advantage has linked to the theoretical performance advantage which was analysed in Chapter 6.

In summary, we have established that exploiting spatially diverse voltage measurements to produce rapid frequency estimates is not only a viable strategy for rapid frequency estimation in dynamic smart grids, but also a physically meaningful tool to deal with a variety of system and observation noises.

Conclusion to Part III

This chapter concludes the third and final part of the thesis. In this part, three-phase signals have been analysed in the context of new signal processing perspectives in Chapter 7, while in Chapter 8, the algorithms that were developed and explored in Parts I and II were applied to the problem of distributed frequency tracking. We have therefore verified both theoretically and through simulation examples that the distributed widely linear framework is natural and best-suited for the applications of fast frequency tracking in low-inertia grids.

Case Study		Origin of Performance Advantage	
1	Noise Suppression	"Contractive" nature of the diffusion coefficients which reduces the steady-state error of the D-ACEKF.	Remark 6.3 , Remark 5.5
2	Voltage Sags	Ability of the widely linear model to model both balanced and unbalanced conditions.	Remark 8.1
3	Frequency Variations	Exponential convergence of the extended Kalman filter allows the frequency to be estimated rapidly.	Theorem 6.1
4	Steady-state error	Diffusion scheme reduces the overall MSE of the individual Kalman filters.	Remark 6.3
5	Real World Case Study	Widely linear modelling and exponential convergence of the D-ACEKF.	Remark 8.1, Theorem 6.1

Table 8.4: *Summary of case studies.*

Chapter 9
Conclusions & Future Work

> No man who values originality will ever be original. But try to tell the truth as you see it, try to do any bit of work as well as it can be done for the work's sake, and what men call originality will come unsought.
>
> *C.S. Lewis*

9.1 Conclusion

This thesis addressed the problem of distributed adaptive filtering for frequency estimation in the electricity grid. The key motivator for this work is the fact that the low inertia grids in the future will require faster frequency estimation algorithms that are robust to a wide range of signal conditions. Drawing upon ideas in complex-valued statistical signal processing and distributed estimation theory, we were able to propose a class of rapid frequency estimators that are suitable for the requirements of smart grid. In the process, several new results in the fields of complex-valued adaptive filtering, distributed adaptive filtering, power system transforms and frequency estimation were introduced. This chapter presents a brief overview of the contributions of the thesis and outlines the directions for future work.

- **Chapter 1: Introduction.** In the first chapter, the distributed frequency estimation problem was was motivated from the perspective of low inertia grids which experience rapid frequency excursions. This context was introduced by reviewing the origins of the AC system frequency, the current frequency regulation paradigm and the anticipated issues with low inertia grids. The introduction also very briefly suggested how recent ideas in complex-valued statistical signal processing and distributed adaptive filtering motivated the casting of the frequency estimation problem into a complex-valued distributed adaptive filtering setting. The details of these ideas were elaborated in the subsequent chapters.

- **Chapter 2: Estimation in the Complex Domain.** This chapter served to provide a background on complex-valued statistical signal processing relevant to the contributions of this thesis. In particular, the chapter introduced the notion of noncircularity of complex-valued random variables and how that effects the estimation of general complex-valued signals. This was explained through widely linear estimation theory and widely linear AR modelling. Adaptive counterparts of the widely linear estimation methodology, referred to as augmented complex adaptive filters were then introduced, together with their statistical properties. To enable the analyses of augmented complex adaptive filters in subsequent chapters (e.g. Chapter 3 and Chapter 5), a matrix factorisation method, termed the approximate uncorrelating transform (AUT) , was also introduced.

- **Chapter 3: Low Complexity Adaptive Filters.** Chapters 3 and 4 presented new contributions to the field of complex-valued adaptive filtering. In particular, a framework for designing low-complexity complex-valued adaptive filters was proposed in Chapter 3. This was accomplished by decomposing the mean square error (MSE) in estimating a complex-valued signal into the MSEs related to estimating the real and imaginary parts of the signal and minimising each MSE independently. This resulted in two separate subfilters, the complex least mean square–r and –i (CLMSr, CLMSi) which are combined to give the dual-channel CLMS (DC-CLMS). The DC-CLMS requires half the computational requirements of the augmented CLMS (ACLMS), while preserving the performance advantage and physical meaning of the ACLMS. The mean convergence of the DC-CLMS was analysed using the AUT introduced in Chapter 2.

- **Chapter 4: Adaptive Tracking of Complex Circularity.** Based on the physical meaning of complex noncircularity, this chapter established a link between the circularity quotient of a complex random variable to the MSE estimation of the complex-random variable from its conjugate. This link was used to propose a simple CLMS-based single-tap adaptive filter which was able to track the circularity quotient of a signal in real-time. The proposed circularity tracker was used in a wireless communication setting to detect rectilinear signals.

- **Chapter 5: Diffusion Complex Least Mean Square.** The focus of the next two chapters shifted towards distributed adaptive filtering strategies. In this chapter (Chapter 5), the ACLMS originally introduced in Chapter 2, was extended to a distributed setting using the so-called "diffusion strategy. The main contribution of the chapter was the mean square analysis of the D-ACLMS algorithm which was accomplished through the use of a proposed similarity conjecture. The proposed analysis removed the need to make a restrictive simplifying assumption on the step-sizes of the algorithm, while maintaining the mathematical tractability and the physical meaning of the problem.

- **Chapter 6: Diffusion Complex Extended Kalman Filter.** This chapter developed a diffusion augmented complex extended Kalman filter (D-ACEKF) for the distributed estimation of non-linear state space models. The D-ACEKF generalised the D-ACLMS introduced in Chapter 5 for non-linear state space representations, while also catering for widely linear models and noncircular signals. The flexibility of non-linear state space models for a wide range of estimation tasks (including frequency estimation in low inertia grids) was the main motivation for the development of the D-ACEKF. The chapter also provided a brief review of current distributed Kalman filtering algorithms and convergence analysis of the D-ACEKF.

- **Chapter 7: Modern View of Three-Phase Transforms.** In this chapter, we re-examined well-known three-phase transforms commonly employed in the power engineering community, by giving them a modern signal processing perspective. The symmetrical, Clarke and Park transforms were derived as a DFT, PCA and FM demodulation schemes respectively. These interpretations were facilitated using complex-domain representations of the transforms. The Clarke and Park transforms were shown to yield non-optimal out-

puts for unbalanced systems operating at off-nominal frequencies. To resolve this issue, we also proposed an adaptive three-phase transform which was able to yield outputs that resemble the outputs of a balanced system, regardless of the operating condition of the system.

- **Chapter 8: Distributed Frequency Tracking Case Studies.** This chapter validated the proposed distributed estimation algorithms from Chapter 5 and Chapter 6 for a range of frequency estimation task. We first developed the linear, widely linear and non-linear models for three-phase signals, drawing upon ideas introduced in Chapter 7. Real-world data in the form of wide-area measurements of three-phase voltages during a fault was also used to validate the algorithms.

9.2 Future Work

While we have demonstrated the potential of using distributed frequency tracking in smart grids, many opportunities for extending the scope of this thesis are yet to be explored. In this section, we shall briefly outline some pertinent ideas and survey several questions that allow for the generalisation of the concepts explored thus far.

Incorporation of topographical information. The performance of distributed frequency estimation algorithms can be enhanced by incorporating relevant *a priori* information about the network topography within the state space formulation. For example, information about the impedance between each node and the level of inertia in a particular region were shown to influence the frequency deviation [13]. Therefore, exploiting such information would serve to improve the estimation accuracy of the proposed algorithms. This work shall also serve to extend the real world case study in Chapter 8 to incorporate more busbars to reflect a wide area distribution network.

Frequency and fault prediction. The rapid frequency excursions in low inertia grids also necessitate the prediction of faults and their corresponding frequency deviations. In addition to estimating the frequency, predicting any frequency excursions will help grid operators anticipate unexpected events. This, in turn, can be fed into predictive control schemes that automatically make adjustments to stabilise the system. Besides the ideas presented in this thesis, prediction tasks draw upon the advancements in machine learning techniques like recurrent neural networks [153] and kernel methods [170, 171].

Distributed state estimation. The distributed state space formulation of the D-ACEKF in Chapter 6 can be extended to solve a wider range problems in the electricity grid besides frequency estimation. The most pertinent of these problems is that of distributed state estimation. Grid operators require accurate estimation of the state of the electricity grid observed from measurements of active/reactive power flows, voltage magnitudes, and phase angles from wide-area measurement devices [27]. The complex-valued and multi-modal natures of the relevant quantities indicate that the D-ACEKF introduced in the thesis is a prime candidate to solve this problem.

Joint phasor estimation and fault detection estimation. Closely related to the state estimation problem is the development of algorithms for the estimation of phasors (voltage magni-

tude and angles) [172]. These estimation algorithms are required for the next generation phasor measurement units (PMU) which are required to adhere to the recent IEEE measurement and protection standards [173]. The widely linear model-based PMU algorithms also enable the joint estimation of system frequency and imbalance conditions [24].

Convergence analysis for general stochastic gradient algorithms. The similarity conjecture proposed in Chapter 5 was shown to simplify the mean square analysis of the D-CLMS and D-ACLMS algorithms. We would next like to generalise this analysis to cater for all distributed stochastic gradient algorithms that are based on the diffusion strategy [97]. This would be an important contribution as the similarity conjecture enables the derivation of a performance bound for the distributed case based on the analysis of a single-node algorithm. This would also extend the work to distributed classification and fault detection schemes by virtue of the generality of the stochastic gradient framework.

Asynchronous distributed adaptive filters. A key limitation in the current framework for distributed adaptive filters is that they operate under the condition which imposes the synchronisation of the measurement, adaptation and communication steps. This condition is a theoretical simplification which is rarely reflected in real-world sensor networks. Moreover, current methods that take into account the asynchronicity start with the synchronised setting and impose random link failures on the communication between the nodes [174]. To this end, distributed adaptive algorithms have to reformulated in an asynchronous setting where it is possible to design least squares based adaptive filters which operate optimally in an asynchronous setting [175, 176]. Performance bounds describing the trade-off between the communication requirements, computational complexity and estimation accuracy will be an integral part of this future research direction.

Connections with Aumann's agreement theorem. The consensus-type diffusion strategies discussed in this thesis have clear connections with the modelling of rational agents which is widespread in economics, philosophy and artificial intelligence [177]. These fields study the behaviour of rational agents that update their prior beliefs on certain states based on mutual observations of each others states. For example, a seminal result in economics known as Aumann's agreement theorem states that two rational Bayesian agents initialised with the same priors but which update their beliefs based on different observations should converge to the same posterior when they both are made aware of each others posterior [178]. This result can be extended to the diffusion Kalman filter in Chapter 6, so that each node should achieve consensus on their state estimates by just sharing their state estimates and not their collective observations. Exploring Aumann's agreement theorem under the distributed adaptive filtering setting would have many implications in the communication and computational complexity of the distributed algorithms [177]. This idea was explored in our recent work, where we introduced a diffusion Kalman filter which only exchanged state estimates and their corresponding state error covariance matrices [179]. The proposed diffusion Kalman filter greatly reduced the communications requirements of the network while maintaining near-optimal performance.

Extensions to biomedical signal processing. The proposed distributed frequency estimation framework can be reformulated for the adaptive acquisition of foetal electrocardiogram (fECG) signals [180]. Formulating the fECG extraction problem in a sensor network framework will allow the redundancy present in multi-lead ECG data to be exploited so as to reliably

identify and extract fECG from artefact-contaminated recordings. Parallels between exploiting wide area voltage measurements in the smart grid and multi-channel ECG data can be drawn here, whereby spatial diversity can be exploited to remove unwanted noise sources (e.g. maternal ECG, muscle and movement artefact) from the desired fECG. This shall be an important contribution as it will enable the prediction and subsequent prevention of stillbirths that occur in approximately 1 in every 200 live births in the U.K. [181, 182].

9.3 Summary of the Connections Between Concepts

An overarching theme of the contributions of this thesis is the unification of various concepts encountered in statistical signal processing. Through intuitive derivations, we have bridged the gap between various concepts in signal processing, communications and power engineering. Table 9.1 presents a brief overview of these connections.

	Concepts		**Thesis Chapter**
1	Adaptive filtering in $\mathbb{R}^2$	Widely linear filtering in $\mathbb{C}$	Chapter 3
2	Wiener solution for estimating a signal from its complex conjugate	Circularity quotient of a signal	Chapter 4
3	Symmetrical transform	Spatial discrete Fourier transform (DFT)	Chapter 7
4	Clarke transform	Principal component analysis (PCA)	Chapter 7
5	Park transform	Frequency demodulation (FM)	Chapter 7
6	Least mean square (LMS)	Kalman filter	Appendix A
7	Recursive DFT	FM Demodulation	Appendix B

Table 9.1: *Connections between various concepts introduced in the thesis.*

Appendix A

Intrinsic Relationship Between the Kalman Filter and LMS

A.1 Perspective

The Kalman filter and the least mean square (LMS) adaptive filter are two of the most popular adaptive estimation algorithms which are often used interchangeably in a number of statistical signal processing applications. They are typically treated as separate entities, the former as a realisation of the optimal Bayesian estimator and the latter as a recursive solution to the optimal Wiener filtering problem. In this appendix, we consider a system identification framework within which we develop a joint perspective on Kalman filtering and LMS-type algorithms, achieved through the analysis of the degrees of freedom necessary for optimal stochastic descent adaptation. This approach permits the introduction of Kalman filters without any notion of Bayesian statistics which may be beneficial for many communities which do not rely on Bayesian methods [183, 184].

There are several, and not immediately patent, aspects of common thinking between gradient descent and recursive state-space estimators. Because of their non-obvious or awkward nature, these are often overlooked. It is hoped that the framework presented in this article, with the seamless transition between least mean square and Kalman filters, will provide a straightforward and unifying platform for understanding the geometry of learning and optimal parameter selection in these approaches. In addition, the material may be useful in lecture courses in statistical signal processing, or indeed, as interesting reading for the intellectually curious and generally knowledgeable reader.

A.2 Problem Formulation

We consider a generic system identification setting

$$y_k = \boldsymbol{x}_k^{\mathsf{T}} \boldsymbol{w}_k^{\mathrm{o}} + \eta_k, \tag{A.1}$$

where the aim to estimate the unknown system parameter vector, $\boldsymbol{w}_k^{\mathrm{o}}$ (weight vector), which characterises the system in (A.1) from observations, y_k, corrupted by noise, η_k. This parameter vector can be fixed, that is $\boldsymbol{w}_k^{\mathrm{o}} = \boldsymbol{w}^{\mathrm{o}}$, or time-varying as in (A.1), while $\boldsymbol{x}_k$ is the input vector and η_k denotes a zero-mean white Gaussian process with variance $\sigma_n^2 = \mathbb{E}\left\{\eta_k^2\right\}$. For simplicity,

we assume that all signals are real valued.

To assist a joint discussion of state space and regression-type models, Table A.1 lists the terms commonly used across different communities for the variables in the system identification paradigm in (A.1).

Area	y_k	$\boldsymbol{x}_k$	$\boldsymbol{w}_k^{\mathrm{o}}$
Adaptive filtering	Desired signal	Input *or* regressor	True weight
Kalman filtering	Observation	Measurement	State vector
Machine learning	Target	Features	Hypothesis parameters

Table A.1: *Terminology used in different communities.*

In this section we address the deterministic and time-invariant case, $\boldsymbol{w}_k^{\mathrm{o}} = \boldsymbol{w}^{\mathrm{o}}$, while Section A.5 deals with a stochastic and time-varying system parameter vector (general Kalman filter).

A.2.1 Performance Evaluation Criteria

Consider observations from an unknown deterministic system

$$y_k = \boldsymbol{x}_k^{\mathsf{T}} \boldsymbol{w}^{\mathrm{o}} + \eta_k. \tag{A.2}$$

We desire to estimate the true parameter vector $\boldsymbol{w}^{\mathrm{o}}$ recursively, based on the existing weight vector estimate $\boldsymbol{w}_{k-1}$ and the observed and input signals, that is, $\widehat{\boldsymbol{w}}^{\mathrm{o}} = \boldsymbol{w}_k = f(\boldsymbol{w}_{k-1}, y_k, \boldsymbol{x}_k)$. Notice that $\boldsymbol{w}_{k-1}, y_k, \boldsymbol{x}_k$ are related through the output error

$$\varepsilon_k = y_k - \boldsymbol{x}_k^{\mathsf{T}} \boldsymbol{w}_{k-1}. \tag{A.3}$$

Performance of statistical learning algorithms is typically evaluated based on the mean square error (MSE) criterion, which is defined as the output error power and is given by

$$\mathrm{MSE} = \xi_k \stackrel{\mathrm{def}}{=} \mathbb{E}\left\{\varepsilon_k^2\right\}. \tag{A.4}$$

Since our goal is to estimate the true system parameters, it is natural to also consider the weight error vector

$$\tilde{\boldsymbol{w}}_k \stackrel{\mathrm{def}}{=} \boldsymbol{w}^{\mathrm{o}} - \boldsymbol{w}_k, \tag{A.5}$$

and its contribution to the output error, given by

$$\varepsilon_k = \boldsymbol{x}_k^{\mathsf{T}} \tilde{\boldsymbol{w}}_{k-1} + \eta_k. \tag{A.6}$$

Without loss in generality, we here treat $\boldsymbol{x}_k$ as a deterministic process although in adaptive filtering convention it is assumed to be a zero-mean stochastic process with covariance matrix $\mathbf{R} = \mathbb{E}\left\{\boldsymbol{x}_k \boldsymbol{x}_k^{\mathsf{T}}\right\}$. Our assumption conforms with the Kalman filtering literature where the vector $\boldsymbol{x}_k$ is often deterministic (and sometimes even time-invariant). Replacing the output error from (A.6) into (A.4) gives

$$\xi_k = \mathbb{E}\left\{(\boldsymbol{x}_k^{\mathsf{T}} \tilde{\boldsymbol{w}}_{k-1} + \eta_k)^2\right\} = \boldsymbol{x}_k^{\mathsf{T}} \mathbf{M}_{k-1} \boldsymbol{x}_k + \sigma_n^2 \tag{A.7a}$$

$$\stackrel{\text{def}}{=} \xi_{\text{ex},k} + \xi_{\text{min}}, \tag{A.7b}$$

where $\mathbf{M}_{k-1} \stackrel{\text{def}}{=} \mathbb{E}\left\{\tilde{\boldsymbol{w}}_{k-1} \tilde{\boldsymbol{w}}_{k-1}^{\mathsf{T}}\right\}$ is the symmetric and positive semi-definite weight error covariance matrix and the noise process n_k is assumed to be statistically independent from all other variables. Therefore, for every recursion step, k, the corresponding MSE denoted by ξ_k comprises two terms: (i) the time-varying excess mean square error (EMSE), $\xi_{\text{ex},k}$, which reflects the misalignment between the true and estimated weights (function of the performance of the estimator), and (ii) the observation noise power, $\xi_{\text{min}} = \sigma_n^2$, which represents the minimum achievable mean square error (for $\boldsymbol{w}_k = \boldsymbol{w}^{\text{o}}$) and is independent of the performance of the estimator.

Our goal is to evaluate the performance of a learning algorithm in identifying the true system parameters, $\boldsymbol{w}^{\text{o}}$, and a more insightful measure of how closely the estimated weights, $\boldsymbol{w}_k$, have approached the true weights, $\boldsymbol{w}^{\text{o}}$, is the mean square deviation (MSD), which represents the power of the weight error vector and is given by

$$\text{MSD} = J_k \stackrel{\text{def}}{=} \mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_k\|^2\right\} = \mathbb{E}\left\{\tilde{\boldsymbol{w}}_k^{\mathsf{T}} \tilde{\boldsymbol{w}}_k\right\} = \mathsf{Tr}[\mathbf{M}_k]. \tag{A.8}$$

Observe that the MSD is related to the MSE in (A.7a) through the weight error covariance matrix, $\mathbf{M}_k = \mathbb{E}\left\{\tilde{\boldsymbol{w}}_k \tilde{\boldsymbol{w}}_k^{\mathsf{T}}\right\}$.

> **Contribution: Connecting the LMS and Kalman Filter**
> The contribution of this appendix is the novel derivation of the Kalman filter via the optimal learning rate of the LMS.

A.3 Optimal Learning Gain for Stochastic Gradient Algorithms

The LMS algorithm employs stochastic gradient descent to approximately minimise the MSE in (A.4) through a recursive estimation of the optimal weight vector, $\boldsymbol{w}^{\text{o}}$ in (A.2), in the form $\boldsymbol{w}_k = \boldsymbol{w}_{k-1} - \mu_k \nabla_{\boldsymbol{w}} \mathbb{E}\left\{\varepsilon_k^2\right\}$. Based on the instantaneous estimate $\mathbb{E}\left\{\varepsilon_k^2\right\} \approx \varepsilon_k^2$, the LMS solution is then given by [185]

$$\text{LMS}: \qquad \boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \Delta \boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \mu_k \boldsymbol{x}_k \varepsilon_k. \tag{A.9}$$

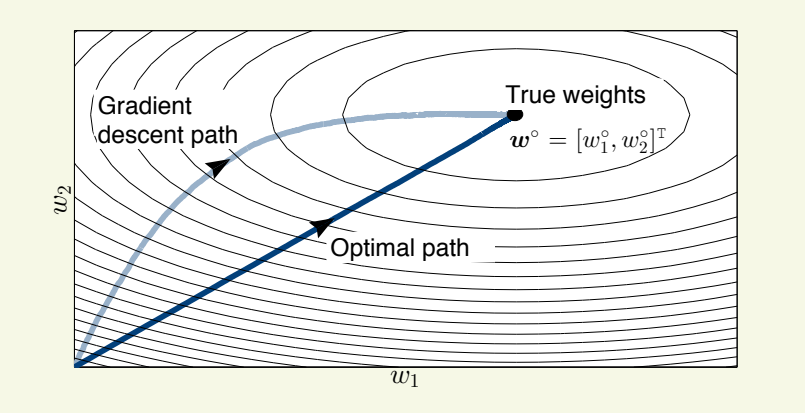

Figure A.1: Mean trajectories of an ensemble of noisy single-realisation gradient descent paths for correlated data. The LMS path, produced based on (A.9), is locally optimal but globally slower converging than the optimal path.

The parameter μ_k is a possibly time-varying positive step-size which controls the magnitude of the adaptation steps the algorithm takes; for fixed system parameters this can be visualised as a trajectory along the error surface – the MSE plot evaluated against the weight vector, $\xi_k(\boldsymbol{w})$. Notice that the weight update $\Delta\boldsymbol{w}_k = \mu_k \boldsymbol{x}_k \varepsilon_k$ has the same direction as the input signal vector $\boldsymbol{x}_k$, which makes the LMS sensitive to outliers and noise in data. Figure A.1 illustrates the geometry of learning of gradient descent approaches for correlated data (elliptical contours of the error surface) – gradient descent performs locally optimal steps but has no means to follow the globally optimal shortest path to the solution, $\boldsymbol{w}^{\circ}$. It is therefore necessary to control both the direction and magnitude of adaptation steps for an algorithm to follow the shortest, optimal, path to the global minimum of error surface, $\xi(\boldsymbol{w}^{\circ})$.

The first step towards Kalman filters is to introduce more degrees of freedom by replacing the scalar step-size, μ_k, with a positive definite learning gain matrix, $\mathbf{G}_k$, so as to control both the magnitude and direction of the gradient descent adaptation, and thus follow the optimal path in Figure A.1. In this way, the weight update recursion in (A.9) now generalises to

$$\boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \mathbf{G}_k \boldsymbol{x}_k \varepsilon_k. \tag{A.10}$$

Unlike standard gradient-adaptive step-size approaches which minimise the MSE via $\partial\xi_k/\partial\mu_k$ [186, 187], our aim is to introduce an optimal step-size (and learning gain) into the LMS based on the direct minimisation of the MSD in (A.8). For convenience, we consider a general recursive weight estimator

$$\boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \boldsymbol{g}_k \varepsilon_k, \tag{A.11}$$

which represents both (A.9) and (A.10), where the gain vector

$$\boldsymbol{g}_k \overset{\text{def}}{=} \begin{cases} \mu_k \boldsymbol{x}_k, & \text{for the conventional LMS in (9),} \\ \mathbf{G}_k \boldsymbol{x}_k, & \text{for a general LMS in (10).} \end{cases} \tag{A.12}$$

To minimise the MSD, given by $J_k = \mathbb{E}\left\{\|\tilde{\boldsymbol{w}}_k\|^2\right\} = \mathsf{Tr}[\mathbf{M}_k]$, we first establish the weight error vector recursion for the general LMS by subtracting $\boldsymbol{w}^{\text{o}}$ from both sides of (A.11) and replacing the output error with $\varepsilon_k = \boldsymbol{x}_k^{\mathsf{T}} \tilde{\boldsymbol{w}}_{k-1} + \eta_k$, to give

$$\tilde{\boldsymbol{w}}_k = \tilde{\boldsymbol{w}}_{k-1} - \boldsymbol{g}_k \boldsymbol{x}_k^{\mathsf{T}} \tilde{\boldsymbol{w}}_{k-1} - \boldsymbol{g}_k \eta_k. \tag{A.13}$$

The recursion for the weight error covariance matrix, $\mathbf{M}_k$, is then established upon post-multiplying both sides of (A.13) by their respective transposes and applying the statistical expectation operator $\mathbb{E}\{\cdot\}$ to both sides, to yield

$$\begin{aligned} \mathbf{M}_k = \mathbb{E}\left\{\tilde{\boldsymbol{w}}_k \tilde{\boldsymbol{w}}_k^{\mathsf{T}}\right\} = \mathbf{M}_{k-1} - \left(\mathbf{M}_{k-1} \boldsymbol{x}_k \boldsymbol{g}_k^{\mathsf{T}} + \boldsymbol{g}_k \boldsymbol{x}_k^{\mathsf{T}} \mathbf{M}_{k-1}\right) \\ + \boldsymbol{g}_k \boldsymbol{g}_k^{\mathsf{T}} \left(\boldsymbol{x}_k^{\mathsf{T}} \mathbf{M}_{k-1} \boldsymbol{x}_k + \sigma_n^2\right). \end{aligned} \tag{A.14}$$

Using the well known matrix trace identities, $\mathsf{Tr}[\mathbf{M}_{k-1} \boldsymbol{x}_k \boldsymbol{g}_k^{\mathsf{T}}] = \mathsf{Tr}[\boldsymbol{g}_k \boldsymbol{x}_k^{\mathsf{T}} \mathbf{M}_{k-1}] = \boldsymbol{g}_k^{\mathsf{T}} \mathbf{M}_{k-1} \boldsymbol{x}_k$ and $\mathsf{Tr}[\boldsymbol{g}_k \boldsymbol{g}_k^{\mathsf{T}}] = \boldsymbol{g}_k^{\mathsf{T}} \boldsymbol{g}_k = \|\boldsymbol{g}_k\|^2$, the MSD evolution, $J_k = \mathsf{Tr}[\mathbf{M}_k]$, is obtained as

$$J_k = J_{k-1} - 2\boldsymbol{g}_k^{\mathsf{T}} \mathbf{M}_{k-1} \boldsymbol{x}_k + \|\boldsymbol{g}_k\|^2 \left(\boldsymbol{x}_k^{\mathsf{T}} \mathbf{M}_{k-1} \boldsymbol{x}_k + \sigma_n^2\right). \tag{A.15}$$

A.3.1 Optimal scalar step-size for LMS

The standard optimal step-size approach to the LMS aims at achieving $\varepsilon_{k+1|k} = y_k - \boldsymbol{x}_k^{\mathsf{T}} \boldsymbol{w}_k = 0$, where the *a posteriori* error, $\varepsilon_{k+1|k}$, is obtained using the updated weight vector, $\boldsymbol{w}_k$, and the current input, $\boldsymbol{x}_k$. The solution is known as the normalised LMS (NLMS), given by (for more detail see [188])

$$\text{NLMS}: \qquad \boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \frac{1}{\|\boldsymbol{x}_k\|^2} \boldsymbol{x}_k \varepsilon_k. \tag{A.16}$$

The effective LMS-type step-size, $\mu_k = 1/\|\boldsymbol{x}_k\|^2$, is now time-varying and data-adaptive. In practice, to stabilise the algorithm a small positive step-size ρ_k can be employed, to give $\mu_k = \rho_k / \|\boldsymbol{x}_k\|^2$. The NLMS is therefore conformal with the LMS, whereby the input vector, $\boldsymbol{x}_k$, is normalised by its norm, $\|\boldsymbol{x}_k\|^2$ (input data power).

To find the optimal scalar step-size for the LMS in (A.9) which minimises the mean square

deviation, we shall first substitute the gain $\boldsymbol{g}_k = \mu_k \boldsymbol{x}_k$ into (A.15), to give the MSD recursion

$$J_k = J_{k-1} - 2\mu_k \underbrace{\boldsymbol{x}_k^{\mathsf{T}} \mathbf{M}_{k-1} \boldsymbol{x}_k}_{\xi_{ex,k}} + \mu_k^2 \|\boldsymbol{x}_k\|^2 (\underbrace{\boldsymbol{x}_k^{\mathsf{T}} \mathbf{M}_{k-1} \boldsymbol{x}_k + \sigma_n^2}_{\xi_k}). \tag{A.17}$$

The optimal step-size which minimises MSD is then obtained by solving for μ_k in (A.17) via $\partial J_k / \partial \mu_k = 0$, to yield [189]

$$\mu_k = \frac{\boldsymbol{x}_k^{\mathsf{T}} \mathbf{M}_{k-1} \boldsymbol{x}_k}{\|\boldsymbol{x}_k\|^2 \left(\boldsymbol{x}_k^{\mathsf{T}} \mathbf{M}_{k-1} \boldsymbol{x}_k + \sigma_n^2\right)} = \underbrace{\frac{1}{\|\boldsymbol{x}_k\|^2}}_{\text{normalisation}} \underbrace{\frac{\xi_{\text{ex},k}}{\xi_k}}_{\text{correction}} . \tag{A.18}$$

Remark A.1. *In addition to the NLMS-type* normalisation *factor,* $1/\|\boldsymbol{x}_k\|^2$*, the optimal LMS step-size in* (A.18) *includes the* correction *term,* $\xi_{ex,k}/\xi_k < 1$*, a ratio of the excess mean square error,* $\xi_{ex,k}$*, to the overall MSE,* ξ_k*. A large deviation from the true system weights causes a large* $\xi_{ex,k}/\xi_k$ *and fast weight adaptation (cf. slow adaptation for a small* $\xi_{ex,k}/\xi_k$*). This also justifies the use of a small step-size* ρ_k *in practical NLMS algorithms, such as that in Eq.* (A.28a) *in Section A.7.*

A.3.2 From LMS to Kalman Filter

The optimal LMS step-size in (A.18) aims to minimise the MSD at every time instant, however, it only controls the magnitude of gradient descent steps (see Figure A.1). To find the optimal learning gain which controls simultaneously both the magnitude and direction of the gradient descent in (A.10), we start again from the MSD recursion (restated from (A.15))

$$J_k = J_{k-1} - 2\boldsymbol{g}_k^{\mathsf{T}} \mathbf{M}_{k-1} \boldsymbol{x}_k + \|\boldsymbol{g}_k\|^2 \left(\boldsymbol{x}_k^{\mathsf{T}} \mathbf{M}_{k-1} \boldsymbol{x}_k + \sigma_n^2\right).$$

The optimal learning gain vector, $\boldsymbol{g}_k$, is then obtained by solving the above MSD for $\mathbf{g}_k$, via $\partial J_k / \partial \boldsymbol{g}_k = \mathbf{0}$, to give

$$\boldsymbol{g}_k = \frac{\mathbf{M}_{k-1}}{\boldsymbol{x}_k^{\mathsf{T}} \mathbf{M}_{k-1} \boldsymbol{x}_k + \sigma_n^2} \boldsymbol{x}_k = \frac{\mathbf{M}_{k-1}}{\xi_k} \boldsymbol{x}_k = \mathbf{G}_k \boldsymbol{x}_k. \tag{A.19}$$

This optimal gain vector is precisely the Kalman gain [46], while the Kalman gain matrix, $\mathbf{G}_k$, represents a ratio between the weight error covariance, $\mathbf{M}_{k-1}$, and the mean square error, ξ_k. A substitution into the update for $\mathbf{M}_k$ in (A.14) yields a Kalman filter which estimates the time-invariant and deterministic weights, $\boldsymbol{w}^{\text{o}}$, as outlined in Algorithm A.1.

Remark A.2. *For* $\sigma_n^2 = 1$*, the Kalman filtering equations in Algorithm A.1 are identical to the recursive least squares (RLS) algorithm. In this way, this note complements the classic paper by Sayed and Kailath [167] which establishes a relationship between the RLS and the Kalman filter.*

Algorithm A.1. Kalman filter for deterministic states

At each time instant $k > 0$, based on measurements $\{y_k, \boldsymbol{x}_k\}$

1: Compute the optimal learning gain (Kalman gain):

$$\boldsymbol{g}_k = \mathbf{M}_{k-1}\boldsymbol{x}_k \Big/ \left(\boldsymbol{x}_k^{\mathsf{T}}\mathbf{M}_{k-1}\boldsymbol{x}_k + \sigma_n^2\right)$$

2: Update the weight vector estimate:

$$\boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \boldsymbol{g}_k(y_k - \boldsymbol{x}_k^{\mathsf{T}}\boldsymbol{w}_{k-1})$$

3: Update the weight error covariance matrix:

$$\mathbf{M}_k = \mathbf{M}_{k-1} - \boldsymbol{g}_k\boldsymbol{x}_k^{\mathsf{T}}\mathbf{M}_{k-1}$$

A.4 Scalar Covariance Update

An additional insight into our joint perspective on Kalman and LMS algorithms is provided for independent and identically distributed (IID) system weight error vectors, whereby the diagonal weight error covariance matrix is given by $\mathbf{M}_{k-1} = \sigma_{P,k-1}^2\mathbf{I}$, while the Kalman gain, $\boldsymbol{g}_k$, in (A.19) now becomes

$$\boldsymbol{g}_k = \frac{\sigma_{P,k-1}^2}{\sigma_{P,k-1}^2\boldsymbol{x}_k^{\mathsf{T}}\boldsymbol{x}_k + \sigma_n^2}\,\boldsymbol{x}_k = \frac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|^2 + \varepsilon_k}, \tag{A.20}$$

where $\varepsilon_k \overset{\text{def}}{=} \sigma_n^2/\sigma_{P,k-1}^2$ denotes the regularisation parameter and $\sigma_{P,k-1}^2$ is the estimated weight error vector variance.

Remark A.3. *A physical interpretation of the regularisation parameter ε_k is that it models our confidence level in the current weight estimate, $\boldsymbol{w}_k$, via a ratio of the algorithm-independent minimum MSE, $\xi_{min} = \sigma_n^2$, and the algorithm-specific weight error variance, $\sigma_{P,k-1}^2$. The more confident we are in current weight estimates, the greater the value of ε_k and the smaller the magnitude of the weight update, $\Delta\boldsymbol{w}_k = \mathbf{g}_k\varepsilon_k$.*

To complete the derivation, since $\mathbf{M}_k = \sigma_{P,k}^2\mathbf{I}$ and thus $\mathsf{Tr}[\mathbf{M}_k] = M\sigma_{P,k}^2$, the MSD recursion in (A.15) now becomes

$$\sigma_{P,k}^2 = \sigma_{P,k-1}^2 - \frac{\|\boldsymbol{x}_k\|^2}{M(\|\boldsymbol{x}_k\|^2 + \varepsilon_k)}\sigma_{P,k-1}^2. \tag{A.21}$$

The resulting hybrid "Kalman-LMS" algorithm is given in Algorithm A.2.

Remark A.4. *The form of the LMS algorithm outlined in Algorithm A.2 is identical to the class of generalised normalised gradient descent (GNGD) algorithms in [187, 190] which update the regularisation*

parameter, ε_k, using stochastic gradient descent. More recently, Algorithm A.2 was derived independently in [191] as an approximate probabilistic filter for linear Gaussian data and is referred to as the probabilistic LMS.

Algorithm A.2. A hybrid Kalman-LMS algorithm

At each time instant $k > 0$, based on measurements $\{y_k, \boldsymbol{x}_k\}$

1: Compute the confidence level (regularisation parameter):

$$\varepsilon_k = \sigma_n^2 / \sigma_{P,k-1}^2$$

2: Update the weight vector estimate:

$$\boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \frac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|^2 + \varepsilon_k}(y_k - \boldsymbol{x}_k^{\mathsf{T}} \boldsymbol{w}_{k-1})$$

3: Update the weight error variance:

$$\sigma_{P,k}^2 = \sigma_{P,k-1}^2 - \frac{\|\boldsymbol{x}_k\|^2}{M(\|\boldsymbol{x}_k\|^2 + \varepsilon_k)} \sigma_{P,k-1}^2$$

A.5 From Optimal LMS to General Kalman Filter

To complete the joint perspective on the LMS and Kalman filters, we now consider a general case of a time-varying and stochastic weight vector $\boldsymbol{w}_k^{\mathrm{o}}$ in (A.1), to give

$$\boldsymbol{w}_{k+1}^{\mathrm{o}} = \mathbf{F}_k \boldsymbol{w}_k^{\mathrm{o}} + \boldsymbol{q}_k, \qquad \boldsymbol{q}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_s), \tag{A.22a}$$

$$y_k = \boldsymbol{x}_k^{\mathsf{T}} \boldsymbol{w}_k^{\mathrm{o}} + \eta_k, \qquad \eta_k \sim \mathcal{N}(0, \sigma_n^2). \tag{A.22b}$$

The evolution of the true weight vector $\boldsymbol{w}_k^{\mathrm{o}}$ is governed by a known state transition matrix, $\mathbf{F}_k$, while the uncertainty in the state transition model is represented by a temporally white state noise vector, $\boldsymbol{q}_k$, with covariance $\mathbf{C}_s = \mathbb{E}\{\boldsymbol{q}_k \boldsymbol{q}_k^{\mathsf{T}}\}$, which is uncorrelated with observation noise η_k. The known weight vector evolution in (A.22a) admits the *update* of the current state estimate, $\boldsymbol{w}_{k|k}$, and the *prediction* of the next state, $\boldsymbol{w}_{k+1|k}$, in an LMS-like fashion as

$$\boldsymbol{w}_{k|k} = \boldsymbol{w}_{k|k-1} + \boldsymbol{g}_k(y_k - \boldsymbol{x}_k^{\mathsf{T}} \boldsymbol{w}_{k|k-1}), \tag{A.23a}$$

$$\boldsymbol{w}_{k+1|k} = \mathbf{F}_k \boldsymbol{w}_{k|k}, \tag{A.23b}$$

where $\boldsymbol{g}_k$ in (A.23a) is the Kalman gain. Figure A.2 illustrates that, unlike the standard LMS or deterministic Kalman filter in Algorithm A.1, the general Kalman filter in (A.23a)–(A.23b) employs its prediction step in (A.23b) to track the time-varying error surface, a "frame of reference" for optimal adaptation.

The update steps (indicated by the index $k|k$) and the prediction steps (index $k+1|k$) for all

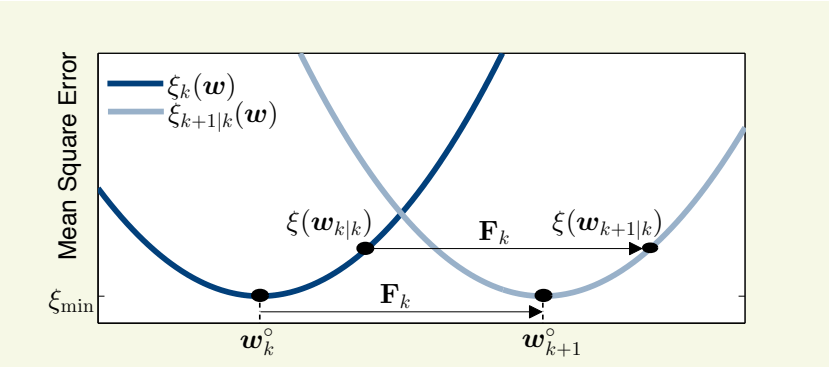

Figure A.2: The time-varying state transition in (A.22a) results in a time-varying MSE surface. For clarity, the figure considers a scalar case and state noise, q_k, is omitted. Within the Kalman filter, the prediction step in (A.23b) preserves the relative position of $w_{k+1|k}$ with respect to the evolved true state, w_{k+1}°.

the quantities involved are defined below as

$$\begin{aligned}
\tilde{\boldsymbol{w}}_{k|k} &\overset{\text{def}}{=} \boldsymbol{w}_k^{\circ} - \boldsymbol{w}_{k|k}, \quad \mathbf{M}_{k|k} \overset{\text{def}}{=} \mathbb{E}\left\{\tilde{\boldsymbol{w}}_{k|k}\tilde{\boldsymbol{w}}_{k|k}^{\mathsf{T}}\right\}, \\
\tilde{\boldsymbol{w}}_{k+1|k} &\overset{\text{def}}{=} \boldsymbol{w}_{k+1}^{\circ} - \boldsymbol{w}_{k+1|k} = \mathbf{F}_k\tilde{\boldsymbol{w}}_{k|k} + \boldsymbol{q}_k, \\
\mathbf{M}_{k+1|k} &\overset{\text{def}}{=} \mathbb{E}\left\{\tilde{\boldsymbol{w}}_{k+1|k}\tilde{\boldsymbol{w}}_{k+1|k}^{\mathsf{T}}\right\} = \mathbf{F}_k\mathbf{M}_{k|k}\mathbf{F}_k^{\mathsf{T}} + \mathbf{C}_s.
\end{aligned} \tag{A.24}$$

Similarly to (A.13)–(A.17), the Kalman gain is derived based on the weight error vector recursion, obtained by subtracting the optimal time-varying $\boldsymbol{w}_k^{\circ}$ from the state transition in (A.23a), as

$$\tilde{\boldsymbol{w}}_{k|k} = \tilde{\boldsymbol{w}}_{k|k-1} - \boldsymbol{g}_k\boldsymbol{x}_k^{\mathsf{T}}\tilde{\boldsymbol{w}}_{k|k-1} - \boldsymbol{g}_k\eta_k, \tag{A.25}$$

so that the evolution of the weight error covariance becomes

$$\begin{aligned}
\mathbf{M}_{k|k} &\overset{\text{def}}{=} \mathbb{E}\left\{\tilde{\boldsymbol{w}}_{k|k}\tilde{\boldsymbol{w}}_{k|k}^{\mathsf{T}}\right\} \\
&= \mathbf{M}_{k|k-1} - \mathbf{M}_{k|k-1}\boldsymbol{x}_k\boldsymbol{g}_k^{\mathsf{T}} - \boldsymbol{g}_k\boldsymbol{x}_k^{\mathsf{T}}\mathbf{M}_{k|k-1} \\
&\quad + \boldsymbol{g}_k\boldsymbol{g}_k^{\mathsf{T}}\left(\boldsymbol{x}_k^{\mathsf{T}}\mathbf{M}_{k|k-1}\boldsymbol{x}_k + \sigma_n^2\right).
\end{aligned} \tag{A.26}$$

Finally, the Kalman gain, $\boldsymbol{g}_k$, which minimises the MSD, $J_{k|k} = \mathsf{Tr}[\mathbf{M}_{k|k}]$, is obtained as (*cf.* optimal LMS in (A.19))

$$\boldsymbol{g}_k = \frac{\mathbf{M}_{k|k-1}}{\boldsymbol{x}_k^{\mathsf{T}}\mathbf{M}_{k|k-1}\boldsymbol{x}_k + \sigma_n^2}\boldsymbol{x}_k = \mathbf{G}_k\boldsymbol{x}_k. \tag{A.27}$$

The whole procedure is summarised in Algorithm A.3.

Algorithm A.3. General Kalman filter

At each time instant $k > 0$, based on measurements $\{y_k, \boldsymbol{x}_k\}$

1: Compute the optimal learning gain (Kalman gain):

$$\boldsymbol{g}_k = \mathbf{M}_{k|k-1}\boldsymbol{x}_k \Big/ \left(\boldsymbol{x}_k^{\mathsf{T}}\mathbf{M}_{k|k-1}\boldsymbol{x}_k + \sigma_n^2\right)$$

2: Update the weight vector estimate:

$$\boldsymbol{w}_{k|k} = \boldsymbol{w}_{k|k-1} + \boldsymbol{g}_k(y_k - \boldsymbol{x}_k^{\mathsf{T}}\boldsymbol{w}_{k|k-1})$$

3: Update the weight error covariance matrix:

$$\mathbf{M}_k = \mathbf{M}_{k-1} - \boldsymbol{g}_k\boldsymbol{x}_k^{\mathsf{T}}\mathbf{M}_{k-1}$$

4: Predict the next state (weight) vector:

$$\boldsymbol{w}_{k+1|k} = \mathbf{F}_k\boldsymbol{w}_{k|k}$$

5: Predict the weight error covariance matrix:

$$\mathbf{M}_{k+1|k} = \mathbf{F}_k\mathbf{M}_{k|k}\mathbf{F}_k^{\mathsf{T}} + \mathbf{C}_s$$

Remark A.5. *Steps 1–3 in Algorithm A.3 are identical to the deterministic Kalman filter which was derived starting from the LMS and is described in Algorithm A.1. The distinguishing difference is in Steps 4–5 which cater for the time-varying and stochastic general system weights. Therefore, the fundamental principles of the Kalman filter can be considered through optimal adaptive step-size LMS algorithms.*

A.6 Conclusions

We have employed "optimal gain" as a mathematical lens to examine variants of the LMS algorithms and Kalman filters. This perspective enabled us to create a framework for unification of these two main classes of adaptive recursive online estimators. A close examination of the relationship between the two standard performance evaluation measures, namely the mean square error (MSE) and mean square deviation (MSD), allowed us to link up in an intuitive way the geometry of learning of Kalman filters and LMS, in both deterministic and stochastic system identification settings. The Kalman filtering algorithm is then derived in an LMS-type fashion via the optimal learning gain matrix, without having to resort to probabilistic approaches [192].

Such a conceptual insight also permits seamless migration of ideas from the state space based Kalman filters to the LMS adaptive linear filters and vice versa, and provides a platform for practical applications and nonlinear extensions [153]. It is our hope that this framework of examination of these normally disparate areas will both demystify recursive estimation for educational purposes [193], and further empower practitioners with enhanced intuition and freedom in algorithmic design for the manifold applications.

A.7 Variants of the LMS

To illustrate the generality of results in Sections A.3 and A.4, consider the normalised LMS (NLMS) and the regularised NLMS (also termed $\varepsilon-$NLMS), given by

$$\text{NLMS:} \quad \boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \rho_k \frac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|^2}\varepsilon_k, \tag{A.28a}$$

$$\varepsilon-\text{NLMS:} \quad \boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \frac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|^2 + \varepsilon_k}\varepsilon_k, \tag{A.28b}$$

where ρ_k is a step-size and ε_k a regularisation factor. Based on (A.17)-(A.18), the optimal values for ρ_k and ε_k can be found as

$$\rho_k = \frac{\boldsymbol{x}_k^{\mathsf{T}}\mathbf{M}_{k-1}\boldsymbol{x}_k}{\boldsymbol{x}_k^{\mathsf{T}}\mathbf{M}_{k-1}\boldsymbol{x}_k + \sigma_n^2}, \qquad \varepsilon_k = \frac{\|\boldsymbol{x}_k\|^2\sigma_n^2}{\boldsymbol{x}_k^{\mathsf{T}}\mathbf{M}_{k-1}\boldsymbol{x}_k}. \tag{A.29}$$

Upon substituting ρ_k and ε_k from (A.29) into their respective weight update recursions in (A.28a) and (A.28b), we arrive at

$$\boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \frac{\boldsymbol{x}_k^{\mathsf{T}}\mathbf{M}_{k-1}\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|^2\left(\boldsymbol{x}_k^{\mathsf{T}}\mathbf{M}_{k-1}\boldsymbol{x}_k + \sigma_n^2\right)}\boldsymbol{x}_k\varepsilon_k, \tag{A.30}$$

for both the NLMS and $\varepsilon-$NLMS, which is identical to the LMS with the optimal step-size in (A.18). Therefore, the minimisation of the mean square deviation with respect to the parameter: (i) μ_k in the LMS, (ii) ρ_k in the NLMS, and (iii) ε_k in the ε-NLMS, yields exactly the same algorithm, which is intimately related to the Kalman filter, as shown in Table A.2 and indicated by the expression for the Kalman gain, $\boldsymbol{g}_k$.

Algorithm	Gain vector	Optimal gain vector
Kalman filter	$\boldsymbol{g}_k$	$\dfrac{\mathbf{M}_{k\|k-1}\boldsymbol{x}_k}{\boldsymbol{x}_k^{\mathsf{T}}\mathbf{M}_{k\|k-1}\boldsymbol{x}_k + \sigma_n^2}$
LMS	$\mu_k\boldsymbol{x}_k$	$\dfrac{\boldsymbol{x}_k^{\mathsf{T}}\mathbf{M}_{k-1}\boldsymbol{x}_k}{\boldsymbol{x}_k^{\mathsf{T}}\mathbf{M}_{k-1}\boldsymbol{x}_k + \sigma_n^2}\dfrac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|^2}$
NLMS	$\rho_k\frac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|^2}$	
$\varepsilon-$NLMS	$\frac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|^2+\varepsilon_k}$	which equals $\boldsymbol{x}_k^{\mathsf{T}}\boldsymbol{g}_k\dfrac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|^2}$

Table A.2: *Summary of optimal gain vectors. The optimal step-sizes for the LMS-type algorithms are linked to the* a priori *variant of the Kalman gain vector, $\boldsymbol{g}_k$, since $\mathbf{M}_{k|k-1} = \mathbf{M}_{k-1}$ for deterministic and time-invariant system weight vectors.*

Appendix B

Equivalence of Classical Frequency Tracking Techniques

Appendix Overview

The problem of tracking the system frequency is ubiquitous in power systems. However, despite numerous comparative studies of various algorithms, the underlying links and commonalities between frequency tracking methods are often overlooked. To this end, we show that the treatment of two of the best known frequency tracking algorithms, the recursive discrete Fourier transform (DFT) and the fixed frequency demodulation technique, can be unified, whereby the former can be interpreted as a special case of the latter.

B.1 Introduction

The problem of tracking the power system frequency has been of practical and theoretical interest for decades, as variations in system frequency indicate a mismatch between the supply and demand of electricity [24]. The area is undergoing a resurgence of interest as future low inertia grids, which experience rapid frequency excursions, call for faster and more accurate frequency tracking algorithms [6]. Extensive benchmarking of various frequency tracking methods does exist [142, 194, 195], however, owing to the lack of a coherent and unifying framework, the underlying commonalities between various frequency tracking methods are often overlooked.

Our aim is to show that the two best known frequency tracking algorithms, the recursive discrete Fourier transform (DFT) and the fixed frequency (de-)modulation (FM) technique, can be treated in a unified way. Specifically, we show that the recursive DFT based frequency tracker can be interpreted as a special case of an FM demodulation scheme whereby the carrier frequency is the fundamental frequency of the power system (e.g. 50 Hz or 60 Hz).

B.2 Problem Formulation

Consider discrete time observations, x_k, of a complex-valued sinusoid, s_k, embedded in a white noise process η_k, given by

$$\begin{aligned} x_k &= s_k + \eta_k, \\ s_k &= |A|e^{j(\omega_\circ k+\phi_k)} + |B|e^{-j(\omega_\circ k+\phi_k)}, \end{aligned} \tag{B.1}$$

where $\omega_\circ = 2\pi\frac{f_\circ}{f_s}$ is the fundamental frequency, f_s the sampling frequency, and k the time instant. The real-valued amplitudes are, $|A|$ and $|B|$, and the time-varying phase, ϕ_k, are assumed to vary considerably more slowly than the fundamental frequency $\omega_\circ$.

The signal model in (B.1) represents a generalisation of single-frequency sinusoids and degenerates into the well-studied cisoid, $s_k = |A|e^{j(\omega_\circ k+\phi_k)}$, for $|B| = 0$, while for $|B| = |A|$, it simplifies into a real valued sinusoid, $s_k = 2|A|\cos(\omega_\circ k + \phi_k)$. The signal structure in (B.1) is therefore quite general and is also observed as the output of the $\alpha\beta$ transform in unbalanced three-phase power systems [24].

The instantaneous frequency, ω_k, of the signal s_k in (B.1) is then given by

$$\omega_k = \omega_\circ + (\phi_k - \phi_{k-1}). \tag{B.2}$$

Such a decomposition of the power system frequency into the fundamental frequency component, $\omega_\circ$, and rate of change of phase, $(\phi_k - \phi_{k-1})$, is convenient as the system frequency, $f_\circ$, is fixed and known to be either 50 Hz or 60 Hz. The estimation task is therefore to track the instantaneous deviation from the fundamental frequency.

B.2.1 Rate of Change of the DFT Phase

The *de facto* standard frequency estimation algorithm in power systems is based on tracking the frequency of the synchronous phasor obtained from the recursion for the fundamental DFT component [5]. To begin, consider the discrete Fourier transform (DFT) of a signal $(x_n)_{n=0}^{N-1}$ is given by

$$\vec{X}_{N-1}[k] = \frac{1}{N}\sum_{n=0}^{N-1} x_n e^{-j\frac{2\pi}{N}kn}, \tag{B.3}$$

where the subscript $(N-1)$ is used to indicate the latest sample, x_{N-1}, used in the DFT computation. The fundamental DFT component $\vec{X}_{N-1}[1]$ is therefore

$$\vec{X}_{N-1}[1] = \frac{1}{N}\sum_{n=0}^{N-1} x_n e^{-j\frac{2\pi}{N}n}, \tag{B.4}$$

where the frequency bin index, "[1]", shall now be omitted for clarity. The fundamental DFT taken at a different window of the data, e.g. $(x_n)_{n=r}^{r+N-1}$, is therefore

$$\vec{X}_{r+N-1} = \frac{1}{N}\sum_{n=0}^{N-1} x_{n+r} e^{-j\frac{2\pi}{N}n}. \tag{B.5}$$

The DFT operation in (B.5) is referred to as the sliding window DFT and the quantity $\vec{X}_{r+N-1}$ is the rotating phasor of the signal x_n. The rotating phasor, $\vec{X}_{N-1+r}$, can be converted into a stationary phasor as

$$\bar{X}_{r+N-1} \overset{\text{def}}{=} \vec{X}_{r+N-1} e^{-j\frac{2\pi}{N}r} = \frac{1}{N}\sum_{n=0}^{N-1} x_{n+r} e^{-j\frac{2\pi}{N}(n+r)}. \tag{B.6}$$

Notice that the difference between the stationary phasor in (B.6) and the rotating phasor in (B.5) is the time index in the complex exponential and that the operations (B.3) – (B.5) are analogous to filtering the sequence x_n with a finite impulse response (FIR) filter. To obtain a recursive implementation of the stationary phasor in (B.6), two consecutive window indices can be expanded as

$$\bar{X}_{r+N-1} = \frac{1}{N}\left[x_r e^{-j\frac{2\pi}{N}r} + x_{r+1} e^{-j\frac{2\pi}{N}(r+1)} + \cdots + x_{r+N-1} e^{-j\frac{2\pi}{N}(r+N-1)}\right], \tag{B.7}$$

$$\bar{X}_{r+N} = \frac{1}{N}\left[x_{r+1} e^{-j\frac{2\pi}{N}(r+1)} + \cdots + x_{r+N-1} e^{-j\frac{2\pi}{N}(r+N-1)} + x_{r+N} e^{-j\frac{2\pi}{N}(r+N)}\right]. \tag{B.8}$$

Subtracting the $\bar{X}_{r+N-1}$ in (B.7) from the $\bar{X}_{r+N}$ phasor in (B.8) removes the overlapping terms, thus yielding

$$\bar{X}_{r+N} - \bar{X}_{r+N-1} = \frac{1}{N}\left[x_{r+N} e^{-j\frac{2\pi}{N}(r+N)} - x_r e^{-j\frac{2\pi}{N}r}\right]. \tag{B.9}$$

Since $e^{-j\frac{2\pi}{N}(r+N)} = e^{-j\frac{2\pi}{N}r} e^{-j\frac{2\pi}{N}N} = e^{-j\frac{2\pi}{N}r}$, the stationary phasor $\bar{X}_r$ can be computed recursively as

$$\bar{X}_{r+N} = \bar{X}_{r+N-1} + \frac{1}{N}\left(x_{N+r} - x_r\right) e^{-j\frac{2\pi}{N}r}. \tag{B.10}$$

Output of the Stationary DFT

The output of the stationary DFT from (B.6), can be expressed as

$$\bar{X}_{r+N-1} = \frac{1}{N}\sum_{n=0}^{N-1} x_{n+r} e^{-j\omega_\circ(n+r)}, \tag{B.11}$$

since the fundamental frequency is related to the number of samples as $\omega_\circ = \frac{2\pi}{N}$. Substituting a general complex valued sinusoid given by

$$x_k = Ae^{j\omega k} + Be^{-j\omega k}, \tag{B.12}$$

where $\omega = \omega_\circ + \Delta\omega$, into the stationary phasor definition in (B.11) yields

$$\begin{aligned}
\bar{X}_{r+N-1} &= \frac{1}{N}\sum_{n=0}^{N-1}\left(Ae^{j\omega(n+r)} + Be^{-j\omega(n+r)}\right)e^{-j\omega_\circ(n+r)} \\
&= \frac{1}{N}\sum_{n=0}^{N-1} Ae^{j(\omega-\omega_\circ)(n+r)} + \frac{1}{N}\sum_{n=0}^{N-1} Be^{-j(\omega+\omega_\circ)(n+r)} \\
&= \frac{1}{N}\sum_{n=0}^{N-1} Ae^{j\Delta\omega(n+r)} + \frac{1}{N}\sum_{n=0}^{N-1} Be^{-j(2\omega_\circ+\Delta\omega)(n+r)} \\
&= \left(\frac{A}{N}\sum_{n=0}^{N-1} e^{j\Delta\omega n}\right)e^{j\Delta\omega r} + \left(\frac{B}{N}\sum_{n=0}^{N-1} e^{-j(2\omega_\circ+\Delta\omega)n}\right)e^{-j(2\omega_\circ+\Delta\omega)r} \\
&= \tilde{A}e^{j\Delta\omega r} + \tilde{B}e^{-j(2\omega_\circ+\Delta\omega)r}.
\end{aligned} \tag{B.13}$$

where the complex valued amplitudes $\tilde{A}$ and $\tilde{B}$ are given by

$$\tilde{A} = \frac{A}{N}\sum_{n=0}^{N-1} e^{j\Delta\omega n}, \quad \text{and} \quad \tilde{B} = \frac{B}{N}\sum_{n=0}^{N-1} e^{-j(2\omega_\circ+\Delta\omega)n}. \tag{B.14}$$

Using the well known identity

$$\sum_{n=0}^{N-1} e^{j\theta n} = \frac{1-e^{j\theta N}}{1-e^{j\theta}} = e^{j\frac{\theta}{2}(N-1)}\frac{\sin(N\theta/2)}{\sin(\theta/2)},$$

yields the expressions

$$\tilde{A} = Ae^{j\frac{\Delta\omega}{2}(N-1)}\left(\frac{1}{N}\frac{\sin(N\Delta\omega/2)}{\sin(\Delta\omega/2)}\right) \overset{\Delta\omega\to 0}{=} Ae^{j\frac{\Delta\omega}{2}(N-1)} \tag{B.15}$$

$$\tilde{B} = Be^{-j\frac{2\omega_\circ+\Delta\omega}{2}(N-1)}\left(\frac{1}{N}\frac{\sin(N(2\omega_\circ+\Delta\omega)/2)}{\sin((2\omega_\circ+\Delta\omega)/2)}\right) \overset{\Delta\omega\to 0}{=} 0, \tag{B.16}$$

for small values of $\Delta\omega$. Therefore the output of the stationary DFT from (B.11) and (B.13) is given by

$$\bar{X}_{r+N-1} \approx \tilde{A}e^{j\Delta\omega r}. \tag{B.17}$$

Multiplying the conjugate of the previous stationary DFT sample in (B.17) yields the freqeuncy deviation $\Delta\omega$ in the form

$$\bar{X}_{r+N}\bar{X}^*_{r+N-1} = |\tilde{A}|^2 e^{j\Delta\omega} \implies \Delta\omega = \arg[\bar{X}_{r+N}\bar{X}^*_{r+N-1}]$$

For further notational compactness, using the variable $k = N + r$, results in the recursive DFT based frequency estimation method given by

$$X_k = X_{k-1} + \frac{1}{N}\left(x_k - x_{k-N}\right) e^{-j\frac{2\pi}{N}k}, \tag{B.18a}$$

where the DFT window length N, is chosen as a ratio of the fundamental frequency and the sampling frequency, that is

$$N = f_s/f_\circ. \tag{B.18b}$$

From (B.2), the instantaneous frequency estimate can therefore be obtained from the rate of change of the single-bin DFT phase angle as

$$\hat{\omega}_{\text{DFT},k} = \omega_\circ + \arg[X_k X^*_{k-1}], \tag{B.18c}$$

where $\arg[\cdot]$ denotes the argument operator. In practice, to obtain smoother estimates, post-processing (filtering) is carried out in the form of fitting a linear or quadratic model to the phase angle sequence, $\arg[X_k]$, obtained from (B.18a).

B.2.2 Fixed Frequency Demodulation

Frequency modulation (FM) is a fundamental concept in communications; the standard demodulation scheme based on the heterodyne principle is given by [140, 195]

$$y_k = x_k e^{-j\omega k} \tag{B.19a}$$

$$u_k = \text{LPF}[y_k] \tag{B.19b}$$

$$\hat{\omega}_{\text{FM},k} = \omega + \arg[u_k u^*_{k-1}]. \tag{B.19c}$$

The step in (B.19a) is known as the heterodyne or demixing step, ω is a known carrier frequency and is chosen to be the fundamental power system frequency ($\omega = \omega_\circ$), while the operator $\text{LPF}[\cdot]$ in (B.19b) denotes a low-pass filtering operation performed by any suitable lowpass filter with a cut-off frequency $\omega_{\text{cut}} < \omega$. The demodulated frequency is then obtained in (B.19c) from the rate of change of the phase angles of the low-pass filtered signal u_k. The block diagram of a FM demodulator is shown in Figure B.1.

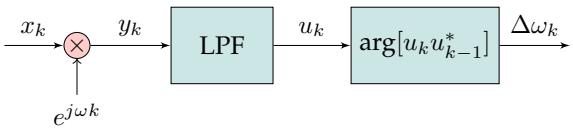

Figure B.1: Block diagram of a general fixed frequency demodulation scheme.

To understand this scheme, we express (B.19a) as

$$y_k = x_k e^{-j\omega_\circ k} = A_k e^{j\phi_k} + B_k e^{-j(2\omega_\circ k+\phi_k)} + \eta_k e^{-j\omega_\circ k},$$

Then, without loss in generality, a low-passed filtered ($\omega_{\text{cut}} < \omega_\circ$) version of y_k is

$$u_k = A_k e^{j\phi_k}, \tag{B.20}$$

since the other two components in y_k have frequencies larger than the cut-off frequency. The low pass filtered output, u_k, can be subsequently used to derive the frequency deviation as $\Delta\omega = \phi_k - \phi_{k-1} = \angle[u_k u_{k-1}^*]$ as shown in the FM frequency estimation step in (B.19c).

> **Contribution: Connecting the Recursive DFT and FM Demodulation**
> The rest of the appendix presents the contribution in unifying the recursive DFT and FM demodulator which are the two most popular frequency estimators in the context of power systems.

B.3 Equivalence between the recursive DFT and FM demodulation

The intrinsic relationship between the recursive DFT in (B.18a) – (B.18c) and the demodulation scheme in (B.19a) – (B.19c) can now be derived without any assumption on the measurement x_k. First, consider again the recursive DFT in (B.18a),

$$X_k = X_{k-1} + \frac{1}{N}(x_k - x_{k-N})e^{-j\omega_\circ k}, \tag{B.21}$$

where, based on the choice of the data length in (B.18b), $\omega_\circ = 2\pi f_\circ / f_{\text{s}} = 2\pi/N$. Next, observe that the terms $x_k e^{-j\omega_\circ k}$ and $x_{k-N} e^{-j\omega_\circ k}$ in (B.21) are related to the demodulated signal y_k in (B.19a) through

$$y_k = x_k e^{-j\omega_\circ k}, \qquad y_{k-N} = x_{k-N} e^{-j\omega_\circ k}, \tag{B.22}$$

which yields the relationship

$$X_k = X_{k-1} + \frac{1}{N}(y_k - y_{k-N}). \tag{B.23}$$

The difference equation in (B.23) represents a low-pass filtering operation on the signal y_k, implemented with an infinite impulse response (IIR) filter with the transfer function [196]

$$H(z) = \frac{1}{N}\frac{1 - z^{-N}}{1 - z^{-1}}. \tag{B.24}$$

Remark #1: If the low pass filter, LPF$[\cdot]$, used in step (B.19b), is chosen to have the transfer function in (B.24), then the output, u_k, in (B.19b) is identical to the synchronised phasor, X_k, in (B.18a). Since the choice of the low-pass filter block in demodulation allows for any suitable low-pass structure, the recursive DFT can be interpreted as a special case of the fixed frequency demodulation scheme.

B.4 Illustrative Example

The interpretation in Remark #1 gives the designers of phasor measurement unit (PMU) algorithms the freedom to choose low-pass filters with characteristics that are favourable to the application at hand, without having to resort to the strict definition of the synchronised phasor in (B.18a).

To illustrate this advantage, consider the magnitude response of the recursive DFT transfer function in (B.24), with $N = f_s/f_\circ = 20$, $f_s = 1$ kHz and $f_\circ = 50$ Hz. Figure B.2 shows that the filter in (B.24) does not exhibit steep roll-off. In contrast, a more desirable magnitude and phase response of a third-order Butterworth filter with the cutoff frequency, $\omega_{\text{cut}} = 2$ Hz, is also shown in Figure B.2.

The benefits of having a filter with sharp roll-off can be illustrated through the following frequency tracking task. The signal $x_k = \cos(2\pi\frac{f}{f_s}k) + \eta_k$, with a signal-to-noise ratio of 50dB and frequency $f = 50$ Hz, undergoes a step-change to 50.5 Hz after 0.5 s. For a sampling frequency of $f_s = 1$ kHz, the instantaneous frequency of the signal x_k was estimated using: (i) the FM demodulator in (B.19a) – (B.19c) with a Butterworth filter (frequency response shown in Figure B.2) and (ii) the recursive DFT frequency estimator based on (B.18a) – (B.18c) and for completeness (iii) the ACEKF based frequency tracking algorithm introduced in Chapter 8.

Figure B.3 shows that both methods were able to accurately track the nominal frequency of 50Hz. However, for off-nominal frequencies, the recursive DFT performed poorly and required further post processing. The problem can be avoided by selecting an appropriate low-pass filter, without having to rely on the filter in (B.24).

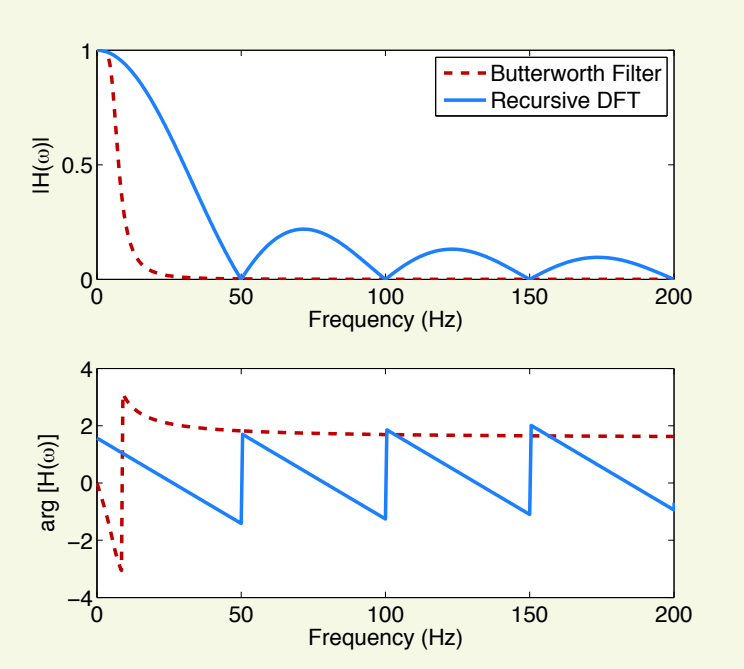

Figure B.2: Magnitude and phase responses of a third-order Butterworth filter with cuttoff frequency $\omega_{\text{cut}} = 2$ Hz, and the Recursive DFT filter in (B.24).

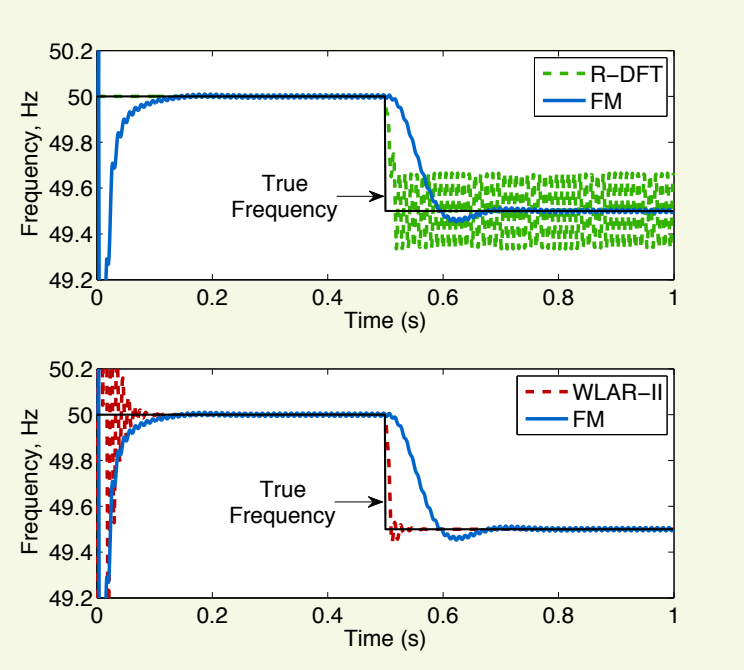

Figure B.3: *Top:* Frequency estimates from the recursive DFT and FM demodulator; *Bottom:* Comparison with ACEKF which employs the WLAR-II model in Table 8.2.

B.5 Conclusion

We have shown that recursive DFT based frequency tracking is a special case of an FM demodulation scheme with a demodulating frequency of 50 Hz or 60 Hz. The so achieved unification of the two standard frequency estimation schemes promises much enhanced flexibility in the design of PMU algorithms and opens avenues for further work on the optimal design and analysis of frequency trackers.

Appendix C

Cramer-Rao Lower Bound and Maximum Likelihood Estimation

Appendix Overview

The Cramer-Rao lower bound (CRLB) for the frequency estimate from the Clarke voltage contaminated by white Gaussian noise is derived. Furthermore, a maximum likelihood estimation (MLE) scheme for the frequency of the Clarke voltage is also derived for completeness. Within the MLE analysis it is found that the ML frequency estimate is the maximiser of an augmented periodogram, which is the sum of the standard periodogram of the data and that of their complex conjugates. To reduce the computational complexity of the algorithm, we show that the MLE frequency estimate can be formed by maximising the standard periodogram alone. It is shown that as desired, the proposed MLE method approaches the theoretical CRLB which is also verified by simulations. Part of this work has been submitted for publication [40].

C.1 Background: Cramer-Rao Lower Bound

Consider a complex-valued signal, $x_k \in \mathbb{C}$, consisting of a deterministic signal, $s_k(\boldsymbol{\theta}) \in \mathbb{C}$, parametrised by a real-valued parameter vector, $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_M]^{\mathsf{T}} \in \mathbb{R}^{M\times 1}$, and corrupted by a zero-mean white Gaussian noise, η_k, so that

$$x_k = s_k(\boldsymbol{\theta}) + \eta_k, \quad k = 0, 1, \ldots, N-1, \tag{C.1}$$

Collecting the samples x_k, $s_k(\boldsymbol{\theta})$ and η_k into $N \times 1$ column vectors, that is

$$\boldsymbol{x} = \begin{bmatrix} x_0, & x_1, & \ldots & , x_{N-1} \end{bmatrix}^{\mathsf{T}}, \boldsymbol{s}(\boldsymbol{\theta}) = \begin{bmatrix} s_0, & s_1, & \ldots & , s_{N-1} \end{bmatrix}^{\mathsf{T}}, \boldsymbol{\eta} = \begin{bmatrix} \eta_0, & \eta_1, & \ldots & , \eta_{N-1} \end{bmatrix}^{\mathsf{T}}$$

gives a vectorised version of the signal in (C.1) as

$$\boldsymbol{x} = \boldsymbol{s}(\boldsymbol{\theta}) + \boldsymbol{\eta}. \tag{C.2}$$

Since the noise process, $\boldsymbol{\eta}$, is a complex-valued Gaussian, its joint-probability density func-

tion (pdf) is given by [43]

$$\mathsf{pdf}(\boldsymbol{\eta}) = \frac{1}{\pi^N \det(\boldsymbol{\Sigma})} \exp\left\{-\boldsymbol{\eta}^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\eta}\right\}, \tag{C.3}$$

where $\boldsymbol{\Sigma} \overset{\text{def}}{=} \mathbb{E}\left\{\boldsymbol{\eta}\boldsymbol{\eta}^{\mathsf{H}}\right\}$ is the covariance of the noise vector. Using the signal model in (C.2), it can be seen that $\boldsymbol{\eta} = \boldsymbol{x} - \boldsymbol{s}(\boldsymbol{\theta})$. Therefore, the joint pdf for the data $\boldsymbol{x}$, which is parametrised by $\boldsymbol{\theta}$ is given by

$$\mathsf{pdf}(\boldsymbol{x}; \boldsymbol{\theta}) = \frac{1}{\pi^N \det(\boldsymbol{\Sigma})} \exp\left\{-(\boldsymbol{x} - \boldsymbol{s}(\boldsymbol{\theta}))^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{s}(\boldsymbol{\theta}))\right\}. \tag{C.4}$$

If the noise sequence, $\boldsymbol{\eta}$, is independent and identically distributed (IID), its covariance matrix is diagonal, so that $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$, which simplifies the pdf in (C.4) into [138, Ch. 3]

$$\mathsf{pdf}(\boldsymbol{x}; \boldsymbol{\theta}) = \frac{1}{(\pi\sigma^2)^N} \exp\left\{-\frac{1}{\sigma^2} \sum_{k=0}^{N-1} |x_k - s_k(\boldsymbol{\theta})|^2\right\}. \tag{C.5}$$

For many applications, the log-likelihood function of the pdf in (C.5) is more convenient to work with and is defined as

$$\ln \mathsf{pdf}(\boldsymbol{x}; \boldsymbol{\theta}) = -\ln(\pi\sigma^2)^N - \frac{1}{\sigma^2} \sum_{k=0}^{N-1} |x_k - s_k(\boldsymbol{\theta})|^2. \tag{C.6}$$

C.1.1 CRLB

The Cramer-Rao lower bound (CRLB) gives the minimum achievable variance over all unbiased estimations of $\boldsymbol{\theta}$, given the pdf of the signal, $\mathsf{pdf}(\boldsymbol{x}; \boldsymbol{\theta})$. For the signal model defined in (C.2), the CRLBs for the parameters within the vector $\boldsymbol{\theta}$ are given by

$$\mathrm{var}(\hat{\theta}_i) \geq \left[\mathbf{F}^{-1}(\boldsymbol{\theta})\right]_{ii}, \tag{C.7}$$

where the matrix $\mathbf{F}(\boldsymbol{\theta}) \in \mathbb{R}^{M \times M}$ is referred to as the Fisher information matrix (FIM) with elements [138, Ch. 3]

$$\mathbf{F}_{\ell m} = -\mathbb{E}\left\{\frac{\partial^2 \ln \mathsf{pdf}(\boldsymbol{x}, \boldsymbol{\theta})}{\partial\theta_\ell \, \partial\theta_m}\right\}. \tag{C.8}$$

The FIM measures the curvature of the log-likelihood function of the data in (C.6), which is inversely proportional to the spread (i.e. variance) of the estimates. Applying the partial derivatives into the log-likelihood function in (C.6), yields the Fisher information matrix elements in

a more convenient form

$$\mathbf{F}_{\ell m} = \frac{2}{\sigma^2}\mathsf{Re}\left\{\sum_{k=0}^{N-1}\frac{\partial s_k}{\partial \theta_\ell}\frac{\partial s_k^*}{\partial \theta_m}\right\}. \tag{C.9}$$

where it can be seen that the FIM is symmetrical, i.e. $\mathbf{F}_{\ell m} = \mathbf{F}_{m\ell}$.

C.2 CRLB for Frequency Estimates

For the problem of frequency estimation, recall the complex-valued Clarke signal voltage embedded in an IID white Gaussian noise process, η_k, such that

$$x_k = s_k(\boldsymbol{\theta}) + \eta_k \tag{C.10}$$

$$s_k(\boldsymbol{\theta}) = Ae^{j\omega k} + Be^{-j\omega k}, \tag{C.11}$$

where $A = |A|e^{j\phi_A}$ refers to the positive sequence phasor, $B = |B|e^{j\phi_B}$, the negative sequence phasor and $\omega = 2\pi fT$ the angular frequency with f the fundamental power system frequency and T the sampling period. The real-valued unknown parameter vector that is to be estimated is consequently given by

$$\boldsymbol{\theta} = [|A|, |B|, \phi_A, \phi_B, \omega]^{\mathsf{T}}. \tag{C.12}$$

To compute the elements of the FIM, we first compute the partial derivatives of the noise-free signal, s_k, with respect to the parameters θ_i, so that

$$\frac{\partial s_k}{\partial |A|} = e^{j\phi_A}e^{j\omega k}, \qquad \frac{\partial s_k}{\partial |B|} = e^{j\phi_B}e^{-j\omega k}, \tag{C.13}$$

$$\frac{\partial s_k}{\partial \phi_A} = jAe^{j\omega k}, \qquad \frac{\partial s_k}{\partial \phi_B} = jBe^{-j\omega k}, \tag{C.14}$$

$$\frac{\partial s_k}{\partial \omega} = jk\left(Ae^{j\omega k} - Be^{-j\omega k}\right). \tag{C.15}$$

Upon substituting the partial derivatives into the definition of the FIM in (C.9), the diagonal

elements are computed as

$$\begin{aligned}
\mathbf{F}_{11} &= \frac{2}{\sigma^2}\mathsf{Re}\left\{\sum_{k=0}^{N-1}\frac{\partial s_k}{\partial|A|}\frac{\partial s_k^*}{\partial|A|}\right\} = \frac{2}{\sigma^2}\sum_{k=0}^{N-1}1 = \frac{2}{\sigma^2}N \\
\mathbf{F}_{22} &= \frac{2}{\sigma^2}\mathsf{Re}\left\{\sum_{k=0}^{N-1}\frac{\partial s_k}{\partial|B|}\frac{\partial s_k^*}{\partial|B|}\right\} = \frac{2}{\sigma^2}\sum_{k=0}^{N-1}1 = \frac{2}{\sigma^2}N \\
\mathbf{F}_{33} &= \frac{2}{\sigma^2}\mathsf{Re}\left\{\sum_{k=0}^{N-1}\frac{\partial s_k}{\partial\phi_A}\frac{\partial s_k^*}{\partial\phi_A}\right\} = \frac{2}{\sigma^2}\sum_{k=0}^{N-1}|A|^2 = \frac{2}{\sigma^2}|A|^2N \\
\mathbf{F}_{44} &= \frac{2}{\sigma^2}\sum_{k=0}^{N-1}\frac{\partial s_k}{\partial\phi_B}\frac{\partial s_k^*}{\partial\phi_B} = \frac{2}{\sigma^2}\sum_{k=0}^{N-1}|B|^2 = \frac{2}{\sigma^2}|B|^2N \\
\mathbf{F}_{55} &= \frac{2}{\sigma^2}\mathsf{Re}\left\{\sum_{k=0}^{N-1}\frac{\partial s_k}{\partial\omega}\frac{\partial s_k^*}{\partial\omega}\right\} = \frac{2}{\sigma^2}\sum_{k=0}^{N-1}k^2(|A|^2+|B|^2-2|A||B|\cos(2\omega k+\Delta_{AB})) \\
&= \frac{2}{\sigma^2}|A|^2+|B|^2\frac{N(N-1)(2N-1)}{6},
\end{aligned}$$

while the off-diagonal elements are found to be

$$\begin{aligned}
\mathbf{F}_{35} = \mathbf{F}_{53} &= \frac{2}{\sigma^2}\mathsf{Re}\left\{\sum_{k=0}^{N-1}\frac{\partial s_k}{\partial\phi_A}\frac{\partial s_k^*}{\partial\omega}\right\} = \frac{2}{\sigma^2}\mathsf{Re}\left\{\sum_{k=0}^{N-1}k(|A|^2-AB^*e^{2j\omega k})\right\} \\
&= \frac{2}{\sigma^2}|A|^2\frac{N(N-1)}{2} \\
\mathbf{F}_{45} = \mathbf{F}_{54} &= \frac{2}{\sigma^2}\mathsf{Re}\left\{\sum_{k=0}^{N-1}\frac{\partial s_k}{\partial\phi_B}\frac{\partial s_k^*}{\partial\omega}\right\} = \frac{2}{\sigma^2}\mathsf{Re}\left\{\sum_{k=0}^{N-1}-k(|B|^2-AB^*e^{-2j\omega k})\right\} \\
&= -\frac{2}{\sigma^2}|B|^2\frac{N(N-1)}{2}.
\end{aligned}$$

Therefore, the FIM for the parameters for the data is given by

$$\mathbf{F}(\boldsymbol{\theta}) = \frac{2}{\sigma^2}\begin{bmatrix} N & 0 & 0 & 0 & 0 \\ 0 & N & 0 & 0 & 0 \\ 0 & 0 & |A|^2N & 0 & |A|^2\frac{N(N-1)}{2} \\ 0 & 0 & 0 & |B|^2N & -|B|^2\frac{N(N-1)}{2} \\ 0 & 0 & |A|^2\frac{N(N-1)}{2} & -|B|^2\frac{N(N-1)}{2} & (|A|^2+|B|^2)\frac{N(N-1)(2N-1)}{6}. \end{bmatrix}. \tag{C.16}$$

The CRLB is then found according to (C.7) upon extracting the diagonal elements of the inverse of the FIM in (C.16).

C.3 Maximum Likelihood Frequency Estimation

The maximum likelihood estimate (MLE) of the parameter vector $\boldsymbol{\theta}$ is given by the estimate that maximises the likelihood of obtaining the data vector, that is [138]

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} = \underset{\boldsymbol{\theta}}{\text{argmax}} \ln \text{pdf}(\boldsymbol{x}, \boldsymbol{\theta}). \tag{C.17}$$

Since the likelihood function for the IID Gaussian data is given in (C.6), the MLE scheme can be expressed as

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} = \underset{\boldsymbol{\theta}}{\text{argmax}} - \sum_{k=0}^{N-1} |x_k - s_k(\boldsymbol{\theta})|^2 . \tag{C.18}$$

Note that due to the negative sign, the maximisation procedure in can be replaced by a minimisation where

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} = \underset{\boldsymbol{\theta}}{\text{argmin}} \underbrace{\sum_{k=0}^{N-1} |x_k - s_k(\boldsymbol{\theta})|^2}_{J_{\text{MLE}}} . \tag{C.19}$$

Using the vector signal model in (C.2), the MLE cost function J_{MLE} can be written as

$$J_{\text{MLE}}(\boldsymbol{\theta}) = \|\boldsymbol{x} - \boldsymbol{s}(\boldsymbol{\theta})\|^2, \tag{C.20}$$

where the signal model, for the frequency estimation problem at hand is given by

$$\boldsymbol{s} = A\boldsymbol{e} + B\boldsymbol{e}^*,$$

with A and B the complex-valued phasors and the exponential vector $\boldsymbol{e}$ is a function of the system frequency ω

$$\boldsymbol{e} \overset{\text{def}}{=} \begin{bmatrix} 1, & e^{j\omega}, & \dots & , e^{j\omega(N-1)} \end{bmatrix}^{\mathsf{T}} . \tag{C.21}$$

The MLE cost function in (C.20) now becomes

$$J_{\text{MLE}}(A, B, \omega) = \|\boldsymbol{x} - A\boldsymbol{e} + B\boldsymbol{e}^*\|^2, \tag{C.22}$$

and is a function of the complex-valued phasors A, B and a real valued frequency, ω.

We first find the MLE for the phasors A and B, using

$$\frac{\partial J}{\partial A^*} = -\boldsymbol{e}^{\mathsf{H}}(\boldsymbol{x} - A\boldsymbol{e} - B\boldsymbol{e}^*) = 0 \tag{C.23}$$

$$= -\boldsymbol{e}^{\mathsf{H}}\boldsymbol{x} + A\boldsymbol{e}^{\mathsf{H}}\boldsymbol{e} + B\boldsymbol{e}^{\mathsf{H}}\boldsymbol{e}^* = 0, \tag{C.24}$$

Noting that $\boldsymbol{e}^{\mathsf{H}}\boldsymbol{e} = N$ and $\boldsymbol{e}^{\mathsf{T}}\boldsymbol{e} = (\boldsymbol{e}^{\mathsf{H}}\boldsymbol{e}^*)^* \approx 0$ gives the MLE estimate as

$$\hat{A}_{\text{MLE}} = \frac{1}{N}\boldsymbol{e}^{\mathsf{H}}\boldsymbol{x}. \tag{C.25}$$

Applying the same principle for the negative sequence phasor, B, gives

$$\hat{B}_{\text{MLE}} = \frac{1}{N}\boldsymbol{e}^{\mathsf{T}}\boldsymbol{x}. \tag{C.26}$$

Now, substituting the MLE estimates of A and B into the original cost function yields the cost function only as a function of the frequency, ω, that is

$$J_{\text{MLE}} = \|\boldsymbol{x} - \frac{1}{N}\boldsymbol{e}^{\mathsf{H}}\boldsymbol{x}\boldsymbol{e} + \frac{1}{N}\boldsymbol{e}^{\mathsf{T}}\boldsymbol{x}\boldsymbol{e}^*\|^2. \tag{C.27}$$

Upon expansion, the cost function becomes

$$J_{\text{MLE}} = \|\boldsymbol{x}\|^2 - \frac{1}{N}\left(|\boldsymbol{e}^{\mathsf{H}}\boldsymbol{x}|^2 + |\boldsymbol{e}^{\mathsf{H}}\boldsymbol{x}^*|^2\right). \tag{C.28}$$

Since minimising (C.28) is equivalent to maximising the last two terms in (C.28), the MLE for the frequency, ω is given by

$$\hat{\omega}_{\text{MLE}} = \arg\max_{\omega} \frac{1}{N}\left(|\boldsymbol{e}^{\mathsf{H}}\boldsymbol{x}|^2 + |\boldsymbol{e}^{\mathsf{H}}\boldsymbol{x}^*|^2\right). \tag{C.29}$$

Note that $\frac{1}{N}|\boldsymbol{e}^{\mathsf{H}}\boldsymbol{x}|^2$ is the well-known periodogram of the data $\boldsymbol{x}$, where [48]

$$P(\omega) = \frac{1}{N}|\boldsymbol{e}^{\mathsf{H}}\boldsymbol{x}|^2 = \frac{1}{N}\left|\sum_{k=0}^{N-1} x_k e^{-j\omega k}\right|^2, \tag{C.30}$$

while $\frac{1}{N}|\boldsymbol{e}^{\mathsf{H}}\boldsymbol{x}^*|^2$ can be considered the "conjugate periodogram" where

$$\tilde{P}(\omega) = \frac{1}{N}|\boldsymbol{e}^{\mathsf{H}}\boldsymbol{x}^*|^2 = \frac{1}{N}\left|\sum_{k=0}^{N-1} x_k^* e^{-j\omega k}\right|^2. \tag{C.31}$$

The MLE scheme in (C.29) is reminiscent of the classic result by Rife & Boorstyn which states that the MLE frequency estimator is the periodogram maximiser [197]. However, the

difference in (C.29) is that for a general unbalanced voltage, both the periodogram and the conjugate periodogram needs to be maximised. We therefore refer to the ML frequency estimation scheme as the augmented periodogram maximiser.

We next show that maximising the augmented periodogram is equivalent to maximising either the original or the conjugate periodogram. To this end, let us denote the exponential vector with the true signal frequency as $\boldsymbol{e}_\circ$, so that the observed signal takes the form

$$\boldsymbol{x} = A\boldsymbol{e}_\circ + B\boldsymbol{e}_\circ^* + \boldsymbol{\eta}.$$

Substituting the model for the signal $\boldsymbol{x}$ into the MLE cost function in (C.29) gives

$$J_{\text{MLE}}(\omega) = \frac{1}{N}\left(|\boldsymbol{e}^{\mathsf{H}}(A\boldsymbol{e}_\circ + B\boldsymbol{e}_\circ^* + \boldsymbol{\eta})|^2 + |\boldsymbol{e}^{\mathsf{H}}(A^*\boldsymbol{e}_\circ^* + B^*\boldsymbol{e}_\circ + \boldsymbol{\eta}^*)|^2\right) \tag{C.32}$$

$$= \underbrace{\frac{1}{N}|A\boldsymbol{e}^{\mathsf{H}}\boldsymbol{e}_\circ + B\boldsymbol{e}^{\mathsf{H}}\boldsymbol{e}_\circ^* + \boldsymbol{e}^{\mathsf{H}}\boldsymbol{\eta})|^2}_{P(\omega)} + \underbrace{\frac{1}{N}|(A^*\boldsymbol{e}^{\mathsf{H}}\boldsymbol{e}_\circ^* + B^*\boldsymbol{e}^{\mathsf{H}}\boldsymbol{e}_\circ + \boldsymbol{e}^{\mathsf{H}}\boldsymbol{\eta}^*)|^2}_{\tilde{P}(\omega)}, \tag{C.33}$$

The MLE cost function, i.e. the augmented periodogram, is maximised when $\boldsymbol{e} = \boldsymbol{e}_\circ$ which is true when the estimated frequency is equal the actual frequency. However notice that the maximisation of either the periodogram, $P(\omega)$, or the conjugate periodogram, $\tilde{P}(\omega)$, also yields the correct frequency. Therefore, the MLE frequency estimate can be performed by the periodogram maximiser, originally proposed for single-tone sinusoids as [197]

$$\begin{aligned}\hat{\omega}_{\text{MLE}} &= \arg\max_\omega \frac{1}{N}|\boldsymbol{e}^{\mathsf{H}}\boldsymbol{x}|^2 \\ &= \arg\max_\omega \frac{1}{N}\left|\sum_{k=0}^{N-1} x_k e^{-j\omega k}\right|^2.\end{aligned} \tag{C.34}$$

C.3.1 Maximisation Scheme

The proposed maximisation scheme follows a standard periodogram maximisation where a coarse search is followed by a finer search [197]. Since the periodogram can be obtained by a discrete Fourier transform (DFT) for frequencies sampled on a grid, we are able to use the frequency that corresponds to the maximum DFT bin as an initial guess to the maximisation procedure, so that,

$$\hat{\omega}_{\text{coarse}} = \max|\text{DFT}\{\boldsymbol{x}\}|^2. \tag{C.35}$$

A finer search is performed by a Newton-Raphson method given by

$$\hat{\omega}_{m+1} = \hat{\omega}_m - \left.\frac{\nabla_\omega J}{\nabla_\omega^2 J}\right|_{\omega=\hat{\omega}_m}, \tag{C.36}$$

where the gradient, $\nabla_\omega J$, and Hessian, $\nabla^2_\omega J$ are found to be

$$\nabla_\omega J = \frac{\partial |\boldsymbol{e}^{\mathsf{H}}\boldsymbol{x}|^2}{\partial \omega} = 2\mathsf{Re}\left\{j\boldsymbol{e}^{\mathsf{H}}\boldsymbol{x}\boldsymbol{e}^{\mathsf{T}}\mathbf{D}\boldsymbol{x}^*\right\} \tag{C.37}$$

$$\nabla^2_\omega J = \frac{\partial^2 |\boldsymbol{e}^{\mathsf{H}}\boldsymbol{x}|^2}{\partial \omega^2} = 2(|\boldsymbol{e}^{\mathsf{T}}\mathbf{D}\boldsymbol{x}^*|^2 - \mathsf{Re}\left\{\boldsymbol{e}^{\mathsf{H}}\boldsymbol{x}\boldsymbol{e}^{\mathsf{T}}\mathbf{D}^2\boldsymbol{x}^*\right\}), \tag{C.38}$$

and the Newton-Raphson procedure is (C.36) is initialised with the coarse estimate from (C.35), $\hat{\omega}_0 \leftarrow \hat{\omega}_{\text{coarse}}$. The iterative maximisation procedure is therefore given by

$$\hat{\omega}_{m+1} = \hat{\omega}_m - \frac{\mathsf{Re}\left\{j\boldsymbol{e}_m^{\mathsf{H}}\boldsymbol{x}\boldsymbol{e}_m^{\mathsf{T}}\mathbf{D}\boldsymbol{x}^*\right\}}{|\boldsymbol{e}_m^{\mathsf{T}}\mathbf{D}\boldsymbol{x}^*|^2 - \mathsf{Re}\left\{\boldsymbol{e}_m^{\mathsf{H}}\boldsymbol{x}\boldsymbol{e}_m^{\mathsf{T}}\mathbf{D}^2\boldsymbol{x}^*\right\}}, \tag{C.39}$$

where the vector of exponential elements, $\boldsymbol{e}_m = \left(e^{j\hat{\omega}_m k}\right)_{k=0}^{N-1}$, is a function of the estimated frequency $\hat{\omega}_m$ at iteration m. The stopping criterion is chosen based on the absolute difference between successive estimates $|\hat{\omega}_{m+1} - \hat{\omega}_m| < \epsilon$.

To illustrate that the MLE estimate proposed in (C.39), it is applied to to the frequency estimation task on the Clarke-voltage which is corrupted with various levels of white Gaussian noise. Figure C.1 shows the steady-state error of the ML estimator in (C.39) achieves the CRLB while the performance of the ACEKF (introduced in Chapter 6) plateaus at the 50 dB SNR level[1]. This conforms with previous results comparing the performance of frequency estimators [141].

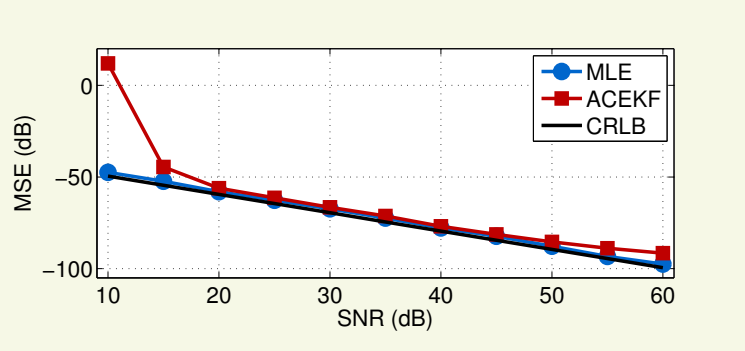

Figure C.1: Steady-state mean square error (MSE) of the maximum likelihood estimator (MLE), augmented complex extended Kalman filter (ACEKF), and Cramer-Rao lower bound (CRLB) at different noise levels.

[1]Note that ACEKF was configured to operate in a stationary environment by setting its state noise covariance matrix, $\bar{\mathbf{C}} = \mathbf{0}$, for a fair comparison with the ML estimator.

Appendix D

Rapid Frequency Response for Low Inertia Grids

Appendix Overview

This appendix describes a potential application of distributed frequency estimation for the provision of synthetic inertia by wind turbines. This is critical for future grids where less inertia will be available to support sudden changes in the supply-demand profile in the system. We exploit the use of spatially diverse frequency estimates to modulate the power output of wind turbines so as to provide synthetic inertia support. This has important implications not only over the dynamics of the power system but also over the sizing of the grid interface of the wind turbines. A 2850-bus model of the Great Britain (GB) electricity grid was used to illustrate the benefit of the proposed idea. This work is currently being prepared for publication [41].

D.1 Background: Synthetic Inertia From Wind Turbines

Future low inertia grids are envisaged to experience rapid frequency excursions and the consequent stability issues. A possible solution to this issue is the provision of rapid frequency response (RFR) in the form of "synthetic inertia" from wind power generation systems [198, 199, 200, 201]. This rapid frequency response would emulate the behaviour of synchronous generators, which inherently exchange power with the network following a frequency excursion event.

To illustrate this issue, consider a frequency event which was modelled using a detailed 2850-bus model of the GB electrical system in the DIgSILENT PowerFactory simulation software [202]. Figure D.1 shows a representation of the model which is divided into three-regions that correspond to Scottish Hydroelectric, Scottish Power, and National Grid [202].

The model was configured to include wind farms that were able to provide synthetic inertia during a frequency event. To simulate a contingency, a nuclear power plant was tripped at the 2 s time instant which caused a 2.2.GW loss in the supply. Figure D.2 shows that without the inertia support, the average frequency of the system rapidly declines after the fault. Figure D.2 also shows that when wind farms are able to provide synthetic inertia support, both the rate of change of frequency (RoCoF) and the minimum frequency (frequency nadir) are made less severe.

Although the mechanisms to provide rapid frequency response from wind turbines have

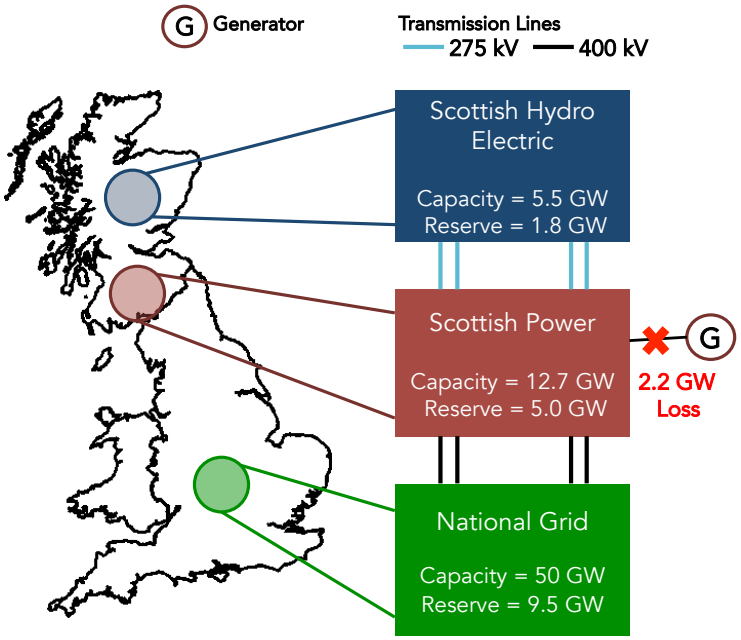

Figure D.1: Great Britain transmission grid model.

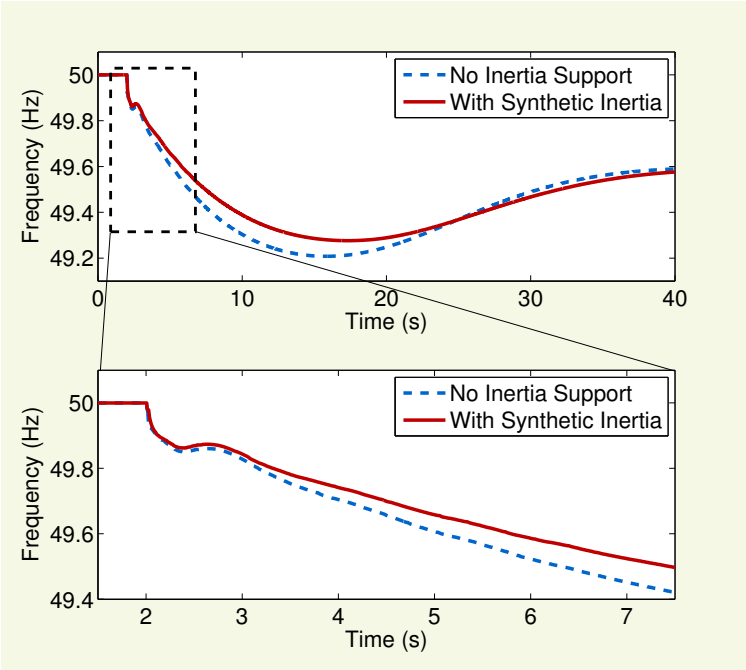

Figure D.2: Average frequency in the network with and without inertia support from wind farms.

received much attention in the power and energy community [203, 204, 205, 206], several issues remain unanswered [207]. The first concern is the possibility that inertia support could amplify different forms of noise and interfere with other controllers due to unreliable measurements of the frequency. Even if the grid is synchronous, transient phenomena may cause short term discrepancies between the frequencies seen in far away buses [199, 208].

Figure D.3 shows the range of frequency measurements from 2850 buses in England, Scotland and Wales after the 2.2 GW generator trip. The overlay of frequency measurements in the GB network in Figure D.3 shows that individual frequency measurements exhibit oscillations which need to be filtered before they are used as inputs to the synthetic inertia support schemes. Although this can be accomplished using low-pass filters, the inherent delays introduced by the filters, typically in the range of hundreds of milliseconds [207, 209], make them undesirable for RFR schemes.

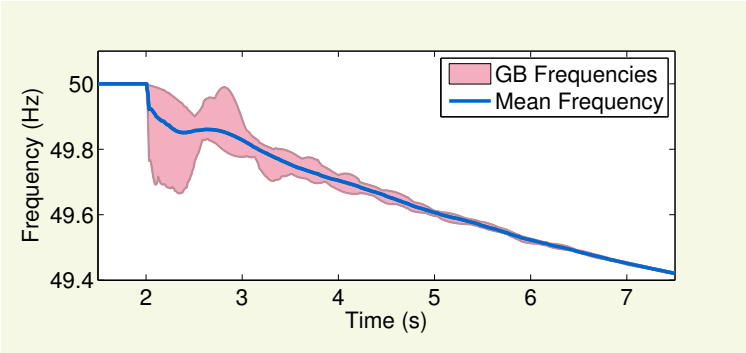

Figure D.3: Overlay of frequency measurements in the GB network. Rapid oscillations in individual measurement can be filtered out by spatial averaging (mean frequency).

Secondly, if wind-turbines were to exactly replicate the inertia of individual synchronous generators, the output power required in the sub-second time-scale could be much higher than the headroom required for nominal operating conditions. This is caused by large frequency swings experienced during the first few seconds after the inception of a fault, as seen in Figure D.3. While electrical machines can easily withstand currents much greater than their rated values over short periods of time, power electronic converters that interface the wind turbines have very modest short-term ratings. Therefore, the speed at which a wind turbine reacts to a frequency event can have a significant impact on how its power electronic converter is sized.

D.2 Proposed Method: Spatial Averaging for Aggregated Synthetic Inertia

In this work, we propose that the two challenges of having more reliable frequency estimates and having more economical power ratings for inverters can be tackled by emulating different levels of *aggregate inertial response* of synchronous generators. This is achieved by spatial filtering of wide-area frequency measurements in order to attenuate spurious artefacts and local

transient oscillations.

Consider a network of N nodes where each node represents a frequency measurement at a certain bus. The frequency estimate at the node i and time-instant k is denoted by $\omega_{i,k}$. The spatially averaged frequency estimate is therefore given by

$$\bar{\omega}_k = \sum_{\ell=1}^{N} a_\ell \, \omega_{\ell,k}, \tag{D.1}$$

where a_ℓ are the (possibly time-varying) weighting coefficients used to obtain the averaged frequency, $\bar{\omega}_k$. An essential condition for the spatial filtering operation in (D.1) is that the weighting coefficients a_ℓ satisfy the condition $\sum_{\ell=1}^{N} a_\ell = 1$.

The choice of the optimal weighting coefficients for the averaging operation in (D.1) is out of the scope of this work. However, for completeness, several weighting schemes ranging from a simple average i.e. $a_\ell = 1/N \; \forall \ell$ to a strategy based on *a priori* knowledge of the electricity grid or the measurement statistics can be employed. In general, weighting coefficients based on the network topology take the form, $a_i = \left(\sum_{\ell=1}^{N} \gamma_\ell\right)^{-1} \gamma_i$, where γ_i represents physical or statistical information about the busbar (node) i. For example, γ_i could be the amount of inertia available in the region surrounding bus i, [208], or confidence in the frequency measurements as indicated by the noise variance, σ_i^2, i.e. $\gamma_i = 1/\sigma_i^2$ [86].

Furthermore, the spatially averaged frequency measurements enable wind turbines to emulate the aggregate inertial response of the electrical machines in a network as opposed matching the power output of any individual machine. Figure D.4 shows the power output (synthetic inertia) from a certain wind turbine which uses different frequency inputs in its controller. Observe that using conventional frequency estimates requires the wind turbines to ramp up their power output within a period of two seconds (dashed yellow line and red line). On the other hand, by exploiting the availability of spatially diverse frequency measurements, we are able to obtain a spatially averaged frequency estimate which in turn can be used to modulate the power output of the wind turbine (blue line).

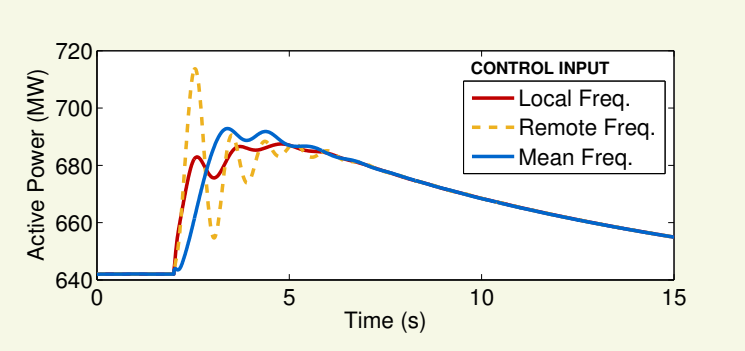

Figure D.4: Power output from wind turbines that provide synthetic inertia. Without spatial filtering, the wind turbines have to provide a much higher power output to emulate synchronous generators.

References

[1] A. Ipakchi and F. Albuyeh, "Grid of the future," *IEEE Power and Energy Magazine*, vol. 7, no. 2, pp. 52–62, March 2009. 19

[2] "Meeting the energy challenge: A white paper on energy," Department of Trade and Industry, HM Goverment, United Kingdom, Tech. Rep. CM 7124, May 2007. 19

[3] C. Xu, "High accuracy real-time GPS synchronized frequency measurement device for wide-area power grid monitoring," Ph.D. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA, U.S.A., February 2006. 19

[4] R. Piwko, "Interconnection requirements for variable generation," North American Electric Reliability Corporation (NERC), Atlanta, GA, USA, Tech. Rep., September 2012. 19

[5] A. G. Phadke, J. S. Thorp, and M. G. Adamiak, "A new measurement technique for tracking voltage phasors, local system frequency, and rate of change of frequency," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-102, no. 5, pp. 1025–1038, May 1983. 19, 179

[6] P. Tielens and D. Van Hertem, "The relevance of inertia in power systems," *Renewable and Sustainable Energy Reviews*, vol. 55, pp. 999–1009, 2016. 19, 22, 23, 178

[7] G. Constable and B. Somerville, *A Century of Innovation: Twenty engineering achievements that transformed our lives*. Washington, D.C, USA: Joseph Henry Press, October 2003. 20

[8] T. McNichol, *AC/DC: The Savage Tale of the First Standards War*. Hoboken, NJ, U.S.A: Jossey-Bass, September 2006. 20

[9] B. G. Lamme, "The technical story of the frequencies," *Transactions of the American Institute of Electrical Engineers*, vol. XXXVII, no. 1, pp. 65–89, Jan 1918. 21

[10] G. Neidhofer, "50-Hz frequency [History]," *IEEE Power and Energy Magazine*, vol. 9, no. 4, pp. 66–81, July 2011. 21

[11] T. Ngo, M. Lwin, and S. Santoso, "Steady-state analysis and performance of low frequency AC transmission lines," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3873–3880, Sept 2016. 21

[12] E. L. Owen, "The origins of 60-Hz as a power frequency," *IEEE Industry Applications Magazine*, vol. 3, no. 6, pp. 8, 10, 12–14, Nov 1997. 21

[13] P. M. Ashton, "Exploiting the use of phasor measurement units for enhanced transmission network operation and control," Ph.D. dissertation, Brunel University London, Uxbridge, U.K., 2014. 21, 164

[14] "Electricity Ten Year Statement," National Grid, United Kingdom, Tech. Rep., November 2015. 21

[15] *Role of the BPA in the Pacific Northwest Power Supply System*, 1st ed. Portland, OR, U.S.A.: Bonneville Power Administration, August 1977. 21

[16] C. Seneviratne and C. Ozansoy, "Frequency response due to a large generator loss with the increasing penetration of wind/PV generation: A literature review," *Renewable and Sustainable Energy Reviews*, vol. 57, pp. 659 – 668, 2016. 22, 23

[17] M. H. J. Bollen, *Understanding Power Quality Problems: Voltage Sags and Interruptions*. New York, U.S.A.: Wiley-IEEE Press, September 1999, vol. 3. 22, 23

[18] "System operability framework," National Grid, United Kingdom, Tech. Rep., November 2015. 23

[19] A. von Jouanne and B. Banerjee, "Assessment of voltage unbalance," *IEEE Trans. Power Del*, vol. 16, no. 4, pp. 782–790, Oct 2001. 23, 42, 138, 139

[20] M. H. J. Bollen, I. Y. H. Gu, S. Santoso, M. F. McGranaghan, P. A. Crossley, M. V. Ribeiro, and P. F. Ribeiro, "Bridging the gap between signal and power," *IEEE Signal Processing Magazine*, vol. 26, no. 4, pp. 11–31, 2009. 23

[21] A. Singh and J. C. Principe, "Using correntropy as a cost function in linear adaptive filters," in *Proc. of the IEEE International Joint Conference on Neural Networks*, June 2009, pp. 2950–2955. 23

[22] Y. Xia and D. Mandic, "Widely linear adaptive frequency estimation of unbalanced three-phase power systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 1, pp. 74–83, 2012. 23, 128, 139, 147, 148

[23] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, no. 1, pp. 65 – 78, 2004. 24, 90

[24] Y. Xia, S. Douglas, and D. Mandic, "Adaptive frequency estimation in smart grid applications: Exploiting noncircularity and widely linear adaptive estimators," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 44–54, 2012. 25, 26, 128, 135, 136, 139, 147, 165, 178, 179

[25] M. Bollen and L. Zhang, "Different methods for classification of three-phase unbalanced voltage dips due to faults," *Electric Power Systems Research*, vol. 66, no. 1, pp. 59 – 69, 2003. 26, 141

[26] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 2030–2033, 1995. 26, 36, 39, 41, 58

[27] Y. F. Huang, S. Werner, J. Huang, N. Kashyap, and V. Gupta, "State estimation in electric power grids: Meeting new challenges presented by the requirements of the future grid," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 33–43, Sept 2012. 27, 164

[28] S. Rusitschka, K. Eger, and C. Gerdes, "Smart grid data cloud: A model for utilizing cloud computing in the smart grid domain," in *Proc. of the First IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Oct 2010, pp. 483–488. 27

[29] S. M. Amin and B. F. Wollenberg, "Toward a smart grid: Power delivery for the 21st century," *IEEE Power and Energy Magazine*, vol. 3, no. 5, pp. 34–41, Sept 2005. 27

[30] C. Jahanchahi, S. Kanna, and D. P. Mandic, "Complex dual channel estimation: Cost effective widely linear adaptive filtering," *Signal Processing*, vol. 104, pp. 33–42, November 2014. 27, 28, 83

[31] S. Kanna, S. Douglas, and D. Mandic, "A real time tracker of complex circularity," in *Proc. of the 8th IEEE Sensor Array and Multichannel Signal Process. Workshop (SAM)*, June 2014, pp. 129–132. 28

[32] S. Kanna, M. Xiang, and D. P. Mandic, "Real-time detection of rectilinear sources for wireless communication signals," in *Proc. of the 2015 International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2015, pp. 541–545. 28

[33] S. Kanna, S. Talebi, and D. Mandic, "Diffusion widely linear adaptive estimation of system frequency in distributed power grids," in *Proc. of the IEEE Intl. Energy Conf. (ENERGYCON)*, May 2014, pp. 772–778. 28, 95, 96

[34] S. Kanna, D. H. Dini, Y. Xia, S. Y. Hui, and D. P. Mandic, "Distributed widely linear Kalman filtering for frequency estimation in power networks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 1, no. 1, pp. 45–57, March 2015. 29

[35] D. Mandic, S. Kanna, and S. Douglas, "Mean square analysis of the CLMS and ACLMS for non-circular signals: The approximate uncorrelating transform approach," in *Proc. of the IEEE Intl. Conf. on Acoust. Speech and Signal Process. (ICASSP)*, April 2015, pp. 3531–3535. 29, 105

[36] S. Kanna and D. P. Mandic, "Steady-state behavior of general complex-valued diffusion LMS strategies," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 722–726, May 2016. 29

[37] ——, "Modern view of three-phase transforms," *Submitted to IEEE Power Electronic Magazine*, September 2016. 30

[38] D. P. Mandic, S. Kanna, and A. G. Constantinides, "On the intrinsic relationship between the least mean square and Kalman filters [Lecture Notes]," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 117–122, Nov 2015. 30

[39] S. Kanna, A. G. Constantinides, and D. P. Mandic, "The equivalence of FM demodulation and recursive DFT for frequency tracking in power systems," *Submitted to International Journal of Electrical Power & Energy Systems*, 2016. 30

[40] Y. Xia, S. Kanna, and D. Mandic, "Maximum likelihood frequency estimation of unbalanced three-phase power signals," *Submitted to IEEE Transactions on Power Systems*, 2016. 33, 187

[41] S. Kanna, Y. Pipelzadeh, A. Junyent-Ferre, T. Green, and D. Mandic, "Aggregating synthetic inertia using wind turbines: A rapid frequency response solution for low inertia grids," *In preparation for IEEE Transactions on Power Systems*, 2016. 33, 195

[42] D. P. Mandic and V. S. L. Goh, *Complex valued nonlinear adaptive filters: Noncircularity, widely linear and neural models*. Wiley, 2009. 35, 46, 57, 65, 82

[43] P. J. Schreier and L. L. Scharf, *Statistical signal processing of complex-valued data: The theory of improper and noncircular signals*. Cambridge University Press, 2010. 35, 36, 37, 39, 57, 188

[44] I. Aizenberg, *Complex-Valued Neural Networks with Multi-Valued Neurons (Studies in Computational Intelligence)*. Springer, 2011. 36

[45] E. Ollila, "On the circularity of a complex random variable," *IEEE Signal Processing Letters*, vol. 15, pp. 841–844, 2008. 36, 37, 73, 78

[46] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000. 39, 120, 172

[47] J. Eriksson, E. Ollila, and V. Koivunen, "Essential statistics and tools for complex random variables," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5400–5408, 2010. 41

[48] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, 1996. 43, 44, 192

[49] J. Navarro-Moreno, J. Moreno-Kayser, R. Fernandez-Alcala, and J. Ruiz-Molina, "Widely linear estimation algorithms for second-order stationary signals," *IEEE Trans. on Signal Process.*, vol. 57, no. 12, pp. 4930–4935, 2009. 44

[50] L. Bottou, "Online learning and stochastic approximations," *On-line Learning in Neural Networks*, vol. 17, no. 9, p. 142, 1998. 45, 46

[51] H. Robbins and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, pp. 400–407, 1951. 45, 46

[52] B. Widrow, J. McCool, and M. Ball, "The complex LMS algorithm," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 719–720, 1975. 46, 60, 75, 82

[53] S. Javidi, M. Pedzisz, S. L. Goh, and D. P. Mandic, "The augmented complex least mean square algorithm with application to adaptive prediction problems," *Proc. 1st IARP Workshop on Cognitive Information Processing*, pp. 54–57, 2008. 47, 61, 82, 94

[54] J. Mazo, "On the independence theory of equalizer convergence," *Bell Systems Technical Journal*, vol. 58, no. 5, pp. 963–966, May-June 1979. 49

[55] A. Knutson and T. Tao, "Honeycombs and sums of Hermitian matrices," *Notices of the AMS*, vol. 48, no. 2, pp. 175–186, 2001. 53

[56] F. Neeser and J. Massey, "Proper complex random processes with applications to information theory," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1293–1302, 1993. 57

[57] K. Kreutz-Delgado, "The complex gradient operator and the $\mathbb{CR}$-calculus," University of California, San Diego, Technical Report, 2006. 57, 61

[58] D. Mandic, S. Still, and S. Douglas, "Duality between widely linear and dual channel adaptive filtering," in *Proc. of the IEEE Intl. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, April 2009, pp. 1729–1732. 57, 60, 63, 64

[59] G. A. N. Fernando, V. H. Nascimiento, and M. T. M. Silva, "Reduced-complexity widely linear adaptive estimation," *Proceedings of the Seventh International Symposium on Wireless Communication Systems*, pp. 399–403, 2010. 57, 63, 64

[60] B. D. H. Brandwood, "A complex gradient operator and its applications in adaptive array theory," *IEE Proceedings: Communications, Radar and Signal Proceedings*, vol. 130, no. 1, pp. 11–16, 1983. 61

[61] A. van den Bos, "Complex gradient and Hessian," *IEE Proc.: Vision, Image and Signal Processing*, vol. 141, no. 6, pp. 380–383, 1994. 61

[62] N. Yousef and A. Sayed, "A unified approach to the steady-state and tracking analyses of adaptive filters," *IEEE Transactions on Signal Processing*, vol. 49, no. 2, pp. 314 –324, 2001.

[63] D. Mandic, Y. Xia, and S. Douglas, "Steady state analysis of the CLMS and augmented CLMS algorithms for noncircular complex signals," in *Conference Record of the Forty Fourth Conference on Signals, Systems and Computers (ASILOMAR)*, 2010, pp. 1635–1639. 71

[64] P. Schreier, L. Scharf, and A. Hanssen, "A generalized likelihood ratio test for impropriety of complex signals," *IEEE Signal Processing Letters*, vol. 13, no. 7, pp. 433–436, 2006. 73, 83

[65] A. Walden and P. Rubin-Delanchy, "On testing for impropriety of complex-valued Gaussian vectors," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 825–834, 2009. 73, 83

[66] J.-P. Delmas, A. Oukaci, and P. Chevalier, "Asymptotic distribution of GLR for impropriety of complex signals," in *Proc. of the IEEE Intl. Conf. on Acoust. Speech and Signal Process. (ICASSP)*, 2010, pp. 3594–3597. 73

[67] E. Ollila and V. Koivunen, "Adjusting the generalized likelihood ratio test of circularity robust to non-normality," in *Proc. of the 10th IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC '09)*, 2009, pp. 558–562. 73

[68] E. Ollila, V. Koivunen, and H. Poor, "A robust estimator and detector of circularity of complex signals," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, 2011, pp. 3620–3623. 73

[69] B. Jelfs, D. P. Mandic, and S. C. Douglas, "An adaptive approach for the identification of improper complex signals," *Signal Processing*, vol. 92, no. 2, pp. 335 – 344, 2012. 74

[70] D. Mandic, S. Javidi, S. Goh, A. Kuh, and K. Aihara, "Complex-valued prediction of wind profile using augmented complex statistics," *Renewable Energy*, vol. 34, no. 1, pp. 196 – 201, 2009. 79

[71] P. Chevalier, J.-P. Delmas, and A. Oukaci, "Optimal widely linear MVDR beamforming for noncircular signals," in *Proc. of the IEEE Intl. Conf. on Acoust. Speech and Signal Process. (ICASSP)*, April 2009, pp. 3573–3576. 82

[72] A. H. Sayed, *Fundamentals of Adaptive Filtering*. John Wiley & Sons, 2003. 82

[73] S. Douglas, "Widely-linear recursive least-squares algorithm for adaptive beamforming," in *Proc. of the IEEE Intl. Conf. on Acoust. Speech and Signal Process. (ICASSP)*, April 2009, pp. 2041–2044. 82

[74] P. Chevalier, J. P. Delmas, and A. Oukaci, "Properties, performance and practical interest of the widely linear MMSE beamformer for non-rectilinear signals," *Signal Processing*, vol. 97, pp. 269–281, 2014. 82, 83

[75] C. Hellings, M. Koller, and W. Utschick, "An impropriety test based on block-skew-circulant matrices," in *Proc. of the 19th International ITG Workshop on Smart Antennas*, March 2015, pp. 1–6. 83

[76] J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. on Signal Process.*, vol. 44, no. 12, pp. 3017–3030, Dec 1996. 83, 84

[77] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic Press, 2010. 83

[78] H. Li and T. Adali, "Algorithms for complex ML ICA and their stability analysis using Wirtinger calculus," *IEEE Trans. on Signal Process.*, vol. 58, no. 12, pp. 6156–6167, 2010. 84

[79] J.-F. Cardoso and T. Adali, "The maximum likelihood approach to complex ICA," in *Proc. of the IEEE Intl. Conf. on Acoust. Speech and Signal Process. (ICASSP)*, May 2006, pp. 673–676. 84

[80] S.-I. Amari and A. Cichocki, "Adaptive blind signal processing-neural network approaches," *Proc. of the IEEE*, vol. 86, no. 10, pp. 2026–2048, 1998. 84

[81] A. Cichocki and S.-I. Amari, *Adaptive blind signal and image processing: Learning algorithms and applications*. John Wiley & Sons, 2002. 84

[82] C.-Y. Chong and S. P. Kumar, "Sensor networks: Evolution, opportunities, and challenges," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1247–1256, Aug 2003. 89

[83] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov 2010. 89

[84] I. D. Schizas, G. Mateos, and G. B. Giannakis, "Distributed LMS for consensus-based in-network adaptive processing," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2365–2382, June 2009. 89, 90

[85] I. D. Schizas, G. B. Giannakis, S. I. Roumeliotis, and A. Ribeiro, "Consensus in ad hoc WSNs with noisy links–Part II: Distributed estimation and smoothing of random signals," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1650–1666, April 2008. 89, 90

[86] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1035 –1048, March 2010. 90, 93, 96, 198

[87] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974. 91

[88] S. Kar and J. M. F. Moura, "Consensus + innovations distributed inference over networks: Cooperation and sensing in networked systems," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 99–109, May 2013. 91

[89] V. Blondel, J. Hendrickx, A. Olshevsky, and J. Tsitsiklis, "Convergence in multiagent coordination, consensus, and flocking," in *Proc of the 44th IEEE Conf. on Decision and Control and European Control Conference. CDC-ECC '05.*, Dec 2005, pp. 2996–3000. 91, 101

[90] A. H. Sayed, "Diffusion Adaptation over Networks," *Academic Press Library in Signal Processing*, vol. 3, pp. 323–454, 2014. 92, 95, 99, 100, 103, 104, 108

[91] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3122 –3136, July 2008. 93, 95, 96

[92] S. S. Stankovic, M. S. Stankovic, and D. M. Stipanovic, "Decentralized parameter estimation by consensus based stochastic approximation," in *Proc. of the 46th IEEE Conference on Decision and Control*, Dec 2007, pp. 1535–1540. 93

[93] Y. Xia, D. P. Mandic, and A. H. Sayed, "An adaptive diffusion augmented CLMS algorithm for distributed filtering of noncircular complex signals," *IEEE Signal Processing Letters*, vol. 18, no. 11, pp. 659 –662, Nov. 2011. 93, 94, 95, 96

[94] X.R. Li, Y. Zhu, J. Wang, and C. Han, "Optimal linear estimation fusion: Part I: Unified fusion rules," *IEEE Transactions on Information Theory*, vol. 49, no. 9, pp. 2192 – 2208, Sep. 2003. 95

[95] R. Bru, L. Elsner, and M. Neumann, "Convergence of infinite products of matrices and inner-outer iteration schemes," *Electronic Transactions on Numerical Analysis*, vol. 2, no. 3, pp. 183 –193, Dec. 1994. 95

[96] F. S. Cattivelli and A. H. Sayed, "Diffusion strategies for distributed Kalman filtering and smoothing," *IEEE Transactions on Automatic Control*, vol. 55, no. 9, pp. 2069 –2084, Sept. 2010. 95, 107, 108, 115

[97] A. H. Sayed, "Adaptive networks," *Proceedings of the IEEE*, vol. 102, no. 4, pp. 460–497, April 2014. 46, 95, 108, 165

[98] J. Chen, C. Richard, and A. Sayed, "Diffusion LMS over multitask networks," *IEEE Transactions on Signal Processing*, vol. 63, no. 11, pp. 2733–2748, June 2015. 96

[99] R. Arablouei, S. Werner, Y.-F. Huang, and K. Dogancay, "Distributed least mean-square estimation with partial diffusion," *IEEE Transactions on Signal Processing*, vol. 62, no. 2, pp. 472–484, Jan 2014. 96

[100] R. Abdolee and B. Champagne, "Centralized adaptation for parameter estimation over wireless sensor networks," *IEEE Communications Letters*, vol. 19, no. 9, pp. 1624–1627, Sept 2015. 96

[101] C. Jiang, Y. Chen, and K. Liu, "Distributed adaptive networks: A graphical evolutionary game-theoretic view," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5675–5688, Nov 2013. 96, 104

[102] X. Zhao and A. Sayed, "Performance limits for distributed estimation over LMS adaptive networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5107–5124, Oct 2012. 96

[103] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, 2012. 98

[104] A. Uncini, *Fundamentals of Adaptive Signal Processing*, ser. Signals and Communication Technology. Cham, Switzerland: Springer International Publishing, 2015. 102

[105] S. Douglas and M. Rupp, "Convergence issues in the LMS adaptive filter," in *The DSP Handbook*, V. Madisetti and D. Williams, Eds. Boca Raton, FL, CRC/IEEE Press, 1998, ch. 19, pp. 19.1 – 19.21. 102

[106] M. T. M. Silva, V. H. Nascimento, and J. Arenas-Garcia, "A transient analysis for the convex combination of two adaptive filters with transfer of coefficients," in *Proc. of the IEEE Intl. Conf. on Acoust. Speech and Signal Process.*, March 2010, pp. 3842–3845. 102

[107] J. Fernandez-Bes, L. A. Azpicueta-Ruiz, J. Arenas-García, and M. T. Silva, "Distributed estimation in diffusion networks using affine least-squares combiners," *Digital Signal Processing*, vol. 36, pp. 1 – 14, 2015. 102

[108] J. Fernandez-Bes, L. A. Azpicueta-Ruiz, M. T. M. Silva, and J. Arenas-Garcia, "Improved least-squares-based combiners for diffusion networks," in *Proc. of the Tenth International Symposium on Wireless Communication Systems (ISWCS 2013)*, Aug 2013, pp. 1–5. 102

[109] F. Cattivelli and A. H. Sayed, "Distributed nonlinear Kalman filtering with applications to wireless localization," in *Proc. of the IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, March 2010, pp. 3522–3525. 107

[110] H. R. Hashemipour, S. Roy, and A. J. Laub, "Decentralized structures for parallel Kalman filtering," *IEEE Transactions on Automatic Control*, vol. 33, no. 1, pp. 88–94, 1988. 107

[111] J. Speyer, "Computation and transmission requirements for a decentralized linear-quadratic-Gaussian control problem," *IEEE Transactions on Automatic Control*, vol. 24, no. 2, pp. 266–269, Apr 1979. 107

[112] B. S. Rao and H. F. Durrant-Whyte, "Fully decentralised algorithm for multisensor Kalman filtering," *IEE Proceedings D - Control Theory and Applications*, vol. 138, no. 5, pp. 413–420, Sept 1991. 107

[113] R. Olfati-Saber, "Distributed Kalman filtering for sensor networks," in *Proc. of the 46th IEEE Conference on Decision and Control*, Dec. 2007, pp. 5492 –5498. 107, 115

[114] R. Olfati-Saber, "Kalman-consensus filter: Optimality, stability, and performance," in *Proc. of the 48th IEEE Conference on Decision and Control*, Dec 2009, pp. 7036–7042. 107

[115] Z. Hidayat, R. Babuska, B. D. Schutter, and A. Nunez, "Decentralized Kalman filter comparison for distributed-parameter systems: A case study for a 1D heat conduction process," in *Proc. of the 16th IEEE Conference on Emerging Technologies Factory Automation (ETFA)*, Sept 2011, pp. 1–8. 107

[116] S. Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6217–6234, Dec 2012. 108

[117] D. H. Dini and D. P. Mandic, "Cooperative adaptive estimation of distributed noncircular complex signals," in *Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Nov 2012, pp. 1518–1522. 108

[118] D. H. Dini and D. P. Mandic, "A class of widely linear complex Kalman filters," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 5, pp. 775 –786, May 2012. 109, 115

[119] T. Karvonen, "Stability of linear and non-linear Kalman filters," Master's thesis, Department of Mathematics and Statistics, University of Helsinki, Dec 2014. 109, 119, 123

[120] L. Ljung, "Asymptotic behavior of the extended kalman filter as a parameter estimator for linear systems," *IEEE Transactions on Automatic Control*, vol. 24, no. 1, pp. 36–50, Feb 1979. 109, 115, 116

[121] D. H. Dini and D. P. Mandic, "Widely linear complex extended Kalman filters," in *Proc. of the Sensor Signal Processing for Defence Conference*, Sept 2011, pp. 1 – 5. 110, 111

[122] S. Kar and J. Moura, "Gossip and distributed Kalman filtering: Weak consensus under weak detectability," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1766 –1784, April 2011. 115

[123] K. Reif, S. Gunther, E. Yaz, and R. Unbehauen, "Stochastic stability of the discrete-time extended Kalman filter," *IEEE Transactions on Automatic Control*, vol. 44, no. 4, pp. 714–728, Apr 1999. 117, 118, 119, 121, 122, 123

[124] J. Deyst, "Correction to "conditions for asymptotic stability of the discrete minimum-variance linear estimator"," *IEEE Transactions on Automatic Control*, vol. 18, no. 5, pp. 562–563, October 1973. 118

[125] J. Deyst and C. Price, "Conditions for asymptotic stability of the discrete minimum-variance linear estimator," *IEEE Transactions on Automatic Control*, vol. 13, no. 6, pp. 702–705, Dec 1968. 118

[126] Y. Song and J. W. Grizzle, "The extended Kalman filter as a local asymptotic observer for nonlinear discrete-time systems," *Journal of Mathematical Systems, Estimation and Control*, vol. 5, no. 1, pp. 59–78, 1995. 118

[127] C. L. Fortescue, "Method of symmetrical co-ordinates applied to the solution of polyphase networks," *Transactions of the American Institute of Electrical Engineers*, vol. 37, no. 2, pp. 1027–1140, 1918. 127, 142

[128] E. Clarke, *Circuit Analysis of A.C. Power Systems*. New York: Wiley, 1943. 127, 141, 142

[129] W. C. Duesterhoeft, M. W. Schulz, and E. Clarke, "Determination of instantaneous currents and voltages by means of alpha, beta, and zero components," *Transactions of the American Institute of Electrical Engineers*, vol. 70, no. 2, pp. 1248–1255, 1951. 127

[130] N. D. Rao and H. N. R. Rao, "Study of symmetrical and related components through the theory of linear vector spaces," *Proceedings of the Institution of Electrical Engineers*, vol. 113, no. 6, pp. 1057–1062, June 1966. 127

[131] G. C. Paap, "Symmetrical components in the time domain and their application to power network calculations," *IEEE Transactions on Power Systems*, vol. 15, no. 2, pp. 522–528, May 2000. 127, 128, 137

[132] R. H. Park, "Two-reaction theory of synchronous machines generalized method of analysis – Part I," *Transactions of the American Institute of Electrical Engineers*, vol. 48, no. 3, pp. 716–727, July 1929. 127, 130, 141, 142

[133] M. Canteli, A. Fernandez, L. Eguiluz, and C. Estebanez, "Three-phase adaptive frequency measurement based on Clarke's transformation," *IEEE Transactions on Power Delivery*, vol. 21, no. 3, pp. 1101–1105, 2006. 128

[134] D. Barbosa, U. C. Netto, D. V. Coury, and M. Oleskovicz, "Power transformer differential protection based on clarke's transform and fuzzy systems," *IEEE Transactions on Power Delivery*, vol. 26, no. 2, pp. 1212–1220, April 2011. 128

[135] E. W. Kimbark, "Two-phase co-ordinates of a three-phase circuit," *Transactions of the American Institute of Electrical Engineers*, vol. 58, no. 12, pp. 894–910, Dec 1939. 128

[136] Y. Xia, K. Wang, W. Pei, and D. P. Mandic, "A balancing voltage transformation for robust frequency estimation in unbalanced power systems," in *Proc. of the Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Dec 2014, pp. 1–6. 128, 139, 140, 141

[137] J. Shlens, "A tutorial on principal component analysis," *CoRR*, vol. abs/1404.1100, 2014. 132

[138] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall International, Inc, 1993. 133, 188, 191

[139] M. H. J. Bollen and I. Y. H. Gu, *Signal processing of power quality disturbances*. John Wiley & Sons, 2006, vol. 30. 135

[140] M. Akke, "Frequency estimation by demodulation of two complex signals," *IEEE Transactions on Power Delivery*, vol. 12, no. 1, pp. 157–163, Jan 1997. 138, 182

[141] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal–ii: Algorithms and applications," *Proceedings of the IEEE*, vol. 80, no. 4, pp. 540–568, Apr 1992. 144, 194

[142] P. Kootsookos, "A review of the frequency estimation and tracking problems," in *Tech. Rep.* Australia: Cooperative Research Centre (CRC) for Robust and Adaptive Systems, Defence Science and Technology Organisation (DSTO), February 1999. 144, 151, 153, 178

[143] R. Prony, "Essai experimental et analytique, etc." *Journal de l'Ecole Polytechnique*, vol. 1, no. 2, pp. 24 – 76, 1795. 145

[144] Y. Xia, Z. Blazic, and D. Mandic, "Complex-valued least squares frequency estimation for unbalanced power systems," *IEEE Transactions on Instrumentation and Measurement*, vol. PP, no. 99, pp. 1–1, 2014. 146

[145] T. Lobos and J. Rezmer, "Real-time determination of power system frequency," *IEEE Transactions on Instrumentation and Measurement*, vol. 46, no. 4, pp. 877 –881, Aug 1997. 146

[146] P. Dash, D. Swain, A. Routray, and A. Liew, "An adaptive neural network approach for the estimation of power system frequency," *Electric Power Systems Research*, vol. 41, no. 3, pp. 203 – 210, 1997. 146

[147] H. So and P. Ching, "Adaptive algorithm for direct frequency estimation," *IEE Proceedings Radar, Sonar and Navigation*, vol. 151, no. 6, pp. 359–364, Dec 2004. 146, 148, 151

[148] D. H. Dini and D. P. Mandic, "Widely linear modeling for frequency estimation in unbalanced three-phase power systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 2, pp. 353–363, 2013. 147, 149, 153

[149] L. J. Griffiths, "Rapid measurement of digital instantaneous frequency," *IEEE Transactions on Acoustics, Speech and Signal Process.*, vol. 23, no. 2, pp. 207–222, Apr 1975. 147

[150] J. Zeidler, E. Satorius, D. Chabries, and H. Wexler, "Adaptive enhancement of multiple sinusoids in uncorrelated noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 3, pp. 240–254, Jun 1978. 147

[151] D. B. Rao and S.-Y. Kung, "Adaptive notch filtering for the retrieval of sinusoids in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 4, pp. 791–802, Aug 1984. 147

[152] A. Nehorai, "A minimal parameter adaptive notch filter with constrained poles and zeros," *IEEE Transactions on Acoustics, Speech and Signal Process.*, vol. 33, no. 4, pp. 983–996, Aug 1985. 147

[153] D. P. Mandic and J. A. Chambers, *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures, and Stability*. Wiley, 2001. 147, 164, 176

[154] A. Pradhan, A. Routray, and A. Basak, "Power system frequency estimation using least mean square technique," *IEEE Transactions on Power Delivery*, vol. 20, no. 3, pp. 1812–1816, 2005. 148

[155] C. Kelly and S. Gupta, "Discrete-time demodulation of continuous-time signals," *IEEE Transactions on Information Theory*, vol. 18, no. 4, pp. 488–493, Jul 1972. 148, 151

[156] K. Nishiyama, "A nonlinear filter for estimating a sinusoidal signal and its parameters in white noise: On the case of a single sinusoid," *IEEE Transactions on Signal Processing*, vol. 45, no. 4, pp. 970–981, Apr 1997. 149, 153

[157] P. K. Dash, A. K. Pradhan, and G. Panda, "Frequency estimation of distorted power system signals using extended complex Kalman filter," *IEEE Transactions on Power Delivery*, vol. 14, no. 3, pp. 761–766, Jul 1999. 149

[158] A. Routray, A. K. Pradhan, and K. P. Rao, "A novel Kalman filter for frequency estimation of distorted signals in power systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 51, no. 3, pp. 469–479, Jun 2002. 149, 153

[159] P. K. Dash, R. K. Jena, G. Panda, and A. Routray, "An extended complex Kalman filter for frequency measurement of distorted signals," *IEEE Transactions on Instrumentation and Measurement*, vol. 49, no. 4, pp. 746–753, Aug 2000. 150, 153

[160] C. H. Huang, C. H. Lee, K. J. Shih, and Y. J. Wang, "Frequency estimation of distorted power system signals using a robust algorithm," *IEEE Transactions on Power Delivery,*, vol. 23, no. 1, pp. 41 –51, Jan. 2008. 150

[161] S. P. Talebi, S. Kanna, and D. P. Mandic, "A non-linear state space frequency estimator for three-phase power systems," in *Proc. of the International Joint Conference on Neural Networks (IJCNN)*, July 2015, pp. 1–7. 150

[162] A. A. Girgis and T. L. D. Hwang, "Optimal estimation of voltage phasors and frequency deviation using linear and non-linear Kalman filtering: Theory and limitations," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-103, no. 10, pp. 2943–2951, Oct 1984. 150, 153

[163] A. A. Girgis and W. L. Peterson, "Adaptive estimation of power system frequency deviation and its rate of change for calculating sudden power system overloads," *IEEE Transactions on Power Delivery*, vol. 5, no. 2, pp. 585–594, Apr 1990. 150

[164] V. V. Terzija, "Improved recursive Newton-type algorithm for power system relaying and measurement," *IEE Proceedings - Generation, Transmission and Distribution*, vol. 145, no. 1, pp. 15–20, Jan 1998. 150

[165] ——, "Improved recursive Newton-type algorithm for frequency and spectra estimation in power systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 52, no. 5, pp. 1654–1659, Oct 2003. 150

[166] D. P. Bertsekas, "Incremental least squares methods and the extended kalman filter," in *Proc. of the 33rd IEEE Conference on Decision and Control*, vol. 2, Dec 1994, pp. 1211–1214 vol.2. 151

[167] A. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Signal Processing Magazing*, vol. 11, no. 3, pp. 18–60, July 1994. 151, 172

[168] "Instrument transformers. General requirements," IEEE Standards, Tech. Rep. BS EN 61869-1:2009, September 2009. 154

[169] M. Bollen, "Characterisation of voltage sags experienced by three-phase adjustable-speed drives," *IEEE Transactions on Power Delivery*, vol. 12, no. 4, pp. 1666–1671, 1997. 154

[170] F. A. Tobar, "Kernel-based adaptive estimation: Multidimensional and state-space approaches," Ph.D. dissertation, Imperial College London, London, U.K., August 2014. 164

[171] F. A. Tobar, S. Y. Kung, and D. P. Mandic, "Multikernel least mean square algorithm," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 2, pp. 265–277, Feb 2014. 164

[172] A. Phadke, "Synchronized phasor measurements in power systems," *IEEE Computer Applications in Power*, vol. 6, no. 2, pp. 10–15, 1993. 165

[173] A. G. Phadke and B. Kasztenny, "Synchronized phasor and frequency measurement under transient conditions," *IEEE Transactions on Power Delivery*, vol. 24, no. 1, pp. 89–95, Jan 2009. 165

[174] X. Zhao and A. H. S. ed, "Asynchronous diffusion adaptation over networks," in *Proc. of the 20th European Signal Processing Conference (EUSIPCO)*, Aug 2012, pp. 86–90. 165

[175] L. Xiao, S. Boyd, and S. Lai, "A space-time diffusion scheme for peer-to-peer least-squares estimation," in *Proc. of the 5th International Conference on Information Processing in Sensor Networks*, April 2006, pp. 168–176. 165

[176] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Transactions on Automatic Control*, vol. 31, no. 9, pp. 803 – 812, sep 1986. 165

[177] S. Aaronson, "The complexity of agreement," in *Proc. of the 37th Annual ACM Symposium on Theory of Computing*. ACM, May 2005, pp. 634–643. 165

[178] R. J. Aumann, "Agreeing to disagree," *The Annals of Statistics*, vol. 4, no. 6, pp. 1236–1239, 1976. 165

[179] S. P. Talebi, S. Kanna, and D. P. Mandic, "Diffusion Kalman filtering with implicit measurement exchanges," in *Submitted to Proc. of the IEEE Intl. Conf. on Acoust. Speech and Signal Process. (ICASSP)*, August 2016, pp. 1–5. 165

[180] N. V. Thakor and Y. S. Zhu, "Applications of adaptive filtering to ecg analysis: noise cancellation and arrhythmia detection," *IEEE Transactions on Biomedical Engineering*, vol. 38, no. 8, pp. 785–794, Aug 1991. 165

[181] R. Horton and U. Samarasekera, "Stillbirths: Ending an epidemic of grief," *The Lancet*, vol. 387, no. 10018, pp. 515–516, Jan 2016. 166

[182] P. Montaldo, S. S. Pauliah, P. J. Lally, L. Olson, and S. Thayyil, "Cooling in a low-resource environment: Lost in translation," *Seminars in Fetal and Neonatal Medicine*, vol. 20, no. 2, pp. 72–79, Apr 2015. 166

[183] D. Simon, *Optimal state estimation: Kalman, H-infinity, and nonlinear approaches*. Wiley, 2006. 167

[184] D. Williams, *Probability with Martingales*. Cambridge University Press, 1991. 167

[185] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Prentice–Hall, 1985. 169

[186] V. Mathews and Z. Xie, "A stochastic gradient adaptive filter with gradient adaptive step size," *IEEE Transactions on Signal Processing*, vol. 41, no. 6, pp. 2075–2087, June 1993. 170

[187] S. Douglas, "Generalized gradient adaptive step sizes for stochastic gradient adaptive filters," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, May 1995, pp. 1396–1399. 170, 173

[188] ——, "A family of normalized LMS algorithms," *IEEE Signal Processing Letters*, vol. 1, no. 3, pp. 49–51, 1994. 171

[189] C. Guimaraes Lopes and J. Bermudez, "Evaluation and design of variable step size adaptive algorithms," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 6, 2001, pp. 3845–3848. 172

[190] D. Mandic, "A generalized normalized gradient descent algorithm," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 115–118, February 2004. 173

[191] J. Fernandez-Bes, V. Elvira, and S. Van Vaerenbergh, "A probabilistic least-mean-squares filter," in *of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015. 174

[192] R. Faragher, "Understanding the basis of the Kalman filter via a simple and intuitive derivation," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 128–132, 2012. 176

[193] J. Humpherys and J. West, "Kalman filtering with Newton's method," *IEEE Control Systems Magazine*, vol. 30, no. 6, pp. 49–51, 2010. 176

[194] D. W. P. Thomas and M. S. Woolfson, "Evaluation of frequency tracking methods," *IEEE Trans. Power Del.*, vol. 16, no. 3, pp. 367–371, Jul 2001. 178

[195] M. M. Begovic, P. M. Djuric, S. Dunlap, and A. G. Phadke, "Frequency tracking in power networks in the presence of harmonics," *IEEE Transactions on Power Delivery*, vol. 8, no. 2, pp. 480–486, Apr 1993. 178, 182

[196] E. Jacobsen and R. Lyons, "The sliding DFT," *IEEE Signal Processing Magazine*, vol. 20, no. 2, pp. 74–80, Mar 2003. 184

[197] D. C. Rife and R. R. Boorstyn, "Single-tone parameter estimation from discrete-time observations," *IEEE Transactions on Information Theory*, vol. 20, no. 5, pp. 591–598, 1974. 192, 193

[198] J. Ekanayake and N. Jenkins, "Comparison of the response of doubly fed and fixed-speed induction generator wind turbines to changes in network frequency," *IEEE Transactions on Energy Conversion*, vol. 19, no. 4, pp. 800–802, Dec 2004. 195

[199] J. Morren, J. Pierik, and S. W. de Haan, "Inertial response of variable speed wind turbines," *Electric Power Systems Research*, vol. 76, no. 11, pp. 980 – 987, 2006. 195, 197

[200] J. Morren, S. W. De Haan, W. L. Kling, and J. Ferreira, "Wind turbines emulating inertia and supporting primary frequency control," *IEEE Transactions on Power Systems*, vol. 21, no. 1, pp. 433–434, 2006. 195

[201] G. Ramtharan, J. B. Ekanayake, and N. Jenkins, "Frequency support from doubly fed induction generator wind turbines," *Renewable Power Generation, IET*, vol. 1, no. 1, pp. 3–9, 2007. 195

[202] Y. Pipelzadeh, B. Chaudhuri, and T. C. Green, "Role of western HVDC link in stability of future great britain (gb) transmission system," in *Proc. of the IEEE Power Energy Society General Meeting*, July 2015, pp. 1–5. 195

[203] J. MacDowell, S. Dutta, M. Richwine, S. Achilles, and N. Miller, "Serving the future: Advanced wind generation technology supports ancillary services," *IEEE Power and Energy Magazine*, vol. 13, no. 6, pp. 22–30, 2015. 197

[204] I. D. Margaris, S. A. Papathanassiou, N. D. Hatziargyriou, A. D. Hansen, and P. Sørensen, "Frequency control in autonomous power systems with high wind power penetration," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 2, pp. 189–199, 2012. 197

[205] Y. Wang, G. Delille, H. Bayem, X. Guillaud, and B. Francois, "High wind power penetration in isolated power systems?assessment of wind inertial and primary frequency responses," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2412–2420, 2013. 197

[206] Y. Pipelzadeh, B. Chaudhuri, and T. C. Green, "Inertial response from remote offshore wind farms connected through VSC-HVDC links: A communication-less scheme," in *Proc. of the IEEE Power and Energy Society General Meeting*, July 2012, pp. 1–6. 197

[207] V. Y. Singarao and V. S. Rao, "Frequency responsive services by wind generation resources in united states," *Renewable and Sustainable Energy Reviews*, vol. 55, pp. 1097 – 1108, 2016. 197

[208] S. Norris and K. Maleka, "Event detection: Smart frequency control," Alstom Grid (Psymetrix), United Kingdom, Tech. Rep. NG-EFCC-SPEC-043, October 2015. 197, 198

[209] I. M. Sanz, B. Chaudhuri, A. Junyent-Ferre, and G. Strbac, "Distributed vs. Concentrated rapid frequency response provision in future Great Britain system," in *Proc. of the IEEE Power Engineering Society General Meeting, 2016*, July 2016, pp. 1–1. 197