A STATE-SPACE APPROACH TO DYNAMIC TOMOGRAPHY

BY

MARK DAVID BUTALA

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2010

Urbana, Illinois

Doctoral Committee:

Associate Professor Farzad Kamalabadi, Chair
Professor Yoram Bresler
Associate Professor Yuguo Chen
Professor Douglas L. Jones
Professor Pierre Moulin

ABSTRACT

The statistical inference of a hidden Markov random process is a problem encountered in numerous signal processing applications including dynamic tomography. In dynamic tomography, the goal is to form images of an object that changes in time from its projection measurements. This work focuses on the case where the object's temporal evolution is significant and governed by a physical model. Solar tomography, the remote sensing problem concerned with the reconstruction of the dynamic solar atmosphere, has served as the motivating application throughout the development of the dissertation.

The proposed state-space formulation provides a natural and general statistical framework for the systematic tomographic reconstruction of dynamic objects when faced with inevitable measurement and modeling uncertainties. In addition, the dissertation offers signal processing methods that scale to meet the computational demands of high-dimensional state estimation problems such as dynamic tomography. Major contributions include a rigorous characterization of the convergence of the ensemble Kalman filter, a new method for ensemble Kalman smoothing and theory regarding its convergence, the first four-dimensional reconstruction of electron density in the solar atmosphere, a new method for dynamic tomography called the Kalman-Wiener filter that has the same computational complexity as filtered backprojection, and a means for characterizing the spatial-temporal resolution of dynamic reconstructions posed under the state-space formulation.

For Mom and Dad, whose support through the years meant everything

ACKNOWLEDGMENTS

The research reported in this dissertation has been supported in part by grants from the National Science Foundation (NSF) and National Aeronautics and Space Administration. In addition, support through fellowships has included an NSF Graduate Research Fellowship, a University of Illinois at Urbana-Champaign (UIUC) Department of Electrical and Computer Engineering Distinguished Fellowship, a UIUC Computational Science and Engineering Fellowship, and a Massachusetts Institute of Technology Lincoln Laboratories Graduate Fellowship. The aforementioned support has sustained this work and it could not have been completed otherwise.

Throughout its development, the dissertation has benefited from numerous helpful suggestions, comments, and critiques from colleagues. Group meetings with Prof. Farzad Kamalabadi, Prof. Yuguo Chen, Dr. Richard Frazin, Russell Hewett, and Jonghyun Yun were especially helpful. Discussions with office mates, technical and otherwise, served as sanity checks, motivators, and stress relievers. I especially thank Dr. Jason Fuemmeler, Dr. Romina Nikoukar, and (soon to be Dr.) Serge Minin for all of these interactions over the years.

I owe an immense debt of gratitude to Farzad. My personal path underwent several unexpected twists during my graduate career. When times were darkest, he shed light. I thank Caitlin for toughing it out throughout the doctoral process and for providing and nurturing the greatest gift imaginable.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Motivation

The statistical inference of a hidden Markov random process is a problem encountered in many applications, including radar tracking of multiple targets [1], dynamic solar tomography [2], geophysical data assimilation [3], and economic forecasting [4]. These problems can be expressed in state-space form and recursive methods have been developed in the Bayesian framework to solve the resultant state estimation problem. For certain applications, such as most forms of medical tomography, the state may change by only a small amount during the measurement interval and can be regarded as static. Dynamic medical tomography methods do exist for cases where the underlying object exhibits significant quasi-periodic (cardiac cycle) [5], [6] or rigid-body (patient movement) [7]–[9] motion. A review of dynamic medical tomography can be found in [10].

This dissertation is concerned with the more general problem of reconstructing a time-varying object from its projections when the state evolution is significant and governed by a complicated physical model, a problem often encountered in remote sensing. In such applications, the unknown object is typically a volumetric image and its voxel representation can have a computationally overwhelming number of degrees

of freedom. While the focus is on remote sensing applications, note that the methods developed in this dissertation are applicable to any high-dimensional state estimation problem provided it exhibits the necessary structure.

The primary driving force behind this research has been the image formation problem of solar tomography [2], [11]–[13]. The goal in this application is to reconstruct the three-dimensional solar atmosphere from two-dimensional remotely sensed projected images of Thomson scattered sunlight. Solar rotation provides the diversity of view angles necessary for tomography, but the rate of rotation is extremely slow relative to the highly dynamic and sometimes chaotic solar atmosphere. Two weeks are required to observe the Sun over a 180° rotation and existing methods, almost all of which assume that the Sun is static, will sometimes fail and produce reconstructions plagued by artifacts [13]. As such, solar tomography provides an extremely challenging, and hence highly motivating, application given its potentially huge dimensionality (e.g., on the order of 10^5 unknowns in [13]), long measurement interval of at least two weeks, huge data volume (the recently launched Solar Dynamics Observatory measures 1.5 TB of images per day), and complicated physical dynamic model [2]. Dynamic biomedical tomography is another ideal application for our methods which deserves special attention because of the potential direct human impact. Joint work with George et al. [14] reconstructs a dynamic cardiac phantom with one of the high-dimensional state estimation methods developed in this dissertation.

A natural framework for solving dynamic tomography problems is to model the unknown object as a Markov random process and then estimate its properties given the measurements. If minimum mean square error (MMSE) estimates of the unknown object are desired, then it is necessary to compute the mean of the posterior distribution of the unknown object conditioned on the measurements, an extremely difficult problem when the posterior distribution is complicated. The particle filter (PF) [15]–

[19] has emerged as a viable state estimation method when faced with analytically unwieldy posterior distributions. Although attractive because of its generality, standard PF methods are computationally intractable for cases where the state dimension is large [20]–[22]. High-dimensional PF methods [23] have been developed when the state can be conditionally divided into a large number of slowly varying or narrowly distributed components plus a much smaller set of more dynamic or broadly distributed components. The standard PF has been applied to electrical capacitance tomography [24], but it was necessary in that work to represent the unknown by only a few parameters on a spline basis for computational tractability.

If the dynamic tomography problem can be posed under the linear state-space model, then the classic Kalman filter (KF) [25]–[27] may be used to recursively compute linear MMSE (LMMSE) state estimates. However, the KF must store and operate on the state estimate error covariance matrix. The number of elements in this matrix grows quadratically with the state dimension and becomes intractably massive for large problems. The KF has solved several dynamic tomography problems [28]–[32], but always when the state dimension is relatively small.

Several high-dimension state estimation methods have been developed based on various approximations to the KF. Four-dimensional variational data assimilation (4D-VAR) [33], [34] assumes that the state transition model is deterministic and that the prior error covariance can be parametrized. As a result, 4D-VAR cannot systematically adapt when the state dynamics deviate from the model as it has no model uncertainty information. Motivated by the finite spatial correlation lengths of typical processes, the band-limited or banded KF [35]–[37] eliminates all bands in the error covariance matrix beyond some prescribed distance from the diagonal, thereby reducing computational expense. However, as mentioned in Section 4.4.1, truncation of the error covariance matrix destroys its positive definite structure, leading to

numerical instabilities. Other similar work [38], [39] truncates the inverse of the error covariance. Another related technique for high-dimensional dynamic tomography involves the minimization of a spatially and temporally regularized quadratic cost function [40]–[43] and can be equivalent to the KF under certain conditions [44], and just as computationally demanding. Another approach to decrease the computational cost of the KF is to apply state reduction [39], [45]–[47]. This strategy is application dependent and is not pursued here, though such efforts could complement this work and further reduce computational costs.

1.2 Methodology

The first proposed approach to dynamic tomography formulates the problem under the linear state-space model and then estimates the state as a function of time with the ensemble Kalman filter (EnKF) [3], [48]–[50]. The EnKF is a Monte Carlo approximation to the KF that is similar to the PF. The EnKF recursively processes an ensemble of samples such that their average approximates the KF estimate. However, when the state dimension is large, the EnKF is tractable only when the ensemble size is kept to a minimum, thereby trading off computational expense for increased uncertainty in the approximate LMMSE estimates. The application of localization or covariance tapering [51], [52] is an effective tool for mitigating issues associated with small ensembles for certain problems by trading off some bias in the estimates for significant reductions in variance. To date, the EnKF has solved a tomographic problem in oceanography [53], but without localization, and in our recent work [14] on dynamic biomedical tomography.

Unfortunately, covariance tapering introduces a state estimate bias that can be difficult to characterize analytically [50]. In addition, Monte Carlo state estimates are also corrupted by sample error [50]. The second proposed approach to dynamic

tomography also exploits problem structure for high-dimensional tractability, but, in contrast to the EnKF, directly in the signal model instead of in the state estimation algorithm [54]. We show that the class of problems with a wide-sense stationary (WSS) random field [55] state, line integral measurements, a linear shift-invariant (LSI) state dynamic model, and WSS additive noise exhibits problem structure that yields a relatively simple optimal spatial-temporal state estimator which we call the Kalman-Wiener filter (KWF). This new state estimator has several advantages: it is free of approximation bias and sample error and has the same computational complexity as the classic filtered back-projection algorithm [56].

1.3 Contribution

The ultimate contribution of this dissertation is a general statistical framework for the inference of high-dimensional hidden Markov random processes. High-dimensional problems are challenging because of their computational demands. However, the ever increasing speed and capability of computers guarantee that the high-dimensional problems of today will one day be solvable using classic methods. On the other hand, continued advancements in sensor technology provide ever increasing data volumes at unprecedented spatial and temporal resolution and spectral diversity. As a rule, the overall dimensionality of a problem must increase to take full advantage of an expanding data volume. The goal of this research is the development of new statistical signal processing methods that scale well with problem dimension to address the computational challenges of the high-dimensional state estimation problems of today and the future.

This dissertation offers several specific novel contributions to the fields of signal processing and statistics. First, the results in Section 5.1 and Section 5.2 (and appearing in [50]) are the first covariance tapered EnKF dynamic tomographic recon-

structions, and the joint work with George et al. [14] is the first use of the EnKF for dynamic medical tomography. Second, no rigorous characterization of EnKF convergence existed prior to the proof in Appendix A (and [50]). The proof is important for the following reasons: it establishes a theoretical basis for the validity of the EnKF by showing that it converges to a well defined limit, provides proof that the unlocalized EnKF is a Monte Carlo KF algorithm, and explicitly identifies the bias introduced by covariance tapering. Third, the ensemble Kalman smoother (EnKS) developed in Section 4.6 is a new method for high-dimensional state smoothing. As demonstrated by the results, state smoothing can yield estimates with greater fidelity without much additional computational expense. The development of the EnKS includes a theoretical characterization of its convergence which is of equal importance as the convergence results derived for the EnKF. Next, the KWF is, to the best of our knowledge, a new method for the inference of WSS spatial random fields. The KWF algorithm reduces the computational complexity of certain dynamic tomography problems to that of filtered back-projection. Lastly, we extend the concept of the local impulse response [57] to problems expressed under the linear dynamic state-space signal model. The local impulse response provides a means for characterizing the achievable spatial-temporal resolution in dynamic tomography and is useful for understanding the type of features that can be reliably reconstructed given the measurements and signal model parameters.

The dissertation also offers the following novel contributions from a scientific perspective. First, the temperature forward model in Section 3.4 (and appearing in [2]) is compactly expressed using a Kronecker matrix product, a significant simplification over previous models. Second, the 4-D reconstruction of solar coronal electron density presented in Section 7.5 (and appearing in [58]) is the first of its kind. Such reconstructions will immediately impact any future joint research with our collaborators

Cohen et al. [59] on the physical modeling of processes in the solar corona. The time evolving solar plume reconstruction of Barbey et al. [60] appears to be the closest related work in the literature, but their signal model is tailored specifically for solar plumes and their method would be unsuitable for global electron density or temperature reconstruction. Finally, the state-space framework for dynamic reconstruction of the solar atmosphere is, in and of itself, novel to the heliophysics community. Our approach was first suggested in [32] and then completed in [2] with the development of a viable high-dimensional state estimation method.

1.4 Summary

The remainder of the dissertation is organized in the following manner. First, signal models for framing dynamic signal processing problems are given in Chapter 2. Next, forward models for tomography are developed in Chapter 3. Then, Chapter 4 develops computationally tractable state estimation methods for high-dimensional state estimation applications such as dynamic tomography. The methods are evaluated in Chapter 5, which presents the results of several numerical experiments designed to test the performance of the state estimation methods in dynamic tomography scenarios. Chapter 6 extends the concept of the local impulse response to spatial-temporal problems expressed under the linear state-space dynamic model. Next, Chapter 7 focuses solely on solar tomography and ends with the first 4-D reconstruction of the solar atmosphere. Conclusions and future research directions are given in Chapter 8.

CHAPTER 2

SIGNAL MODELS

This chapter forms the mathematical foundation for the remainder of the dissertation. Two broad signal model classes are considered: static and dynamic. First, we define the linear static model which assumes the unknown does not change over the measurement interval. Three dynamic models are then defined, each ranging in terms of generality, from nonlinear, to linear, and finally to linear wide-sense stationary.

2.1 Linear Static Signal Model

Each measurement in the linear static model is given by

$$\boldsymbol{y}_i = \boldsymbol{H}_i \, \boldsymbol{x} + \boldsymbol{v}_i \tag{2.1}$$

where the M-dimensional vector $\boldsymbol{y}_i$ contains the measurements at the integer time index i, the matrix $\boldsymbol{H}_i$ is the linear measurement operator, the N-dimensional vector $\boldsymbol{x}$ is the unknown, and $\boldsymbol{v}_i$ accounts for measurement noise. The available measurements may then be combined to form the linear system of equations

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_I \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_1 \\ \vdots \\ \boldsymbol{H}_I \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} \boldsymbol{v}_1 \\ \vdots \\ \boldsymbol{v}_I \end{bmatrix} \tag{2.2}$$

or, equivalently,

$$\boldsymbol{y}_{1:I} = \boldsymbol{H}_{1:I}\,\boldsymbol{x} + \boldsymbol{v}_{1:I} \tag{2.3}$$

where I is the total number of measurements. The covariance of the the unknown is denoted by

$$\boldsymbol{R}_{\boldsymbol{x}} \triangleq \mathrm{Cov}(\boldsymbol{x}) \triangleq \mathbb{E}\big(\boldsymbol{x}\,\boldsymbol{x}^T\big) - \mathbb{E}(\boldsymbol{x})\;\mathbb{E}(\boldsymbol{x})^T \tag{2.4}$$

where $\mathbb{E}(\cdot)$ is the expectation operator and the covariance of the zero mean measurement noise is denoted by

$$\boldsymbol{R}_{\boldsymbol{v}_{1:I}} \triangleq \mathrm{Cov}(\boldsymbol{v}_{1:I}) = \mathbb{E}\big(\boldsymbol{v}_{1:I}\,\boldsymbol{v}_{1:I}^T\big)\,. \tag{2.5}$$

2.2 Dynamic Signal Models

2.2.1 Nonlinear model

Under the discrete-time hidden Markov model (HMM), the unknown is the Markov random process $\{\boldsymbol{x}_i,\, i \in \mathbb{N}\}$. The observable measurements are denoted by the random process $\{\boldsymbol{y}_i,\, i \in \mathbb{N}\}$. The state and measurement processes are fully described by the probability density functions (PDFs)

$$p_{\boldsymbol{X}_1}(\boldsymbol{x}_1) \tag{2.6}$$

$$f_i(\boldsymbol{x}_{i+1}|\boldsymbol{x}_i) \tag{2.7}$$

$$h_i(\boldsymbol{y}_i|\boldsymbol{x}_i) \tag{2.8}$$

where $p_{\boldsymbol{X}_1}(\boldsymbol{x}_1)$ is the prior PDF on the initial state $\boldsymbol{x}_1$, $f_i(\boldsymbol{x}_{i+1}|\boldsymbol{x}_i)$ characterizes the state transition, and $h_i(\boldsymbol{y}_i|\boldsymbol{x}_i)$ is the forward model.

2.2.2 Linear model

A less general form of the discrete-time HMM is given by the state-space equations

$$\boldsymbol{x}_{i+1} = \boldsymbol{F}_i \, \boldsymbol{x}_i \; + \boldsymbol{u}_i \tag{2.9}$$

$$\boldsymbol{y}_i = \boldsymbol{H}_i \, \boldsymbol{x}_i + \boldsymbol{v}_i \tag{2.10}$$

where $\boldsymbol{F}_i$ models the linear component of the state transition and the vector $\boldsymbol{u}_i$ accounts for uncertainty in the state transition model.

The state and measurement noise random processes are mutually uncorrelated, zero mean, and have known covariances

$$\mathbb{E}\big(\boldsymbol{u}_i \, \boldsymbol{u}_j^T\big) = \boldsymbol{Q}_i \, \delta_{ij} \tag{2.11}$$

$$\mathbb{E}\big(\boldsymbol{v}_i \, \boldsymbol{v}_j^T\big) = \boldsymbol{R}_i \, \delta_{ij} \tag{2.12}$$

where δ_{ij} is the Kronecker delta function. In addition, the unknown object at the initial time index has known prior mean

$$\boldsymbol{\mu}_1 \triangleq \mathbb{E}(\boldsymbol{x}_1) \tag{2.13}$$

and covariance

$$\boldsymbol{\Pi}_1 \triangleq \mathrm{Cov}(\boldsymbol{x}_1) \tag{2.14}$$

and is uncorrelated with the state and measurement noise.

2.2.3 Linear wide-sense stationary model

The last HMM is given by the state-space equations

$$x_{i+1}(\boldsymbol{r}) = \{F_i\, x_i\}(\boldsymbol{r}) + u_i(\boldsymbol{r}) \tag{2.15}$$

$$y_i(\boldsymbol{s}) = \{H_i\, x_i\}(\boldsymbol{s}) + v_i(\boldsymbol{s}) \tag{2.16}$$

where the notation $\{F_i\, x_i\}(\boldsymbol{r})$ denotes the result of applying the linear operator F_i to x_i at coordinate $\boldsymbol{r} \in \mathbb{R}^d$. Sometimes we will write $\{F_i\, x_i\}$ when the coordinate is not important. The state-transition operator F_i is restricted to be a linear shift-invariant (LSI) filter, meaning

$$\{F_i\, x_i\}(\boldsymbol{r}) = \int_{\mathbb{R}^d} f_i(\boldsymbol{r} - \boldsymbol{\rho})\, x_i(\boldsymbol{\rho})\, d\boldsymbol{\rho} \triangleq \{f_i \star x_i\}(\boldsymbol{r}) \tag{2.17}$$

where f_i is the impulse response and $\star$ denotes convolution. This class of state-transition operators includes:

- The identity operator which results in a purely random walk state transition model.
- Rigid-body motion, such as translation and rotation.
- State dynamics described by linear spatially invariant partial differential equations, including diffusion and advection. Such dynamic models are suitable in a variety of physical scenarios.

Both (2.15) and (2.16) include an additive noise term that accounts for model uncertainties. We delay discussing the statistical models for the noise until the end of the subsection so that required notation can first be defined.

Each lowercase symbol in (2.15) and (2.16) is a wide-sense stationary (WSS) spatial random field [55], [61]. For example, the state at time index i is the spatial

random field denoted $(x_i(\boldsymbol{r}) : \boldsymbol{r} \in \mathbb{R}^d)$ where $\boldsymbol{r}$ is the spatial coordinate. The first two moments of a WSS spatial random field X satisfy: (1) the mean is constant, meaning $\mu_X(\boldsymbol{r}) \triangleq \mathbb{E}[x(\boldsymbol{r})] = \mu_X$, and (2) the auto-correlation

$$R_X(\boldsymbol{r}_1, \boldsymbol{r}_2) \triangleq \mathbb{E}[x(\boldsymbol{r}_1)\, x(\boldsymbol{r}_2)] \tag{2.18}$$

is a function only of spatial separation, meaning $R_X(\boldsymbol{r}_1, \boldsymbol{r}_2) = R_X(\boldsymbol{r}_1 - \boldsymbol{r}_2)$. Two random fields X and Y are jointly WSS if the cross-correlation

$$R_{X,Y}(\boldsymbol{r}, \boldsymbol{s}) \triangleq \mathbb{E}[x(\boldsymbol{r})\, y(\boldsymbol{s})] \tag{2.19}$$

is a function only of spatial separation. The d-dimensional Fourier transform of the auto-correlation is the power spectrum denoted $S_X(\boldsymbol{f})$ where $\boldsymbol{f} \in \mathbb{R}^d$ is the spatial frequency coordinate.

All spatial random fields in this work are members of the Hilbert space $\mathsf{X} = \mathcal{L}^2(\Omega, F, P)$, the zero-mean and finite variance spatial random fields [62]. An inner product on X is

$$\langle x(\boldsymbol{r}), y(\boldsymbol{s}) \rangle = R_{X,Y}(\boldsymbol{r}, \boldsymbol{s}). \tag{2.20}$$

The induced norm is

$$\|x(\boldsymbol{r})\| = \sqrt{\mathbb{E}[x(\boldsymbol{r})^2]} = \sqrt{R_X(\boldsymbol{r}, \boldsymbol{r})}. \tag{2.21}$$

Also included in the signal model (2.15) and (2.16) are the power spectrum of the initial state $S_{X_1}(\boldsymbol{f})$ and the power spectrum of the state and measurement noise $S_{U_i}(\boldsymbol{f})$ and $S_{V_i}(\boldsymbol{f})$. In addition, the state and measurement noise are uncorrelated

with the initial state, i.e.,

$$\langle x_1(\boldsymbol{r}),\, u_i(\boldsymbol{\rho})\rangle = 0 \qquad \text{and} \qquad \langle x_1(\boldsymbol{r}),\, v_i(\boldsymbol{s})\rangle = 0 \qquad \forall\, i \tag{2.22}$$

and the state and measurement noise are uncorrelated, i.e.,

$$\langle u_i(\boldsymbol{r}),\, v_j(\boldsymbol{s})\rangle = 0 \qquad \forall\, i,\, j. \tag{2.23}$$

The following lemma details the basic orthogonality properties of the linear WSS signal model.

Lemma 2.1: Under the linear WSS signal model defined by (2.15) and (2.16), the statistical assumptions (2.22) and (2.23) imply

$$\text{(i)} \quad \text{For } i \geq j, \qquad u_i(\boldsymbol{r}) \perp x_j(\boldsymbol{\rho}) \qquad \text{and} \qquad v_i(\boldsymbol{s}) \perp x_j(\boldsymbol{r}) \tag{2.24}$$

$$\text{(ii)} \quad \text{For } i > j, \qquad u_i(\boldsymbol{r}) \perp y_j(\boldsymbol{s}) \qquad \text{and} \qquad v_i(\boldsymbol{s}) \perp y_j(\boldsymbol{\sigma}) \tag{2.25}$$

$$\text{(iii)} \quad \text{For } i = j, \qquad u_i(\boldsymbol{r}) \perp y_i(\boldsymbol{s}) \qquad \text{and} \qquad \langle v_i(\boldsymbol{r}),\, y_i(\boldsymbol{s})\rangle = R_{V_i}(\boldsymbol{r} - \boldsymbol{s}) \tag{2.26}$$

for all $\boldsymbol{r}$, $\boldsymbol{\rho} \in \mathbb{R}^d$ and for all $\boldsymbol{s}$, $\boldsymbol{\sigma} \in \Theta^{\perp}$.

Proof. The proof is a straightforward extension to the proof of Lemma 5.3.1 in [27] concerning the basic orthogonality properties of the linear state-space model. □

CHAPTER 3

FORWARD MODELS FOR TOMOGRAPHY

The goal of tomography is to reconstruct a volumetric structure given its projection measurements. Tomography is most widely known for its use in radiology to noninvasively diagnose disease. As shown later in the chapter, the imaging methodology also pertains to the volumetric reconstruction of the physical parameters of the Sun's inner atmosphere given a series of remotely sensed images.

The chapter begins with the definition of the X-ray transform, a mathematical model for parallel beam tomography where each measurement consists of a set of parallel line integrals of an unknown object. The discretized X-ray transform will later be used as the forward model in several of the numerical experiments in Chapter 5. Next, we define the related exterior and hollow X-ray transforms. These transforms are relevant to the forward models for electron density and temperature of the Sun's inner solar atmosphere defined at the end of the chapter.

3.1 X-ray Transform

The X-ray transform [56] is defined by

$$\{H_{\boldsymbol{\theta}_i}\, x_i\}(\boldsymbol{s}) \triangleq \int_{\mathbb{R}} x_i(\boldsymbol{s} + l\boldsymbol{\theta}_i)\, dl. \tag{3.1}$$

Figure 3.1 depicts the geometric parameters of (3.1) in the three-dimensional (3-D) case where $d = 3$, the unit vector $\boldsymbol{\theta}_i \in \mathsf{S}^{d-1}$ is parallel to the line of integration, the set $\mathsf{S}^{d-1} = \{\boldsymbol{r} \in \mathbb{R}^d : \|\boldsymbol{r}\|_2 = 1\}$ is the unit sphere in $\mathbb{R}^d$, the vector $\boldsymbol{s} \in \Theta_i^{\perp}$ is orthogonal to $\boldsymbol{\theta}_i$, the vector space $\Theta_i^{\perp} = \{\boldsymbol{z} : \boldsymbol{z} \perp \boldsymbol{\theta}_i\}$ defines the projection plane, and the vector $\boldsymbol{s} + l\boldsymbol{\theta}_i \in \mathbb{R}^d$ parametrizes the line of integration as a function of the scalar $l \in \mathbb{R}$. Note that the X-ray transform (3.1) involves the stochastic integration of a spatial random field which exists in the mean square Riemann sense if and only if [55]

$$\int_{\mathbb{R}} \int_{\mathbb{R}} R_{X_i}(\boldsymbol{s} + l\boldsymbol{\theta}_i, \boldsymbol{s} + l'\boldsymbol{\theta}_i)\, dl\, dl' < \infty \tag{3.2}$$

where $R_{X_i}(\boldsymbol{r}, \boldsymbol{r}') \triangleq \mathbb{E}[x_i(\boldsymbol{r})\, x_i(\boldsymbol{r}')]$ is the autocorrelation function of the spatial random field x_i. If the spatial random field x_i is WSS, then (3.1) exists in the mean square Riemann sense if and only if $R_{X_i}(\boldsymbol{r})$ is continuous at $\boldsymbol{r} = \boldsymbol{0}$ [63].

The continuous formulation of the X-ray transform (3.1) will be used in Section 4.5.4. The discrete version can be used in the spatially discrete signal models defined in Chapter 2. The discretization begins with the choice of a finite basis representation for the spatial random field x_i. Then, because the X-ray transform is linear and the measurements are discrete in practice, the discrete form of the X-ray transform is the $M \times N$ matrix $\boldsymbol{H}_{\boldsymbol{\theta}_i}$ where M is the number of measurements at time index i and N is the number of basis functions.

3.2 Exterior and Hollow X-ray Transforms

As implied by its name, the X-ray transform is useful for modeling the measurement process in certain medical imaging scenarios [56]. Below, we define the related exterior and hollow X-ray transforms that are used later to develop forward models for the electron density and temperature in the solar tomography application.

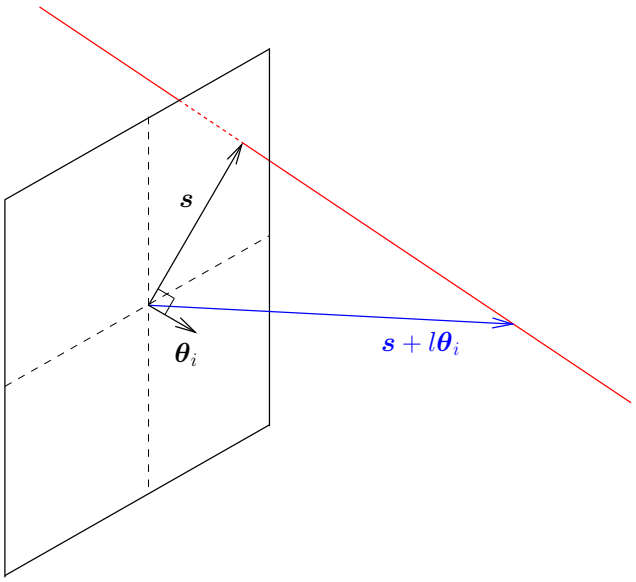

Figure 3.1 Diagram of the three-dimensional X-ray transform geometric parameters. The unit vector $\boldsymbol{\theta}_i$ is normal to the projection plane. The vector $\boldsymbol{s}$ is orthogonal to $\boldsymbol{\theta}_i$ and is the position within the projection plane. The blue vector $\boldsymbol{s}+l\boldsymbol{\theta}_i$ is the position along the line-of-sight in red.

The exterior X-ray transform [64], [65] is defined by

$$\{H^a_{\boldsymbol{\theta}_i}\, x_i\}(\boldsymbol{s}) = \begin{cases} \int_{\mathbb{R}} x_i(\boldsymbol{s} + l\boldsymbol{\theta}_i)\, dl, & \|s\|_2 \geq a \\ \text{undefined}, & \text{otherwise} \end{cases} \tag{3.3}$$

and is identical to the X-ray transform (3.1) except that lines-of-sight must be exterior to the d-dimensional sphere with radius a. The discrete representation of (3.3) is denoted by the $M \times N$ matrix $\boldsymbol{H}^a_{\boldsymbol{\theta}_i}$.

The hollow X-ray transform is defined by

$$\{H^{0-a}_{\boldsymbol{\theta}_i}\}(\boldsymbol{s}) = \int_{\mathcal{L}(\boldsymbol{s},\, a)} x_i(\boldsymbol{s} + l\boldsymbol{\theta}_i)\, dl \tag{3.4}$$

where

$$\mathcal{L}(\boldsymbol{s},\, a) = \begin{cases} \left[\sqrt{a^2 - \|\boldsymbol{s}\|_2^2},\, \infty\right), & \|\boldsymbol{s}\|_2 < a \\ \mathbb{R}, & \text{otherwise} \end{cases} \tag{3.5}$$

and we denote its discrete form by the $M \times N$ matrix $\boldsymbol{H}_{\boldsymbol{\theta}_i}^{0\text{–}a}$. The hollow X-ray transform is also similar to (3.1) except that lines-of-sight terminate at the surface of the d-dimensional sphere with radius a when $\|\boldsymbol{s}\|_2 < a$.

3.3 Electron Density

The physical relationship between the intensity of remotely observed polarized light and the density of free electrons in the corona integrated along the measurement line-of-sight (LOS) is given by [11], [66]

$$y_i(\boldsymbol{\theta}_i,\, \boldsymbol{s}) = C \int_{\mathcal{L}} H(\boldsymbol{s},\, \boldsymbol{s} + l\,\boldsymbol{\theta}_i)\, x_i(\boldsymbol{s} + l\,\boldsymbol{\theta}_i)\, dl + v_i(\boldsymbol{\theta}_i,\, \boldsymbol{s}). \tag{3.6}$$

The function $y_i(\boldsymbol{\theta}_i,\, \boldsymbol{s})$ is the polarized brightness (pB) measured at time index i from a point-of-view defined by $\boldsymbol{\theta}_i$ and $\boldsymbol{s}$. The function x_i is the electron density at a specified point in the stellar atmosphere at a distance l along the line-of-sight and the set $\mathcal{L}$ is the range of l corresponding to the ray from the observer to infinity. The scalar C is a combination of several physical constants necessary to balance the physical units of the equation. The Thomson scattering physics are encapsulated in the known function $H(\boldsymbol{s},\, \boldsymbol{s} + l\,\boldsymbol{\theta})$ [11], [66]. Lastly, the function $v_i(\boldsymbol{\theta}_i,\, \boldsymbol{s})$ is the instrument-dependent measurement noise.

The statistics of the measurement noise $v_i(\boldsymbol{\theta}_i,\, \boldsymbol{s})$ are Poisson because scattering governs the emission of pB light from the corona and the light sensors are most often charged-coupled devices (CCDs). In practice, the true statistics of the measurement noise may only be crudely specified given that coronagraphs are very complicated

optical instruments and various systematic artifacts may persist even after calibration. Often, only the variance of the noise for each pixel is known.

Measurements of pB are typically in the form of images. Each pixel of such an image is the intensity of polarized light focused on a sensor, usually a cell of a CCD camera. It will later be convenient to represent such an image by the vector $\boldsymbol{y}_i$, with each of its M elements equal to a single pixel intensity of the pB image. The vector $\boldsymbol{y}_i$ is related to the continuous forward model (3.6) through

$$(\boldsymbol{y}_i)_m = \int_{\mathcal{S}_m} y_i(\boldsymbol{\theta}_i,\, \boldsymbol{s})\, d\boldsymbol{s} \tag{3.7}$$

where $(\cdot)_m$ is the mth element of the vector argument and $\mathcal{S}_m$ is the set of lines-of-sight that are focused onto the mth image pixel.

Closed-form analytical solutions to Fredholm integral equations like (3.6) are possible only when the kernel H is of a simple form. However, the Thomson scattering kernel in (3.6) is too complicated and the forward model must be discretized to approximate the solution through numerical techniques. A general expression for the necessary finite basis expansion is

$$x_i(\boldsymbol{r}) \approx \sum_{n=1}^{N} (\boldsymbol{x}_i)_n\, \phi_n(\boldsymbol{r}) \tag{3.8}$$

where $\boldsymbol{r}$ is the 3-D position vector, N is the total number of basis functions, $\phi_n(\cdot)$ is the nth basis function, and the vector $\boldsymbol{x}_i$ has N components. The corona is an amorphous plasma with unknown correlation structure and it is therefore difficult to form a low-dimensional basis that faithfully captures the corona. However, research exists in characterizing the corona with consistent, but time independent, low-dimensional models [67], [68]. Faced with such difficulties, a voxel basis, an orthogonal set of

3-D volume elements that span the reconstructed image domain, represents a general choice that makes few prior assumptions. The choice of the basis functions has a significant impact on the reconstructions as all features in the solar atmosphere that cannot be expressed as a linear combination of basis functions are lost. The unfortunate byproduct of a voxel basis is that N will typically be immense. For example, dividing the reconstructed image domain into $100 \times 100 \times 100$ voxels results in $N = 10^6$.

The finite basis expansion results in the linear system

$$\boldsymbol{y}_i = \boldsymbol{H}_i \, \boldsymbol{x}_i + \boldsymbol{v}_i \tag{3.9}$$

where the measurement operator $\boldsymbol{H}_i$ is an $M \times N$ matrix. Each element of the matrix $\boldsymbol{H}_i$ is given by

$$[\boldsymbol{H}_i]_{(m,\,n)} \triangleq C \int_{\mathcal{L}} \int_{\mathcal{S}_m} H(\boldsymbol{s},\, \boldsymbol{s} + l\,\boldsymbol{\theta}_i)\, \phi_n(\boldsymbol{s} + l\,\boldsymbol{\theta}_i)\, d\boldsymbol{s}\, dl \tag{3.10}$$

where $[\cdot]_{(m,\,n)}$ is the $(m,\, n)$th component of the matrix argument. Note that the measurement operator can be broken into two more easily understood operators

$$\boldsymbol{H}_i = \boldsymbol{H}^a_{\boldsymbol{\theta}_i}\, \boldsymbol{\Lambda}. \tag{3.11}$$

The matrix $\boldsymbol{H}^a_{\boldsymbol{\theta}_i}$ is the discrete representation of the exterior X-ray transform (3.3) and the $N \times N$ diagonal matrix $\boldsymbol{\Lambda}$ accounts for the spatial dependence of the scattering function in (3.6). As previously discussed, the full statistics of the measurement noise vector $\boldsymbol{v}_i$ may not be known.

3.4 Temperature

Differential emission measure (DEM) analysis is useful for determining the temperature structure of the projected solar atmosphere. DEM tomography (DEMT) [69] combines DEM analysis and tomography to reconstruct the 3-D electron temperature distribution. The forward model is

$$y_i(\boldsymbol{\theta}_i,\, \boldsymbol{s},\, \lambda) = C \int_{\mathcal{L}} \int_0^\infty \psi(\lambda,\, T)\, x_i(\boldsymbol{s} + l\boldsymbol{\theta}_i,\, T)\, dT\, dl + v_i(\boldsymbol{\theta}_i,\, \boldsymbol{s},\, \lambda). \tag{3.12}$$

Identically to (3.6), the vectors $\boldsymbol{\theta}_i$ and $\boldsymbol{s}$ define the point-of-view of the observer relative to the Sun, the vector $\boldsymbol{s} + l\boldsymbol{\theta}_i$ specifies a 3-D spatial location along an individual measurement line-of-sight, and the scalar C ensures that physical units balance. The function $y_i(\boldsymbol{\theta}_i,\, \boldsymbol{s},\, \lambda)$ is the extreme ultraviolet (EUV) or X-ray intensity measured at wavelength λ at time index i and $v_i(\boldsymbol{\theta}_i,\, \boldsymbol{s}, \lambda)$ is the measurement noise. The function $x_i(\boldsymbol{s} + l\boldsymbol{\theta}_i,\, T)$ is the local DEM at temperature T. The function $\psi(\lambda,\, T)$ depends on the atomic physics of the emission process in the stellar plasma and may be calculated with tools such as CHIANTI [70]. As discussed in Section 3.3, the full statistics of the measurement noise vector $\boldsymbol{v}_i$ may not be known.

Like pB, typical measurements of the EUV or X-ray corona are in the form of images. The M pixels of an EUV or X-ray image at a particular wavelength can be represented by the vector

$$[\boldsymbol{Y}_i]_{(m,\,k)} = \int_{\mathcal{S}_m} \int_0^\infty \Phi_k(\lambda)\, y_i(\boldsymbol{\theta}_i,\, \boldsymbol{s},\, \lambda)\, d\lambda\, d\boldsymbol{s} \tag{3.13}$$

where the mth element of the kth column of the $M \times K$ matrix $\boldsymbol{Y}_i$ is the mth pixel in the kth frequency band measured at time index i. Again, $\mathcal{S}_m$ is the set of lines-of-sight that are focused onto the mth pixel. The function $\Phi_k(\lambda)$ is the frequency response of

the instrument in the kth frequency band.

The temperature forward model (3.12) may be discretized by choosing both a spatial and temperature basis

$$x_i(\boldsymbol{r},\, T) = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} [\boldsymbol{X}_i]_{(n_1,\, n_2)}\ \phi_{n_1}(\boldsymbol{r})\, \varphi_{n_2}(T) \tag{3.14}$$

where $\boldsymbol{r}$ is the 3-D spatial coordinate, $\boldsymbol{X}_i$ is an $N_1 \times N_2$ matrix with $(n_1,\, n_2)$th component equal to the local DEM at the n_1th spatial coordinate and n_2th temperature, $\phi_{n_1}(\cdot)$ is the n_1th spatial basis function, and $\varphi_{n_2}(\cdot)$ is the n_2th temperature basis function. As argued in Section 3.3, it is difficult to determine a faithful low-dimensional basis for the 3-D structure of the coronal electron density and the same is true for the local DEM. DEMT research has so far focused on voxel spatial and uniform interval temperature bases [69].

The measurements $\boldsymbol{Y}_i$ are related to the elements of the matrix unknown $\boldsymbol{X}_i$ by

$$\begin{aligned}[\boldsymbol{Y}_i]_{(m,\, k)} = &\sum_{n_1=1}^{N_1} \int_{\mathcal{S}_m} \int_{\mathcal{L}} \phi_{n_1}(\boldsymbol{s} + l\, \boldsymbol{\theta}_i)\, dl\, d\boldsymbol{s} \\ &\times \sum_{n_2=1}^{N_2} \int_0^\infty \int_0^\infty \Phi_k(\lambda)\, \psi(\lambda,\, T)\, \varphi_{n_2}(T)\ [\boldsymbol{X}_i]_{(n_1,\, n_2)}\ d\lambda\, dT + [\boldsymbol{V}_i]_{(m,\, k)}\end{aligned} \tag{3.15}$$

where the $(m,\, k)$th element of the $M \times K$ matrix $\boldsymbol{V}_i$ is the measurement noise of the mth pixel in the kth frequency band. Note that the terms in the sum over n_2 in (3.15) only operate on the columns of the matrix $\boldsymbol{X}_i$ for a fixed row. Similarly, the terms in the sum over n_1 not appearing in the sum over n_2 operate only on the rows of the matrix $\boldsymbol{X}_i$ for a fixed column. Thus, it is possible to separate these two operations. To do so, first define each element of the $K \times N_2$ matrix $\boldsymbol{\Psi}$ by

$$[\boldsymbol{\Psi}]_{(k,\, n_2)} = \int_0^\infty \int_0^\infty \Phi_k(\lambda)\, \psi(\lambda,\, T)\, \varphi_{n_2}(T)\, d\lambda\, dT \tag{3.16}$$

and note that the matrix $\boldsymbol{\Psi}$ corresponds to the column-wise operations on $\boldsymbol{X}_i$ in (3.15). Also note the function $\psi(\lambda, T)$ may only be evaluated at discrete points by CHIANTI and the integrals in (3.16) must be evaluated numerically. Secondly, define the $M \times N_1$ matrix $\boldsymbol{H}_{\boldsymbol{\theta}_i}^{0\text{–}a}$ by

$$\left[\boldsymbol{H}_{\boldsymbol{\theta}_i}^{0\text{–}a}\right]_{(m,\,n_1)} = \int_{\mathcal{S}_m} \int_{\mathcal{L}} \phi_{n_1}(\boldsymbol{s} + l\,\boldsymbol{\theta}_i)\,dl\,d\boldsymbol{s}. \tag{3.17}$$

The matrix $\boldsymbol{H}_{\boldsymbol{\theta}_i}^{0\text{–}a}$ corresponds to row-wise operations on $\boldsymbol{X}_i$ in (3.15) and is the discretized representation of the hollow X-ray transform (3.4). Finally, (3.15) can be compactly expressed as

$$\boldsymbol{Y}_i = \boldsymbol{H}_{\boldsymbol{\theta}_i}^{0\text{–}a}\,\boldsymbol{X}_i\,\boldsymbol{\Psi}^T + \boldsymbol{V}_i. \tag{3.18}$$

A matrix identity involving the Kronecker matrix product [71] may be applied to (3.18) to give

$$\mathrm{vec}(\boldsymbol{Y}_i) = \left(\boldsymbol{\Psi} \otimes \boldsymbol{H}_{\boldsymbol{\theta}_i}^{0\text{–}a}\right) \mathrm{vec}(\boldsymbol{X}_i) + \mathrm{vec}(\boldsymbol{V}_i) \tag{3.19}$$

where the operator $\mathrm{vec}(\cdot)$ stacks the columns of the matrix argument into a vector and $\otimes$ is the Kronecker matrix product. Note that (3.19) is now in the form of a linear system of equations, cf. (3.9).

CHAPTER 4

HIGH-DIMENSIONAL STATE ESTIMATION

The signal models in Section 2.2 provide a general statistical framework for dynamic tomography. This chapter develops viable state estimation algorithms for solving high-dimensional statistical inference problems under these models. While dynamic tomography is the focus, we emphasize that the methods developed in this chapter are applicable to any problem that can be expressed under a signal model in Section 2.2 and exhibits the necessary problem structure.

This chapter first addresses the filtering problem. Filtered estimates are causal in the sense that they depend only on current and past observations in time and are therefore ideal for online processing. We consider several methods for computing filtered estimates: linear minimum mean square error (LMMSE) estimation under the static signal model and the particle filter (PF), Kalman filter (KF), ensemble Kalman filter (EnKF), and Kalman-Wiener filter (KWF), all under a dynamic model. Each of these methods provides a different trade-off between signal model generality and computational tractability. We note that the dynamic methods can also compute predictive estimates to forecast the future evolution of the unknown in time. The chapter concludes with methods for smoothed estimation where each estimate depends on all available observations regardless of time. Smoothed estimates are ideal for retrospective analysis of a complete data set.

4.1 Linear Minimum Mean Square Error Estimator

Under the static signal model defined in Section 2.1, the LMMSE estimator of the unknown is given by [27]

$$\widehat{\boldsymbol{x}} = \mathbb{E}(\boldsymbol{x}) + \left(\boldsymbol{R}_{\boldsymbol{x}}^{-1} + \boldsymbol{H}_{1:I}^{T}\, \boldsymbol{R}_{\boldsymbol{v}_{1:I}}^{-1}\, \boldsymbol{H}_{1:I}\right)^{-1} \boldsymbol{H}_{1:I}^{T}\, \boldsymbol{R}_{\boldsymbol{v}_{1:I}}^{-1} \left[\boldsymbol{y}_{1:I} - \boldsymbol{H}_{1:I}\, \mathbb{E}(\boldsymbol{x})\right]. \tag{4.1}$$

If $\boldsymbol{R}_{\boldsymbol{x}}^{-1} = \lambda\, \boldsymbol{D}^{T} \boldsymbol{D}$, then the above estimator can be expressed through the following optimization problem:

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left\| \boldsymbol{y}_{1:I} - \boldsymbol{H}_{1:I}\, \boldsymbol{x} \right\|_{\boldsymbol{R}_{\boldsymbol{v}_{1:I}}^{-1}}^{2} + \lambda \left\| \boldsymbol{D}\left(\boldsymbol{x} - \mathbb{E}(\boldsymbol{x})\right) \right\|_{2}^{2} \tag{4.2}$$

where the weighted vector norms are defined by $\|\boldsymbol{z}\|_{\boldsymbol{A}}^{2} \triangleq \boldsymbol{z}^{T} \boldsymbol{A}\, \boldsymbol{z}$ with $\boldsymbol{A}$ positive definite. Note that (4.2) establishes the connection to a deterministic weighted least squares formulation [27]. In this work, the matrix $\boldsymbol{D}$ is a discrete approximation to a spatial derivative and penalizes roughness in the solution $\widehat{\boldsymbol{x}}$ to the degree specified by the regularization parameter λ. Iterative methods such as conjugate gradient [72] can efficiently solve (4.2) when the matrices $\boldsymbol{H}_{1:I}$, $\boldsymbol{R}_{\boldsymbol{v}_{1:I}}$, and $\boldsymbol{D}$ are large but sparse.

4.2 Particle Filter

Under HMMs, state estimation problems are solved by first choosing a statistical objective such as the minimum mean square error (MMSE) criterion and then applying a suitable state estimation algorithm. Filtered MMSE estimates are denoted $\widehat{\boldsymbol{x}}_{i|i}^{\star} \triangleq \mathbb{E}(\boldsymbol{x}_i \,|\, \boldsymbol{y}_{1:i})$ where $\boldsymbol{y}_{1:i}$ is the set of available measurements $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_i\}$. Except for a few special cases, the posterior distribution $p(\boldsymbol{x}_i|\boldsymbol{y}_{1:i})$ necessary to compute the MMSE estimate $\widehat{\boldsymbol{x}}_{i|i}^{\star}$ cannot be derived under the general HMM defined in Sec-

tion 2.2.1. The particle filter (PF) [15]–[19] circumvents this problem by instead operating on a simpler proposal distribution denoted $p(\boldsymbol{x}_i|\boldsymbol{x}_{i-1}, \boldsymbol{y}_i)$, which, when properly weighted, approximates the posterior. However, the PF suffers from particle degeneracy [20]–[22] especially when the state dimension is large. As a result, the PF becomes computationally intractable and, in general, is unsuitable for dynamic tomography.

The posterior PDF may be derived under the linear state-space model (2.9)-(2.10) when the initial state prior, state noise, and measurement noise are jointly Gaussian. In such special cases, the PF may use an optimal proposal PDF that minimizes the variance of the importance weights [21], thereby minimizing degeneracy. The optimal proposal PF algorithm [21] under (2.9)-(2.10) is initialized by generating L samples from the initial prior PDF with uniform importance weight

$$\boldsymbol{x}_0^l \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Pi}_1) \qquad\qquad w_0^l = \frac{1}{L} \tag{4.3}$$

where $\boldsymbol{x}_0^l$ is the lth initial sample, i.i.d. indicates the samples are independent and identically distributed, $\mathcal{N}(\boldsymbol{m}, \boldsymbol{\Theta})$ denotes the Gaussian PDF with mean $\boldsymbol{m}$ and covariance $\boldsymbol{\Theta}$, and w_0^l is the lth initial importance weight. Then, the following steps are recursively applied for each measurement $\boldsymbol{y}_i$. First, the optimal proposal PDF is sampled L times

$$\boldsymbol{x}_i^l \overset{\text{i.i.d.}}{\sim} p(\boldsymbol{x}_i|\boldsymbol{x}_{i-1}^l, \boldsymbol{y}_i) = \mathcal{N}(\boldsymbol{m}_i^l, \boldsymbol{\Theta}_i) \tag{4.4}$$

where

$$\boldsymbol{\Theta}_i \triangleq (\boldsymbol{H}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{H}_i + \boldsymbol{Q}_{i-1}^{-1})^{-1} \tag{4.5}$$

$$\boldsymbol{m}_i^l \triangleq \boldsymbol{\Theta}_i (\boldsymbol{Q}_{i-1}^{-1} \boldsymbol{F}_{i-1} \boldsymbol{x}_{i-1}^l + \boldsymbol{H}_i^T \boldsymbol{R}_i^{-1} \boldsymbol{y}_i) \tag{4.6}$$

and, for notational convenience, $\boldsymbol{Q}_0 \triangleq \boldsymbol{\Pi}_1$ and $\boldsymbol{F}_0 \triangleq \boldsymbol{I}$. Then, the importance weights are updated with

$$w_i^l = w_{i-1}^l \mathcal{N}(\boldsymbol{y}_i;\, \boldsymbol{H}_i \boldsymbol{F}_{i-1} \boldsymbol{x}_{i-1}^l,\, \boldsymbol{H}_i \boldsymbol{Q}_{i-1} \boldsymbol{H}_i^T + \boldsymbol{R}_i) \tag{4.7}$$

where $\mathcal{N}(\boldsymbol{x};\, \boldsymbol{m},\, \boldsymbol{\Theta})$ denotes the value of the $\mathcal{N}(\boldsymbol{m},\, \boldsymbol{\Theta})$ PDF evaluated at $\boldsymbol{x}$. Finally, the importance weights are normalized

$$\mu_i^l = \frac{w_i^l}{\sum_{l=1}^L w_i^l} \tag{4.8}$$

and approximate filtered or posterior MMSE estimates are computed as the weighted average

$$\bar{x}_{i|i}^\star = \sum_{l=1}^L \mu_i^l \boldsymbol{x}_i^l. \tag{4.9}$$

We emphasize that the processing and storage requirements of the optimal proposal PF are at least as great as the Kalman filter discussed below. For example, the matrix $\boldsymbol{\Theta}_i$ (4.5) is $N \times N$ and its storage becomes computationally intractable when the state dimension is large. As empirically demonstrated in Section 5.1, the optimal proposal PF can require a very large number of particles L to produce faithful approximations $\bar{x}_{i|i}^\star$ even for a small dynamic tomography problem.

4.3 Kalman Filter

High-dimensional state estimation is computationally intractable under the general HMM and the remainder of the dissertation focuses on the linear and linear WSS state-space models. This section reviews the Kalman filter (KF) [25], [26], a recursive algorithm for computing linear MMSE (LMMSE) state estimates denoted $\widehat{\boldsymbol{x}}_{i|j}$ under the linear state-space model (2.9)-(2.10).

The KF algorithm begins with the initial prior estimate $\widehat{\boldsymbol{x}}_{1|0} = \boldsymbol{\mu}_1$ and estimate error covariance $\boldsymbol{P}_{1|0} = \boldsymbol{\Pi}_1$. The estimate error covariance at time index i given the measurements $\boldsymbol{y}_{1:j}$ is denoted

$$\boldsymbol{P}_{i|j} \triangleq \mathbb{E}\big[(\boldsymbol{x}_i - \widehat{\boldsymbol{x}}_{i|j})\,(\boldsymbol{x}_i - \widehat{\boldsymbol{x}}_{i|j})^T\big]\,. \tag{4.10}$$

The KF then applies the following two recursive steps for each measurement $\boldsymbol{y}_i$. The first step is the measurement update given by

$$\boldsymbol{K}_i = \boldsymbol{P}_{i|i-1}\,\boldsymbol{H}_i^T\,(\boldsymbol{H}_i\,\boldsymbol{P}_{i|i-1}\,\boldsymbol{H}_i^T + \boldsymbol{R}_i)^{-1} \tag{4.11}$$

$$\widehat{\boldsymbol{x}}_{i|i} = \widehat{\boldsymbol{x}}_{i|i-1} + \boldsymbol{K}_i\,(\boldsymbol{y}_i - \boldsymbol{H}_i\,\widehat{\boldsymbol{x}}_{i|i-1}) \tag{4.12}$$

$$\boldsymbol{P}_{i|i} = \boldsymbol{P}_{i|i-1} - \boldsymbol{K}_i\,\boldsymbol{H}_i\,\boldsymbol{P}_{i|i-1} \tag{4.13}$$

where $\boldsymbol{K}_i$ is commonly referred to as the Kalman gain. The second step is the time update:

$$\widehat{\boldsymbol{x}}_{i+1|i} = \boldsymbol{F}_i\,\widehat{\boldsymbol{x}}_{i|i} \tag{4.14}$$

$$\boldsymbol{P}_{i+1|i} = \boldsymbol{F}_i\,\boldsymbol{P}_{i|i}\,\boldsymbol{F}_i^T + \boldsymbol{Q}_i. \tag{4.15}$$

The KF measurement update can be expressed as the optimization problem [73]

$$\widehat{\boldsymbol{x}}_{i|i} = \arg\min_{\boldsymbol{x}_i} \|\boldsymbol{y}_i - \boldsymbol{H}_i\,\boldsymbol{x}_i\|^2_{\boldsymbol{R}_i^{-1}} + \|\boldsymbol{x}_i - \widehat{\boldsymbol{x}}_{i|i-1}\|^2_{\boldsymbol{P}_{i|i-1}^{-1}}\,. \tag{4.16}$$

The optimization problem (4.16) shows that the posterior estimate $\widehat{\boldsymbol{x}}_{i|i}$ represents a trade-off between agreement with the measurement $\boldsymbol{y}_i$ weighted by its uncertainty given by the measurement covariance $\boldsymbol{R}_i$ in the first term and agreement with the prior estimate $\widehat{\boldsymbol{x}}_{i|i-1}$ weighted by the prior uncertainty $\boldsymbol{P}_{i|i-1}$ in the second term.

Appendix B.2 shows how an additional quadratic penalty term can be appended to (4.16) for the purpose of Tikhonov regularization [74], [75].

The Kalman gain (4.11) is a function of the prior estimate error covariance $\boldsymbol{P}_{i|i-1}$, a symmetric positive definite matrix with $N(N+1)/2 \approx N^2/2$ unique components that can quickly become computationally unwieldy for volumetric image formation. For example, if the unknown object is represented by $128 \times 128 \times 128$ voxels, then 8 TB of memory is necessary to store $\boldsymbol{P}_{i|i-1}$ using 32-bit floating point elements. We note that both the state transition operator $\boldsymbol{F}_i$ and state error covariance $\boldsymbol{Q}_i$ are $N \times N$ matrices and the measurement operator $\boldsymbol{H}_i$ is $M \times N$ and, as a result, each of these matrices also becomes unwieldy when the state dimension N is large. This is not an issue in practice for most remote sensing applications because $\boldsymbol{F}_i$, $\boldsymbol{Q}_i$, and $\boldsymbol{H}_i$ are typically sparse, banded, or have some other simplifying structure.

4.4 Ensemble Kalman Filter

The ensemble Kalman filter (EnKF) [3], [48]–[50] is a Monte Carlo approximation to the KF developed to find approximate LMMSE estimates when the state dimension N is large enough that the storage of the KF estimate error covariance becomes computationally intractable. The general idea is to efficiently update an ensemble of samples such that the ensemble sample mean approximates the LMMSE state estimate. The sample mean is denoted

$$\widetilde{\boldsymbol{x}}_{i|j} \triangleq \frac{1}{L} \sum_{l=1}^{L} \widetilde{\boldsymbol{x}}_{i|j}^{l} \tag{4.17}$$

where $\widetilde{\boldsymbol{x}}_{i|j}^{l}$ is the lth member of the ensemble and L is the total number of ensemble members. The EnKF algorithm begins by sampling each initial prior ensemble

member based on the prior information with

$$\widetilde{\boldsymbol{x}}_{1|0}^{l} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{\mu}_1,\, \boldsymbol{\Pi}_1). \tag{4.18}$$

In the EnKF measurement update, the sample Kalman gain is computed with

$$\widetilde{\boldsymbol{K}}_i = \left[(\boldsymbol{C}_i \circ \widetilde{\boldsymbol{P}}_{i|i-1})\, \boldsymbol{H}_i^T\right] \left[\boldsymbol{H}_i\, (\boldsymbol{C}_i \circ \widetilde{\boldsymbol{P}}_{i|i-1})\, \boldsymbol{H}_i^T + \boldsymbol{R}_i\right]^{-1} \tag{4.19}$$

where $\boldsymbol{C}_i$ is the covariance taper matrix discussed separately in Section 4.4.1, $\circ$ is the Hadamard or element-by-element matrix product [71], and $\widetilde{\boldsymbol{P}}_{i|i-1}$ is the prior sample error covariance

$$\widetilde{\boldsymbol{P}}_{i|i-1} \triangleq \frac{1}{L-1} \sum_{l=1}^{L} (\widetilde{\boldsymbol{x}}_{i|i-1}^{l} - \widetilde{\boldsymbol{x}}_{i|i-1})\, (\widetilde{\boldsymbol{x}}_{i|i-1}^{l} - \widetilde{\boldsymbol{x}}_{i|i-1})^T \tag{4.20}$$

and, as discussed below, $\widetilde{\boldsymbol{P}}_{i|i-1}$ is never explicitly computed or stored in the implementation of the EnKF. Then, each posterior ensemble member is computed with

$$\widetilde{\boldsymbol{x}}_{i|i}^{l} = \widetilde{\boldsymbol{x}}_{i|i-1}^{l} + \widetilde{\boldsymbol{K}}_i\, (\boldsymbol{y}_i^l - \boldsymbol{H}_i\, \widetilde{\boldsymbol{x}}_{i|i-1}^{l}) \tag{4.21}$$

where each vector $\boldsymbol{y}_i^l$ is sampled independently from $\mathcal{N}(\boldsymbol{y}_i,\, \boldsymbol{R}_i)$. In the EnKF time update, each posterior ensemble member is computed with

$$\widetilde{\boldsymbol{x}}_{i+1|i}^{l} = \boldsymbol{F}_i\, \widetilde{\boldsymbol{x}}_{i|i}^{l} + \boldsymbol{u}_i^l \tag{4.22}$$

where each vector $\boldsymbol{u}_i^l$ is an independent sample from $\mathcal{N}(\boldsymbol{0},\, \boldsymbol{Q}_i)$. The additional randomness introduced by the samples $\boldsymbol{y}_i^l$ and $\boldsymbol{u}_i^l$ may seem counterproductive, but is necessary for the prior sample error covariance to be correct on average and match

the KF error covariance [48]. The end result of the EnKF is the set of approximate LMMSE estimates $\widetilde{\boldsymbol{x}}_{i|i}$ for each time index i. The sense in which the EnKF estimates approximate the LMMSE estimates and the fact that the approximation improves as the ensemble size L increases is addressed by Theorem 4.1 in Section 4.4.2.

The computational advantage of the EnKF results from the use of the sample error covariance $\widetilde{\boldsymbol{P}}_{i|i-1}$ in (4.19). Whereas the KF requires the storage of the $N \times N$ matrix $\boldsymbol{P}_{i|i-1}$ to compute the Kalman gain (4.11), the EnKF never explicitly computes the sample error covariance $\widetilde{\boldsymbol{P}}_{i|i-1}$. The equation for the sample Kalman gain $\widetilde{\boldsymbol{K}}_i$ (4.19) has been written to emphasize that only the $N \times M$ matrix $(\boldsymbol{C}_i \circ \widetilde{\boldsymbol{P}}_{i|i-1})\,\boldsymbol{H}_i^T$ is necessary for the computation of $\widetilde{\boldsymbol{K}}_i$ and, in the severely underdetermined case considered in this chapter ($M \ll N$), the storage requirement is relatively small. As discussed in Appendix B.1, it is always possible to process the components of the measurement $\boldsymbol{y}_i$ sequentially ($M = 1$), and the storage requirement for this intermediate result is then relatively negligible. Thus, in terms of computational storage, the largest expense of the EnKF is the LN storage elements necessary for the ensemble.

To minimize storage and processing requirements, a small ensemble size L is desirable. The application of the covariance taper $\boldsymbol{C}_i$ in (4.19) is critical for reducing the required ensemble size for large-dimensional problems and is discussed separately in Section 4.4.1. Ideally, a well-informed choice of L would be based on a theoretical error analysis of the EnKF. The best attempt to date is given in [76] in which the difficulties associated with such an analysis are noted and several key analytical results, including the bias in the sample Kalman gain and the expected error between the true and sample estimate error covariance matrices, are derived under the assumption that $\boldsymbol{H}_i^T \boldsymbol{H}_i = \boldsymbol{I}$ and $\boldsymbol{R}_i = \sigma^2 \boldsymbol{I}$. As noted in [76], it is always possible to apply a linear transformation to the state-space equations (2.9) and (2.10) to ensure that

these conditions apply. However, the results then hold for the transformed system and must be transformed back for interpretation, a process that is only possible if $\boldsymbol{H}_i$ is invertible. An operational method for choosing the ensemble size is demonstrated in Section 5.1.

4.4.1 Covariance tapering

The ensemble size L represents a fundamental trade-off in the EnKF: the larger L the closer the EnKF estimates are to the LMMSE estimates due to reduced sample error but at the cost of increased storage and processing requirements. A serious concern is that the sample error may be significant for sizable ensembles let alone the computationally desirable situation when the ensemble size is small. For example, the EnKF may be expected to perform poorly even at the break-even point $L \approx N/2$ where the storage requirements for the KF and EnKF are the same. In this case there are only $N/2$ available samples of the prior distribution from which to estimate the approximately $N^2/2$ unique components of the prior estimate error covariance matrix and a significant amount of sample error is unavoidable especially when the state dimension N is large.

Covariance tapering [76]–[78] can improve small sample size covariance estimates by incorporating prior information regarding the structure of the population covariance. As with all shrinkage estimators, covariance tapering introduces a systematic bias into the estimator while simultaneously aiming to reduce its variance. For certain applications, the implication is that covariance tapering can provide biased yet superior sample covariance estimates in terms of mean square error.

The EnKF prior sample error covariance $\widetilde{\boldsymbol{P}}_{i|i-1}$ is tapered by the matrix $\boldsymbol{C}_i$ in the computation of the sample Kalman gain (4.19) in the hope of ultimately reducing the EnKF estimate mean square error. Though the best choice of the taper matrix

$\boldsymbol{C}_i$ is problem dependent, two properties it should satisfy are clear. First, the taper matrix must be positive definite, which then ensures that the product $\boldsymbol{C}_i \circ \widetilde{\boldsymbol{P}}_{i|i-1}$ is also positive definite by the Schur product theorem [71]. Also note that covariance tapering is different from covariance truncation. Covariance truncation is equivalent to tapering with a banded binary matrix, a matrix that is equal to 1 along prescribed bands and 0 otherwise. A banded binary matrix is generally indefinite and will not preserve positive definiteness when used as a taper matrix, resulting in algorithmic numerical instabilities. Second, from a practical standpoint, a sparse taper matrix results in numerous algorithmic simplifications that greatly reduce the computational cost of the EnKF.

Some work has been done to characterize the taper matrix that results in the greatest reduction in EnKF sample error [76], but, as noted in that work, it is difficult to optimize over the class of positive definite matrices. A good choice for the taper matrix incorporates prior knowledge of the error covariance matrix structure. Typical physical dynamic systems are correlated over only a finite spatial distance. A taper matrix that suppresses unphysical long correlations and is positive definite, sparse, and easily computed is given in [79]. These taper matrices are constructed through self convolution where $\boldsymbol{C}_i = \boldsymbol{B}_i^T \boldsymbol{B}_i$ and $\boldsymbol{B}_i$ is the $K \times N$ matrix that represents convolution with a specified finite impulse response filter. Figure 4.1 (left) shows a taper matrix constructed in this manner that eliminates all error correlation beyond a distance of 3 pixels in the tapered sample error covariance $\boldsymbol{C}_i \circ \widetilde{\boldsymbol{P}}_{i|i-1}$.

The form of the taper matrix in Figure 4.1 (left) is justified in the following example. Figure 4.1 (right) shows the KF prior error covariance $\boldsymbol{P}_{i|i-1}$ for an 8×8 version of the dynamic tomography example studied in Section 5.2. Since the EnKF sample error covariance $\widetilde{\boldsymbol{P}}_{i|i-1}$ approximates $\boldsymbol{P}_{i|i-1}$ (the exact sense is defined in Section 4.4.2), $\widetilde{\boldsymbol{P}}_{i|i-1}$ will have similar structure as $\boldsymbol{P}_{i|i-1}$ in Figure 4.1 (right). A

large error correlation in $\widetilde{\boldsymbol{P}}_{i|i-1}$ beyond 3 pixels is due to sample error because the corresponding element of $\boldsymbol{P}_{i|i-1}$ in Figure 4.1 (right) is relatively small. The example taper matrix in Figure 4.1 (left) eliminates such erroneous correlations in the product $\boldsymbol{C}_i \circ \widetilde{\boldsymbol{P}}_{i|i-1}$.

4.4.2 Convergence

To date, the convergence of the EnKF has not been fully addressed. The best attempt is in [76] where it is mentioned that the EnKF without covariance tapering ($\boldsymbol{C}_i = \mathbf{1}$, where $\mathbf{1}$ is the matrix with all elements equal to 1), henceforth the unlocalized EnKF, converges to the KF as the ensemble size L increases. However, [76] provides no proof and does not specify the sense of convergence. The theorem below states that the EnKF estimates $\widetilde{\boldsymbol{x}}_{i|i}$ converge in probability to the estimates given by the localized Kalman filter (LKF) defined immediately after the theorem. Note that the LKF is as computationally expensive as the KF and is not intended for use in large-scale state estimation. A corollary to the theorem states that the unlocalized EnKF converges to the KF. The theorem and corollary are important for the following reasons: they establish a theoretical basis for the validity of the EnKF by showing that it converges to a well defined limit, provide proof that the unlocalized EnKF is a Monte Carlo KF algorithm, and suggest a new means for investigating the bias introduced by covariance tapering through the comparison of the KF and LKF.

Theorem 4.1: For each time index i, the EnKF estimates $\widetilde{\boldsymbol{x}}_{i|i}$ converge in probability to the LKF estimates $\widetilde{\boldsymbol{x}}^{\infty}_{i|i}$ defined below, i.e.,

$$\widetilde{\boldsymbol{x}}_{i|i} \xrightarrow{\text{p.}} \widetilde{\boldsymbol{x}}^{\infty}_{i|i} \tag{4.23}$$

in the limit as the ensemble size $L \to \infty$.

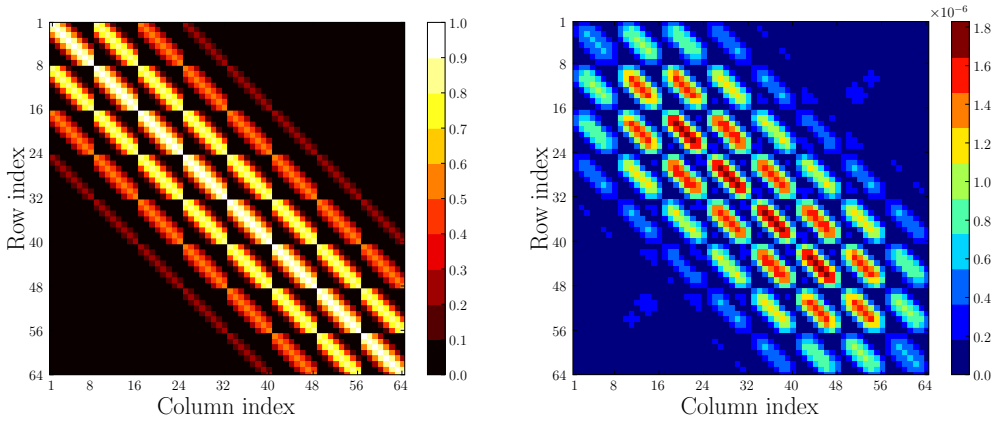

Figure 4.1 An example taper matrix $\boldsymbol{C}_i$ (left) and the absolute value of the KF error covariance $\boldsymbol{P}_{i|i-1}$ for an 8×8 dynamic tomography numerical example (right).

Proof. See Appendix A. □

The LKF is initialized with $\widetilde{\boldsymbol{x}}^{\infty}_{1|0} = \boldsymbol{\mu}_1$ and $\widetilde{\boldsymbol{P}}^{\infty}_{1|0} = \boldsymbol{\Pi}_1$. The LKF measurement update is given by

$$\widetilde{\boldsymbol{K}}^{\infty}_i = (\boldsymbol{C}_i \circ \widetilde{\boldsymbol{P}}^{\infty}_{i|i-1})\, \boldsymbol{H}_i^T \left[\boldsymbol{H}_i\, (\boldsymbol{C}_i \circ \widetilde{\boldsymbol{P}}^{\infty}_{i|i-1})\, \boldsymbol{H}_i^T + \boldsymbol{R}_i\right]^{-1} \tag{4.24}$$

$$\widetilde{\boldsymbol{x}}^{\infty}_{i|i} = \widetilde{\boldsymbol{x}}^{\infty}_{i|i-1} + \widetilde{\boldsymbol{K}}^{\infty}_i\, (\boldsymbol{y}_i - \boldsymbol{H}_i\, \widetilde{\boldsymbol{x}}^{\infty}_{i|i-1}) \tag{4.25}$$

$$\begin{aligned}\widetilde{\boldsymbol{P}}^{\infty}_{i|i} = {} & \widetilde{\boldsymbol{P}}^{\infty}_{i|i-1} - \widetilde{\boldsymbol{K}}^{\infty}_i\, \boldsymbol{H}_i\, \widetilde{\boldsymbol{P}}^{\infty}_{i|i-1} - \widetilde{\boldsymbol{P}}^{\infty}_{i|i-1}\, \boldsymbol{H}_i^T\, (\widetilde{\boldsymbol{K}}^{\infty}_i)^T \\ & + \widetilde{\boldsymbol{K}}^{\infty}_i\, (\boldsymbol{H}_i\, \widetilde{\boldsymbol{P}}^{\infty}_{i|i-1}\, \boldsymbol{H}_i^T + \boldsymbol{R}_i)\, (\widetilde{\boldsymbol{K}}^{\infty}_i)^T.\end{aligned} \tag{4.26}$$

The LKF time update is given by

$$\widetilde{\boldsymbol{x}}^{\infty}_{i+i|i} = \boldsymbol{F}_i\, \widetilde{\boldsymbol{x}}^{\infty}_{i|i} \tag{4.27}$$

$$\widetilde{\boldsymbol{P}}^{\infty}_{i+1|i} = \boldsymbol{F}_i\, \widetilde{\boldsymbol{P}}^{\infty}_{i|i}\, \boldsymbol{F}_i^T + \boldsymbol{Q}_i. \tag{4.28}$$

Corollary 4.2: For each time index i, the unlocalized ($\boldsymbol{C}_i = \mathbf{1}$) EnKF estimates $\widetilde{\boldsymbol{x}}_{i|i}$

converge in probability to the LMMSE estimates $\widehat{\boldsymbol{x}}_{i|i}$, i.e.,

$$\widetilde{\boldsymbol{x}}_{i|i} \xrightarrow{\text{p.}} \widehat{\boldsymbol{x}}_{i|i} \tag{4.29}$$

in the limit as the ensemble size $L \to \infty$.

Proof. When no covariance taper is applied, the LKF and KF are equivalent. Then, apply the result of Theorem 4.1. □

The LKF measurement update (4.25) may alternatively be expressed as the optimization problem

$$\widetilde{\boldsymbol{x}}_{i|i}^{\infty} = \arg\min_{\boldsymbol{x}_i} \|\boldsymbol{y}_i - \boldsymbol{H}_i\boldsymbol{x}_i\|^2_{\boldsymbol{R}_i^{-1}} + \|\boldsymbol{x}_i - \widetilde{\boldsymbol{x}}_{i|i-1}^{\infty}\|^2_{(\boldsymbol{C}_i \circ \boldsymbol{P}_{i|i-1}^{\infty})^{-1}} \tag{4.30}$$

which, when compared to (4.16), shows that covariance tapering alters the prior error covariance used in the measurement update. The covariance taper bias will be small when $\boldsymbol{C}_i \circ \boldsymbol{P}_{i|i-1} \approx \boldsymbol{P}_{i|i-1}$ because (4.16) and (4.30) will almost be equal. Note that this condition is applicable in the dynamic tomography example illustrated in Figure 4.1.

4.5 Kalman-Wiener Filter

The optimal state estimation algorithm under the signal model in Section 2.2.3 is derived in this section. The first three subsections contain intermediate results that will ultimately be used in Section 4.5.3 to derive the optimal spatial-temporal estimator. Some of these results are classic but are included for completeness. The final subsection derives the optimal state estimation algorithm when the measurements are noisy tomographic projections.

4.5.1 Orthogonal projector

The orthogonal projector to a subspace of a Hilbert space is used throughout the remainder of this section to guarantee optimality. We summarize the relevant properties in the following theorem.

Theorem 4.3: If S is a complete subspace of the Hilbert space X, then there exists a unique continuous linear operator $\mathcal{P}_{\mathsf{S}}$ called the orthogonal projector that has the property

$$\|\{\mathcal{P}_{\mathsf{S}}\, x\} - x\| \leq \|y - x\| \tag{4.31}$$

for all $x \in \mathsf{X}$ and $y \in \mathsf{S}$ and equality holds if and only if $y = \{\mathcal{P}_{\mathsf{S}}\, x\}$. Furthermore, the orthogonal projector satisfies

$$\{\mathcal{P}_{\mathsf{S}}\, x\} - x \in \mathsf{S}^{\perp} \qquad \text{and} \qquad \{\mathcal{P}_{\mathsf{S}}\, x\} \in \mathsf{S} \tag{4.32}$$

for all $x \in \mathsf{X}$.

Proof. See [62]. □

An important property of the orthogonal projector $\mathcal{P}_{\mathsf{S}}$ results when the subspace S can be factored into two orthogonal subspaces.

Lemma 4.4: Let A and B be orthogonal closed subspaces of the Hilbert space X. Then $\mathcal{P}_{\mathsf{A}\oplus\mathsf{B}} = \mathcal{P}_{\mathsf{A}} + \mathcal{P}_{\mathsf{B}}$ where $\mathsf{A} \oplus \mathsf{B}$ is the direct sum [62] of A and B.

Proof. See [62]. □

This lemma will later be used to derive a recursive estimator in Section 4.5.3.

4.5.2 Optimal spatial estimator

The Wiener filter is the minimum mean square error (MMSE) optimal estimation algorithm for WSS random fields. The following theorem summarizes the properties of the Wiener filter that will later be used in Section 4.5.3.

Theorem 4.5: If X and Y are zero mean random fields, Y is WSS, and X and Y are jointly WSS, then the Wiener filter is the MMSE estimator in the class of all linear filters. This is stated mathematically as

$$\|\widehat{x}(\boldsymbol{r}) - x(\boldsymbol{r})\| \leq \left\| \int_{\mathsf{S}} k(\boldsymbol{r}, \boldsymbol{s})\, y(\boldsymbol{s})\, d\boldsymbol{s} - x(\boldsymbol{r}) \right\| \tag{4.33}$$

where $\widehat{x}(\boldsymbol{r})$ is the output of the Wiener filter, $k(\boldsymbol{r}, \boldsymbol{s})$ is a finite energy function, and equality holds if and only if $k(\boldsymbol{r}, \boldsymbol{s})$ is the Wiener filter impulse response. Furthermore, the Wiener filter impulse response is shift-invariant and has frequency response

$$\{\mathcal{F}\, k\}(\boldsymbol{f}) = \frac{S_{X,Y}(\boldsymbol{f})}{S_Y(\boldsymbol{f})} \tag{4.34}$$

where $\mathcal{F}$ is the Fourier transform.

Proof. See [27] for a proof for the case $d = 1$ or [80] for the multidimensional case $d > 1$. □

The following lemma shows that the Wiener filter is an orthogonal projector, justifying its use as the optimal spatial filter in Section 4.5.3.

Lemma 4.6: The Wiener filter is the orthogonal projector of the random field X onto the WSS random field Y, meaning

$$\{\mathcal{P}_{\text{span}(y)}\, x\}(\boldsymbol{r}) = \{k \star y\}(\boldsymbol{r}) \tag{4.35}$$

where span(y) is the subspace spanned by Y.

Proof. Note that (4.33) is a particular case of the inequality (4.31). Then, because equality in (4.33) is satisfied if and only if $\widehat{x}(\boldsymbol{r}) = \{k \star y\}(\boldsymbol{r})$, the Wiener filter is the orthogonal projector. □

4.5.3 Optimal recursive spatial-temporal estimator

The derivation of the Kalman-Wiener filter (KWF) will proceed in three steps in a manner analogous to the derivation of the Kalman filter [27]: initialization, measurement update, and time update. The initialization step consists of identifying the initial prior information for use in the first measurement update. In this case, the given initial prior information consists of the prior estimate $\widehat{x}_{1|0}(\boldsymbol{r}) = 0$ and prior error power spectrum $S_{\ddot{X}_{1|0}}(\boldsymbol{f}) = S_{X_1}(\boldsymbol{f})$ where

$$\ddot{x}_{i|i-1}(\boldsymbol{r}) \triangleq x_i(\boldsymbol{r}) - \widehat{x}_{i|i-1}(\boldsymbol{r}) \tag{4.36}$$

is the prior error.

The KWF measurement update is summarized in the following theorem.

Theorem 4.7: The KWF estimate measurement update is given by

$$\widehat{x}_{i|i}(\boldsymbol{t}) = \widehat{x}_{i|i-1}(\boldsymbol{t}) + \{k_i \star e_i\}(\boldsymbol{t}) \tag{4.37}$$

where $e_i(\boldsymbol{s}) \triangleq y_i(\boldsymbol{s}) - \{H_i\, \widehat{x}_{i|i-1}\}(\boldsymbol{s})$ is the innovation and the frequency response of the Wiener filter is

$$K_i(\boldsymbol{f}) \triangleq \{\mathcal{F}\, k_i\}(\boldsymbol{f}) = \frac{S_{\ddot{X}_{i|i-1},\, \ddot{Z}_{i|i-1}}(\boldsymbol{f})}{S_{\ddot{Z}_{i|i-1}}(\boldsymbol{f}) + S_{V_i}(\boldsymbol{f})} \tag{4.38}$$

where $\ddot{z}(\boldsymbol{s}) \triangleq \{H_i\, \ddot{x}_{i|i-1}\}(\boldsymbol{s})$, provided $\ddot{x}_{i|i-1}$ and $\ddot{z}_{i|i-1}$ are jointly WSS, $\ddot{z}_{i|i-1}$ is WSS, and the measurement noise v_i is WSS. Furthermore, the measurement update on the error power spectrum is given by

$$\begin{aligned} S_{\ddot{X}_{i|i}}(\boldsymbol{f}) = S_{\ddot{X}_{i|i-1}}(\boldsymbol{f}) - 2\,\Re\big[K_i(\boldsymbol{f})\, S_{\ddot{X}_{i|i-1},\,\ddot{Z}_{i|i-1}(\boldsymbol{f})}\big] \\ + |K_i(\boldsymbol{f})|^2\, S_{\ddot{Z}_{i|i-1}}(\boldsymbol{f}) + |K_i(\boldsymbol{f})|^2\, S_{V_i}(\boldsymbol{f}) \end{aligned} \tag{4.39}$$

where $\Re(\cdot)$ is the real part of the complex argument.

Proof. First, we derive the estimate measurement update (4.37). Given (2.22) and (2.23), the innovation has the following properties:

$$\operatorname{span}(y_{1:j}) = \operatorname{span}(e_{1:j}) \qquad \text{and} \qquad \langle e_i(\boldsymbol{s}),\, e_j(\boldsymbol{\sigma})\rangle = 0 \text{ for all } i \neq j. \tag{4.40}$$

The above properties and Lemma 4.4 give the result

$$\begin{aligned} \widehat{x}_{i|i}(\boldsymbol{r}) &= \{\mathcal{P}_{\operatorname{span}(e_{1:i-1})}\, x_i\}(\boldsymbol{r}) + \{\mathcal{P}_{\operatorname{span}(e_i)}\, x_i\}(\boldsymbol{r}) \\ &= \widehat{x}_{i|i-1}(\boldsymbol{r}) + \{\mathcal{P}_{\operatorname{span}(e_i)}\, x_i\}(\boldsymbol{r}). \end{aligned} \tag{4.41}$$

In addition, Lemma 4.6 can be applied to (4.41) to give

$$\widehat{x}_{i|i}(\boldsymbol{r}) = \widehat{x}_{i|i-1}(\boldsymbol{r}) + \{k_i \star e_i\}(\boldsymbol{r}) \tag{4.42}$$

where the frequency response of the Wiener filter is given by

$$K_i(\boldsymbol{f}) \triangleq \{\mathcal{F}\, k_i\}(\boldsymbol{f}) = \frac{S_{X_i,\,E_i}(\boldsymbol{f})}{S_{E_i}(\boldsymbol{f})}. \tag{4.43}$$

We now show that (4.43) is equivalent to (4.38). By definition, the cross-correlation

between the state random field x_i and innovation e_i is given by

$$\begin{aligned} R_{X_i,\,E_i}(\boldsymbol{r}-\boldsymbol{r}') &= \mathbb{E}[x_i(\boldsymbol{r})\, e_i(\boldsymbol{r}')] \\ &= \mathbb{E}\big[x_i(\boldsymbol{r})\, \ddot{z}_{i|i-1}(\boldsymbol{r}')\big] + \mathbb{E}[x_i(\boldsymbol{r})\, v_i(\boldsymbol{r}')] = \mathbb{E}\big[x_i(\boldsymbol{r})\, \ddot{z}_{i|i-1}(\boldsymbol{r}')\big] \end{aligned} \tag{4.44}$$

and the last step holds because of (2.24). Then,

$$R_{X_i,\,E_i}(\boldsymbol{r}-\boldsymbol{r}') = \mathbb{E}\big[\ddot{x}_{i|i-1}(\boldsymbol{r})\, \ddot{z}_{i|i-1}(\boldsymbol{r}')\big] + \mathbb{E}\big[\widehat{x}_{i|i-1}(\boldsymbol{r})\, \ddot{z}_{i|i-1}(\boldsymbol{r}')\big] \tag{4.45}$$

and, from (4.32),

$$\ddot{x}_{i|i-1}(\boldsymbol{r}) \in \operatorname{span}(y_{1:i-1})^{\perp} \qquad \text{and} \qquad \widehat{x}_{i|i-1}(\boldsymbol{r}) \in \operatorname{span}(y_{1:i-1}) \tag{4.46}$$

which implies

$$R_{X_i,\,E_i}(\boldsymbol{r}-\boldsymbol{r}') = \mathbb{E}\big[\ddot{x}_{i|i-1}(\boldsymbol{r})\, \ddot{z}_{i|i-1}(\boldsymbol{r}')\big]\,. \tag{4.47}$$

Then, because $\ddot{x}_{i|i-1}$ and $\ddot{z}_{i|i-1}$ are assumed jointly WSS,

$$S_{X_i,\,E_i}(\boldsymbol{f}) = S_{\ddot{X}_{i|i-1},\,\ddot{Z}_{i|i-1}}(\boldsymbol{f}). \tag{4.48}$$

Finally, the auto-correlation of the innovation is by definition

$$R_{E_i}(\boldsymbol{s}-\boldsymbol{s}') = \mathbb{E}\big\{[\ddot{z}_{i|i-1}(\boldsymbol{s}) - v_i(\boldsymbol{s})][\ddot{z}_{i|i-1}(\boldsymbol{s}') - v_i(\boldsymbol{s}')]\big\}\,. \tag{4.49}$$

The equation above can be simplified by first noting that

$$\ddot{x}_{i|i-1}(\boldsymbol{r}) \in \operatorname{span}(x_1,\, u_{1:i-1},\, v_{1:i-1}) \tag{4.50}$$

and, as a consequence,

$$\ddot{z}_{i|i-1}(\boldsymbol{r}) \in \text{span}(x_1,\, u_{1:i-1},\, v_{1:i-1}). \tag{4.51}$$

The above imply that

$$\ddot{x}_{i|i-1} \perp v_i \qquad \text{and} \qquad \ddot{z}_{i|i-1} \perp v_i \tag{4.52}$$

by also noting properties (2.22) and (2.23). Then, the innovation has the simplified power spectrum

$$S_{E_i}(\boldsymbol{f}) = S_{\ddot{Z}_{i|i-1}}(\boldsymbol{f}) + S_{V_i}(\boldsymbol{f}) \tag{4.53}$$

because $\ddot{z}_{i|i-1}$ and v_i are assumed WSS and (4.52).

In the final step, we derive the recursive update on the error power spectrum (4.39). First, we expand (4.37) as

$$\ddot{x}_{i|i}(\boldsymbol{r}) = \ddot{x}_{i|i-1}(\boldsymbol{r}) - \{k_i \star \ddot{z}_{i|i-1}\}(\boldsymbol{r}) - \{k_i \star v_i\}(\boldsymbol{r}). \tag{4.54}$$

Then, by definition, the error auto-correlation is given by

$$R_{\ddot{X}_{i|i}}(\boldsymbol{r} - \boldsymbol{r}') = \mathbb{E}\big[\ddot{x}_{i|i}(\boldsymbol{r})\, \ddot{x}_{i|i}(\boldsymbol{r}')\big]\,. \tag{4.55}$$

The above relationship and (4.52) can be used to derive the recursive power spectrum update

$$\begin{aligned} S_{\ddot{X}_{i|i}}(\boldsymbol{f}) = S_{\ddot{X}_{i|i-1}}(\boldsymbol{f}) - K_i^*(\boldsymbol{f})\, S_{\ddot{X}_{i|i-1},\, \ddot{Z}_{i|i-1}(\boldsymbol{f})} - K_i(\boldsymbol{f})\, S_{\ddot{Z}_{i|i-1},\, \ddot{X}_{i|i-1}}(\boldsymbol{f}) \\ + |K_i(\boldsymbol{f})|^2\, S_{\ddot{Z}_{i|i-1}}(\boldsymbol{f}) + |K_i(\boldsymbol{f})|^2\, S_{V_i}(\boldsymbol{f}) \end{aligned} \tag{4.56}$$

where $(\cdot)^*$ is the complex conjugation operator. $\square$

The KWF time update is summarized in the following theorem.

Theorem 4.8: The KWF estimate time update is given by

$$\widehat{x}_{i+1|i}(\boldsymbol{r}) = \{f_i \star \widehat{x}_{i|i}\}(\boldsymbol{r}) \tag{4.57}$$

where f_i is the impulse response of the state-transition operator as in (2.17). The KWF error power spectrum time update is given by

$$S_{\ddot{x}_{i+1|i}}(\boldsymbol{f}) = |F_i(\boldsymbol{f})|^2 \, S_{\ddot{X}_{i|i}}(\boldsymbol{f}) + S_{U_i}(\boldsymbol{f}) \tag{4.58}$$

where the function $F_i(\boldsymbol{f}) \triangleq \{\mathcal{F} f_i\}(\boldsymbol{f})$ is the state-transition frequency response and S_{U_i} is the power spectrum of the WSS state noise u_i.

Proof. The MMSE optimal one-step prediction of the state is by definition

$$\begin{aligned} \widehat{x}_{i+1|i}(\boldsymbol{r}) &\triangleq \{P_{\text{span}(e_{1:i})} \, x_{i+1}\}(\boldsymbol{r}) \\ &= \{P_{\text{span}(e_{1:i})} \, \{f_i \star x_i\}\}(\boldsymbol{r}) + \{P_{\text{span}(e_{1:i})} \, u_i\}(\boldsymbol{r}). \end{aligned} \tag{4.59}$$

Note that $\{P_{\text{span}(e_{1:i})} \, u_i\}(\boldsymbol{r}) = 0$ by (2.26). Then, we have

$$\begin{aligned} \{P_{\text{span}(e_{1:i})} \, \{f_i \star x_i\}\}(\boldsymbol{r}) &= \left\{P_{\text{span}(e_{1:i})} \int_{\mathbb{R}^d} f_i(\boldsymbol{r} - \boldsymbol{r}') \, x_i(\boldsymbol{r}') \, d\boldsymbol{r}'\right\}(\boldsymbol{r}) \\ &= \int_{\mathbb{R}^d} f_i(\boldsymbol{r} - \boldsymbol{r}') \{P_{\text{span}(e_{1:i})} \, x_i\}(\boldsymbol{r}') \, d\boldsymbol{r}' \\ &= \{f_i \star \widehat{x}_{i|i}\}(\boldsymbol{r}) \end{aligned} \tag{4.60}$$

where the second step holds because the orthogonal projector is a linear operator. $\square$

4.5.4 KWF tomographic projection measurement update

This subsection focuses on the case where the prior error $\ddot{x}_{i|i-1}$ and measurement noise $v_{\boldsymbol{\theta}_i}$ are WSS spatial random fields and the measurements are noisy tomographic projections as in

$$y_{\boldsymbol{\theta}_i}(\boldsymbol{s}) = \{H_{\boldsymbol{\theta}_i}\, x_i\}(\boldsymbol{s}) + v_{\boldsymbol{\theta}_i}(\boldsymbol{s}) \tag{4.61}$$

with x_i and $v_{\boldsymbol{\theta}_i}$ uncorrelated. Under these conditions, the KWF measurement update detailed in Theorem 4.7 can be derived in closed form. This derivation proceeds by first stating the general statistical properties of a WSS spatial random field and its noisy tomographic projection. These properties are the d-dimensional extension of Jain's work [81], [82] which only considers the 2-D case. Next, the KWF measurement update is derived in closed form under the above conditions. The subsection concludes with a corollary that shows that this KWF measurement update can be implemented by the filtered backprojection algorithm.

We begin with the derivation of the cross-correlation between a WSS spatial random field and its noisy tomographic projection.

Lemma 4.9: The cross-correlation between a WSS random field x and its noisy tomographic projection $y_{\boldsymbol{\theta}_i}$ is given by

$$R_{X_i, Y_{\boldsymbol{\theta}_i}}(\boldsymbol{\tau}) = \{H_{\boldsymbol{\theta}_i}\, R_{X_i}\}(\boldsymbol{\tau}),\ \boldsymbol{\tau} \in \Theta_i^{\perp} \tag{4.62}$$

and note that the random fields x_i and $y_{\boldsymbol{\theta}_i}$ are jointly WSS.

Proof. We have

$$\begin{aligned} R_{X_i, Y_{\boldsymbol{\theta}_i}}(\boldsymbol{t},\, \boldsymbol{s}) &= \mathbb{E}[x_i(\boldsymbol{t})\, y_{\boldsymbol{\theta}_i}(\boldsymbol{s})] \\ &= \int_{\mathbb{R}} \mathbb{E}[x_i(\boldsymbol{t})\, x_i(\boldsymbol{s} + l\boldsymbol{\theta}_i)]\, dl + \mathbb{E}[x_i(\boldsymbol{t})\, v_{\boldsymbol{\theta}_i}(\boldsymbol{s})] \end{aligned}$$

$$= \int_{\mathbb{R}} R_{X_i}(\boldsymbol{t} - \boldsymbol{s} - l\boldsymbol{\theta}_i)\, dl$$
$$= \{H_{\boldsymbol{\theta}_i}\, R_{X_i}\}(\boldsymbol{t} - \boldsymbol{s}). \qquad \square$$

The following theorem is the well-known projection slice theorem and will be useful in deriving the cross power spectrum between a WSS spatial random field and its noisy tomography projection below. We repeat the theorem for the sake of completeness.

Theorem 4.10: A function x and its noisy projection $\{H_{\boldsymbol{\theta}}\, x\}$ are related in the frequency domain by

$$\{\mathcal{F}\,\{H_{\boldsymbol{\theta}}\, x\}\}(\boldsymbol{f}) = \{\mathcal{F}\, x\}(\boldsymbol{f}) \tag{4.63}$$

where $\mathcal{F}$ is the Fourier transform operator and $\boldsymbol{f} \in \Theta^{\perp}$.

Proof. See the proof to Theorem 1.1 in [56]. $\square$

The following corollary is a direct consequence of Lemma 4.9 and Theorem 4.10.

Corollary 4.11: The cross power spectrum between a WSS random field x_i and its noisy projection $y_{\boldsymbol{\theta}_i}$ is given by

$$S_{X_i, Y_{\boldsymbol{\theta}_i}}(\boldsymbol{f}) = S_{X_i}(\boldsymbol{f}),\ \ \boldsymbol{f} \in \Theta_i^{\perp}. \tag{4.64}$$

Proof. We have

$$\begin{aligned} S_{X_i, Y_{\boldsymbol{\theta}_i}}(\boldsymbol{f}) &\triangleq \{\mathcal{F}\, R_{X_i, Y_{\boldsymbol{\theta}_i}}\}(\boldsymbol{f}) \\ &= \{\mathcal{F}\,\{H_{\boldsymbol{\theta}_i}\, R_{X_i}\}\}(\boldsymbol{f}) \\ &= \{\mathcal{F}\, R_{X_i}\}(\boldsymbol{f}) \triangleq S_{X_i}(\boldsymbol{f}) \end{aligned}$$

where the last step is a consequence of Theorem 4.10. □

We now derive the auto-correlation function for a noisy tomographic projection.

Lemma 4.12: The auto-correlation of a noisy projection $y_{\boldsymbol{\theta}_i}$ is given by

$$R_{Y_{\boldsymbol{\theta}_i}}(\boldsymbol{\tau}) = \{H_{\boldsymbol{\theta}_i}\, R_{X_i, Y_{\boldsymbol{\theta}_i}}\}(\boldsymbol{\tau}) + R_{V_{\boldsymbol{\theta}_i}}(\boldsymbol{\tau}),\ \boldsymbol{\tau} \in \Theta_i^{\perp}. \tag{4.65}$$

Proof. We have

$$\begin{aligned} R_{Y_{\boldsymbol{\theta}_i}}(\boldsymbol{s},\, \boldsymbol{\sigma}) &\triangleq \mathbb{E}[y_{\boldsymbol{\theta}_i}(\boldsymbol{s})\, y_{\boldsymbol{\theta}_i}(\boldsymbol{\sigma})] \\ &= \int_{\mathbb{R}} \mathbb{E}[x_i(\boldsymbol{s} + l\boldsymbol{\theta}_i)\, y_{\boldsymbol{\theta}_i}(\boldsymbol{\sigma})]\, dl + \mathbb{E}[v_{\boldsymbol{\theta}_i}(\boldsymbol{s})\, v_{\boldsymbol{\theta}_i}(\boldsymbol{\sigma})] \\ &= \int_{\mathbb{R}} R_{X_i, Y_{\boldsymbol{\theta}_i}}(\boldsymbol{s} + l\boldsymbol{\theta}_i - \boldsymbol{\sigma})\, dl + R_{V_{\boldsymbol{\theta}_i}}(\boldsymbol{s} - \boldsymbol{\sigma}) \\ &= \{H_{\boldsymbol{\theta}_i}\, R_{X_i, Y_{\boldsymbol{\theta}_i}}\}(\boldsymbol{s} - \boldsymbol{\sigma}) + R_{V_{\boldsymbol{\theta}_i}}(\boldsymbol{s} - \boldsymbol{\sigma}). \end{aligned}$$ □

The following corollary addresses the power spectrum of a noisy tomographic projection and follows directly as a consequence of Lemma 4.12.

Corollary 4.13: The power spectrum of a noisy projection $y_{\boldsymbol{\theta}_i}$ is given by

$$S_{Y_{\boldsymbol{\theta}_i}}(\boldsymbol{f}) = S_{X_i}(\boldsymbol{f}) + S_{V_{\boldsymbol{\theta}_i}}(\boldsymbol{f}),\ \boldsymbol{f} \in \Theta_i^{\perp}. \tag{4.66}$$

Proof. We have

$$\begin{aligned} S_{Y_{\boldsymbol{\theta}_i}}(\boldsymbol{f}) &\triangleq \{\mathcal{F}\, R_{Y_{\boldsymbol{\theta}_i}}\}(\boldsymbol{f}) \\ &= \{\mathcal{F}\, \{H_{\boldsymbol{\theta}_i}\, R_{X_i, Y_{\boldsymbol{\theta}_i}}\}\}(\boldsymbol{f}) + \{\mathcal{F}\, R_{V_{\boldsymbol{\theta}_i}}\}(\boldsymbol{f}) \\ &= S_{X_i, Y_{\boldsymbol{\theta}_i}}(\boldsymbol{f}) + S_{V_{\boldsymbol{\theta}_i}}(\boldsymbol{f}) = S_{X_i}(\boldsymbol{f}) + S_{V_{\boldsymbol{\theta}_i}}(\boldsymbol{f}) \end{aligned}$$

where the last step has made use of Theorem 4.10 and (4.64). □

We now summarize the relevant Wiener filter for estimating a WSS spatial random field given a noisy tomographic projection measurement in the following theorem. Note that the final result is also expressed in the notation used in Section 4.5.3.

Theorem 4.14: The impulse response of the Wiener filter for estimating a WSS random field x_i given its noisy projection $y_{\boldsymbol{\theta}_i}$ is

$$k_i(\boldsymbol{t}) = \frac{1}{|\mathsf{S}^{n-2}|}\{H^*_{\boldsymbol{\theta}_i}\{r \star w_i\}\}(\boldsymbol{t}) \tag{4.67}$$

where $|\mathsf{S}^{n-1}|$ is the surface area of the n-dimensional unit hypersphere and is equal to

$$|\mathsf{S}^{n-1}| = \frac{2\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)} \tag{4.68}$$

where $\Gamma(\cdot)$ is the Gamma function, $H^*_{\boldsymbol{\theta}_i}$ is the backprojection operator, r is the ramp filter with frequency response

$$R(\boldsymbol{f}) = |\boldsymbol{f}|,\ \boldsymbol{f} \in \Theta_i^{\perp}, \tag{4.69}$$

and w_i is the Wiener filter impulse response for signals in additive uncorrelated noise with frequency response

$$W_i(\boldsymbol{f}) = \frac{S_{X_i}(\boldsymbol{f})}{S_{X_i}(\boldsymbol{f}) + S_{V_{\boldsymbol{\theta}_i}}(\boldsymbol{f})},\ \boldsymbol{f} \in \Theta_i^{\perp}. \tag{4.70}$$

Expressed using the notation of Section 4.5.3, the Wiener filter frequency response is

$$W_i(\boldsymbol{f}) = \frac{S_{\ddot{X}_{i|i-1}}(\boldsymbol{f})}{S_{\ddot{X}_{i|i-1}}(\boldsymbol{f}) + S_{V_i}(\boldsymbol{f})},\ \boldsymbol{f} \in \Theta_i^{\perp}. \tag{4.71}$$

Proof. Given a single noisy tomographic projection $y_{\boldsymbol{\theta}_i}$, the Wiener filter has frequency response

$$K_i(\boldsymbol{f}) = \frac{S_{X_i, Y_{\boldsymbol{\theta}_i}}(\boldsymbol{f})}{S_{Y_{\boldsymbol{\theta}_i}}(\boldsymbol{f})},\ \boldsymbol{f} \in \Theta_i^{\perp} \tag{4.72}$$

which simplifies to

$$K_i(\boldsymbol{f}) = \frac{S_{X_i}(\boldsymbol{f})}{S_{X_i}(\boldsymbol{f}) + S_{V_{\boldsymbol{\theta}_i}}(\boldsymbol{f})} \tag{4.73}$$

by noting (4.64) and (4.66). By definition, the impulse response of the Wiener filter is given by

$$\begin{aligned} k_i(\boldsymbol{t}) &= \int_{\mathbb{R}^d} K_i(\boldsymbol{f})\, e^{j2\pi \boldsymbol{f}\cdot\boldsymbol{t}}\, d\boldsymbol{t} \\ &= \frac{1}{|\mathsf{S}^{n-2}|} \int_{\mathsf{S}^{n-1}} \int_{\Phi_i^{\perp}} |\boldsymbol{f}'|\, K_i(\boldsymbol{f}')\, e^{j2\pi \boldsymbol{f}'\cdot\boldsymbol{t}}\, d\boldsymbol{f}'\, d\boldsymbol{\phi} \end{aligned} \tag{4.74}$$

where the above change to generalized polar coordinates is justified by (VII.2.8) in [56]. Note that the Wiener filter (4.73) updates only on the plane $\boldsymbol{f} \in \Theta_i^{\perp}$ and we may plug this relationship into (4.74) by using a Dirac delta function as follows:

$$\begin{aligned} k_i(\boldsymbol{t}) &= \frac{1}{|\mathsf{S}^{n-2}|} \int_{\mathsf{S}^{n-1}} \int_{\Phi_i^{\perp}} \frac{|\boldsymbol{f}'|\, S_{X_i}(\boldsymbol{f}')\, \delta(\boldsymbol{\theta} - \boldsymbol{\phi})}{S_{X_i}(\boldsymbol{f}') + S_{V_{\boldsymbol{\theta}_i}}}\, e^{j2\pi \boldsymbol{f}'\cdot\boldsymbol{t}}\, d\boldsymbol{f}'\, d\boldsymbol{\phi} \\ &= \frac{1}{|\mathsf{S}^{n-2}|} \int_{\Theta_i^{\perp}} \frac{|\boldsymbol{f}'|\, S_{X_i}(\boldsymbol{f}')}{S_{X_i}(\boldsymbol{f}') + S_{\boldsymbol{\theta}_i}(\boldsymbol{f}')}\, e^{j2\pi \boldsymbol{f}'\cdot\boldsymbol{t}}\, d\boldsymbol{f}' \\ &= \frac{1}{|\mathsf{S}^{n-2}|} \int_{\Theta_i^{\perp}} R(\boldsymbol{f}')\, W_i(\boldsymbol{f}')\, e^{j2\pi \boldsymbol{f}'\cdot \boldsymbol{P}_{\Theta_i^{\perp}} \boldsymbol{t}}\, d\boldsymbol{f}' \end{aligned} \tag{4.75}$$

where the matrix $\boldsymbol{P}_{\Theta_i^{\perp}} = \boldsymbol{I} - \boldsymbol{\theta}_i\, \boldsymbol{\theta}_i^T$ is the orthogonal projector from $\mathbb{R}^d$ to $\Theta_i^{\perp}$. By the convolution property of the Fourier transform, (4.75) simplifies to

$$k_i(\boldsymbol{t}) = \frac{1}{|\mathsf{S}^{n-2}|} \{r \star w_i\}(\boldsymbol{P}_{\Theta_i^{\perp}}\, \boldsymbol{t}). \tag{4.76}$$

Finally, note that the adjoint of the projection operator is the backprojection operator

defined by

$$\{H_{\boldsymbol{\theta}}^{*}\, x\}(\boldsymbol{t}) \triangleq x\big(\boldsymbol{P}_{\Theta^{\perp}}\, \boldsymbol{t}\big). \tag{4.77}$$

□

Next we will develop a corollary to Theorem 4.14 that will show that the Wiener filter (4.67) can be implemented using the standard filtered backprojection algorithm. First we will require the following lemma that is closely related to Theorem 1.3 in [56].

Lemma 4.15: Given functions f and g we have

$$\{\{H_{\boldsymbol{\theta}}^{*}\, g\} \star f\} = \{H_{\boldsymbol{\theta}}^{*}\{g \star \{H_{\boldsymbol{\theta}}\, f\}\}\}. \tag{4.78}$$

Proof. We have

$$\begin{aligned}\{\{H_{\boldsymbol{\theta}}\, g\} \star f\}(\boldsymbol{t}) &= \int_{\mathbb{R}^d} \{H_{\boldsymbol{\theta}}^{*}\, g\}(\boldsymbol{t}-\boldsymbol{\tau})\, f(\boldsymbol{\tau})\, d\boldsymbol{\tau} \\ &= \int_{\mathbb{R}^d} g\big[\boldsymbol{\theta},\, \boldsymbol{P}_{\Theta^{\perp}}\,(\boldsymbol{t}-\boldsymbol{\tau})\big]\, f(\boldsymbol{\tau})\, d\boldsymbol{\tau}\end{aligned}$$

where we have expressed the function g using generalized polar coordinates. Then, making the substitution $\boldsymbol{\tau} = \boldsymbol{s} + l\boldsymbol{\theta}$ with $\boldsymbol{s} \in \Theta^{\perp}$, we obtain

$$\begin{aligned}\{\{H_{\boldsymbol{\theta}}\, g\} \star f\}(\boldsymbol{t}) &= \int_{\Theta^{\perp}} g(\boldsymbol{\theta},\, \boldsymbol{P}_{\Theta^{\perp}}\, \boldsymbol{t} - \boldsymbol{s}) \int_{\mathbb{R}} f(\boldsymbol{s} + l\boldsymbol{\theta})\, dl\, d\boldsymbol{s} \\ &= \int_{\Theta^{\perp}} g(\boldsymbol{\theta},\, \boldsymbol{P}_{\Theta^{\perp}}\, \boldsymbol{t} - \boldsymbol{s})\, \{H_{\boldsymbol{\theta}}\, f\}(\boldsymbol{s})\, d\boldsymbol{s} \\ &= \{g \star \{H_{\boldsymbol{\theta}}\, f\}\}(\boldsymbol{P}_{\Theta^{\perp}}\, \boldsymbol{t}).\end{aligned}$$

□

Finally, the corollary below demonstrates that the KWF measurement update can be accomplished using the filtered backprojection algorithm.

Corollary 4.16: The impulse response of the Wiener filter for estimating a WSS

random field x_i given its noisy projection $y_{\boldsymbol{\theta}_i}$ is

$$k_i(\boldsymbol{t}) = \frac{1}{|\mathsf{S}^{n-2}|}\{\{H^*_{\boldsymbol{\theta}_i}\, w_i\} \star g\}(\boldsymbol{t}) \tag{4.79}$$

where g is the ramp filter with frequency response

$$G(\boldsymbol{f}) = |\boldsymbol{f}|,\ \boldsymbol{f} \in \mathbb{R}^d. \tag{4.80}$$

Proof. We can make use of Lemma 4.15 to reformulate Theorem 4.14 by first noting that the ramp filter r can be written as the projection of another function. In particular, consider the d-dimensional ramp filter (4.80) that, by Theorem 4.10, has the property

$$\{\mathcal{F}\{H_{\boldsymbol{\theta}_i}\, g\}\}(\boldsymbol{f}') = G(\boldsymbol{f}') = R(\boldsymbol{f}'),\ \boldsymbol{f}' \in \Theta_i^{\perp}, \tag{4.81}$$

which implies

$$\{H_{\boldsymbol{\theta}_i}\, g\}(\boldsymbol{s}) = r(\boldsymbol{s}),\ \boldsymbol{s} \in \Theta_i^{\perp}. \tag{4.82}$$

Then we have

$$\begin{aligned} k_i(\boldsymbol{t}) &= \frac{1}{|\mathsf{S}^{n-2}|}\{H^*_{\boldsymbol{\theta}_i}\{r \star w_i\}\}(\boldsymbol{t}) \\ &= \frac{1}{|\mathsf{S}^{n-2}|}\{H^*_{\boldsymbol{\theta}_i}\{\{H_{\boldsymbol{\theta}_i}\}\, g\} \star w_i\}(\boldsymbol{t}) \end{aligned} \tag{4.83}$$

which simplifies to (4.79) by Lemma 4.15. □

4.6 Smoothed Estimation

The remainder of this chapter focuses on the smoothing problem under the linear state-space model of Section 2.2.2. We consider only the Bryson-Frazier Kalman

smoother [27], [83] because it leads to a new ensemble Kalman smoothing algorithm discussed in the final subsection. Furthermore, we only consider the fixed-interval smoothing problem where the goal is to compute smoothed estimates $\widehat{\boldsymbol{x}}_{i:1:I}$ based on all available data. Methods for computing fixed-point and fixed-lag smoothed estimates [27] can also be derived using the techniques we now develop.

4.6.1 Bryson-Frazier Kalman smoother

The Bryson-Frazier Kalman smoother [27], [83] consists of three stages. The first stage processes the complete data set through the Kalman filter described in Section 4.3 and, for each time index i, stores the following quantities to disk for later use: the Kalman gain $\boldsymbol{K}_i$ (4.11), the innovation

$$\boldsymbol{e}_i \triangleq \boldsymbol{y}_i - \boldsymbol{H}_i \, \widehat{\boldsymbol{x}}_{i|i-1}, \tag{4.84}$$

and the innovation covariance

$$\boldsymbol{R}_{e,i} \triangleq \mathrm{Cov}(\boldsymbol{e}_i) = \boldsymbol{R}_i + \boldsymbol{H}_i \, \boldsymbol{P}_{i|i-1} \, \boldsymbol{H}_i^T \tag{4.85}$$

in addition to the filtered estimate $\widehat{\boldsymbol{x}}_{i|i}$ and error covariance $\boldsymbol{P}_{i|i}$. The second stage processes the time-reversed data set to compute the adjoint variable

$$\boldsymbol{\lambda}_{I+1} = \boldsymbol{0} \tag{4.86}$$

$$\boldsymbol{\lambda}_i = (\boldsymbol{I} - \boldsymbol{K}_i \, \boldsymbol{H}_i)^T \, \boldsymbol{F}_i^T \, \boldsymbol{\lambda}_{i+1} + \boldsymbol{H}_i^T \, \boldsymbol{R}_{e,\,i}^{-1} \, \boldsymbol{e}_i \tag{4.87}$$

and its covariance

$$\boldsymbol{\Lambda}_{I+1} = \boldsymbol{0} \tag{4.88}$$

$$\boldsymbol{\Lambda}_i = (\boldsymbol{I} - \boldsymbol{K}_i \boldsymbol{H}_i)^T \boldsymbol{F}_i^T \boldsymbol{\Lambda}_{i+1} \boldsymbol{F}_i (\boldsymbol{I} - \boldsymbol{K}_i \boldsymbol{H}_i) + \boldsymbol{H}_i^T \boldsymbol{R}_{e,i}^{-1} \boldsymbol{H}_i \tag{4.89}$$

through the above backwards recursions. We note that

$$\widehat{\boldsymbol{u}}_{i|1:I} = \boldsymbol{Q}_i \boldsymbol{\lambda}_{i+1} \tag{4.90}$$

and, as a consequence, the adjoint variable is related to the smoothed estimate of the state noise [27]. The final stage combines quantities found in the first two stages to compute the smoothed estimate and error covariance:

$$\widehat{\boldsymbol{x}}_{i|1:I} = \widehat{\boldsymbol{x}}_{i|i} + \boldsymbol{P}_{i|i} \boldsymbol{F}_i^T \boldsymbol{\lambda}_{i+1} \tag{4.91}$$

$$\boldsymbol{P}_{i|1:I} = \boldsymbol{P}_{i|i} - \boldsymbol{P}_{i|i} \boldsymbol{F}_i^T \boldsymbol{\Lambda}_{i+1} \boldsymbol{F}_i \boldsymbol{P}_{i|i}. \tag{4.92}$$

4.6.2 Ensemble Kalman smoother

The Bryson-Frazier smoother is computationally intractable for high-dimensional state estimation problems because its first stage is the Kalman filter. In the following, we develop an ensemble Kalman smoother (EnKS) based on the Bryson-Frazier approach that utilizes covariance tapering and a Monte Carlo methodology to ease the computational burden in a manner similar to the EnKF of Section 4.4. Analogous to the development of the Kalman smoother, methods for ensemble Kalman smoothing were developed not long after the publication of the EnKF. The first was by van Leeuwen and Evensen [84], but that method is limited to relatively low-dimensional problems because it does not utilize computational simplifications like covariance tapering. Evensen and van Leeuwen developed another approach to smoothing geared to nonlinear problems in [85], but, again, the method is not suited to high-dimensional problems. Finally, Khare et al. [86] consider larger problems and

investigate the implications of covariance tapering on ensemble smoothing. However, they do not consider the convergence properties of their approach and neglect the smoothed error covariance.

The EnKS can be thought of as a Monte Carlo approximation to the Bryson-Frazier Kalman smoother that utilizes covariance tapering. The first stage of the EnKS is the EnKF detailed in Section 4.4. At each time index i, the following quantities are stored to disk for later use: the filtered estimate $\widetilde{\boldsymbol{x}}_{i|i}$, the ensemble Kalman gain $\widetilde{\boldsymbol{K}}_i$ (4.19), the ensemble innovation

$$\widetilde{\boldsymbol{e}}_i \triangleq \boldsymbol{y}_i - \boldsymbol{H}_i \widetilde{\boldsymbol{x}}_{i|i-1}, \tag{4.93}$$

the ensemble innovation covariance

$$\widetilde{\boldsymbol{R}}_{e\,i} \triangleq \boldsymbol{R}_i + \boldsymbol{H}_i \left(\boldsymbol{C}_i \circ \widetilde{\boldsymbol{P}}_{i|i-1} \right) \boldsymbol{H}_i^T, \tag{4.94}$$

and the prior and posterior ensembles denoted $\widetilde{\boldsymbol{X}}_{i|i-1}$ and $\widetilde{\boldsymbol{X}}_{i|i}$, respectively, where the lth column of the matrix $\widetilde{\boldsymbol{X}}_{i|j}$ is given by

$$\left[\widetilde{\boldsymbol{X}}_{i|j} \right]_{(:,\,l)} = \widetilde{\boldsymbol{x}}_{i|j}^{l}. \tag{4.95}$$

The second stage of the EnKS involves the following backwards recursion on the ensemble adjoint variable:

$$\widetilde{\boldsymbol{\lambda}}_{I+1} = \boldsymbol{0} \tag{4.96}$$

$$\widetilde{\boldsymbol{\lambda}}_i = \left(\boldsymbol{I} - \widetilde{\boldsymbol{K}}_i \boldsymbol{H}_i \right)^T \boldsymbol{F}_i^T \widetilde{\boldsymbol{\lambda}}_{i+1} + \boldsymbol{H}_i^T \widetilde{\boldsymbol{R}}_{e,\,i}^{-1} \widetilde{\boldsymbol{e}}_i \tag{4.97}$$

and its covariance

$$\widetilde{\boldsymbol{\Lambda}}_{I+1} = \mathbf{0} \tag{4.98}$$

$$\widetilde{\boldsymbol{\Lambda}}_i = \left(\boldsymbol{I} - \widetilde{\boldsymbol{K}}_i \boldsymbol{H}_i\right)^T \boldsymbol{F}_i^T \widetilde{\boldsymbol{\Lambda}}_{i+1} + \widetilde{\boldsymbol{R}}_{e,\,i}^{-1/2} \boldsymbol{Z}_i \tag{4.99}$$

where $[\boldsymbol{Z}_i]_{(:,\,l)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0},\, \boldsymbol{I})$. It is easily shown that

$$\boldsymbol{\Lambda}_i \approx \frac{\widetilde{\boldsymbol{\Lambda}}_i \widetilde{\boldsymbol{\Lambda}}_i^T}{L-1} \tag{4.100}$$

and the approximation is within the sample error when no covariance tapering is applied, i.e., when $\boldsymbol{C}_i = \mathbf{1}$.

The presence of a square root of the ensemble innovation covariance $\widetilde{\boldsymbol{R}}_{e,\,i}^{1/2}$ in (4.99) poses a significant implementation challenge especially when covariance tapering is involved. Any hope to find a square root factor to (4.94) rests on factoring the prior tapered ensemble covariance. To find this square root factor, first note that

$$\boldsymbol{C}_i \circ \widetilde{\boldsymbol{P}}_{i|i-1} = \left(\boldsymbol{C}_i^{1/2}\, \boldsymbol{C}_i^{T/2}\right) \circ \left(\widetilde{\boldsymbol{P}}_{i|i-1}^{1/2}\, \widetilde{\boldsymbol{P}}_{i|i-1}^{T/2}\right) \tag{4.101}$$

where $\boldsymbol{A}^{T/2} \triangleq \left(\boldsymbol{A}^{1/2}\right)^T$, $\boldsymbol{C}_i^{T/2}$ is a convolution matrix used in the construction of the covariance taper matrix as mentioned in Section 4.4.1, and

$$\widetilde{\boldsymbol{P}}_{i|i-1}^{1/2} = \frac{\widetilde{\boldsymbol{X}}_{i|i-1}}{\sqrt{L-1}}. \tag{4.102}$$

Square root factors involving Hadamard products such as (4.101) have been addressed in the literature on spatial temporal adaptive processing in radar applications [87], [88]. The following lemma addresses the relevant square root factorization.

Lemma 4.17: Let $\boldsymbol{A}$ be an $M \times N$ matrix and $\boldsymbol{B}$ be an $M \times P$ matrix. Then

$$\big(\boldsymbol{A}\,\boldsymbol{A}^T\big) \circ \big(\boldsymbol{B}\,\boldsymbol{B}^T\big) = \big(\boldsymbol{A}^T \odot \boldsymbol{B}^T\big)^T \big(\boldsymbol{A}^T \odot \boldsymbol{B}^T\big) \tag{4.103}$$

where $\odot$ denotes the Khatri-Rao matrix product defined in Appendix C.

Proof. See the proof to P 6.4.2 in [89]. □

The square root factorization of (4.94) also requires the following lemma which is the basis for numerically stable square root form implementations of the Kalman filter [27].

Lemma 4.18: Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $M \times N$ ($M \leq N$) matrices. Then $\boldsymbol{A}\,\boldsymbol{A}^T = \boldsymbol{B}\,\boldsymbol{B}^T$ if, and only if, there exists an $N \times N$ unitary matrix $\boldsymbol{\Theta}$ such that $\boldsymbol{A} = \boldsymbol{B}\,\boldsymbol{\Theta}$.

Proof. See the proof to Lemma A.5.1 in [27]. □

Finally, the square root factorization of the ensemble innovations covariance (4.94) required in (4.99) is summarized in the following theorem.

Theorem 4.19: Let $\boldsymbol{\mathcal{Q}}_i$ and $\boldsymbol{\mathcal{R}}_i$ be the unitary and the upper triangular matrices found in the QR decomposition [72] of the matrix

$$\boldsymbol{A}_i = \Big[\boldsymbol{H}_i\big(\boldsymbol{C}_i^{T/2} \odot \widetilde{\boldsymbol{P}}_{i|i-1}^{T/2}\big)^T \quad \boldsymbol{R}_i^{1/2}\Big]^T. \tag{4.104}$$

Then,

$$\boldsymbol{R}_{e,\,i}^{1/2} = [\boldsymbol{\mathcal{R}}_i]_{(1:M,\,:)}^T \tag{4.105}$$

where $[\cdot]_{(1:M,\,:)}$ selects the first M rows of its matrix argument.

Proof. First note that

$$\boldsymbol{A}_i^T \boldsymbol{A}_i = \left[\widetilde{\boldsymbol{R}}_{e,i}^{1/2} \quad \mathbf{0}\right] \left[\widetilde{\boldsymbol{R}}_{e,i}^{1/2} \quad \mathbf{0}\right]^T \tag{4.106}$$

by Lemma 4.17. Then, by Lemma 4.18, (4.106) implies that there exists a unitary matrix $\boldsymbol{\Theta}_i$ such that

$$\boldsymbol{A}_i^T = \left[\widetilde{\boldsymbol{R}}_{e,i}^{1/2} \quad \mathbf{0}\right] \boldsymbol{\Theta}_i. \tag{4.107}$$

Note that the QR decomposition provides the following factorization:

$$\boldsymbol{A}_i = \boldsymbol{\mathcal{Q}_i} \boldsymbol{\mathcal{R}_i}. \tag{4.108}$$

Finally, note that

$$\boldsymbol{A}_i^T = \boldsymbol{\mathcal{R}_i}^T \boldsymbol{\mathcal{Q}_i}^T \tag{4.109}$$

and $\boldsymbol{\mathcal{Q}_i}^T$ is a unitary matrix. □

The third and final stage of the EnKS computes the ensemble approximation to the smoothed estimate

$$\widetilde{\boldsymbol{x}}_{i|1:I} = \widetilde{\boldsymbol{x}}_{i|i} + \left(\boldsymbol{C}_i' \circ \widetilde{\boldsymbol{P}}_{i|i}\right) \boldsymbol{F}_i^T \widetilde{\boldsymbol{\lambda}}_{i+1} \tag{4.110}$$

and the smoothed error covariance

$$\widetilde{\boldsymbol{P}}_{i|1:I} = \widetilde{\boldsymbol{P}}_{i|i} - \left(\boldsymbol{C}_i' \circ \widetilde{\boldsymbol{P}}_{i|i}\right) \boldsymbol{F}_i^T \left(\boldsymbol{C}_i'' \circ \widetilde{\boldsymbol{\Lambda}}_{i+1}\right) \boldsymbol{F}_i \left(\boldsymbol{C}_i' \circ \widetilde{\boldsymbol{P}}_{i|i}\right) \tag{4.111}$$

where $\boldsymbol{C}_i'$ and $\boldsymbol{C}_i''$ are two additional application dependent covariance taper matrices.

4.6.3 Ensemble Kalman smoother convergence

The following theorem addresses the convergence of the EnKS as the ensemble size L increases. This theorem concerning the asymptotic convergence of the EnKS is important because it demonstrates that the approach converges to a well defined limit which we call the localized Kalman smoother (LKS), shows that the EnKS without covariance tapering is a Monte Carlo Bryson-Frazier smoother, and provides a means for investigating the implications of the covariance taper.

Theorem 4.20: For each time index i, the EnKS estimates $\widetilde{\boldsymbol{x}}_{i|1:I}$ converge in probability to the LKS estimates $\widetilde{\boldsymbol{x}}^{\infty}_{i|1:I}$ defined below, i.e.,

$$\widetilde{\boldsymbol{x}}_{i|1:I} \xrightarrow{\text{p.}} \widetilde{\boldsymbol{x}}^{\infty}_{i|1:I} \tag{4.112}$$

in the limit as the ensemble size $L \to \infty$.

Proof. The proof proceeds in a manner similar to the EnKF convergence proof in Appendix A. Again, Slutsky's theorem plays an important role since the ensemble members in the EnKS are identically distributed but dependent. The proof proceeds by induction, showing that the terms in each stage converge in probability to the corresponding terms in the LKS. The details are omitted because of the close similarity to the proof in Appendix A. □

Like the Bryson-Frazier smoother, the LKS has three stages, the first being the LKF defined by (4.24)-(4.28). The second stage is the backwards recursion for the adjoint variable

$$\widetilde{\boldsymbol{\lambda}}^{\infty}_{I+1} = \boldsymbol{0} \tag{4.113}$$

$$\widetilde{\boldsymbol{\lambda}}^{\infty}_{i} = \left(\boldsymbol{I} - \widetilde{\boldsymbol{K}}^{\infty}_{i}\, \boldsymbol{H}_i\right)^T \boldsymbol{F}_i^T\, \widetilde{\boldsymbol{\lambda}}^{\infty}_{i+1} + \boldsymbol{H}_i^T \left(\widetilde{\boldsymbol{R}}^{\infty}_{e,\,i}\right)^{-1} \widetilde{\boldsymbol{e}}^{\infty}_{i} \tag{4.114}$$

where $\widetilde{\boldsymbol{K}}_i^{\infty}$ is the localized Kalman gain (4.24) and

$$\widetilde{\boldsymbol{R}}_{e,\,i}^{\infty} \triangleq \boldsymbol{R}_i + \boldsymbol{H}_i\, \widetilde{\boldsymbol{P}}_{i|i-1}^{\infty}\, \boldsymbol{H}_i^T \tag{4.115}$$

$$\widetilde{\boldsymbol{e}}_i^{\infty} \triangleq \boldsymbol{y}_i - \boldsymbol{H}_i\, \widetilde{\boldsymbol{x}}_{i|i-1}^{\infty}. \tag{4.116}$$

The backwards recursion for the covariance of the adjoint variable is given by

$$\widetilde{\boldsymbol{\Lambda}}_{I+1}^{\infty} = \boldsymbol{0} \tag{4.117}$$

$$\widetilde{\boldsymbol{\Lambda}}_i^{\infty} = \left(\boldsymbol{I} - \widetilde{\boldsymbol{K}}_i^{\infty}\, \boldsymbol{H}_i\right)^T \boldsymbol{F}_i^T\, \widetilde{\boldsymbol{\Lambda}}_{i+1}^{\infty}\, \boldsymbol{F}_i \left(\boldsymbol{I} - \widetilde{\boldsymbol{K}}_i^{\infty}\, \boldsymbol{H}_i\right) + \boldsymbol{H}_i^T \left(\widetilde{\boldsymbol{R}}_{e,\,i}^{\infty}\right)^{-1} \boldsymbol{H}_i. \tag{4.118}$$

The final stage combines information in the first two stages to compute the localized smoothed estimate and error covariance with

$$\widetilde{\boldsymbol{x}}_{i|1:I}^{\infty} = \widetilde{\boldsymbol{x}}_{i|i}^{\infty} + \left(\boldsymbol{C}_i' \circ \widetilde{\boldsymbol{P}}_{i|i}^{\infty}\right) \boldsymbol{F}_i^T\, \widetilde{\boldsymbol{\lambda}}_{i+1|1:I}^{\infty} \tag{4.119}$$

$$\widetilde{\boldsymbol{P}}_{i|1:I}^{\infty} = \widetilde{\boldsymbol{P}}_{i|i}^{\infty} - \left(\boldsymbol{C}_i' \circ \widetilde{\boldsymbol{P}}_{i|i}^{\infty}\right) \boldsymbol{F}_i^T \left(\boldsymbol{C}_i'' \circ \widetilde{\boldsymbol{\Lambda}}_{i+1}^{\infty}\right) \boldsymbol{F}_i \left(\boldsymbol{C}_i' \circ \widetilde{\boldsymbol{P}}_{i|i}^{\infty}\right). \tag{4.120}$$

The following corollary to Theorem 4.20 addresses the convergence of the EnKS when no covariance tapering is applied.

Corollary 4.21: For each time index i, the unlocalized ($\boldsymbol{C}_i = \boldsymbol{C}_i' = \boldsymbol{C}_i'' = \boldsymbol{1}$) EnKS estimates $\widetilde{\boldsymbol{x}}_{i|1:I}$ converge in probability to the LMMSE smoothed estimates $\widehat{\boldsymbol{x}}_{i|1:I}$, i.e.,

$$\widetilde{\boldsymbol{x}}_{i|1:I} \xrightarrow{\text{p.}} \widehat{\boldsymbol{x}}_{i|1:I} \tag{4.121}$$

in the limit as the ensemble size $L \to \infty$.

Proof. When no covariance tapering is applied, the LKS and Bryson-Frazier Kalman smoother of Section 4.6.1 are equivalent. Then, apply the result of Theorem 4.20. □

CHAPTER 5

NUMERICAL EXPERIMENTS

The following set of experiments evaluates the performance of the state estimation algorithms developed in Chapter 4. The first experiment compares the PF, KF, and EnKF in a relatively low-dimensional dynamic tomography scenario. The results corroborate the computational intractability of the PF and the tremendous computational savings offered by the EnKF over the KF. The second experiment focuses on the random walk signal model and its applicability to a dynamic tomography problem with known state dynamics. The next experiment shows that the KF and KWF estimates are essentially equal and that the computational complexity of the KWF is the same as filtered back-projection. The final two experiments consider the EnKS in a 1-D example and demonstrate the convergence of the EnKS to the LKS.

5.1 Low-Dimensional Example

The purpose of the first numerical experiment is to demonstrate the performance of the EnKF in reconstructing a dynamic 2-D object from tomographic measurements when no deterministic model for the state process dynamics is given ($\boldsymbol{F}_i = \boldsymbol{I}$). The purely stochastic model is commonly used when little is known of the system dynamics [44]. The dynamic object $\boldsymbol{x}_i$ at each time index i is a vector representation of an image where each element of $\boldsymbol{x}_i$ corresponds to one image pixel. Each image is a sample of a

random process that consists of the superposition of 256 randomly moving disks in a 2-D field. This dynamic process was chosen because it can be sampled at any spatial and temporal resolution and, like many dynamic tomography problems that arise in remote sensing and other fields, the dynamics of the process are complex and not easily modeled. Two versions of this random process are studied in the experiment. The first is at low resolution with a spatial resolution of 32×32 pixels ($N = 32^2$) and a temporal resolution of 256 frames in a period T, the duration of the random process measured in seconds. The second is at high resolution with 128×128 pixels ($N = 128^2$) and 1024 frames in the same period T. Five images of the process at low resolution are given in Figure 5.1 (left) and the first row of Figure 5.2. The same five images at high resolution are given in Figure 5.3 (left) and the first row of Figure 5.4.

Each measurement $\boldsymbol{y}_i$ consists of M uniformly spaced parallel line integrals of the unknown object $\boldsymbol{x}_i$ at a projection angle θ_i with $M = 46$ and $M = 184$ in the low and high resolution cases. The projection angle θ_i sweeps uniformly through 360° a total of 4 times in the simulation time T in both cases. White Gaussian noise has been added to each measurement resulting in a 40 dB signal-to-noise ratio (SNR) defined

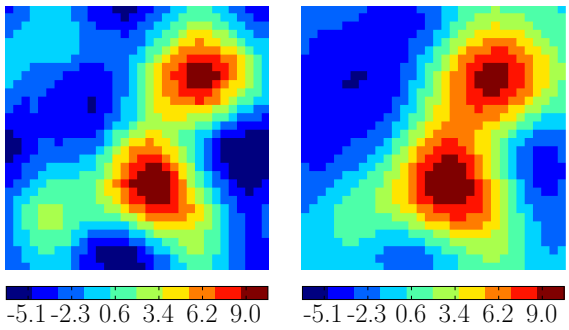

Figure 5.1 The initial state $\boldsymbol{x}_1$ (left) and the initial prior estimate $\boldsymbol{\mu}_1$ (right) given by the back-projected estimate (5.2) at low resolution. The error (5.3) between $\boldsymbol{x}_1$ and $\boldsymbol{\mu}_1$ is 0.49.

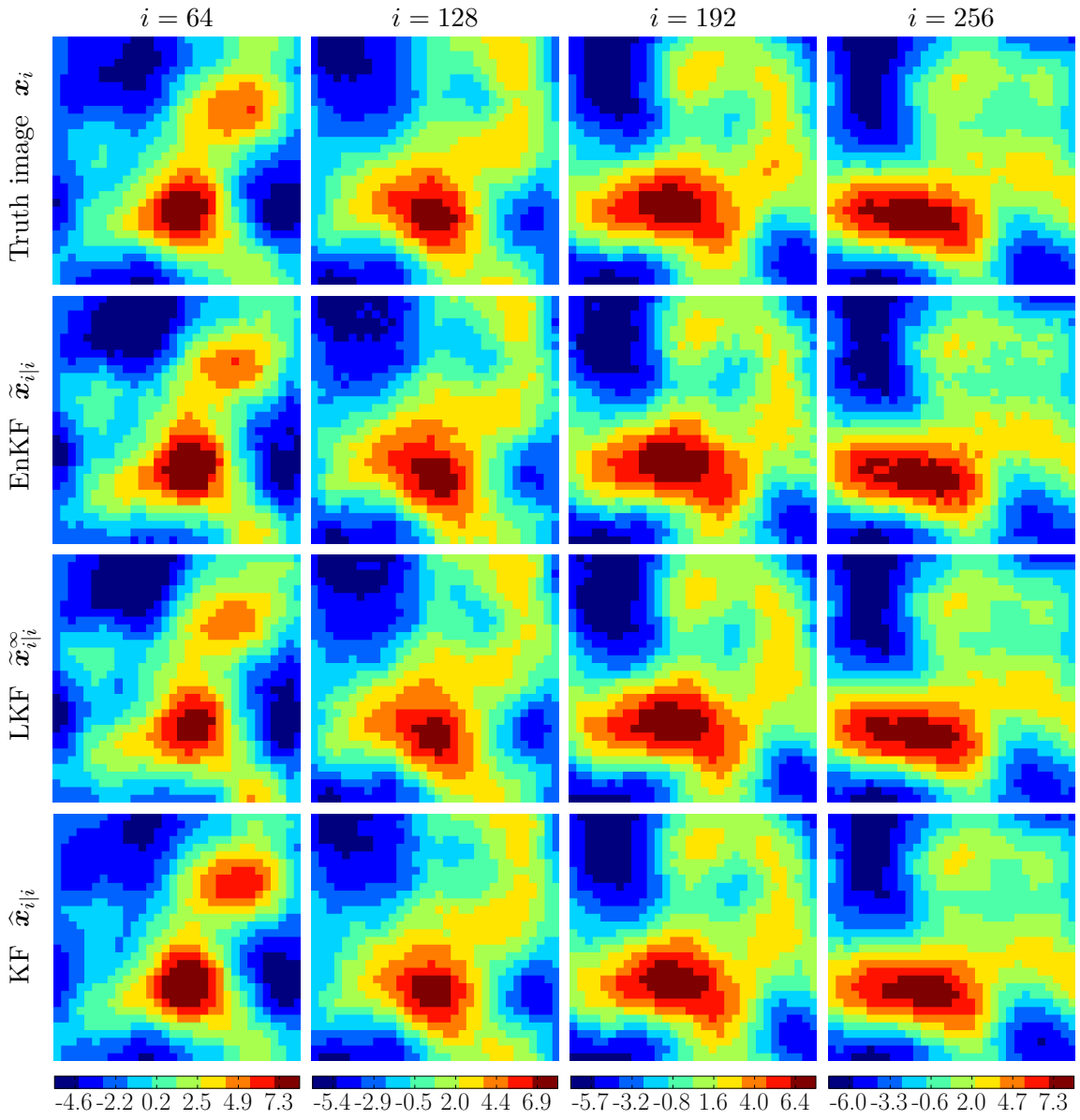

Figure 5.2 The results of the Section 5.1 numerical experiment at low resolution where the vector $\boldsymbol{x}_i$ represents a 32×32 image, each vector $\boldsymbol{y}_i$ contains 46 parallel line integral measurements, and the final time index is $I = 256$. Each column shows the true image and the EnKF, LKF, and KF estimates of that image at a given time index i on the same color scale.

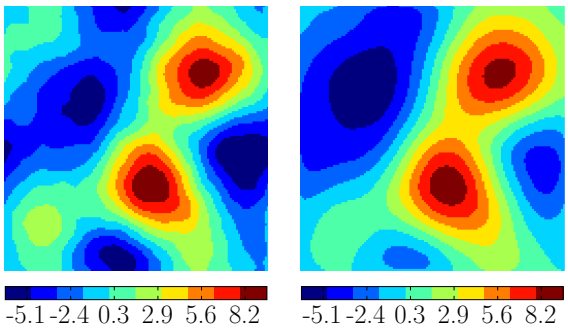

Figure 5.3 The initial state $\boldsymbol{x}_1$ (left) and the initial prior estimate $\boldsymbol{\mu}_1$ (right) given by the backprojected estimate (5.2) at high resolution. The error (5.3) between $\boldsymbol{x}_1$ and $\boldsymbol{\mu}_1$ is 0.45.

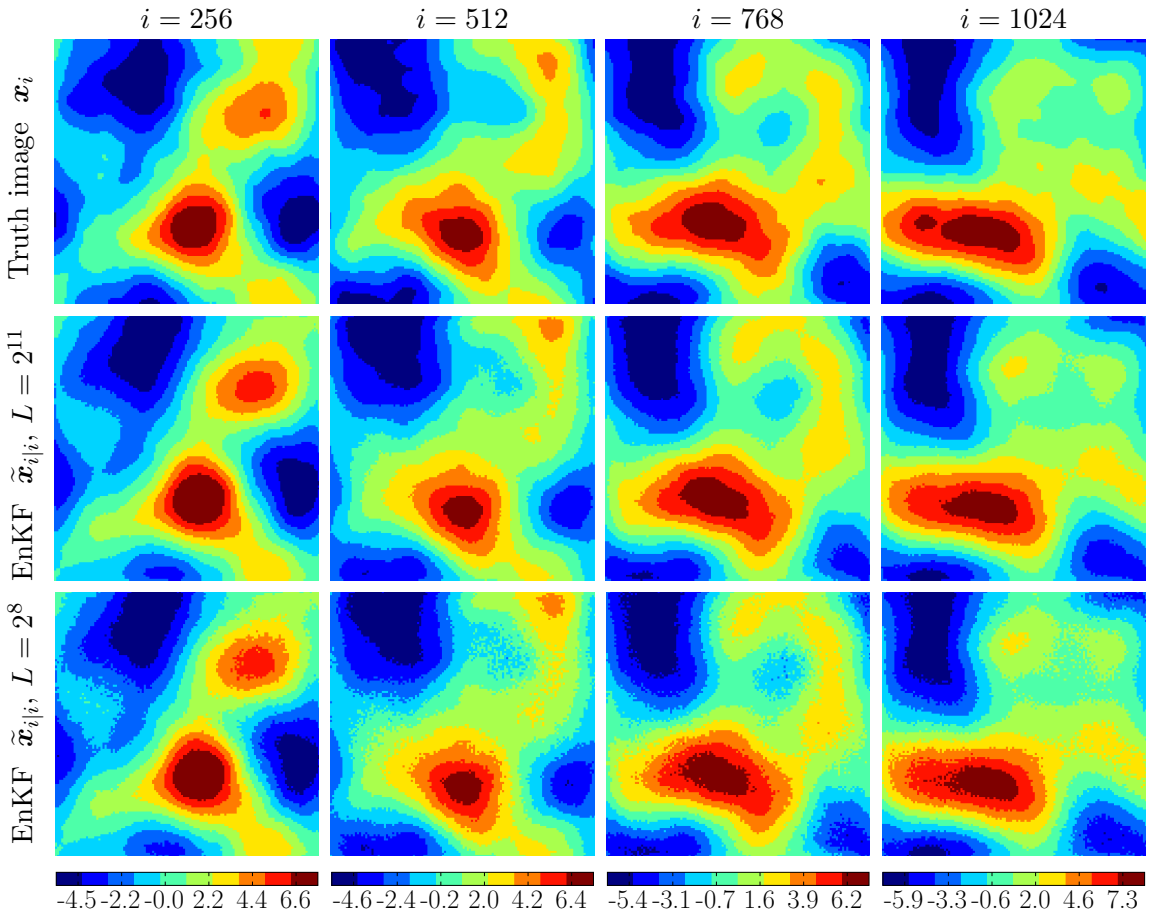

Figure 5.4 The results of the Section 5.1 numerical experiment at high resolution where the vector $\boldsymbol{x}_i$ represents a 128×128 image, each vector $\boldsymbol{y}_i$ contains 184 parallel line integral measurements, and the final time index is $I = 1024$. Each column shows the true image and two EnKF estimates with ensemble sizes $L = 2048$ and $L = 256$ at time index i on the same color scale.

as

$$\text{SNR} \triangleq 10 \log_{10} \frac{\|\boldsymbol{s}\|_2^2}{\|\boldsymbol{v}\|_2^2} \tag{5.1}$$

where

$$\boldsymbol{s} \triangleq \left[(\boldsymbol{H}_1 \boldsymbol{x}_1)^T, \cdots, (\boldsymbol{H}_I \boldsymbol{x}_I)^T \right]^T \qquad \text{and} \qquad \boldsymbol{v} \triangleq \left(\boldsymbol{v}_1^T, \cdots, \boldsymbol{v}_I^T \right)^T$$

and I is the final time index. On average, $\|\boldsymbol{v}\|_2^2 = \sigma_v^2 MI$, where σ_v^2 denotes the measurement noise variance. White measurement noise enables the use of sequential processing and related computational benefits as discussed in Appendix B.1. The low and high resolution noisy measurements are shown in Figures 5.5 and 5.6.

The remaining state-space model parameters are detailed below. Because no simple model describes the dynamics of the random process $\boldsymbol{x}_i$, the state transition matrix is set to $\boldsymbol{F}_i = \boldsymbol{I}$ and the state evolution is modeled as a random walk. The measurement operator $\boldsymbol{H}_i$ is a matrix representation of the Radon transform [56], [90] at projection angle θ_i. Both the state noise covariance $\boldsymbol{Q}_i$ and initial prior estimate error covariance $\boldsymbol{\Pi}_i$ are of a form similar to the taper matrix $\boldsymbol{C}_i$ depicted in Figure 4.1 (left) with correlation decreasing as a function of the spatial distance between pixels and no correlation above a distance of 2 pixels. The measurement noise covariance is $\boldsymbol{R}_i = \sigma_v^2 \boldsymbol{I}$, where σ_v^2 is the measurement noise variance. The prior mean

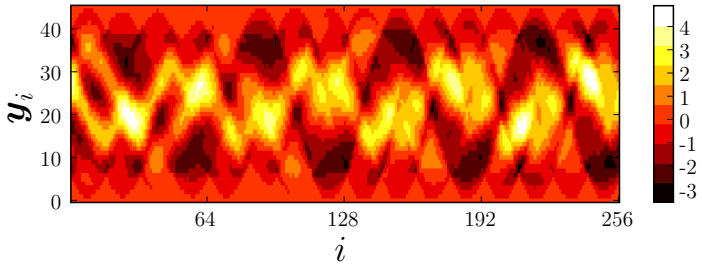

Figure 5.5 The 40 dB SNR low resolution tomographic measurements $\boldsymbol{y}_i$ with SNR defined in (5.1).

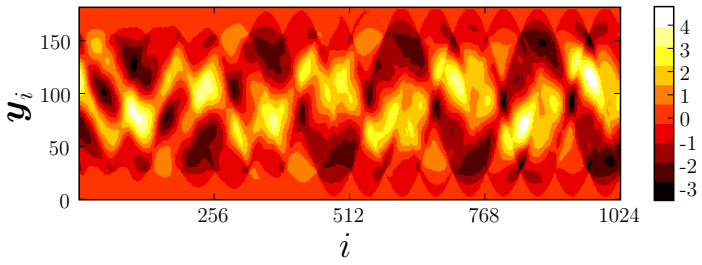

Figure 5.6 The 40 dB SNR high resolution tomographic measurements $\boldsymbol{y}_i$ with SNR defined in (5.1).

of the initial state, the initial guess, is the so-called backprojected image

$$\boldsymbol{\mu}_1 = \frac{1}{\alpha}\, \boldsymbol{H}_{1:J}^T\, \boldsymbol{y}_{1:J} \tag{5.2}$$

where α is a manually tuned scaling factor,

$$\boldsymbol{H}_{1:J} \triangleq \left(\boldsymbol{H}_1^T, \cdots, \boldsymbol{H}_J^T\right)^T \qquad \text{and} \qquad \boldsymbol{y}_{1:J} \triangleq \left(\boldsymbol{y}_1^T, \cdots, \boldsymbol{y}_J^T\right)^T,$$

and $J+1$ is the time index that θ_i is first greater than 180°. The initial state $\boldsymbol{x}_1$ and the backprojected estimate $\boldsymbol{\mu}_1$ are shown in Figures 5.1 and 5.3 for the low and high resolution cases. All estimates computed in this experiment are regularized by the method of augmenting the forward model (2.10) as discussed in Appendix B.2. The regularization matrix is $\boldsymbol{D}_i = (\boldsymbol{D}_x^T, \boldsymbol{D}_y^T)^T$ where $\boldsymbol{D}_x$ and $\boldsymbol{D}_y$ are first-order difference approximations to the spatial derivative operators in the horizontal and vertical directions. The regularization parameter λ_i and the scaling of all the model parameters were manually tuned to produce the best possible estimates. Methods such as cross validation [91], [92] exist to automatically and systematically select such parameters based on available information.

The low resolution experimental results are given in Figure 5.2. The first row shows the dynamic object $\boldsymbol{x}_i$ at four uniformly spaced time indices. The next three

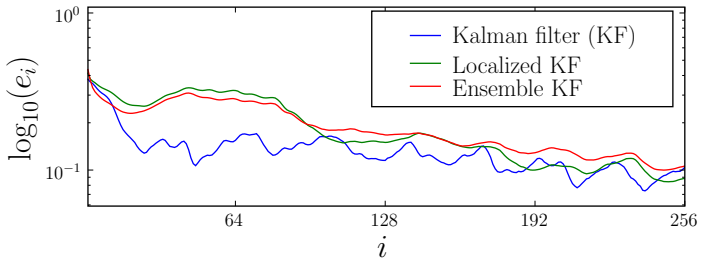

Figure 5.7 A log-lin comparison of the error (5.3) for the results in Figure 5.2.

Table 5.1 Computation run-time (Time) and memory (Mem.).

	Low resolution			High resolution		
Method	L	Time	Mem. (MB)	L	Time	Mem. (MB)
KF/LKF	N/A	39.7 m	2	N/A	(576 d)	512
EnKF	128	26 s	0.5	2048	11.2 h	128
				256	6.0 h	16

rows show the corresponding EnKF, LKF, and KF estimates $\widetilde{\boldsymbol{x}}_{i|i}$, $\widetilde{\boldsymbol{x}}^{\infty}_{i|i}$, and $\widehat{\boldsymbol{x}}_{i|i}$, respectively. The EnKF taper matrix is the extreme case $\boldsymbol{C}_i = \boldsymbol{I}$ in both the low and high resolution scenarios. Figure 5.7 shows the relative error of the three estimation methods defined as

$$e_i \triangleq \frac{\|\boldsymbol{x}_i - \boldsymbol{z}_i\|_2}{\|\boldsymbol{x}_i\|_2} \tag{5.3}$$

with $\boldsymbol{z}_i$ equal to $\boldsymbol{\mu}_1$, $\widetilde{\boldsymbol{x}}_{i|i}$, $\widetilde{\boldsymbol{x}}^{\infty}_{i|i}$, or $\widehat{\boldsymbol{x}}_{i|i}$, depending on context. The computational expense on a 2 GHz workstation is summarized in Table 5.1. Note that even in the low resolution case, the EnKF is far less expensive than the KF and LKF. Though qualitatively and quantitatively different, the EnKF estimates at ensemble size $L = 128$ are comparable in quality to the KF and LKF estimates. The LKF is the asymptotic limit of the EnKF and represents the upper bound on the average achievable quality. At $L = 128$, the EnKF begins to approach this upper estimate quality bound as evidenced by the similarity between the EnKF and LKF estimates in Figure 5.2 and the error in Figure 5.7. When compared to the KF estimates, the LKF estimates

show the effect of the bias introduced by covariance tapering without the additional effect of sample error present in the EnKF estimates and any other bias associated with a finite ensemble size. As shown in Figure 5.7, the effect of the covariance taper bias is apparent, but it does not cause the LKF estimate error to be far worse than the KF estimate error, especially at time indices $i \geq 96$. Note that it may seem surprising that the LKF outperforms the KF in Figure 5.7 at several time indices since the KF is the LMMSE optimal estimation procedure and the LKF is a linear filter. However, the KF is LMMSE optimal only over the average of all realizations of the measurement process and when the parameters assumed in the state-space signal model match the true system parameters. Section 5.2 further explores these issues when the system dynamics are known ($\boldsymbol{F}_i \neq \boldsymbol{I}$). Also, it may be possible to improve upon the relatively poor performance of the LKF at time indices $32 \leq i \leq 96$ by dynamically adjusting model parameters as changes are detected in the estimated system dynamics.

The numerical experiment results at high resolution are shown in Figure 5.4. At this resolution, the KF and LKF become computationally intractable as shown in Table 5.1, where the total computation time of the KF and LKF are listed parenthetically as it was extrapolated from the 13.5 h of processing necessary for the first measurement. Results with smaller ensemble sizes such as $L = 128$ are not shown because the visual quality of the EnKF estimates began to significantly degrade. The results in Figure 5.4 demonstrate that, for this example, only a modest increase in ensemble size is required to faithfully reconstruct the dynamic object even though the state dimension N has increased by a factor 16 relative to the low resolution experiment. Also, Figure 5.4 shows that increasing the ensemble size results in better estimates, but at the cost of greater computational effort.

Figure 5.8 (left) explores the rate of convergence of the EnKF with respect to

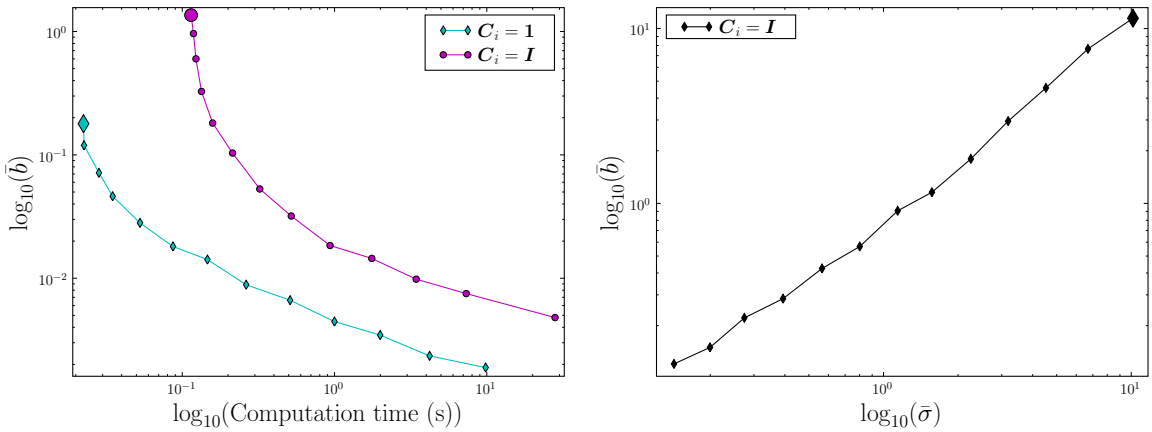

Figure 5.8 A log-log comparison (left) between the average sample bias $\overline{b}$ (5.4) and average computation time, and (right) $\overline{b}$ and average sample standard deviation $\overline{\sigma}$ (5.5). In both plots, the kth marker corresponds to ensemble size $L = 2^{k+1}$ and the first marker is filled (L ranges from 2^2–2^{14}).

ensemble size for an 8×8 pixel and 16 time step sampling of the same dynamic process as above. A total of 64 independent EnKF trial estimates were computed using the same measurements and model parameters. The dependent variable in Figure 5.8 (left) is the average sample bias

$$\overline{b} = \frac{1}{IN} \sum_{i=1}^{I} \sum_{n=1}^{N} \left| (\overline{\boldsymbol{\mu}}_{\widetilde{\boldsymbol{x}}_{i|i}} - \widetilde{\boldsymbol{x}}_{i|i}^{\infty})_n \right| \tag{5.4}$$

where $\overline{\boldsymbol{\mu}}_{\widetilde{\boldsymbol{x}}_{i|i}}$ is the sample mean of the 64 trial estimates and $(\cdot)_n$ denotes the nth component of the vector argument. For large enough ensemble size, the curve is linear with slope approximately $-1/2$ for both extreme ($\boldsymbol{C}_i = \boldsymbol{I}$) and no ($\boldsymbol{C}_i = \boldsymbol{1}$) covariance tapering. In other words, a quadratic increase in computation time is required to halve the bias error in $\widetilde{\boldsymbol{x}}_{i|i}$. Put another way, L^2 additional ensemble members are required to halve the error because computation time and ensemble size L are directly proportional. This rate of convergence is in agreement with other Monte Carlo integration methods [17]. Figure 5.8 (left) is also an empirical verification of Theorem 4.1 and Corollary 4.2. Both curves continue to decrease with slope $-1/2$ as

the ensemble size, and, in direct proportion, the computation time increase. The top curve demonstrates that $\widetilde{\boldsymbol{x}}_{i|i}$ converges to $\widehat{\boldsymbol{x}}_{i|i}$ when there is no localization ($\boldsymbol{C}_i = \mathbf{1}$) and the bottom demonstrates that $\widetilde{\boldsymbol{x}}_{i|i}$ converges to $\widetilde{\boldsymbol{x}}^{\infty}_{i|i}$ when there is localization ($\boldsymbol{C}_i \neq \mathbf{1}$).

Figure 5.8 (right) demonstrates an operational procedure for determining the quality of the EnKF estimates. The curve is based on the same scenario as Figure 5.8 (left) and compares the average sample bias (5.4) to the average sample standard deviation

$$\overline{\sigma} = \left[\frac{1}{I} \sum_{i=1}^{I} \operatorname{trace}\Big(\overline{\boldsymbol{\Sigma}}_{\widetilde{\boldsymbol{x}}_{i|i}} \Big) \right]^{1/2} \tag{5.5}$$

where $\operatorname{trace}(\cdot)$ is the matrix trace and $\overline{\boldsymbol{\Sigma}}_{\widetilde{\boldsymbol{x}}_{i|i}}$ is the sample covariance of the 64 trial estimates. The curve has a slope approximately 1, which implies (5.4) and (5.5) are directly proportional and small $\overline{\sigma}$ implies better estimate quality.

Finally, the optimal proposal PF of Section 4.2 was compared to the EnKF for a coarse 6×6 pixel ($N = 36$) and 8 time step sampling of the same dynamic process as above. The PF resampled [21] the particles whenever the ratio of the number of effective particles [21] to particles was less than half. In this example, the PF required 2048 particles to achieve the same estimate error as the EnKF with only 16 ensemble members when averaged over 64 trials. This example empirically demonstrates the inapplicability of the PF to even low-dimensional dynamic tomography.

5.2 Study of the Random Walk Dynamic Model

The goal of the second numerical experiment is to empirically evaluate the consequences of assuming a purely stochastic state transition model ($\boldsymbol{F}_i = \boldsymbol{I}$). We consider

a dynamic process that evolves according to the classic heat equation

$$\frac{\partial\, x_t(\boldsymbol{s})}{\partial\, t} = k\, \nabla_{\boldsymbol{s}}^2\, x_t(\boldsymbol{s}) \tag{5.6}$$

where $x_t(\boldsymbol{s})$ is the temperature at time t and spatial coordinate $\boldsymbol{s}$, k is the diffusion coefficient, and $\nabla_{\boldsymbol{s}}^2(\cdot)$ is the Laplace operator. Spatially and temporally discretizing (5.6) results in the linear model (2.9) where the state transition operator $\boldsymbol{F}_i$ is an $N \times N$ truncated convolution operator and the state noise $\boldsymbol{u}_i$ accounts for all approximation error. For this experiment, we consider a 16×16 pixel ($N = 16^2$) and 128 time step spatial-temporal discretization.

The measurements are $M = 23$ parallel line integrals of the dynamic object at an angle θ_i that sweeps uniformly through 360° a total of 4 times by the final time index. The remaining model parameters were similar to those in Section 5.1 but scaled for the smaller problem size and new system dynamics. Somewhat different model parameters were used in the two cases (random walk model and true model). Specifically, the regularization parameter λ, scaling of the state noise covariance $\boldsymbol{Q}_i$, and its correlation length were greater for the case of the random walk model ($\boldsymbol{F}_i = \boldsymbol{I}$). In addition, the random walk model ensemble size is double that used in the true model case. These increases are reasonable given the greater estimation uncertainties under the random walk model.

Figure 5.9 shows the EnKF, LKF, and KF estimates of the diffusive process reconstructed from line integral measurements when the state dynamics are specified ($\boldsymbol{F}_i \neq \boldsymbol{I}$) and Figure 5.10 shows the error. The results in Figure 5.11 are computed from the same data set assuming purely stochastic dynamics ($\boldsymbol{F}_i = \boldsymbol{I}$) and Figure 5.12 shows the error. The estimates found when the state dynamics are known are clearly superior to when the dynamics are ignored. This is not surprising; however, the

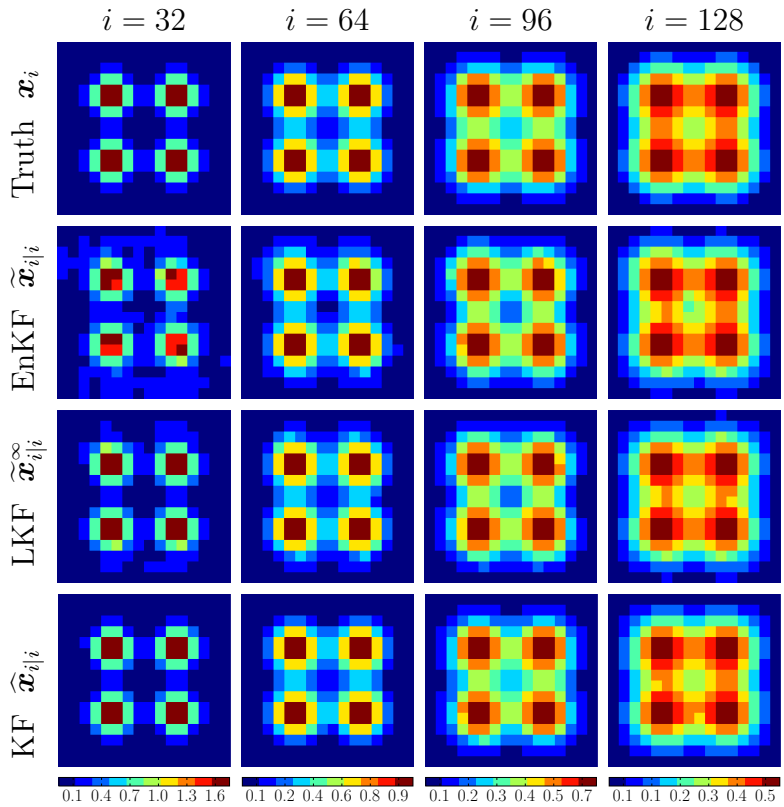

Figure 5.9 The results of the numerical experiment of Section 5.2 comparing the EnKF, LKF, and KF estimates when the system dynamics are given ($\boldsymbol{F}_i \neq \boldsymbol{I}$). The example is a 16×16 pixel tomography problem with 128 total time steps. The estimate error (5.3) is shown in Figure 5.10.

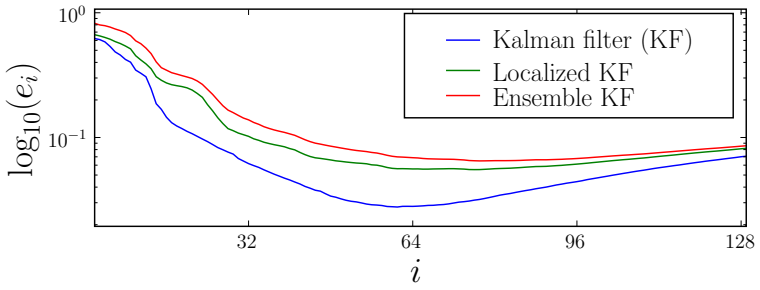

Figure 5.10 The log-lin comparison of the error (5.3) for the results in Figure 5.9.

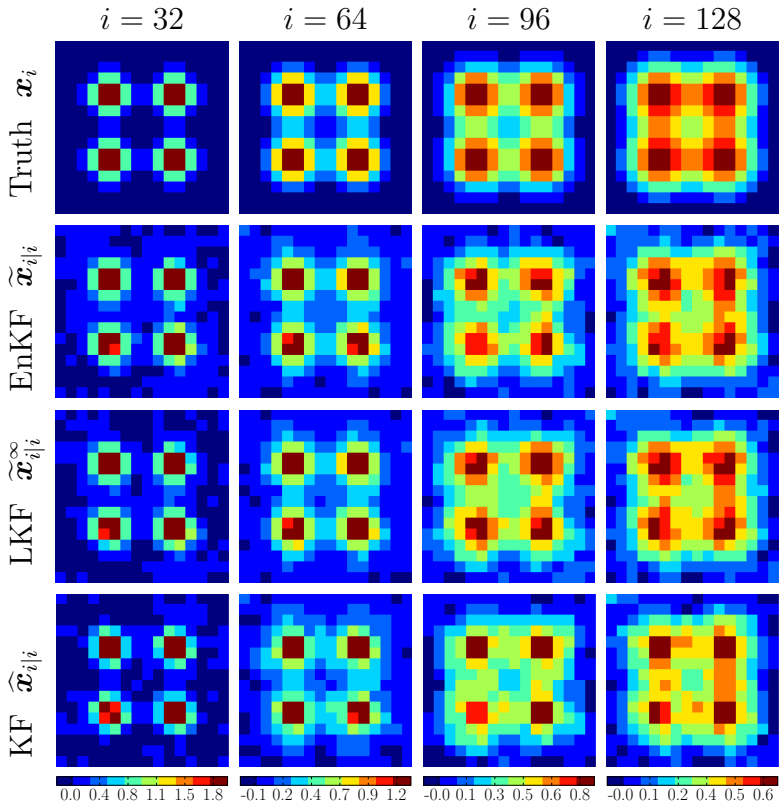

Figure 5.11 The results of the numerical experiment of Section 5.2 comparing the EnKF, LKF, and KF estimates when the system dynamics are given ($\boldsymbol{F}_i \neq \boldsymbol{I}$). The example is a 16×16 pixel tomography problem with 128 total time steps. The estimate error (5.3) is shown in Figure 5.10.

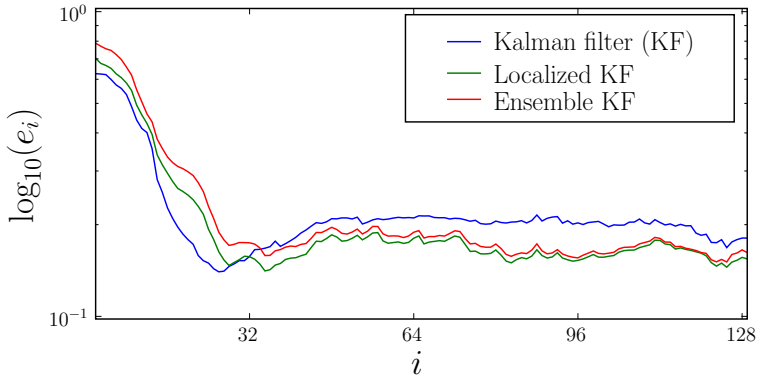

Figure 5.12 The log-lin comparison of the error (5.3) for the results in Figure 5.11.

estimates found when the dynamics are ignored do capture the main features and general morphology. The system dynamics may be unknown in practice and these results demonstrate that dynamic tomography may still be possible, though faithful models of the system dynamics will clearly improve the results. Also, note that the KF outperforms the LKF in Figure 5.10 at all time indices but not in Figure 5.12. This demonstrates that while the KF is LMMSE optimal, a significant mismatch between the state-space model and the true system parameters can result in estimates that are a poor reflection of the true state.

5.3 Kalman-Wiener Filter Examples

The following section details the results of the two experiments that test the KWF in simulation. The first experiment involves an unknown random process generated through LSI filtering of white Gaussian noise with measurements given by further LSI filtering. The experiment is designed to verify that the KWF and KF are nearly equivalent in this case, where the KWF is essentially the frequency domain version of the spatial domain KF. In addition, the $O(N \log N)$ computational scaling of the KWF is demonstrated against the $O(N^3)$ scaling of the KF. The second experiment examines the potential of the KWF in a 2-D dynamic tomography scenario. We demonstrate that the KWF is a filtered backprojection algorithm with a time varying filter.

5.3.1 Case 1: Convolutional measurements

The first numerical experiment involves a 1-D state random process ($d = 1$). The initial state is generated by LSI filtering Gaussian white noise. Each measurement is a noisy Gaussian blur of the state. The expressions for the cross-correlation and auto-correlation functions required for the posterior error power spectrum update (4.39)

are [63]

$$S_{\ddot{X}_{i|i-1},\ddot{Z}_{i|i-1}}(f) = H_i^*(f)\, S_{\ddot{X}_{i|i-1}}(f) \tag{5.7}$$

and

$$S_{\ddot{Z}_{i|i-1}}(f) = |H_i(f)|^2\, S_{\ddot{X}_{i|i-1}}(f). \tag{5.8}$$

The resultant WF in this case is

$$K_i(f) = \frac{H_i^*(f)\, S_{\ddot{X}_{i|i-1}}(f)}{|H_i(f)|^2\, S_{\ddot{X}_{i|i-1}}(f) + S_{V_i}(f)}. \tag{5.9}$$

At each time step, the state randomly circularly shifts one element to the left or right. For this numerical experiment, the KF and KWF state dynamic models are a random walk, i.e., $\boldsymbol{F}_i = \boldsymbol{I}$ where $\boldsymbol{I}$ is the identity matrix and $F_i(f) = 1$. The precise definition of the remaining experimental model parameters is not essential because the goal here is to compare the KWF to the KF and to investigate their scaling without regard to the estimates generated by the methods.

The plot in Figure 5.13 shows the absolute difference between the posterior KF estimate $\widehat{\boldsymbol{x}}_{i|i}$ and KWF estimate $\widehat{x}_{i|i}(\boldsymbol{r})$ at time index 128 for an $N = 127$ dimensional sampling of the above problem. To be precise, the vertical axis of Figure 5.13 shows

$$\left| \left[\widehat{\boldsymbol{x}}_{i|i}\right]_n - \widehat{x}_{i|i}(\boldsymbol{r}_n) \right| \tag{5.10}$$

where $i = 128$ and $\boldsymbol{r}_n$ is the nth component of the discretized 1-D spatial coordinate. Except at the edges (near $n = 1$ and $n = 127$), the KF and KWF estimates are almost equal. The differences at the edges result from the fact that the KF assumes the state has finite spatial support whereas the KWF state support is assumed infinite. These differences are not statistically significant for this problem. For example, the relative

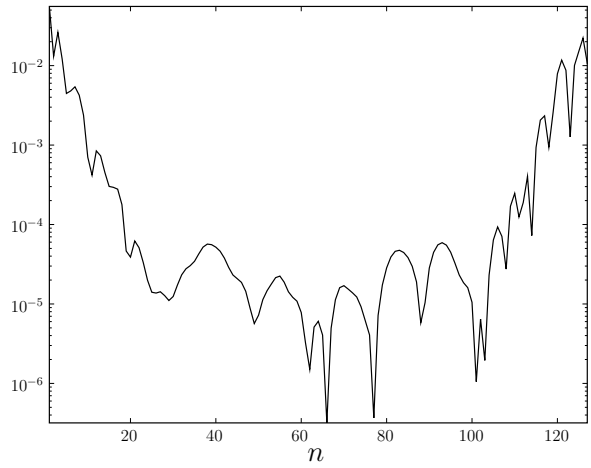

Figure 5.13 A log-lin plot of the absolute difference between KF and KWF estimates (5.10) at the final time index $i = 128$.

empirical mean square errors of the KF and KWF were 0.337 and 0.348, respectively, after averaging over 64 realizations of the problem. Therefore, statistically speaking, the KF and KWF give essentially equal results.

Figure 5.14 shows how the computational expense of the KF and KWF scale as a function of the problem dimension N. Note that the asymptotic slope of the KF curve in Figure 5.14 is 2.92 on the log-log plot and is 1.18 for the KWF. These asymptotic slopes provide empirical proof for the $O(N^3)$ and $O(N \log N)$ computational complexity of the KF and KWF. The KF and KWF also differ in storage complexity. Whereas the KF must store approximately $N^2/2$ elements of the symmetric error covariance matrix [50], the KWF need only store N samples of the error power spectral density function $S_{\ddot{X}_{i|j}}(\boldsymbol{f})$. The resultant storage complexity is therefore $O(N^2)$ for the KF and $O(N)$ for the KWF.

5.3.2 Case 2: 2-D projection measurements

The goal of the second numerical experiment is to reconstruct a randomly translating 2-D Shepp-Logan head phantom ($d = 2$) from noisy projection measurements. We

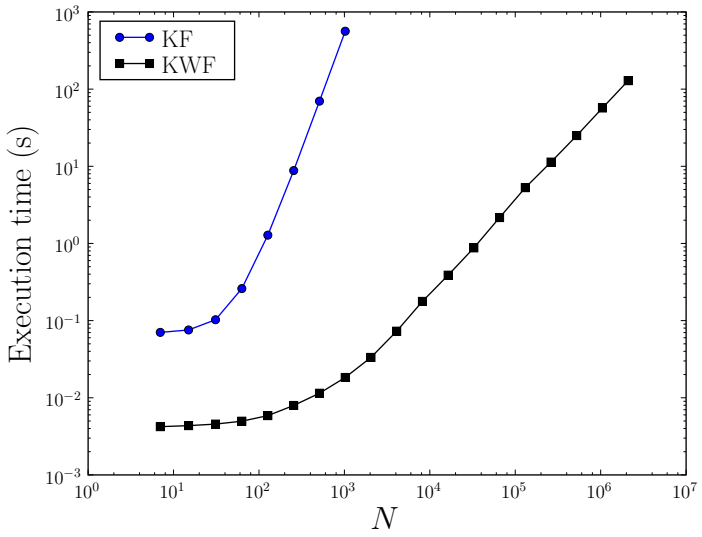

Figure 5.14 A log-log plot of execution time versus problem dimension N comparing the KF and KWF.

consider a 512×512 square pixel sampling of the head phantom state ($N = 512^2$). Each projection is sampled at $n_p = 725$ uniform points. White Gaussian noise is added to each projection at the 40 dB SNR level (5.1). The projections are sampled at 512 uniformly spaced angles over a 180° sweep about the head phantom. Between each projection measurement, the state randomly shifts 0, 1, or -1 pixels in the horizontal direction. Assuming that the images in Figure 5.15 have a physical dimension of 20 cm $\times$ 20 cm (the average human skull is about 20 cm in length), then a single pixel shift corresponds to about 0.4 mm of patient movement.

For this problem, the cross-correlation and auto-correlation functions for the posterior update (4.39) are

$$S_{\ddot{X}_{i|i-1},\ddot{Z}_{i|i-1}}(f) = S_{\ddot{X}_{i|i-1}}(f\cos\theta_i,\ f\sin\theta_i) \tag{5.11}$$

and

$$S_{\ddot{Z}_{i|i-1}}(f) = S_{\ddot{X}_{i|i-1}}(f\cos\theta_i,\ f\sin\theta_i). \tag{5.12}$$

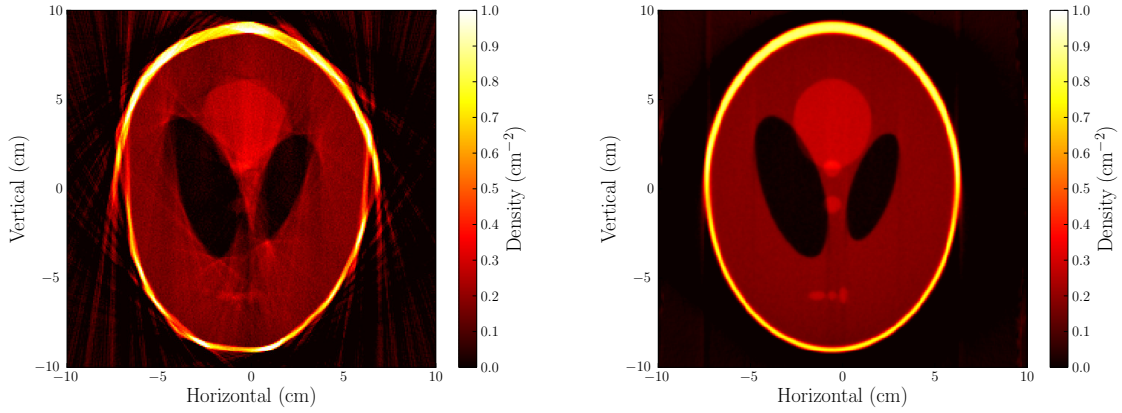

Figure 5.15 The left image shows the filtered backprojection reconstruction of a 40 dB SNR (5.1) head phantom that randomly walks in the horizontal direction while the right image shows KWF reconstruction given perfect knowledge of the movement.

The WF is given by

$$K_i(f) = \frac{|f|\, S_{\ddot{X}_{i|i-1}}(f\cos\theta_i,\ f\sin\theta_i)}{S_{\ddot{X}_{i|i-1}}(f\cos\theta_i,\ f\sin\theta_i) + S_{V_{\theta_i}}(f)}. \tag{5.13}$$

Note that (5.11), (5.12), and (5.13) are the 2-D cases of the more general results developed in Section 4.5. The initial prior error auto-correlation was chosen to be the circularly symmetric exponential function $R_{\ddot{X}_{1|0}}(\theta,\, \tau) = e^{-a|\tau|}$. The correlation length a was hand tuned to minimize reconstruction error. The initial state was assumed mean zero, i.e., $\widehat{x}_{1|0}(\boldsymbol{r}) = 0$.

Figure 5.15 (left) shows the filtered backprojection reconstruction of the moving head phantom. Even the relatively small patient motion in this example yields severe reconstruction artifacts. In contrast, given perfect knowledge of the patient motion, the KWF gives the reconstruction Figure 5.15 (right). However, only imprecise or statistical knowledge of the patient motion is available in practice. Suitable sources for patient motion knowledge include direct patient position measurement and joint estimation of motion by tracking features in the reconstructed images. Such statistical knowledge may be incorporated into the KWF state dynamic model described in

Section 2.2.3. Therefore, given sufficient statistical knowledge of patient motion, the KWF can approach the performance demonstrated in Figure 5.15 (right).

5.4 One-Dimensional Ensemble Kalman Smoother Example

The goal of the following experiment is to evaluate the performance of the EnKS in a low-dimensional 1-D example. The problem dimension is small enough that the EnKS state estimates can be compared to the Kalman smoother (KS) and localized Kalman smoother (LKS) estimates. We show that the KS provides a superior estimate of the state when compared to the KF, that the bias introduced by covariance tapering in the LKS is relatively small, and that the EnKS with a relatively small number of ensemble members provides estimates that are close to the LKS estimates.

The ground truth $\boldsymbol{x}_i$ is the discretized harmonic oscillator depicted in Figure 5.16. The 1-D example is discretized on an $N = 128$ grid and the oscillator passes through one complete period over $I = 32$ time steps. A measurement $\boldsymbol{y}_i$ consists of $M = 64$ direct noisy observations of randomly chosen state elements. In other words, the matrix $\boldsymbol{H}_i$ has M rows, each all zero except for a single randomly chosen column. Each measurement is corrupted by AWGN at the 30 dB SNR (5.1) level and the measurement noise covariance matrix is $\boldsymbol{R}_i = \sigma_v^2 \boldsymbol{I}$ where σ_v^2 is the noise variance. The initial state mean is $\boldsymbol{\mu}_1 = \boldsymbol{0}$ and the initial state covariance $\boldsymbol{\Pi}_1$ is a Toeplitz matrix with three bands indicating correlation with immediate neighbors only. The state transition operator is $\boldsymbol{F}_i = \boldsymbol{I}$, which models the state evolution as a random walk, and the state noise covariance $\boldsymbol{Q}_i$ is a Toeplitz matrix with three bands. Finally, the covariance taper matrix $\boldsymbol{C}_i$ is also a Toeplitz matrix with three bands.

The results of the experiment are depicted in Figure 5.17 which shows the output of the KF, KS, LKF, LKS, EnKF, and EnKS when the ensemble size is $L = 16$. From

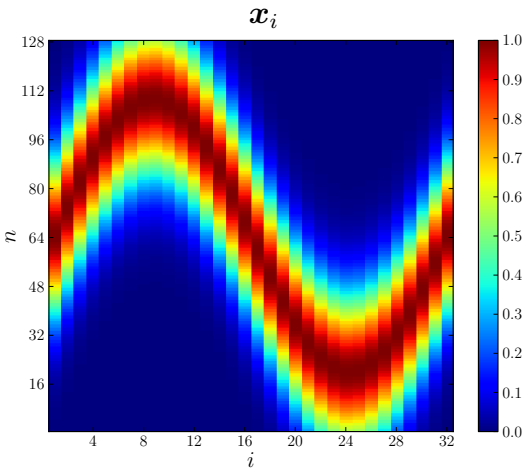

Figure 5.16 The ground truth $\boldsymbol{x}_i$ used in the experiment described in Section 5.4 to investigate the performance of the ensemble Kalman smoother (EnKS).

a qualitative comparison of the filtered and smoothed results, it is clear that taking future observations into account through smoothing provides tremendous benefits in this example. Also, the qualitative similarity between the KF and LKF results, and between the KS and LKS results, demonstrates that the covariance taper does not introduce a large bias in this example. Finally, the similarities between the ensemble methods (EnKF and EnKS) and their limiting solutions (LKF and LKS) demonstrate that the relatively small ensemble size of $L = 16$ is sufficient to obtain qualitatively faithful estimates in this example.

A quantitative comparison is made in Table 5.2 which compares the relative error

$$\text{relerror}(\boldsymbol{x}_{1:I},\, \boldsymbol{y}_{1:I}) \triangleq \left(\frac{\sum_{i=1}^{I} \|\boldsymbol{x}_i - \boldsymbol{y}_i\|_2^2}{\sum_{i=1}^{I} \|\boldsymbol{y}_i\|_2^2} \right)^{1/2} \tag{5.14}$$

between the six estimates shown in Figure 5.17 to the ground truth shown in Figure 5.16. In addition, Table 5.2 provides the relative error between the EnKF and LKF, and between the EnKS and LKS. The bias introduced by covariance tapering can be quantified by comparing the LKF and KF, and the LKS and KS, and is

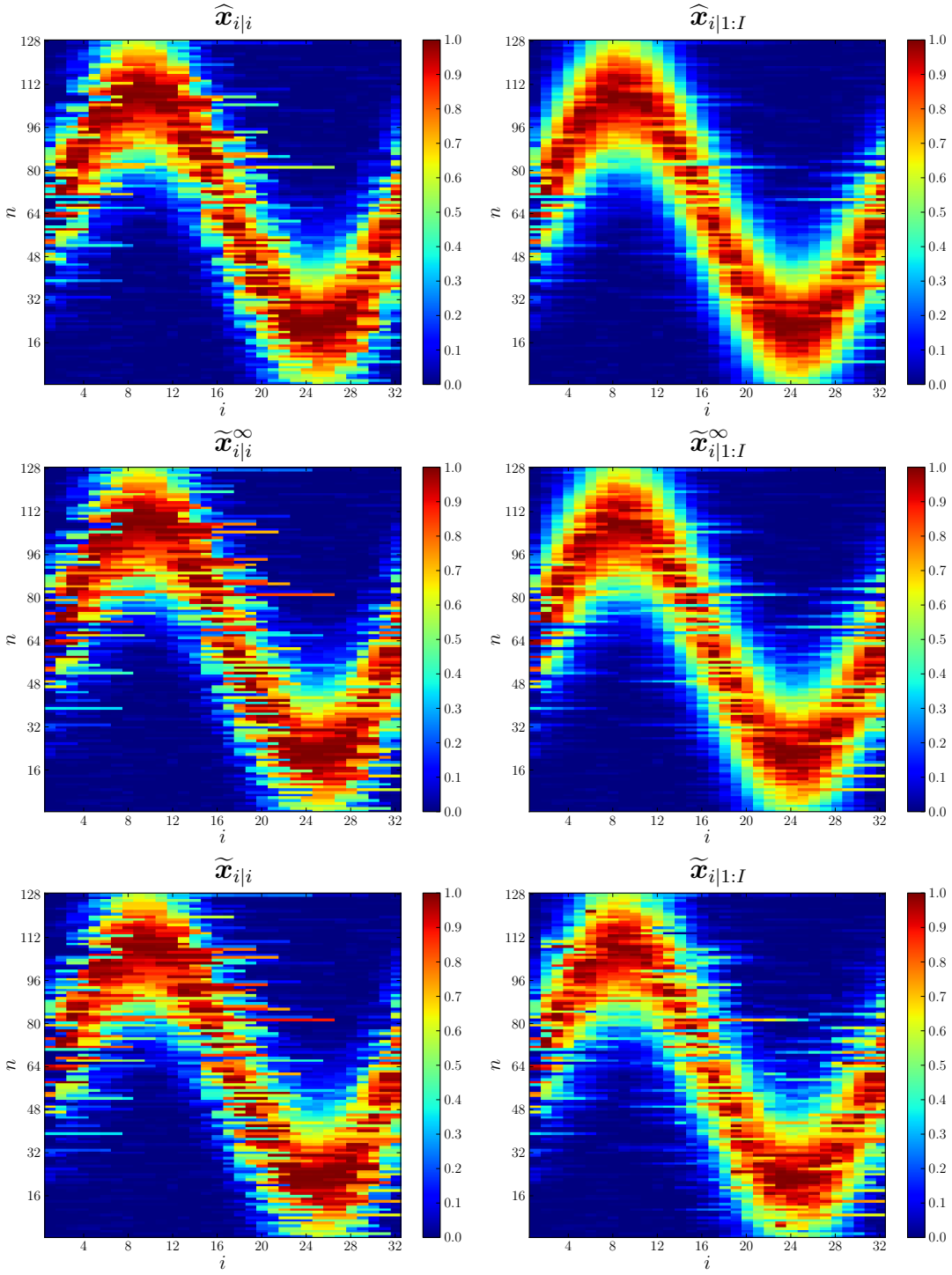

Figure 5.17 A comparison of the Kalman filter ($\widehat{\boldsymbol{x}}_{i|i}$), Kalman smoother ($\widehat{\boldsymbol{x}}_{i|1:I}$), localized Kalman filter ($\widetilde{\boldsymbol{x}}^{\infty}_{i|i}$), localized Kalman smoother ($\widetilde{\boldsymbol{x}}^{\infty}_{i|1:I}$), ensemble Kalman filter ($\widetilde{\boldsymbol{x}}_{i|i}$), and ensemble Kalman smoother ($\widetilde{\boldsymbol{x}}_{i|1:I}$) reconstructions of the harmonic oscillator experiment described in Section 5.4. All results are shown on the same color scale as the ground truth ($\boldsymbol{x}_i$) in Figure 5.16.

Table 5.2 A comparison of the relative error (5.14) between the filtered and smoothed estimates depicted in Figure 5.17 to the ground truth depicted in Figure 5.16. In addition, a comparison of the relative error between the EnKF and its limit as $L \to \infty$, i.e., the LKF, and the EnKS and its limit as $L \to \infty$, i.e., the LKS.

Description	$\boldsymbol{x}_{1:I}$	$\boldsymbol{y}_{1:I}$	relerror($\boldsymbol{x}_{1:I}$, $\boldsymbol{y}_{1:I}$)
KF vs. Truth	$\{\widehat{\boldsymbol{x}}_{i\|i}\}_{i=1}^{I}$	$\{\boldsymbol{x}_i\}_{i=1}^{I}$	0.315
KS vs. Truth	$\{\widehat{\boldsymbol{x}}_{i\|1:I}\}_{i=1}^{I}$	$\{\boldsymbol{x}_i\}_{i=1}^{I}$	0.193
LKF vs. Truth	$\{\widetilde{\boldsymbol{x}}_{i\|i}^{\infty}\}_{i=1}^{I}$	$\{\boldsymbol{x}_i\}_{i=1}^{I}$	0.401
LKS vs. Truth	$\{\widetilde{\boldsymbol{x}}_{i\|1:I}^{\infty}\}_{i=1}^{I}$	$\{\boldsymbol{x}_i\}_{i=1}^{I}$	0.248
EnKF vs. Truth	$\{\widetilde{\boldsymbol{x}}_{i\|i}\}_{i=1}^{I}$	$\{\boldsymbol{x}_i\}_{i=1}^{I}$	0.413
EnKS vs. Truth	$\{\widetilde{\boldsymbol{x}}_{i\|1:I}\}_{i=1}^{I}$	$\{\boldsymbol{x}_i\}_{i=1}^{I}$	0.291
EnKF vs. LKF	$\{\widetilde{\boldsymbol{x}}_{i\|i}\}_{i=1}^{I}$	$\{\widetilde{\boldsymbol{x}}_{i\|i}^{\infty}\}_{i=1}^{I}$	0.077
EnKS vs. LKS	$\{\widetilde{\boldsymbol{x}}_{i\|1:I}\}_{i=1}^{I}$	$\{\widetilde{\boldsymbol{x}}_{i\|1:I}^{\infty}\}_{i=1}^{I}$	0.143

summarized in Table 5.3.

5.5 Ensemble Kalman Smoother Convergence

The following details a straightforward low-dimensional example aimed to empirically verify the EnKS convergence result summarized by Theorem 4.20, which states that the EnKS converges to the LKS as the ensemble size L increases. All parameters of the linear dynamic model of Section 2.2.2 were chosen at random. In particular, the initial state $\boldsymbol{x}_i$ is an $N = 8$ random vector chosen from the distribution $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Pi}_1)$

Table 5.3 Quantification of the bias introduced by covariance tapering for the results depicted in Figure 5.17.

Description	$\boldsymbol{x}_{1:I}$	$\boldsymbol{y}_{1:I}$	relerror($\boldsymbol{x}_{1:I}$, $\boldsymbol{y}_{1:I}$)
LKF vs. KF	$\{\widetilde{\boldsymbol{x}}_{i\|i}^{\infty}\}_{i=1}^{I}$	$\{\widehat{\boldsymbol{x}}_{i\|i}\}_{i=1}^{I}$	0.125
LKS vs. KS	$\{\widetilde{\boldsymbol{x}}_{i\|1:I}^{\infty}\}_{i=1}^{I}$	$\{\widehat{\boldsymbol{x}}_{i\|1:I}\}_{i=1}^{I}$	0.082

where the initial state mean $\boldsymbol{\mu}_i$ is a randomly sampled N element vector and the initial state covariance $\boldsymbol{\Pi}_1$ is a random $N \times N$ symmetric positive definite (SPD) matrix. The state process is sampled randomly according to $\boldsymbol{x}_{i+1} \sim \mathcal{N}(\boldsymbol{F}_i \boldsymbol{x}_i, \boldsymbol{Q}_i)$ where the state transition operator $\boldsymbol{F}_i$ is a random $N \times N$ matrix and the state noise covariance is a random $N \times N$ SPD matrix. Similarly, each measurement is sampled randomly according to $\boldsymbol{y}_i \sim \mathcal{N}(\boldsymbol{H}_i \boldsymbol{x}_i, \boldsymbol{R}_i)$ where the measurement operator $\boldsymbol{H}_i$ is a random $M \times N$ matrix with $M = 6$ and the measurement noise covariance is a $M \times M$ random SPD matrix. The covariance taper matrix $\boldsymbol{C}_i$ is a random $N \times N$ SPD matrix and $\boldsymbol{C}_i = \boldsymbol{C}'_i = \boldsymbol{C}''_i$.

Figure 5.18 depicts the convergence of the EnKS to the LKS with increasing ensemble size L. The plot shows the average EnKS estimate error

$$\overline{x} \triangleq \frac{1}{n_{\text{trial}}} \sum_{n=1}^{n_{\text{trial}}} \text{relerror}\Big(\big\{ \widetilde{\boldsymbol{x}}_{i|1:I}(n) \big\}_{i=1}^{I},\ \big\{ \widetilde{\boldsymbol{x}}_{i|1:I}^{\infty} \big\}_{i=1}^{I} \Big) \tag{5.15}$$

as the ensemble size increases from $L = 2^1$ to $L = 2^{14}$. In this experiment, $n_{\text{trials}} = 64$ is the number of independent trials. The line of best fit has a slope of -0.511 which agrees with the expected $O(L^{-1/2})$ rate of convergence for Monte Carlo algorithms. The empirical convergence of the EnKS error covariance to the LKS error covariance is demonstrated in Figure 5.19. The vertical axis of this plot is the average EnKS error covariance error defined as

$$\overline{P} \triangleq \frac{1}{n_{\text{trial}}} \sum_{n=1}^{n_{\text{trial}}} \text{relerror}\Bigg(\Big\{ \text{vec}\left[\widetilde{\boldsymbol{P}}_{i|1:I}(n) \right] \Big\}_{i=1}^{I},\ \Big\{ \text{vec}\left[\widetilde{\boldsymbol{P}}_{i|1:I}^{\infty} \right] \Big\}_{i=1}^{I} \Bigg) \tag{5.16}$$

where $n_{\text{trial}} = 64$. The horizontal axis is the ensemble size which ranges from $L = 2^1$ to $L = 2^{14}$. The line of best fit in this plot has a slope of -0.503, demonstrating the expected rate of convergence.

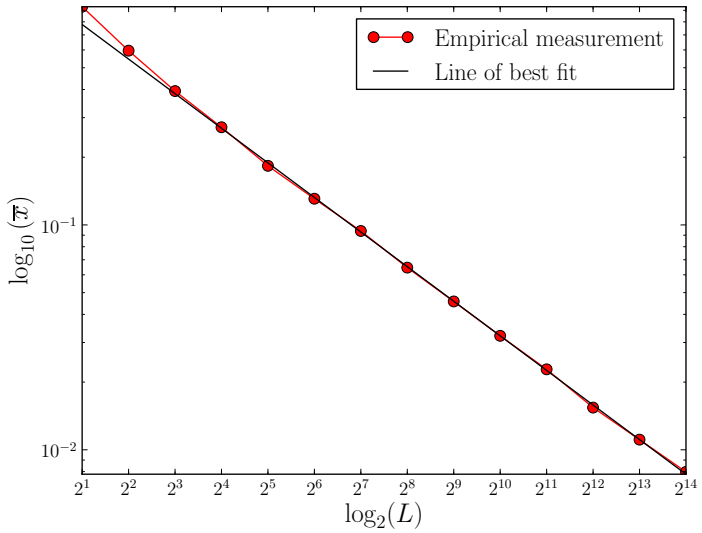

Figure 5.18 A log-log plot of the average EnKS estimate error $\overline{x}$ (5.15) versus ensemble size L. Note that the horizontal axis is on a base-2 scale and the vertical axis is on a base-10 scale.

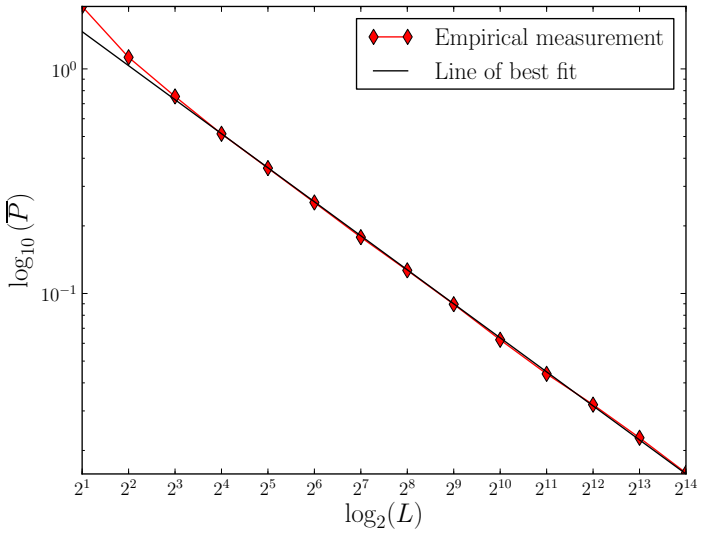

Figure 5.19 A log-log plot of the average EnKS error covariance error $\overline{P}$ (5.16) versus ensemble size L. Note that the horizontal axis is on a base-2 scale and the vertical axis is on a base-10 scale.

CHAPTER 6

SPATIAL-TEMPORAL RESOLUTION

This chapter develops a method to quantify the spatial-temporal resolution of the state estimators defined in Chapter 4. We consider the local impulse response [57] which describes the effect upon the average state estimate reconstruction due to a perturbation at a single spatial-temporal coordinate. Methods exist to ascertain the spatial resolution under the static signal model defined in Section 2.1, e.g., see [57], and the spatial-temporal resolution under the linear dynamic model defined in Section 2.2.2 with the restriction of a deterministic state-transition model where $\boldsymbol{x}_{i+1} = \boldsymbol{F}_i \boldsymbol{x}_i$ and there is no state noise component $\boldsymbol{u}_i$, e.g., see [93], [94]. To the best of our knowledge, Theorem 6.2 is novel because it quantifies the local impulse response under the general fully stochastic signal model of Section 2.2.2.

The chapter begins with an overview of the LMMSE estimator spatial resolution under the static signal model of Section 2.1. Then, we develop a method to quantify the spatial-temporal resolution under the linear dynamic model of Section 2.2.2. Finally, the concept of local impulse response is illustrated by the numerical experiments detailed in the last section.

6.1 Linear Static Model

The local impulse response under the static signal model defined in Section 2.1 is summarized by the following theorem.

Theorem 6.1: The local impulse response [57] for the jth parameter of the estimator (4.1) is given by

$$\boldsymbol{l}^j = \left(\boldsymbol{H}_{1:I}^T \, \boldsymbol{R}_{\boldsymbol{v}_{1:I}}^{-1} \, \boldsymbol{H}_{1:I} + \lambda \, \boldsymbol{D}^T \boldsymbol{D}\right)^{-1} \boldsymbol{H}_{1:I}^T \, \boldsymbol{R}_{\boldsymbol{v}_{1:I}}^{-1} \, \boldsymbol{H}_{1:I} \, \boldsymbol{e}^j \tag{6.1}$$

where

$$\left[\boldsymbol{e}^j\right]_n \triangleq \begin{cases} 1, & j = n \\ 0, & \text{otherwise.} \end{cases} \tag{6.2}$$

Proof. See Section II in [57]. □

6.2 Linear Dynamic Model

The local impulse response for the filtered LMMSE estimator computed by the Kalman filter of Section 4.3 is summarized by the following theorem.

Theorem 6.2: Consider the filtered LMMSE estimator under the linear dynamic signal model defined in Section 2.2.2. The filtered local impulse response for the kth time index and jth parameter is denoted by $\boldsymbol{l}_{i|i}^{j,k}$ and may be computed by processing the following simulated measurements through the Kalman filter with $\boldsymbol{\mu}_1 = \boldsymbol{0}$:

$$\boldsymbol{y}_{1:i} = \boldsymbol{H}_{1:i}^{\mathrm{B}} \, \boldsymbol{e}_k^j \tag{6.3}$$

where

$$\boldsymbol{H}_{1:i}^{\mathrm{B}} \triangleq \mathrm{diag}\big(\boldsymbol{H}_1, \, \ldots, \, \boldsymbol{H}_i\big) \tag{6.4}$$

and the function diag$(\cdot)$ forms the block diagonal matrix with its arguments as the block elements. The unit vector $\boldsymbol{e}_k^j$ is defined by

$$\left[\boldsymbol{e}_k^j\right]_{(l-1)M:\,l\,M} \triangleq \begin{cases} \boldsymbol{e}^j, & l = k \\ \boldsymbol{0}, & \text{otherwise} \end{cases} \tag{6.5}$$

and $[\cdot]_{(l-1)M:\,l\,M}$ selects the lth block of length M from its vector argument.

Proof. First note that the Kalman filter is an affine function of the measurements [26], meaning it may be written in the form

$$\widehat{\boldsymbol{x}}_{i|i} = \boldsymbol{A}_i\, \boldsymbol{y}_{1:i} + \boldsymbol{b}_i \tag{6.6}$$

and, because the Kalman filter computes LMMSE estimates, the above parameters are given by [63]

$$\boldsymbol{A}_i = \mathrm{Cov}(\boldsymbol{x}_i, \boldsymbol{y}_{1:i})\ \mathrm{Cov}(\boldsymbol{y}_{1:i}, \boldsymbol{y}_{1:i})^{-1} \tag{6.7}$$

$$\boldsymbol{b}_i = \mathbb{E}(\boldsymbol{x}_i) - \mathrm{Cov}(\boldsymbol{x}_i, \boldsymbol{y}_{1:i})\ \mathrm{Cov}(\boldsymbol{y}_{1:i}, \boldsymbol{y}_{1:i})^{-1}\ \mathbb{E}(\boldsymbol{y}_{1:i}) \tag{6.8}$$

where

$$\boldsymbol{y}_{1:i} = \boldsymbol{H}_{1:i}^{\mathrm{B}}\, \boldsymbol{x}_{1:i} + \boldsymbol{v}_{1:i} \tag{6.9}$$

with

$$\boldsymbol{x}_{1:i} \triangleq \begin{bmatrix} \boldsymbol{x}_1^T & \cdots & \boldsymbol{x}_i^T \end{bmatrix}^T \tag{6.10}$$

and

$$\boldsymbol{v}_{1:i} \triangleq \begin{bmatrix} \boldsymbol{v}_1^T & \cdots & \boldsymbol{v}_i^T \end{bmatrix}^T. \tag{6.11}$$

Next, we will require the expectation of the estimator conditioned on the state:

$$\begin{aligned}\mathbb{E}\big(\widehat{\boldsymbol{x}}_{i|i}|\boldsymbol{x}_{1:i}\big) &= \boldsymbol{A}_i\ \mathbb{E}(\boldsymbol{y}_{1:i}|\boldsymbol{x}_{1:i}) + \boldsymbol{b}_i \\ &= \boldsymbol{A}_i\,\boldsymbol{H}^{\mathrm{B}}_{1:i}\,\boldsymbol{x}_{1:i} + \boldsymbol{b}_i. \end{aligned} \tag{6.12}$$

The local impulse response is given by [57]

$$\boldsymbol{l}^{j,k}_{i|i} \triangleq \frac{\partial}{\partial\,[\boldsymbol{x}_k]_j}\,\mathbb{E}\big(\widehat{\boldsymbol{x}}_{i|i}|\boldsymbol{x}_{1:i}\big) = \boldsymbol{A}_i\,\boldsymbol{H}^{\mathrm{B}}_{1:i}\,\boldsymbol{e}^j_k \tag{6.13}$$

and note that (6.13) is equal to (6.6) if $\boldsymbol{y}_{1:i} = \boldsymbol{H}^{\mathrm{B}}_{1:i}\,\boldsymbol{e}^j_l$ and $\boldsymbol{b}_i = \boldsymbol{0}$. Finally, it is not difficult to show that

$$\mathbb{E}(\boldsymbol{x}_i) = \boldsymbol{F}_{i-1}\,\boldsymbol{F}_{i-2}\,\cdots\,\boldsymbol{F}_1\,\boldsymbol{\mu}_1 \tag{6.14}$$

and

$$\mathbb{E}(\boldsymbol{y}_{1:i}) = \boldsymbol{H}^{\mathrm{B}}_{1:i}\ \mathbb{E}(\boldsymbol{x}_{1:i}) \tag{6.15}$$

which imply $\boldsymbol{b}_i = \boldsymbol{0}$ if $\boldsymbol{\mu}_1 = \boldsymbol{0}$. □

The local impulse response for smoothed and predictive LMMSE estimators are addressed by the following corollary to Theorem 6.2.

Corollary 6.3: Consider the LMMSE predictor that produces the estimate $\widehat{\boldsymbol{x}}_{i+1|i}$. The predicted local impulse response $\boldsymbol{l}^{j,k}_{i+1|i}$ is given by

$$\boldsymbol{l}^{j,k}_{i+1|i} = \boldsymbol{F}_i\,\boldsymbol{l}^{j,k}_{i|i}. \tag{6.16}$$

Furthermore, consider the LMMSE fixed-interval smoother that produces the estimates $\{\widehat{\boldsymbol{x}}_{i|1:I}\}_{i=1}^{I}$. The smoothed local impulse response $\boldsymbol{l}^{j,k}_{i|1:I}$ can be computed by processing the simulated measurements $\boldsymbol{y}_{1:I} = \boldsymbol{H}^{\mathrm{B}}_{1:I}\,\boldsymbol{e}^j_k$ with the Kalman smoother

defined in Section 4.6.1 with $\boldsymbol{\mu}_1 = \mathbf{0}$.

Proof. First, note that the LMMSE prediction of the state $\boldsymbol{x}_{i+1}$ given the measurements $\boldsymbol{y}_{1:i}$ is $\widehat{\boldsymbol{x}}_{i+1|i} = \boldsymbol{F}_i \, \widehat{\boldsymbol{x}}_{i|i}$. Therefore, following the development in the proof of Theorem 6.2,

$$\boldsymbol{l}_{i+1|i}^{j,\,k} = \boldsymbol{F}_i \, \boldsymbol{A}_i \, \boldsymbol{H}_{1:i}^{\mathrm{B}} \, \boldsymbol{e}_k^j \tag{6.17}$$

which can be computed by applying the Kalman filter time update to $\boldsymbol{l}_{i|i}^{j,\,k}$.

Next, note that the Kalman smoother is an affine function of the measurements and can be written in the form

$$\widehat{\boldsymbol{x}}_{i|1:I} = \boldsymbol{A}_i' \, \boldsymbol{y}_{1:I} + \boldsymbol{b}_i' \tag{6.18}$$

where

$$\boldsymbol{A}_i' = \mathrm{Cov}(\boldsymbol{x}_i, \boldsymbol{y}_{1:I}) \; \mathrm{Cov}(\boldsymbol{y}_{1:I}, \boldsymbol{y}_{1:I})^{-1} \tag{6.19}$$

$$\boldsymbol{b}_i' = \mathbb{E}(\boldsymbol{x}_i) - \mathrm{Cov}(\boldsymbol{x}_i, \boldsymbol{y}_{1:I}) \; \mathrm{Cov}(\boldsymbol{y}_{1:I}, \boldsymbol{y}_{1:I})^{-1} \; \mathbb{E}(\boldsymbol{y}_{1:I}) \, . \tag{6.20}$$

The remainder of the argument then closely mirrors the proof of Theorem 6.2. □

6.3 Illustrative Numerical Example

We now investigate the local impulse response in a numerical experiment with a similar setup to the experiment detailed in Section 5.2. In this experiment, the unknown state is a 16×16 and 48 time step spatial-temporal discretization of the diffusive process described in Section 5.2. Each 40 dB SNR (5.1) measurement consists of $M = 23$ parallel line integrals of the state that uniformly sweeps through 540° over the 48 time steps. The state noise covariance matrix $\boldsymbol{Q}_i$ is a symmetric positive

definite matrix with 9 bands, indicating correlation with immediate neighbors only. The initial state mean is $\boldsymbol{\mu}_1 = \mathbf{0}$ and the initial state covariance is $\boldsymbol{\Pi}_1 = \alpha\left(\boldsymbol{D}^T\boldsymbol{D}\right)$. The parameter α was manually tuned and

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{D}_x \\ \boldsymbol{D}_y \end{bmatrix} \tag{6.21}$$

where $\boldsymbol{D}_x$ and $\boldsymbol{D}_y$ are discrete approximations to the spatial derivative in the horizontal and vertical directions, respectively.

This experiment focuses on the implications of knowing the state transition operator $\boldsymbol{F}_i$ versus assuming the random walk model $\boldsymbol{F}_i = \boldsymbol{I}$. Specifically, we will consider the local impulse response of the Kalman filter and Kalman smoother for these two cases at the 29th time index and spatial location depicted in Figure 6.1. The Kalman filter and smoother estimates in these two cases are first shown in Figure 6.2 over the range of two time indices, one prior and ones after the time index under study, $i = 29$. Then, Figure 6.3 shows these estimates for time indices at an absolute distance of 6 and 12 from time index 29. As expected, the results when the state dynamic model is known are superior to when the random walk model is assumed. Furthermore, the smoothed results are superior to the filtered results. A quantitative evaluation of the error is provided in Table 6.1. The potential improvement in estimate fidelity offered through smoothing is exemplified by the fact that the smoothed estimate computed with a poor state transition model has a smaller relative error than the filtered estimate computed with the correct model.

The local impulse response for the estimators and time indices corresponding with Figure 6.2 is shown in Figure 6.4. Similarly, the local impulse response corresponding to Figure 6.3 is shown in Figure 6.5. All local impulse response functions are

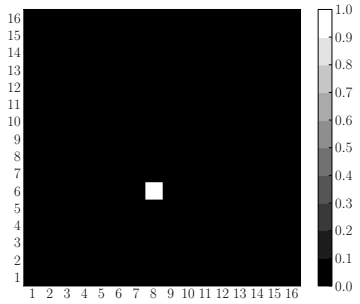

Figure 6.1 The unit vector $\boldsymbol{e}_k^j$ as viewed at time index 29. The unit vector is equal to $\mathbf{0}$ at all other time indices.

Table 6.1 A summary of the relative error (5.14) between the Kalman filter and smoother estimates and the ground truth in the spatial-temporal resolution study. Two cases are considered: when the state transition operator $\boldsymbol{F}_i$ is known and under the assumption of the random walk model ($\boldsymbol{F}_i = \boldsymbol{I}$).

	Relative error	
Estimator	$\boldsymbol{F}_i$ known	$\boldsymbol{F}_i = \boldsymbol{I}$
Kalman filter ($\widehat{\boldsymbol{x}}_{i\|i}$)	0.480	0.596
Kalman smoother ($\widehat{\boldsymbol{x}}_{i\|1:I}$)	0.135	0.356

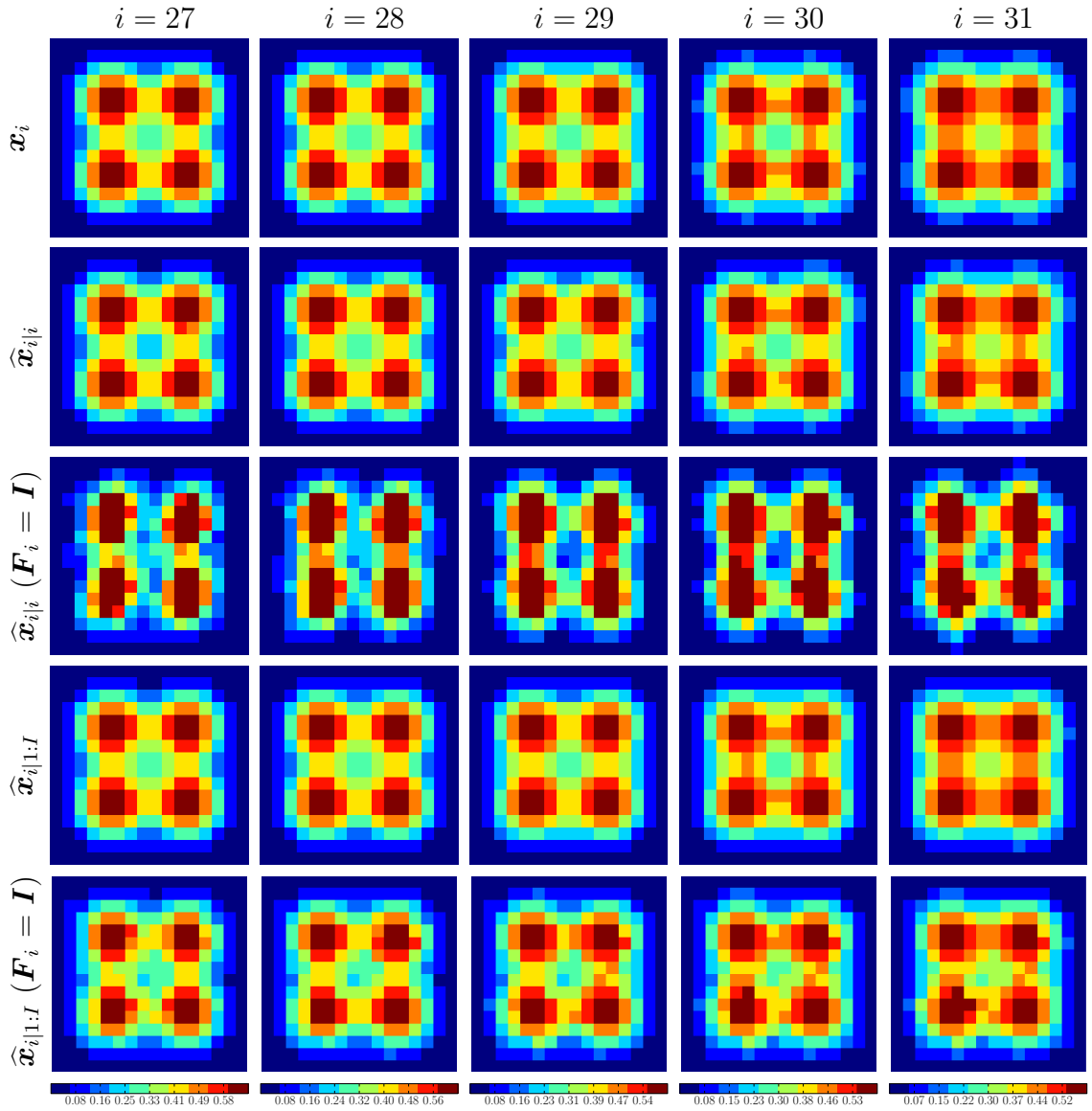

Figure 6.2 The state estimates within 2 time steps of the time index of interest, $i = 29$, in the spatial-temporal resolution study. The first row shows the ground truth at five time indices. The remaining rows show the Kalman filter estimate $\widehat{\boldsymbol{x}}_{i|i}$ under the true and random walk ($\boldsymbol{F}_i = \boldsymbol{I}$) state transition models and the Kalman smoother estimate $\widehat{\boldsymbol{x}}_{i|1:I}$, also for the two choices for $\boldsymbol{F}_i$.

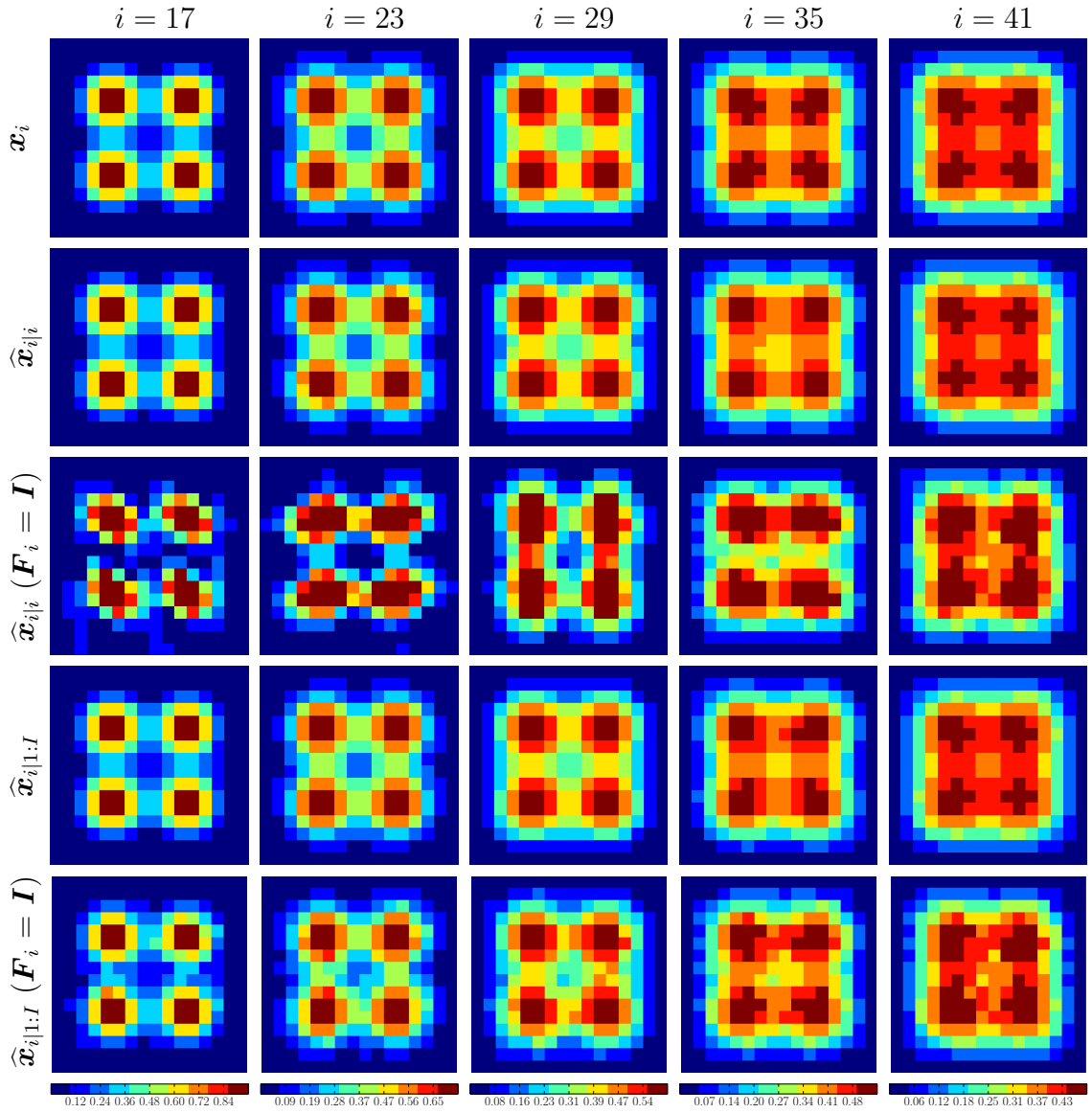

Figure 6.3 The spatial-temporal resolution study state estimates at times ± 6 and ± 12 centered at the time index of interest, $i = 29$. Similar to Figure 6.2, we show the ground truth and Kalman filter and Kalman smoother estimates when the true state transition operator $\boldsymbol{F}_i$ is known and for the random walk model $\boldsymbol{F}_i = \boldsymbol{I}$.

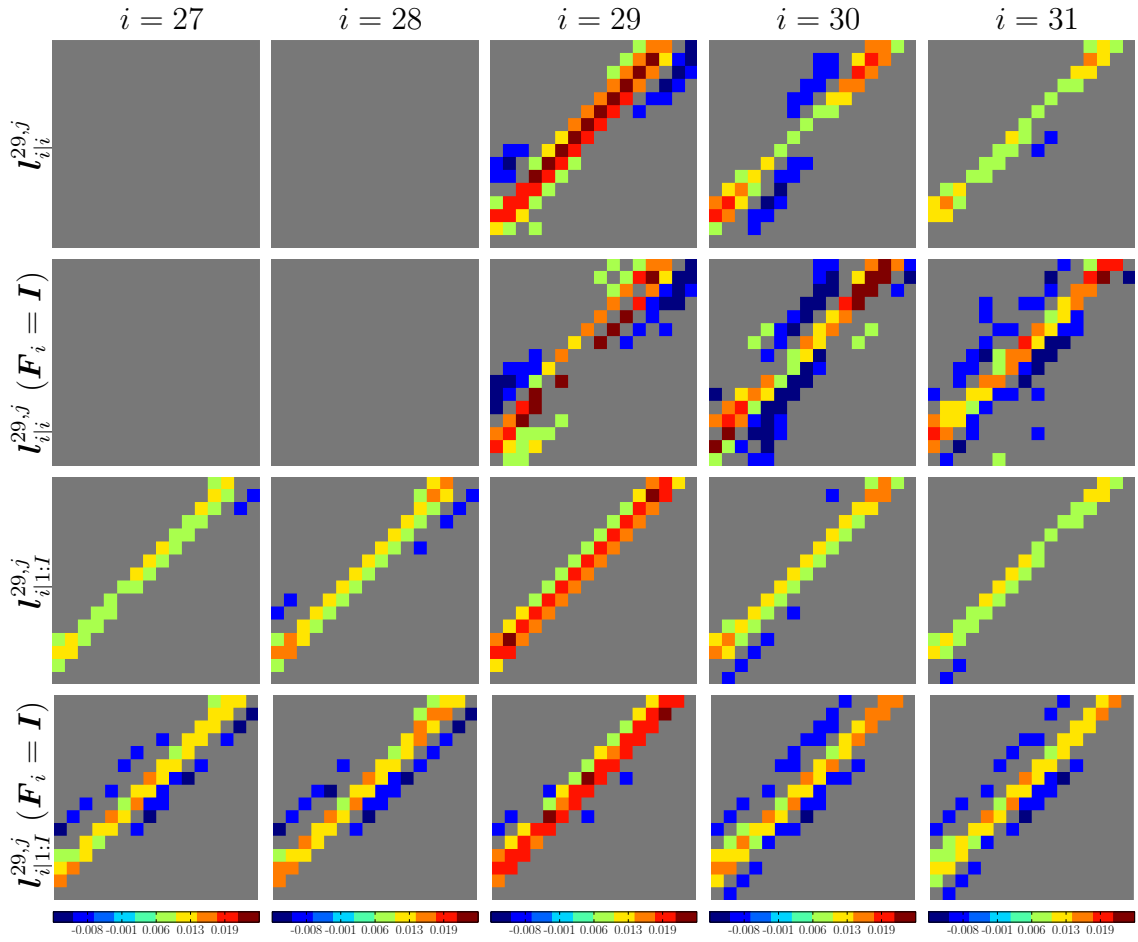

Figure 6.4 The local impulse response for time index 29 and spatial location depicted in Figure 6.1 at the same time indices as Figure 6.2. The first two rows show the filtered local impulse response $\boldsymbol{l}_{i|i}^{29,\,j}$ for two cases: known state transition operator $\boldsymbol{F}_i$ and random walk model $\boldsymbol{F}_i = \boldsymbol{I}$. The last two rows show the smoothed local impulse response $\boldsymbol{l}_{i|1:I}^{29,\,j}$ for the same two cases. For each row, portions shown in gray have an absolute value less than one quarter of the maximum at time index 29. Note that each local impulse response is shown on the same color scale.

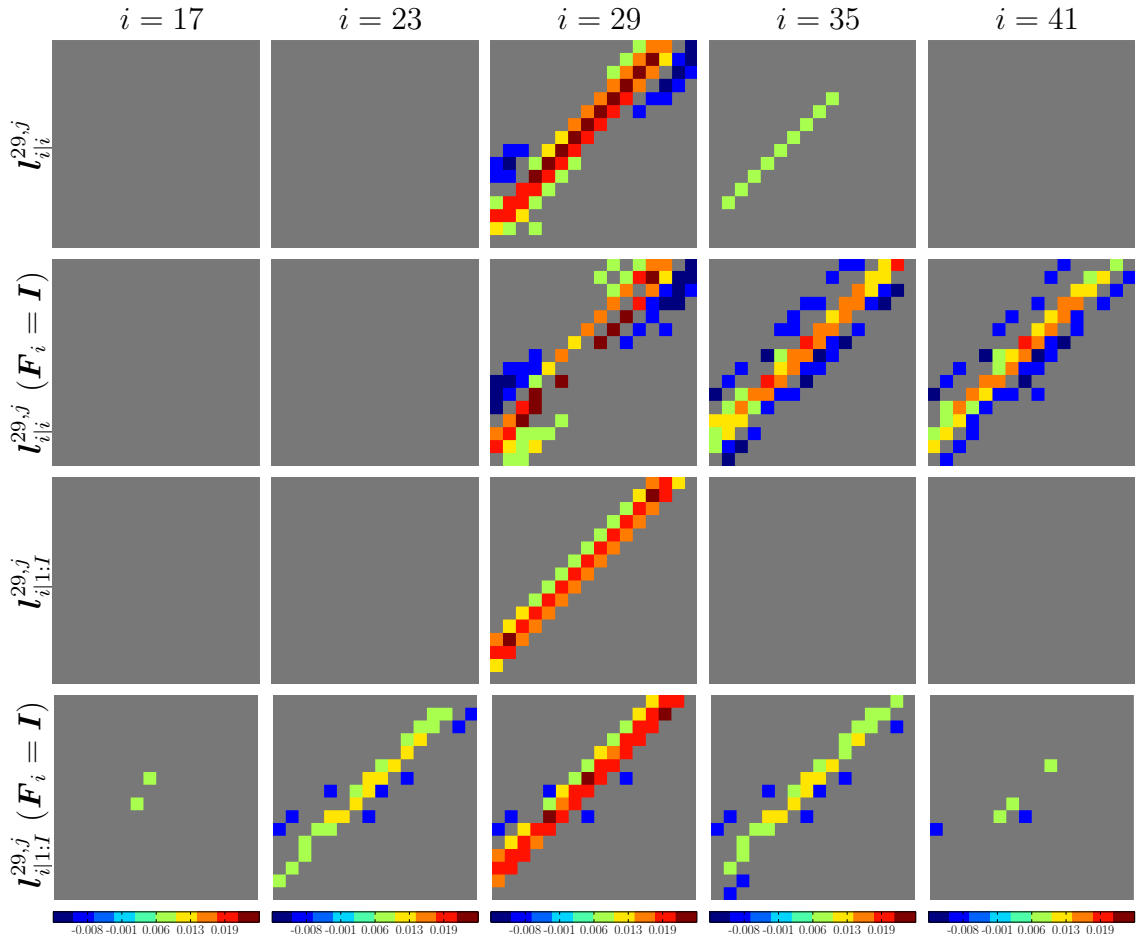

Figure 6.5 The local impulse response for time index 29 and spatial location depicted in Figure 6.1 at the same time indices as Figure 6.3. The first two rows show the filtered local impulse response $\boldsymbol{l}_{i|i}^{29,j}$ for two cases: known state transition operator $\boldsymbol{F}_i$ and random walk model $\boldsymbol{F}_i = \boldsymbol{I}$. The last two rows show the smoothed local impulse response $\boldsymbol{l}_{i|1:I}^{29,j}$ for the same two cases. For each row, portions shown in gray have an absolute value less than one quarter of the maximum at time index 29. Note that each local impulse response is shown on the same color scale.

shown on the same color scale so that they may be compared. Portions of an impulse response function appearing in gray correspond to absolute values less than one quarter of the impulse response maximum. Note that the filtered local impulse response functions are completely gray (in fact, they are uniformly equal to 0) at time indices less than $i = 29$. This is expected because the filtered estimates are causal and, as such, a perturbation cannot affect a filtered estimate prior to its occurrence. Also, note that the smoothed local impulse response functions decay more quickly in time than the filtered functions. In addition, the local impulse response functions under the correct state transition models decay faster than the random walk case. Both observations give another means to understand the differences in the relative error reported in Table 6.1. Qualitatively, the local impulse response functions under the random walk model are less smooth than their counterparts under the correct state transition model, which explains why the random walk state estimates in Figure 6.2 and Figure 6.3 are less smooth than the corresponding state estimates under the correct model. Finally, by comparing the first and last rows of Figure 6.4 and Figure 6.5, note that the smoothed random walk model response has smaller amplitude and is less broad than the filtered correct model response at times near time index 29. Though a full investigation of the spatial-temporal response at all spatial-temporal coordinates would yield a more definitive answer, the previous observation provides some insight as to why the smoothed incorrect model estimate has a smaller relative error than the filtered correct model estimate.

CHAPTER 7

DYNAMIC SOLAR TOMOGRAPHY

Several signal models for tomography were defined in Chapter 2. State inference methods focused on the high-dimensional problems that can arise in dynamic tomography were developed in Chapter 4 and evaluated in the numerical experiments in Chapter 5. This chapter casts the problem of reconstructing the volumetric inner solar atmosphere or corona under the linear dynamic models and demonstrates the use of the ensemble Kalman filter to solve the resultant high-dimensional state estimation problem.

The chapter begins with introductory material that motivates the problem. Then, Section 7.2 identifies the data sources that remotely sense the Sun's corona. Section 7.3 discusses the validity of the signal models used in this study. Finally, Section 7.4 details the model parameters that are later used to form the electron density reconstructions that appear in Section 7.5.

7.1 Introduction

Much is still unknown regarding some of the fundamental physical processes of the Sun. For example, long-standing questions remain regarding the physical mechanisms that give rise to many of the geo-effective phenomena, collectively referred to as "space weather," that may in extreme cases damage satellites in Earth's orbit, cause power

distribution failures, or even harm astronauts [95]. To a large degree, our knowledge of the solar environment has been obtained empirically from a variety of dedicated space-borne and ground-based instruments. The current generation of Sun-observing spacecraft generate a rich and complex suite of observations that provide the scientific community with measurements at unprecedented resolution. New signal processing methods are required to take full advantage of all available information to ultimately address open stellar astrophysical questions. In particular, this chapter develops signal processing methods to empirically reconstruct the 3-D time-varying structure of the solar corona.

The corona [96] is the very sparse plasma that forms the upper atmosphere of the Sun, beginning at about 2000 km above the solar surface. The corona is extremely hot, with temperatures in excess of 10^6 K. Despite more than 50 years of research, the non-thermal processes that heat the corona to such an extreme temperature over the span of the relatively thin (10^2 km) transition region remain unknown [96], though it is clear that the solar magnetic field provides a conduit for energy transport and storage as well as several possible mechanisms for energy release. The solar magnetic field is also central in driving the solar wind, coronal mass ejections (CMEs; occasional violent solar explosions originating in the corona), and accelerating solar energetic particles that all contribute to space weather. The solar magnetic field also has a strong influence on the structure of the corona, a plasma with low β, the ratio of plasma to magnetic pressure. As a consequence, the solar magnetic field traps the structures of the corona.

In principle, tomographic analysis [90], [97] may be used to reconstruct the 3-D corona given image measurements (2-D projections) from multiple points-of-view (POVs) obtained by solar rotation or from spatially distinct sensors [98], [99]. Until recently, measurements of the Sun have been available from only one POV at any given

time and tomographic analysis of the corona has relied on solar rotation to provide the necessary measurement diversity. With a synodic period of 27.3 d, almost 14 d are required to observe the corona from a single POV over the 180° necessary for a full tomographic analysis. The situation has recently changed with the launch of the Solar and Terrestrial Relations Observatory (STEREO) [100] in October, 2006. The mission provides two additional simultaneous POVs from dual satellites, one that leads Earth's orbit by 22°/y and the other that lags by the same amount. However, almost 5 d are still required to view the Sun over 180° when the STEREO satellites are each separated from the Earth by 60°.

This chapter develops state-space framework for reconstructing the time-evolving 3-D structure of the solar atmosphere. Solar tomography has some unique aspects that challenge current state estimation methods, including the dimensionality of the problem, the tomographic nature of the measurements that are corrupted by Poisson noise, nonlinear dynamics described by the equations of magnetohydrodynamics (MHD) [101], [102] that govern the physics of electrically conducting fluids, and the existence of additional prior knowledge, such as the empirical structure of the solar magnetic field, that could also be incorporated into the analysis.

7.2 Data Sources

Several coronagraphs, specialized telescopes for measuring the relatively dim corona, regularly measure polarized brightness (pB) images suitable for electron density tomography as summarized in Table 7.1. One example is the ground-based Mark-IV coronagraph [103] at the Mauna Loa Solar Observatory (MLSO). Two pB images measured by the Mark-IV are shown in Figure 7.1. Space-borne coronagraphs include the Large Angle and Spectrometric Coronagraph Experiment (LASCO) C2 instrument [104] on board the Solar and Heliospheric Observatory (SOHO) [105] and

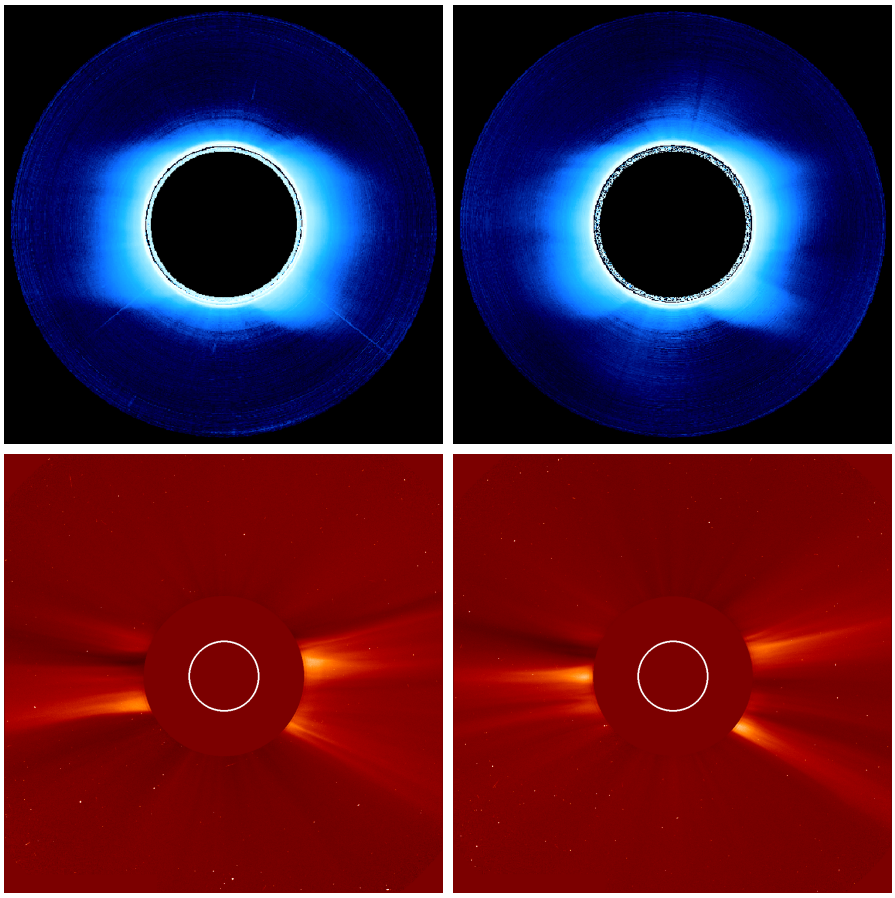

Figure 7.1 Four images of polarized brightness as measured by the Mark-IV coronagraph at the MLSO (first row) and the C2 coronagraph on SOHO (second row). The inner and outer radii of the Mark-IV images range from 1.1–2.85 $R_\odot$, where $R_\odot$ is the solar radius equal to 7×10^5 km, and from 1.5–6 $R_\odot$ for the C2 images. The images in the left and right columns were measured on April 7 and May 4 of 2007, a span of 27 d or one full solar rotation. As a result, the differences between images in each row are due to solar dynamics.

Table 7.1 Polarized brightness white light coronagraphs.

Name	Location	FOV ($R_\odot$)	Sensor pixels	Time res. (images/d)[a]
Mark-IV	MLSO	1.1 – 2.8	720 × 384[b]	1
C2	SOHO	1.5 – 6	1024 × 1024	1 – 6
C3	SOHO	3.7 – 30	1024 × 1024	1
COR1	STEREO	1.4 – 4	2048 × 2048	100
COR2	STEREO	2 – 15	2048 × 2048	100

[a]Temporal resolution is only approximate. Telemetry constraints and SNR requirements can change over time and affect the sensor data rates.

[b]Resolution is in polar coordinates $r \times \phi$. All other sensor spatial resolutions are given in Cartesian coordinates $x \times y$.

the COR1 and COR2 coronagraphs available as part of the Sun Earth Connection Coronal and Heliospheric Investigation (SECCHI) suite of instruments [106] on both STEREO satellites. The best quality measurements of pB are from space-borne coronagraphs that are not affected by sky-scattered light that can plague terrestrial coronagraph measurements.

Temperature reconstruction requires data at multiple wavelengths. Measurements of the corona in the extreme ultraviolet (EUV) are routinely measured by dedicated space-borne instruments as summarized in Table 7.2. Ground-based observations of the corona in the EUV are not possible because Earth's atmosphere absorbs EUV emissions. Relevant instruments include the Extreme Ultraviolet Imaging Telescope (EIT) [107] on SOHO, the EUV imager as part of the SECCHI instruments [106] on STEREO, and the Atmospheric Imaging Assembly (AIA) [108] on the Solar Dynamics Observatory (SDO). Four images of the EUV corona measured by the EIT are shown in Figure 7.2. X-ray measurements are available, for example, from the X-ray Telescope (XRT) [109] on Hinode [110].

Data for the dynamic coronal density reconstructions in Section 7.5 are acquired from the COR1 coronagraph [111], the innermost coronagraph within the SECCHI

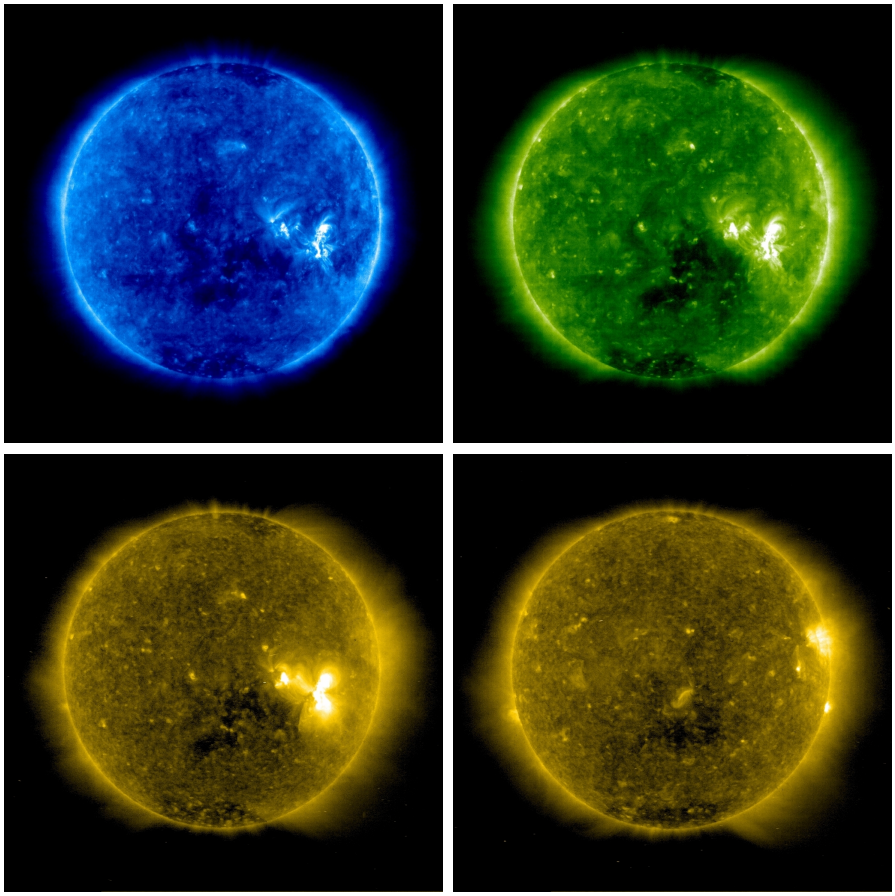

Figure 7.2 Four images of the Sun to 1.5 $R_\odot$ as measured by the EIT on SOHO. Three of the images were measured on May 4, 2008, at different EUV wavelengths: (top left) at 171 Å, (top right) at 195 Å, and (bottom left) at 284 Å. The fourth image is at 284 Å and was measured on April 7, 2008, 27 d prior to the previous images. Differences between the images in the bottom row are due to solar dynamics because the Sun rotates 360°/27 d relative to the Earth.

Table 7.2 Extreme ultraviolet imagers.

Name	Location	FOV ($R_\odot$)	Filter bands (Å)	Sensor pixels	Time res. (images/d)[a]
EIT	SOHO	1 – 1.5	171, 195, 284, 304	1024 × 1024	4
EUVI	STEREO	1 – 1.7	171, 195, 284, 304	2048 × 2048	50
AIA	SDO	1 – 1.3	94, 131, 171, 193, 211 304, 335, 1600, 1700	4096 × 4096	> 8640

[a]Temporal resolution is only approximate. Telemetry constraints and SNR requirements can change over time and affect the sensor data rates.

instrument package. We will henceforth use COR1A to refer to the COR1 coronagraph on the Ahead satellite, and likewise COR1B will henceforth refer to the COR1 coronagraph on the Behind satellite. The COR1 coronagraphs have a field of view of $1.3-4\,R_\odot$. We use data from $1.5-3\,R_\odot$ to form the electron density reconstructions. The lower radius is restricted to avoid systematic instrumental diffraction clearly visible in the COR1 images and is the same as the lower radius chosen in [112]. The upper radius is restricted only to reduce computational costs for this work. Images from polarization angles of 0°, 120°, and 240° are combined to form pB images, calibrated, and background subtracted using the `SECCHI_PREP` (`http://sohowww.nascom.nasa.gov/solarsoft/stereo/secchi/doc/secchi_prep.html`) IDL routine with the so-called monthly-roll background images. Additionally, the calibration routine applies an intercalibration factor to each COR1A and COR1B image [113]. However, we find that the processed COR1 images still contain a few localized outlier values, due most likely to cosmic rays hitting the detectors. We found that the simple rule of discarding COR1 image pixels with values greater than $5 \times 10^{-8}\,B_\odot$ was sufficient to remove all outliers. During the reconstruction process, the images were binned from 1024×1024 to 256×256 pixels to reduce computational costs.

The data used are from the four-week period between 1 February 2008 and 29 February 2008 of Carrington Rotation (CR) numbers 2066 and 2067. This time period was chosen because it is close to a monthly calibration roll and, as result, the background subtraction will best remove instrumental scattered light. In addition, we choose this period to compare our results in Section 7.5 to [112] which also considers CR 2066. During this time period, the spacecraft were separated by approximately 45°. The data set used in this study consists of four images per instrument, per day. The images used were spaced evenly, every six hours. The acquisition times are 00:05:00 UTC, 06:05:00 UTC, 12:05:00 UTC, and 18:05:00 UTC. If no image is

available for a time, that time is skipped. Thus, there are 115 images used in the COR1A data set and 110 used in the COR1B data set.

7.3 Signal Model Suitability

The primary criticism of using the static approach in Section 2.1 for solar tomography is that the solar atmosphere is not static. The bottom row of images in Figure 7.2 demonstrates that the dynamics of the solar corona can be significant over a period of a 27 d full rotation where the differences between the two images are due to changes in the solar atmosphere. Unphysical smearing artifacts may result in the static reconstructions due to such changes [13]. Depending on the situation, the solar atmosphere may change significantly at time scales of even less than a day.

The dynamics of the solar atmosphere are modeled by the nonlinear equations of magnetohydrodynamics (MHD). However, we do not currently have a viable high-dimensional state estimation procedure for problems posed under the general nonlinear dynamic model in Section 2.2.1. Incorporation of MHD models into our high-dimensional linear methods poses two major difficulties. First, while the MHD equations are functions of the electron density and temperature, they also depend, for example, on the energy density, vector velocity, and vector magnetic field at each spatial location in the solar atmosphere. Including these additional parameters into the signal model would significantly increase the dimensionality of the already large problem. Secondly, the dynamic models in Section 2.2.2 require a linearized representation of the MHD equations, a process that is questionable at the time scales under consideration [2]. A less sophisticated model for the state dynamics would only include the differential rotation of the Sun to account for the fact that the solar atmosphere rotates at different rates depending on latitude. A very simple model that is often used in practice [44] is the random walk model where $\boldsymbol{F}_i = \boldsymbol{I}$ and the

evolution of $\boldsymbol{x}_i$ is then assumed to be entirely stochastic. Despite the simplicity of the random walk model, it can be superior to the static model [32].

7.4 Model Parameter Selection

The linear dynamic model of Section 2.2.2 includes measurement and state noise covariance matrices $\boldsymbol{R}_i$ and $\boldsymbol{Q}_i$, the mean of the initial state $\boldsymbol{x}_1$, the initial state error covariance $\boldsymbol{\Pi}_1$, the state dynamic model $\boldsymbol{F}_i$, and the covariance-taper matrix $\boldsymbol{C}$. The operational procedure used to choose each of these model parameters is described below.

First, we use the state transition model $\boldsymbol{F}_i = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix, which results in a purely random walk evolution model for the electron density. The model is often used in practice when a better temporal dynamic model is not known [44], and although it is simple, a dynamic tomographic reconstruction with a random walk model can be superior to a static reconstruction [32]. Ideally, we would incorporate nonlinear magnetohydrodynamics into the state transition operator $\boldsymbol{F}_i$. Such an approach would require the joint estimation of the energy density, velocity, temperature, and magnetic field in addition to electron density, greatly increasing the dimensionality of the already huge problem [2]. Also, the dynamic reconstruction algorithm in Section 4.4 cannot handle nonlinear operators, although methods do exist for adapting the EnKF to such problems [114]. Incorporating a more sophisticated dynamic model into our dynamic reconstruction framework is the subject of ongoing research.

The covariance taper matrix $\boldsymbol{C}$ trades off some bias in the dynamic estimates for a reduction in computational effort as discussed in [50]. Several choices for the taper matrix were considered on a lower resolution reconstruction grid. We found that the electron density reconstructed with the taper matrix $\boldsymbol{C} = \boldsymbol{I}$, which results in the

greatest reduction in computational effort and largest additional bias, was comparable to the electron densities reconstructed with non-diagonal taper matrices.

We assume that the STEREO/SECCHI-COR1 measurement noise is Poisson. A convenient property of a Poisson random variable is that its mean is equal to its variance. We make the common assumption that the mean of the kth measurement at time index i is given by the kth element of $\boldsymbol{y}_i$. With this choice, the measurement covariance $\boldsymbol{R}_i$ is a diagonal matrix with the diagonal equal to $\boldsymbol{y}_i$. This choice does not account for correlations in the measurement noise, but this information, if known, can easily be incorporated into the off-diagonal elements of $\boldsymbol{R}_i$.

The state noise covariance $\boldsymbol{Q}_i$ is constant in time, i.e. $\boldsymbol{Q}_i = \boldsymbol{Q}$. This assumption holds true under the random walk dynamic model if the temporal electron-density variance and correlations are constant. An approximation to $\boldsymbol{Q}$ is found by first estimating the state noise from the static reconstructions by computing the first-order differences

$$\widehat{\boldsymbol{u}}_i = \widehat{\boldsymbol{x}}^S_{i+1} - \widehat{\boldsymbol{x}}^S_i \tag{7.1}$$

where the static reconstructions are the positively constrained regularized estimates given by

$$\widehat{\boldsymbol{x}}^{\mathrm{S}}_i = \underset{\boldsymbol{x}_i \geq \boldsymbol{0}}{\arg\min} \, \|\boldsymbol{y}_{i-I_{180}:i} - \boldsymbol{H}_{i-I_{180}:i}\, \boldsymbol{x}_i\|_2^2 + \lambda \, \|\boldsymbol{D}\, \boldsymbol{x}_i\|_2^2 \tag{7.2}$$

and I_{180} corresponds to the time interval of 180° of solar rotation (about 14 days). To use a similar notation, we will sometimes refer to the EnKF dynamic reconstructions with the following:

$$\widehat{\boldsymbol{x}}^{\mathrm{D}}_i \triangleq \widetilde{\boldsymbol{x}}_{i|i}. \tag{7.3}$$

The nth element on the diagonal of $\boldsymbol{Q}$ is approximated as the temporal variance in the nth element of the state noise estimate $\widehat{\boldsymbol{u}}_i$. Correlations in the state noise can

also be approximated in this manner to determine the non-diagonal elements of $\boldsymbol{Q}$, but we found that a diagonal model for the state noise was sufficient to obtain the results in Section 7.5.

The remaining model parameters were chosen as follows. The spatial derivative matrix $\boldsymbol{D}$ is a first-order difference approximation to the derivative in both azimuth and elevation (but not radius). The same gradient operator is used as the spatial regularizer in the static and dynamic reconstructions. The EnKF initial state mean $\boldsymbol{x}_1$ is the static estimate $\widehat{\boldsymbol{x}}^S_{180}$. The EnKF initial state error covariance is $\boldsymbol{\Pi}_1 = \sigma^2 \boldsymbol{I}$. The initial state error variance σ^2 was increased until the initial prior had only a minor impact on the final dynamic electron density reconstruction. We found that $\sigma^2 = 10^{10}$ was sufficient to minimize the impact of the initial-state mean. Lastly, with the other parameters set, the regularization parameter λ was varied until the electron-density reconstruction was sufficiently smooth. We found that $\lambda = 1.6 \times 10^{-9}$ was a good choice for the static reconstructions and $\lambda = 1.0 \times 10^{-14}$ worked well in the dynamic case. An unsupervised approach to choosing these remaining model parameters could make use of methods such as cross validation [13], [91].

7.5 Electron Density Reconstruction

We now present tomographic reconstructions of electron density during February 2008 computed from COR1A and COR1B observations. The tomographic reconstructions are best when data are available over at least half a solar rotation. To satisfy this condition, all results are shown at dates after 14 February 2008. Each static reconstruction $\widehat{\boldsymbol{x}}^S_i$ (7.2) required, on average, nearly 30 minutes of computation while the complete dynamic reconstruction, i.e., the computation of $\widetilde{\boldsymbol{x}}_{i|i}$ at all time indices i, required about 40 minutes with an ensemble size $L = 32$.

Figure 7.3 shows a representative sample of spherical shell slices of the 3-D electron-

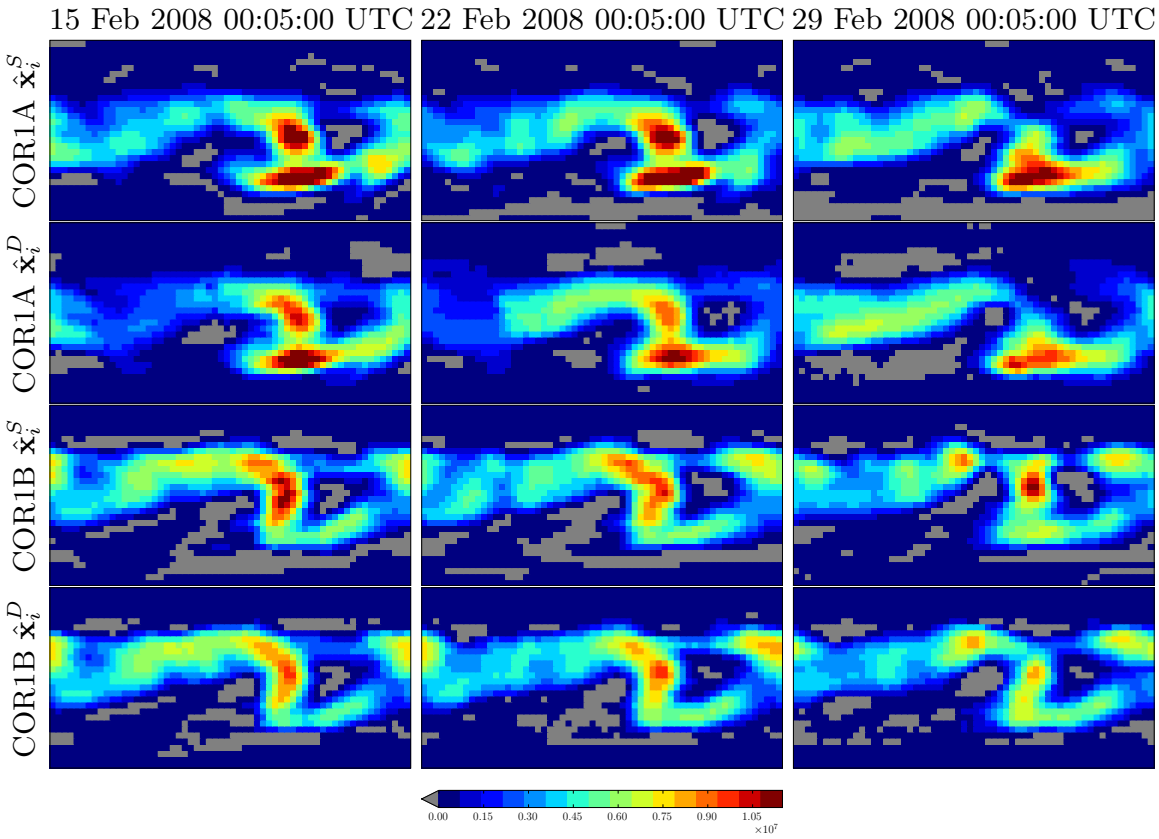

Figure 7.3 Two-dimensional spherical shell slices of reconstructed electron density at radius 1.7 $R_\odot$. Each column shows reconstructed electron density on a particular date. The first row shows the static reconstruction of electron density computed using COR1A pB measurements. The remaining rows show, in order, the dynamic COR1A, static COR1B, and dynamic COR1B electron density reconstructions. The horizontal and vertical axes of each image are the solar longitude and latitude, respectively, in units of degrees ranging from 0° to 360° in the horizontal direction and −90° to 90° in the vertical direction. The electron densities are shown on a common scale as indicated by the color bar in the last row with units of N_e cm^{-3}. All non-positive densities are shown in gray.

density reconstructions. Each spherical shell slice is a latitude versus longitude plot of reconstructed electron density at a constant radius of 1.7 $R_\odot$. Each row of Figure 7.3 shows a COR1A static, COR1A dynamic, COR1B static, or COR1B dynamic electron density reconstruction slice on three dates: 15 February 2008 00:05:00 UTC, 22 February 2008 00:05:00 UTC, and 29 February 2008 00:05:00 UTC.

We observe distinct changes in the structure of the streamer belt as a function of time in all four reconstructions in Figure 7.3. The static reconstructions tend to have larger, more connected non-positive patches that seem to coalesce toward large areas near the poles. The non-positive patches in the dynamic reconstructions tend to cluster around the high density regions in the streamer belt. These observations agree with the assertion in [112] that non-positive patches appear along high density borders. Also, in general, the dynamic reconstructed density is lower, more spread out, and smoother than the static densities. The same effects would occur if the dynamic reconstructions were more significantly spatially regularized than the static reconstructions. This could also imply a larger dynamic relative residual, but this is not the case as discussed below.

In Figure 7.4, we see a 3-D rendering of the coronal electron-density reconstructions on 25 February 2008 18:05:00 UTC. Each of the four reconstructions (COR1A static and dynamic and COR1B static and dynamic) is shown with two isosurfaces. The red isosurface is at a density of 2.5×10^6 N_e cm^{-3} and the blue isosurface shows non-positive patches of reconstructed electron density. In this view, we can see the full 3-D nature of the reconstructed electron density and visualize how the density structure extends throughout the reconstructed volume of the corona. When viewed in 3-D, as opposed to 2-D slices such as in Figure 7.3, it is clear that the higher densities are clustered near the Sun while the non-positive density artifacts are more prominent near the outer boundary of the reconstructed corona. Also, it is clear

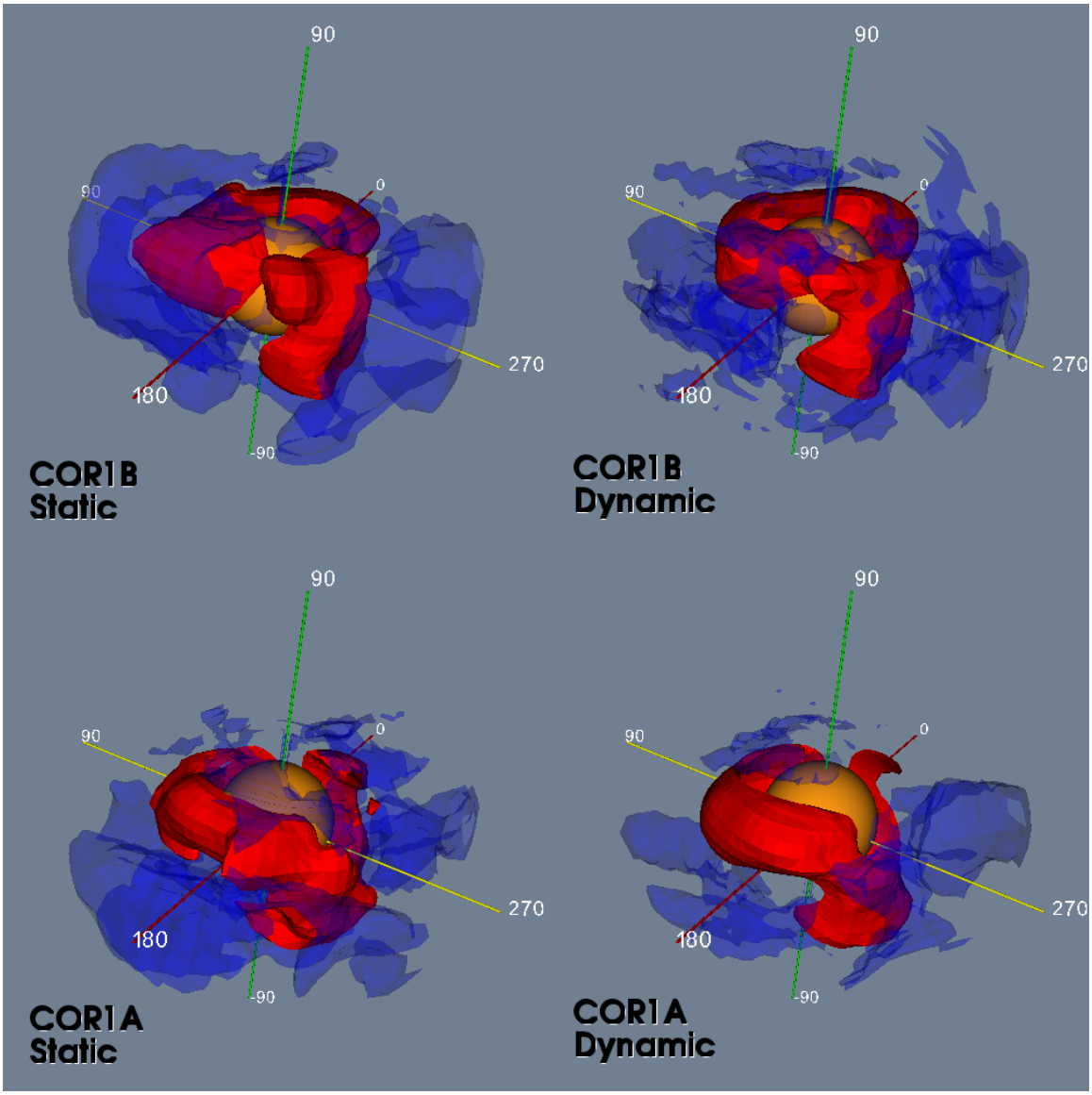

Figure 7.4 Three-dimensional isosurfaces of the four electron-density reconstructions on 25 February 2008 18:05:00 UTC. The red isosurface shows densities at 2.5×10^6 cm^{-3}. The translucent blue surface shows non-positive (non-physical) reconstructed densities. The central orange sphere has radius 1 $R_\odot$ and represents the Sun.

that the non-positive artifacts are spatially much larger in the static reconstructions whereas the dynamic non-positive artifacts are smaller and less connected.

Figure 7.5 examines the agreement between the reconstructed electron density and the observed measurements in a plot of the relative residual defined as

$$r_i^{(S \text{ or } D)} = \frac{\left\| \boldsymbol{y}_i - \boldsymbol{H}_i \widehat{\boldsymbol{x}}_i^{(S \text{ or } D)} \right\|_2}{\|\boldsymbol{y}_i\|_2}. \tag{7.4}$$

The relative residual compares the difference between a pB image and the estimate of that image computed from the forward model and an electron density reconstruction. We note that a small residual indicates good agreement with the measurements but not necessarily good agreement between the reconstruction and the true electron density. Indeed, one major concern with the residual as a performance metric is that it must increase with the inclusion of spatial regularization, but with the trade-off of reduced noise sensitivity and potentially better agreement between the reconstructions and the true coronal electron density [75]. This is exemplified in Figure 7.6 which compares the regularized and unregularized static electron-density reconstructions on 15 February 2008 00:05:00 UTC. We find that the regularized reconstruction is clearly more physically reasonable even though the unregularized reconstruction has a smaller residual. Ultimately, the relative residual shown in Figure 7.5 does not by itself provide definitive proof that the dynamic reconstruction algorithm produces more faithful electron-density reconstructions than the static approach. Instead, the reduction in the relative residual shown in Figure 7.5 coupled with the reduction in nonphysical artifacts discussed below provides evidence to support the conclusion that the dynamic method is better than the static method, but further study is required.

Alternatively, we could evaluate our approach in simulations where the true electron density is known. For reference, note that the static and dynamic approaches

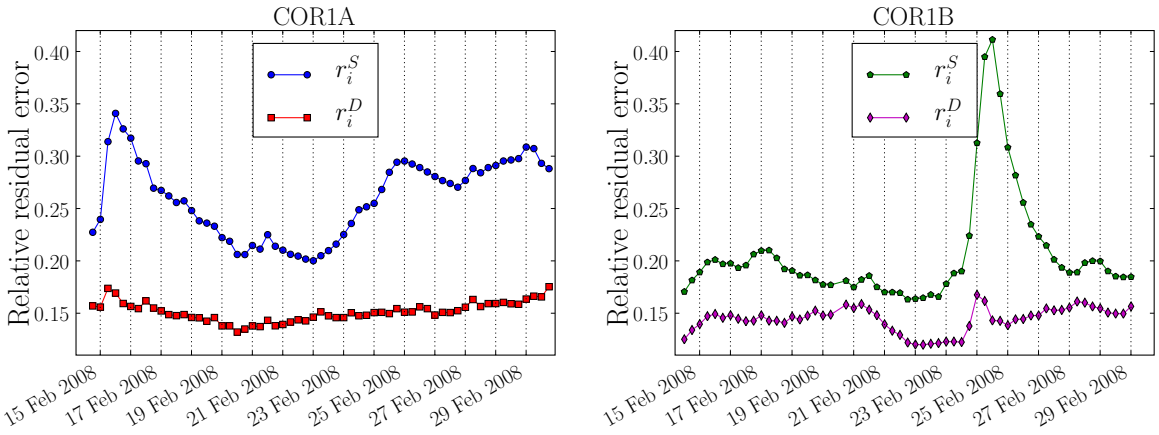

Figure 7.5 The relative static and dynamic residual (defined in (7.4)) for the COR1A (left) and COR1B (right) electron density reconstructions.

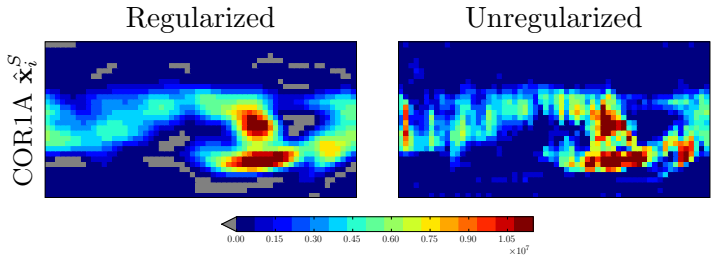

Figure 7.6 A spherical shell slice at radius 1.7 $R_\odot$ of the 15 February 2008 00:05:00 UTC COR1A static electron density reconstruction with regularization (left) and with no regularization (right), i.e. a static reconstruction computed with $\lambda = 0$ in (7.2). Note that the left image is a copy of the upper-left image in Figure 7.3 and is included for easy comparison to the unregularized reconstruction. Both images are shown on the same color scale. The relative residual of the regularized reconstruction is 0.239 and is 0.227 for the unregularized reconstruction.

have been previously evaluated and compared in numerical experiments including the simulated collapse of a magnetized molecular cloud [32], emergence of magnetic flux in the solar corona [2], and diffusion [50]. In the context of this work, a more realistic and thus more compelling comparison of the static and dynamic approaches would consider a simulated time-dependent corona computed using hourly-updated magnetrograms to drive a 3-D magnetohydrodynamic solar wind solution. However, the development of such numerical simulations is an ongoing research effort in the heliospheric modeling community. Though not considered in this work, we plan to evaluate the dynamic and static reconstruction algorithms in more realistic coronal simulations once they are available.

With the above caveats, we now consider the relative residual in Figure 7.5. We see that the dynamic reconstruction typically has half the relative residual of the static reconstruction for both the COR1A and COR1B results. The reduction in the residual could be a result of decreased regularization. However, the results in Figure 7.3 and Figure 7.4 show that the dynamic and static reconstructions have similar qualitative smoothness. The dynamic and static reconstructions also have quantitatively similar smoothness, i.e., $\|\boldsymbol{D}\,\widehat{\boldsymbol{x}}_i^{(S \text{ or } D)}\|_2$ is 1.24×10^7 and 1.14×10^7 in the static and dynamic cases, respectively, for the time index i corresponding to 15 February 2008 00:05:00 UTC. In addition, the relative residuals in the dynamic reconstructions are in fact smaller in comparison to the unregularized static reconstructions. These observations support the conclusion that differences in spatial regularization alone cannot explain the reduction in the dynamic residual. Further, for COR1B, there is a significant jump in the residual starting at 23 February 2008 18:05:00 UTC. Data outliers can cause such jumps, but they have been removed from the pB images prior to computing the electron density reconstructions as discussed in Section 7.2. What is most surprising is that the COR1B dynamic relative residual also exhibits a jump

starting at 23 February 2008 18:05:00 UTC, but it is much less pronounced. One difference between the static and dynamic reconstruction algorithms is that the dynamic method is provided with the measurement noise covariance $\boldsymbol{R}_i$, a measure of data quality. As a result, the dynamic method will be able to compensate for relatively poor quality measurements (assuming $\boldsymbol{R}_i$ is correct) that could severely distort the static electron density reconstructions.

Interestingly, the results presented in the previous figures have shown that the EnKF produces fewer non-positive reconstructed densities than the static method. This can be seen in Figure 7.4, where the dynamic reconstructions have smaller volumes of non-positive artifacts. Figure 7.7 shows the percentage of non-positive voxels as a function of time in the four electron-density reconstructions. We find that the static reconstructions can have as many as twice the number of non-positive voxels. We emphasize that the dynamic reconstruction method produces a more physical result than the static method even though the dynamic algorithm does not constrain the electron density to be positive.

Figure 7.8 shows the average reconstructed density versus altitude above the Sun. The plot shows reconstructed electron density averaged over a spherical shell slice (i.e., a latitude versus longitude slice similar to those in Figure 7.3) at a given radius. All four reconstructions show a similar profile in the falloff of electron density with distance from the Sun, with the COR1A densities generally greater than the COR1B densities and static greater than dynamic. All four reconstructions increase in density at the largest altitudes, a reconstruction artifact common in tomographic imaging when the lines of sight extend to infinity and significant enough density exists outside of the finite reconstruction domain. On the other hand, there seems to be an artifact in the COR1A reconstructions at the lowest altitude that is either not present or not as severe in the COR1B reconstructions. In addition, the electron-density reconstruc-

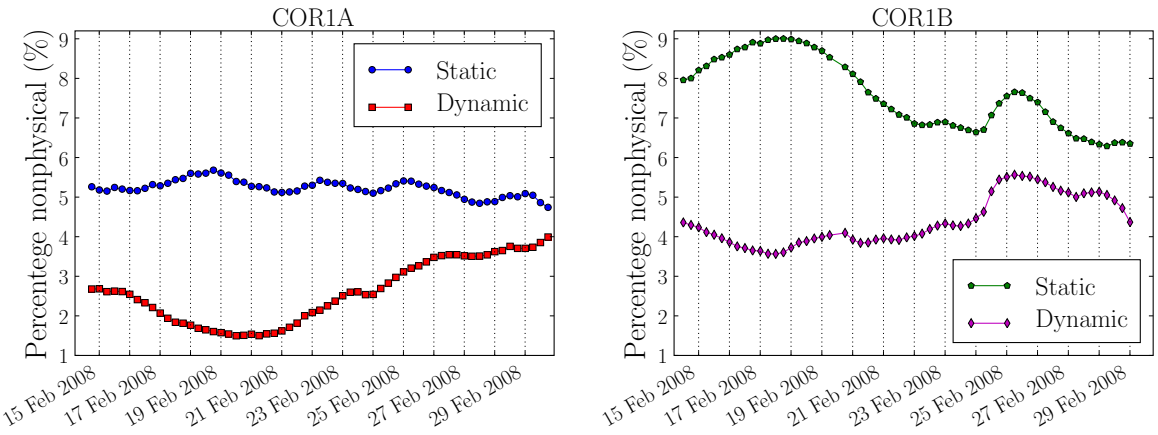

Figure 7.7 The percentage of non-positive reconstructed densities appearing in the COR1A (left) and COR1B (right) electron-density reconstructions.

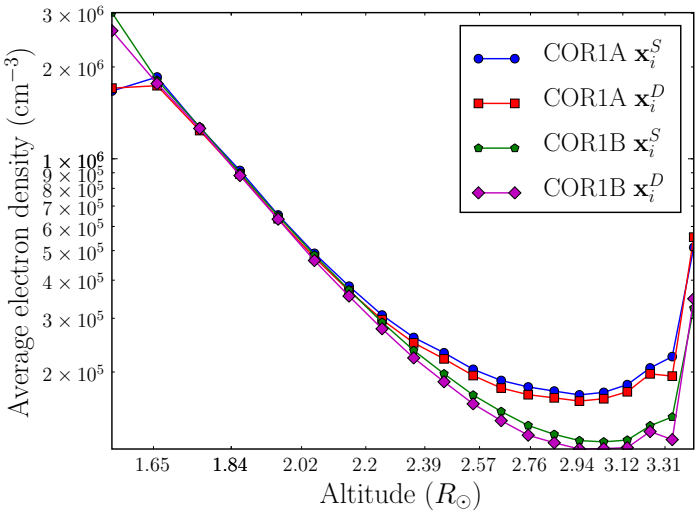

Figure 7.8 A log–log depiction of the reconstructed electron density versus the altitude above the Sun (an altitude of 1 $R_\odot$ is at the photosphere). Each plotted point corresponds to reconstructed electron density averaged over a spherical shell slice at a particular radius. As detailed in the legend, each curve in the figure is associated with a static or dynamic reconstruction (indicated by S or D) based either on COR1A or COR1B measurements.

tions all decay linearly between about 1.7 and 2.2 $R_{\odot}$ on the log–log plot with a fitted slope of about -6, indicating a clear power law in the falloff rate of average electron density versus distance from the Sun.

CHAPTER 8

CONCLUSIONS

This dissertation has developed a state-space formulation to the dynamic tomography problem. The problem has been formulated under several stochastic state-space signal models that offer different trade-offs between generality and ultimately computational tractability. State estimation methods were developed with special attention to computational scalability when faced with high-dimensional problems such as dynamic tomography. The favorable computational aspects of these methods were demonstrated in a series of numerical experiments. This work has also generated theoretical contributions concerning convergence, state smoothing, and spatial-temporal resolution. The final chapter details the first time-dependent reconstruction of the global solar corona and uses the formulation and methods developed in the dissertation.

8.1 Future Work

Continued research extending this work could benefit the statistical signal processing field in a number of ways. This dissertation has focused on estimation under quadratic cost functions, but estimation under the minimax or $\mathcal{H}^\infty$ norm is also possible with the KF [27]. As a result, an adaptation of the EnKF may potentially solve certain minimax estimation problems, opening the path to high-dimensional robust estimation. While this dissertation offers theory regarding expected spatial-temporal

resolution, it has not been applied in an operational setting. A full spatial-temporal resolution characterization of the exterior and hollow X-ray transforms would quantify the achievable limits in solar tomography. We note that theoretical bounds on spatial resolution in tomography [115] could be extended to the dynamic case considered here. Such bounds could help identify the appropriate lag-length for use in fixed-lag smoothing, an approach with lower computational cost than fixed-interval smoothing. Next, the KWF was developed under the proviso that the state dynamic model is linear and shift-invariant. While convenient for the analysis, it is more restrictive than the necessary joint stationarity condition. We note that conditions for the class of linear operators that preserve joint stationarity have been identified [116], [117]. Lastly, a definitive connection between the EnKF and PF does not exist. The similarities between the Monte Carlo methods are clear, but no theory yet relates them. The connection could show how localization principles may be applied to the more general PF and open a path to currently intractable high-dimensional applications.

Several future scientific research directions are also immediately suggested. For one, we have so far considered COR1A and COR1B separately, but tomography improves with additional simultaneous vantage points. A forthcoming paper will address intercalibration issues between COR1A and COR1B and present 4-D electron density reconstructions based on both data sets simultaneously, perhaps combining data from additional coronagraphs depending on further intercalibration issues. Second, the random-walk dynamic model is certainly not the best possible choice for solar tomography. The first step would be to consider the effects of differential rotation which can be incorporated as a spatially varying rotation in the state dynamic model. However, after a number of studies, it is still unclear how the rotation rate depends on latitude and height [118]. We are also currently working towards incorporating MHD physics into our dynamic reconstruction framework. A better dynamic model

will improve our 4-D electron density estimates and, ultimately, enable forecasting of the future electron density state in the solar corona. Further, the solar magnetic field strongly influences the coronal electron density. Models for the solar magnetic-field structure could be a powerful means to further constrain our dynamic tomographic electron density reconstructions [119].

APPENDIX A

CONVERGENCE PROOF

As suggested in [76], Slutsky's theorem [120] is useful in proving that the unlocalized EnKF converges and is repeated here because it is used in the proof of Theorem 4.1.

Theorem A.1 (Slutsky's theorem)**:** Consider two sequences of random vectors denoted $\{\boldsymbol{x}_n,\, n \in \mathbb{N}\}$ and $\{\boldsymbol{y}_n,\, n \in \mathbb{N}\}$, a random vector $\boldsymbol{x}$ with N components, and a constant vector $\boldsymbol{a}$ with M components. Let $f(\boldsymbol{u},\, \boldsymbol{v})$ be a function that maps the M and N dimensional vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ to an $M \times N$ matrix and is continuous at $(\boldsymbol{u},\, \boldsymbol{a})$ for all vectors $\boldsymbol{u}$ in the range of the random vector $\boldsymbol{x}$. Suppose $\boldsymbol{x}_n \xrightarrow{\text{d.}} \boldsymbol{x}$ and $\boldsymbol{y}_n \xrightarrow{\text{p.}} \boldsymbol{a}$ where d. and p. indicate convergence in distribution and probability. Then

$$f(\boldsymbol{x}_n,\, \boldsymbol{y}_n) \xrightarrow{\text{d.}} f(\boldsymbol{x},\, \boldsymbol{a}). \tag{A.1}$$

The proof of Theorem 4.1 follows. The proof will proceed by induction and is a generalization of the proof that the EnKF estimates $\overline{\boldsymbol{x}}_{i|i}$ converge to the LMMSE estimates $\widehat{\boldsymbol{x}}_{i|i}$ [121].

Proof. The first step is to prove the base case by showing that the initial EnKF estimate $\widetilde{\boldsymbol{x}}_{1|0}$ and error covariance $\widetilde{\boldsymbol{P}}_{1|0}$ converge to the LKF estimate $\widetilde{\boldsymbol{x}}^{\infty}_{1|0}$ and error covariance $\widetilde{\boldsymbol{P}}^{\infty}_{1|0}$. The second step is to show that the posterior ensemble mean $\widetilde{\boldsymbol{x}}_{i|i}$ and covariance $\widetilde{\boldsymbol{P}}_{i|i}$ converge to the LKF estimate $\widetilde{\boldsymbol{x}}^{\infty}_{i|i}$ and error covariance $\widetilde{\boldsymbol{P}}^{\infty}_{i|i}$ under

the assumption that the prior ensemble mean and covariance converge, i.e., $\widetilde{\boldsymbol{x}}_{i|i-1} \xrightarrow{\text{p.}} \widetilde{\boldsymbol{x}}^{\infty}_{i|i-1}$ and $\widetilde{\boldsymbol{P}}_{i|i-1} \xrightarrow{\text{p.}} \widetilde{\boldsymbol{P}}^{\infty}_{i|i-1}$. The final step is to show that the EnKF time update modifies the posterior ensemble such that the resultant ensemble sample mean and covariance converge to the prior mean and covariance necessary at the next time step, i.e., $\widetilde{\boldsymbol{x}}_{i+1|i} \xrightarrow{\text{p.}} \widetilde{\boldsymbol{x}}^{\infty}_{i+1|i}$ and $\widetilde{\boldsymbol{P}}_{i+1|i} \xrightarrow{\text{p.}} \widetilde{\boldsymbol{P}}^{\infty}_{i+1|i}$. The inductive proof that $\widetilde{\boldsymbol{x}}_{i|i} \xrightarrow{\text{p.}} \widetilde{\boldsymbol{x}}^{\infty}_{i|i}$ at each time index i is then complete.

Step 1: Because the initial EnKF ensemble is a set of i.i.d. samples (4.18), the sample mean and covariance will converge in probability to the population mean and covariance by the weak law of large numbers (WLLN), i.e.,

$$\widetilde{\boldsymbol{x}}_{1|0} \xrightarrow{\text{p.}} \boldsymbol{\mu}_1 = \widetilde{\boldsymbol{x}}^{\infty}_{1|0} \tag{A.2}$$

$$\widetilde{\boldsymbol{P}}_{1|0} \xrightarrow{\text{p.}} \boldsymbol{\Pi}_1 = \widetilde{\boldsymbol{P}}^{\infty}_{1|0}. \tag{A.3}$$

Step 2: The next step is to show that the EnKF estimate $\widetilde{\boldsymbol{x}}_{i|i}$ converges to the LKF estimate $\widetilde{\boldsymbol{x}}^{\infty}_{i|i}$ under the following assumptions on the sample mean and covariance of the prior ensemble:

$$\widetilde{\boldsymbol{x}}_{i|i-1} \xrightarrow{\text{p.}} \widetilde{\boldsymbol{x}}^{\infty}_{i|i-1} \tag{A.4}$$

$$\widetilde{\boldsymbol{P}}_{i|i-1} \xrightarrow{\text{p.}} \widetilde{\boldsymbol{P}}^{\infty}_{i|i-1}. \tag{A.5}$$

To begin, the posterior samples $\widetilde{\boldsymbol{x}}^{l}_{i|i}$ can be written in terms of the prior samples by expanding (4.21):

$$\widetilde{\boldsymbol{x}}^{l}_{i|i} = \widetilde{\boldsymbol{x}}^{l}_{i|i-1} + \widetilde{\boldsymbol{K}}_i \, \boldsymbol{y}^{l}_i - \widetilde{\boldsymbol{K}}_i \, \boldsymbol{H}_i \, \widetilde{\boldsymbol{x}}^{l}_{i|i-1}. \tag{A.6}$$

The posterior estimate can then be written as a function of the prior samples:

$$\widetilde{\boldsymbol{x}}_{i|i} = \widetilde{\boldsymbol{x}}_{i|i-1} + \widetilde{\boldsymbol{K}}_i \left(\frac{1}{L} \sum_{l=1}^{L} \boldsymbol{y}_i^l \right) - \widetilde{\boldsymbol{K}}_i \, \boldsymbol{H}_i \, \widetilde{\boldsymbol{x}}_{i|i-1}. \tag{A.7}$$

The first term in (A.7) converges in probability to the LKF estimate $\widetilde{\boldsymbol{x}}^{\infty}_{i|i-1}$ by assumption (A.4). The two remaining terms in (A.7) involve the localized sample Kalman gain $\widetilde{\boldsymbol{K}}_i$ (4.19) which must first be characterized before proceeding further. Under the assumption on the prior sample error covariance (A.5) and the fact that the Hadamard matrix product is a linear operation, the term $(\boldsymbol{C}_i \circ \widetilde{\boldsymbol{P}}_{i|i-1}) \, \boldsymbol{H}_i^T$ appearing in (4.19) converges in probability to $(\boldsymbol{C}_i \circ \widetilde{\boldsymbol{P}}^{\infty}_{i|i-1}) \, \boldsymbol{H}_i^T$. Note that the matrix

$$\boldsymbol{H}_i \, (\boldsymbol{C}_i \circ \widetilde{\boldsymbol{P}}_{i|i-1}) \, \boldsymbol{H}_i^T + \boldsymbol{R}_i \tag{A.8}$$

inverted in the computation of (4.19) is positive definite and hence invertible because it is the sum of the positive semidefinite matrix $\boldsymbol{H}_i(\boldsymbol{C}_i \circ \widetilde{\boldsymbol{P}}_{i|i-1})\boldsymbol{H}_i^T$ and the positive definite measurement error covariance matrix $\boldsymbol{R}_i$. Then, $\widetilde{\boldsymbol{K}}_i$ converges in distribution to the localized Kalman gain $\widetilde{\boldsymbol{K}}_i^{\infty}$ by Slutsky's theorem. Also by Slutsky's theorem, the second and third terms in (A.7) converge in distribution to $\widetilde{\boldsymbol{K}}_i^{\infty} \, \boldsymbol{y}_i$ and $\widetilde{\boldsymbol{K}}_i^{\infty} \, \boldsymbol{H}_i \, \widetilde{\boldsymbol{x}}^{\infty}_{i|i-1}$, respectively. Note that $\boldsymbol{y}_i$ and $\widetilde{\boldsymbol{x}}^{\infty}_{i|i-1}$ are nonrandom constant vectors upon the observation of the measurements $\boldsymbol{y}_{1:i}$. Convergence in distribution to a constant also implies convergence in probability to that constant [120]. Therefore, the three terms in (A.7) converge in probability, and, because convergence in probability is closed under addition, the end result is

$$\widetilde{\boldsymbol{x}}_{i|i} \xrightarrow{\text{p.}} \widetilde{\boldsymbol{x}}^{\infty}_{i|i-1} + \widetilde{\boldsymbol{K}}_i^{\infty} \, \boldsymbol{y}_i - \widetilde{\boldsymbol{K}}_i^{\infty} \, \boldsymbol{H}_i \, \widetilde{\boldsymbol{x}}^{\infty}_{i|i-1} = \widetilde{\boldsymbol{x}}^{\infty}_{i|i}. \tag{A.9}$$

Now, consider the convergence of the posterior sample error covariance matrix $\widetilde{\boldsymbol{P}}_{i|i}$

which can be expressed as the sum of four terms by expanding (4.20) (with $j = i-1$) and substituting (A.6) and (A.7) for the posterior samples $\widetilde{\boldsymbol{x}}_{i|i}^{l}$ and sample mean $\widetilde{\boldsymbol{x}}_{i|i}$:

$$\widetilde{\boldsymbol{P}}_{i|i} = \frac{1}{L-1}\Bigg[\boldsymbol{A}_i \sum_{l=1}^{L} \boldsymbol{z}_{i|i-1}^{l} \left(\boldsymbol{z}_{i|i-1}^{l}\right)^T \boldsymbol{A}_i^T + \boldsymbol{A}_i \sum_{l=1}^{L} \boldsymbol{z}_{i|i-1}^{l} \left(\boldsymbol{w}_i^{l}\right)^T \widetilde{\boldsymbol{K}}_i^T + \widetilde{\boldsymbol{K}}_i \sum_{l=1}^{L} \boldsymbol{w}_i^{l} \left(\boldsymbol{z}_{i|i-1}^{l}\right)^T \boldsymbol{A}_i^T + \widetilde{\boldsymbol{K}}_i \sum_{l=1}^{L} \boldsymbol{w}_i^{l} \left(\boldsymbol{w}_i^{l}\right)^T \widetilde{\boldsymbol{K}}_i^T\Bigg]. \quad \text{(A.10)}$$

The following symbols have been introduced above to conserve space: $\boldsymbol{A}_i \triangleq \boldsymbol{I} - \widetilde{\boldsymbol{K}}_i \boldsymbol{H}_i$, $\boldsymbol{z}_{i|i-1}^{l} \triangleq \widetilde{\boldsymbol{x}}_{i|i-1}^{l} - \widetilde{\boldsymbol{x}}_{i|i-1}$, and $\boldsymbol{w}_i^{l} \triangleq \boldsymbol{y}_i^{l} - \frac{1}{L}\sum_{k=1}^{L} \boldsymbol{y}_i^{k}$. The first term of (A.10) converges in distribution to

$$\left(\boldsymbol{I} - \widetilde{\boldsymbol{K}}_i^{\infty} \boldsymbol{H}_i\right) \widetilde{\boldsymbol{P}}_{i|i-1}^{\infty} \left(\boldsymbol{I} - \widetilde{\boldsymbol{K}}_i^{\infty} \boldsymbol{H}_i\right)^T \quad \text{(A.11)}$$

by Slutsky's theorem and the assumption on the prior sample error covariance (A.5). Both the second and third terms involve a summation of the form

$$\frac{1}{L-1} \sum_{l=1}^{L} \boldsymbol{z}_{i|i-1}^{l} \left(\boldsymbol{w}_i^{l}\right)^T \quad \text{(A.12)}$$

or its transpose. The convergence of (A.12) will first be examined component-wise, i.e, consider the (m,n)th component of (A.12) expanded as

$$\frac{1}{L-1}\left(\sum_{l=1}^{L} \left[\boldsymbol{z}_{i|i-1}^{l}\right]_m \left[\boldsymbol{y}_i^{l}\right]_n - \sum_{l=1}^{L} \left[\boldsymbol{z}_{i|i-1}^{l}\right]_m \left[\overline{\boldsymbol{y}}_i\right]_n\right) \quad \text{(A.13)}$$

where $[\cdot]_m$ is the mth component of the vector argument and $\overline{\boldsymbol{y}}_i \triangleq \frac{1}{L}\sum_{l=1}^{L} \boldsymbol{y}_i^{l}$. Note that the second term above simplifies:

$$\sum_{l=1}^{L} \left[\boldsymbol{z}_{i|i-1}^{l}\right]_m \left[\overline{\boldsymbol{y}}_i\right]_n = \left(\sum_{l=1}^{L} \left[\widetilde{\boldsymbol{x}}_{i|i-1}^{l} - \widetilde{\boldsymbol{x}}_{i|i-1}\right]_m\right) \left[\overline{\boldsymbol{y}}_i\right]_n = 0. \quad \text{(A.14)}$$

Now, consider the first term of (A.13)

$$\sum_{l=1}^{L} \left[\boldsymbol{z}_{i|i-1}^{l}\right]_m \left[\boldsymbol{y}_i^l\right]_n = \sum_{l=1}^{L} \left[\boldsymbol{z}_{i|i-1}^{l}\right]_m \left[\boldsymbol{y}_i\right]_n + \sum_{l=1}^{L} \left[\boldsymbol{z}_{i|i-1}^{l}\right]_m \left[\boldsymbol{v}_i^l\right]_n \tag{A.15}$$

where $\boldsymbol{v}_i^l \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0},\, \boldsymbol{R}_i)$. Then note that $\mathbb{E}\Big(\left[\boldsymbol{z}_{i|i-1}^{l}\right]_m\Big) = 0$ because the samples in the prior ensemble $\{\boldsymbol{x}_{i|i-1}^{1},\, \ldots,\, \boldsymbol{x}_{i|i-1}^{L}\}$ are identically distributed. As a result the first term on the right-hand side (RHS) of (A.15) converges to 0, i.e.,

$$\frac{1}{L-1} \sum_{l=1}^{L} \left[\boldsymbol{z}_{i|i-1}^{l}\right]_m \left[\boldsymbol{y}_i\right]_n \xrightarrow{\text{p.}} \mathbb{E}\big(\left[\boldsymbol{z}_{i|i-1}^{l}\right]_m\big) \left[\boldsymbol{y}_i\right]_n = 0 \tag{A.16}$$

by the WLLN. Also note that the summands in the second term on the RHS of (A.15) are mutually uncorrelated because

$$\operatorname{Cov}\Big(\left[\boldsymbol{z}_{i|i-1}^{k}\right]_m \left[\boldsymbol{v}_i^k\right]_n, \left[\boldsymbol{z}_{i|i-1}^{l}\right]_m \left[\boldsymbol{v}_i^l\right]_n\Big) = 0 \tag{A.17}$$

for all $k \neq l$. Then the second term on the RHS of (A.15) also converges to 0, i.e.,

$$\frac{1}{L-1} \sum_{l=1}^{L} \left[\boldsymbol{z}_{i|i-1}^{l}\right]_m \left[\boldsymbol{v}_i^l\right]_n \xrightarrow{\text{p.}} \mathbb{E}\big(\left[\boldsymbol{z}_{i|i-1}^{l}\right]_m\big)\, \mathbb{E}\big(\left[\boldsymbol{v}_i^l\right]_n\big) = 0 \tag{A.18}$$

by the L^2 WLLN [122, pg. 36] and the fact that $\boldsymbol{z}_{i|i-1}^{l}$ and $\boldsymbol{v}_i^l$ are independent. Then the first term on the RHS of (A.13) also converges to 0. Therefore, each component of (A.12) converges to 0. Component-wise convergence then implies that (A.12) converges in probability to $\boldsymbol{0}$, the matrix with all elements equal to 0. The fourth term of (A.10) converges to $\widetilde{\boldsymbol{K}}_i^{\infty} \boldsymbol{R}_i \, (\widetilde{\boldsymbol{K}}_i^{\infty})^T$ by Slutsky's theorem. The first and forth terms of (A.10) converge in distribution to a constant which also implies convergence in probability to that constant. As a result, the convergence of each of the four terms

of (A.10) then implies

$$\widetilde{\boldsymbol{P}}_{i|i} \xrightarrow{\text{p.}} (\boldsymbol{I} - \widetilde{\boldsymbol{K}}_i^{\infty} \boldsymbol{H}_i)\, \widetilde{\boldsymbol{P}}_{i|i-1}^{\infty}\, (\boldsymbol{I} - \widetilde{\boldsymbol{K}}_i^{\infty} \boldsymbol{H}_i)^T + \widetilde{\boldsymbol{K}}_i^{\infty} \boldsymbol{R}_i\, (\widetilde{\boldsymbol{K}}_i^{\infty})^T = \widetilde{\boldsymbol{P}}_{i|i}^{\infty}. \tag{A.19}$$

Step 3: The last step is to show that the EnKF time update results in an ensemble with sample mean $\widetilde{\boldsymbol{x}}_{i+1|i}$ and covariance $\widetilde{\boldsymbol{P}}_{i+1|i}$ that satisfy the assumptions (A.4) and (A.5) at the next time index (time index $i+1$) to complete the inductive chain. First, consider the ensemble sample mean after the EnKF time update (4.22) given by

$$\widetilde{\boldsymbol{x}}_{i+1|i} = \boldsymbol{F}_i\, \widetilde{\boldsymbol{x}}_{i|i} + \frac{1}{L} \sum_{l=1}^{l} \boldsymbol{u}_i^l. \tag{A.20}$$

The first term converges in probability to $\boldsymbol{F}_i \widetilde{\boldsymbol{x}}_{i|i}^{\infty}$ by (A.9). The second term converges in probability to the vector $\mathbf{0}$ by the WLLN. The result is

$$\widetilde{\boldsymbol{x}}_{i+1|i} \xrightarrow{\text{p.}} \boldsymbol{F}_i\, \widetilde{\boldsymbol{x}}_{i|i}^{\infty} = \widetilde{\boldsymbol{x}}_{i+1|i}^{\infty}. \tag{A.21}$$

The ensemble sample covariance after the time update can be expanded as

$$\widetilde{\boldsymbol{P}}_{i+1|i} = \frac{1}{L-1} \Bigg[\boldsymbol{F}_i \sum_{l=1}^{L} \boldsymbol{z}_{i|i}^l\, (\boldsymbol{z}_{i|i}^l)^T \boldsymbol{F}_i^T + \boldsymbol{F}_i \sum_{l=1}^{L} \boldsymbol{z}_{i|i}^l\, (\boldsymbol{u}_i^l - \overline{\boldsymbol{u}}_i)^T + \sum_{l=1}^{L} (\boldsymbol{u}_i^l - \overline{\boldsymbol{u}}_i)\, (\boldsymbol{z}_{i|i}^l)^T \boldsymbol{F}_i^T + \sum_{l=1}^{L} (\boldsymbol{u}_i^l - \overline{\boldsymbol{u}}_i)\, (\boldsymbol{u}_i^l - \overline{\boldsymbol{u}}_i)^T \Bigg] \tag{A.22}$$

with $\boldsymbol{z}_{i|i}^l \triangleq \widetilde{\boldsymbol{x}}_{i|i}^l - \widetilde{\boldsymbol{x}}_{i|i}$, and $\overline{\boldsymbol{u}}_i \triangleq \frac{1}{L} \sum_{l=1}^{L} \boldsymbol{u}_i^l$. The first term converges to $\boldsymbol{F}_i\, \widetilde{\boldsymbol{P}}_{i|i}^{\infty}\, \boldsymbol{F}_i^T$ by (A.19). The independence between the samples $\boldsymbol{z}_{i|i}^l$ and $\boldsymbol{u}_i^k$ implies that the second and third terms of (A.22) converge in probability to the matrix $\mathbf{0}$ by a similar argument used to show that (A.12) converges in probability to the matrix $\mathbf{0}$. The last term converges in probability to the state noise covariance matrix $\boldsymbol{Q}_i$ by the WLLN.

The result is

$$\widetilde{\boldsymbol{P}}_{i+1|i} \xrightarrow{\text{p.}} \boldsymbol{F}_i \widetilde{\boldsymbol{P}}^{\infty}_{i|i} \boldsymbol{F}_i^T + \boldsymbol{Q}_i = \widetilde{\boldsymbol{P}}^{\infty}_{i+1|i}. \tag{A.23}$$

□

APPENDIX B

IMPLEMENTATION DETAILS

B.1 Sequential Processing

The KF can sequentially process the components of the measurement $\boldsymbol{y}_i$ when the measurement noise is white and the noise covariance $\boldsymbol{R}_i$ is a diagonal matrix [26], [27]. The main computational advantage is that the $M \times M$ matrix inversion in the evaluation of the Kalman gain (4.11) is reduced to scalar division. Note that the whitening filter $\boldsymbol{R}_i^{-1/2}$ (the inverse matrix square root of $\boldsymbol{R}_i$) may always be applied to the measurements and sequential processing is always an option. However, $\boldsymbol{R}_i^{-1/2}$ is generally computed by first taking the Cholesky factorization of the measurement covariance $\boldsymbol{R}_i$, an operation of the same computational complexity as matrix inversion [72]. Ultimately, sequential processing is advantageous only when the measurement noise is white, when the measurement noise statistics are constant in time and the whitening filter needs to be computed only once, or when the inverse square root of $\boldsymbol{R}_i$ may be easily computed.

The EnKF, with any taper matrix $\boldsymbol{C}_i$, may also process the measurements sequentially and inherits the same benefits and caveats as the KF with sequential processing. An additional benefit is that the product $(\boldsymbol{C}_i \circ \widetilde{\boldsymbol{P}}_{i|i-1})\,\boldsymbol{H}_i^T$ is then an N element vector in (4.19) and the storage requirement of the intermediate result is

significantly reduced. However, the bias introduced by the covariance taper will increase with sequential processing because the tapering operation is then applied once for each component of $\boldsymbol{y}_i$ instead of once for the complete measurement vector. Block processing of the measurements is an option that can trade off the computational advantage of sequential processing for reduced covariance taper bias.

B.2 Tikhonov Regularization

An additional quadratic penalty term may be appended to (4.16) by augmenting the forward model (2.10). Johns and Mandel are the first to use this methodology in the context of the EnKF [123]. However, they do not cite the earlier work of Baroudi, Kaipio, and Somersalo who first suggest the approach in the context of the KF [28].

The procedure begins by augmenting the forward model in the following manner:

$$\begin{bmatrix} \boldsymbol{y}_i \\ \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_i \\ \boldsymbol{D}_i \end{bmatrix} \boldsymbol{x}_i + \begin{bmatrix} \boldsymbol{v}_i \\ \boldsymbol{w}_i \end{bmatrix} \tag{B.1}$$

which may be written succinctly as

$$\boldsymbol{y}'_i = \boldsymbol{H}'_i \, \boldsymbol{x}_i + \boldsymbol{v}'_i \tag{B.2}$$

where the augmented measurement noise vector $\boldsymbol{v}'_i$ is zero mean, uncorrelated with the initial state $\boldsymbol{x}_1$ and state noise $\boldsymbol{u}_i$, and has covariance

$$\boldsymbol{R}'_i \triangleq \mathbb{E}\big[\boldsymbol{v}'_i \, (\boldsymbol{v}'_i)^T\big] = \begin{bmatrix} \boldsymbol{R}_i & \boldsymbol{0} \\ \boldsymbol{0} & \lambda_i^{-1} \, \boldsymbol{I} \end{bmatrix}. \tag{B.3}$$

The resultant measurement update can be expressed as

$$\widehat{\boldsymbol{x}}_{i|i} = \arg\min_{\boldsymbol{x}_i} \|\boldsymbol{y}_i - \boldsymbol{H}_i\,\boldsymbol{x}_i\|^2_{\boldsymbol{R}_i^{-1}} + \|\boldsymbol{x}_i - \widehat{\boldsymbol{x}}_{i|i-1}\|^2_{\boldsymbol{P}^{-1}_{i|i-1}} + \lambda_i\,\|\boldsymbol{D}_i\,\boldsymbol{x}_i\|_2^2 \tag{B.4}$$

where λ_i is the Tikhonov regularization parameter [74], [75]. With an appropriate choice for $\boldsymbol{D}_i$, a prescribed degree of spatial smoothness can be enforced on the estimates computed with the PF, KF, and EnKF by replacing the model parameters $\boldsymbol{y}_i$, $\boldsymbol{H}_i$, and $\boldsymbol{R}_i$ with the augmented model parameters $\boldsymbol{y}'_i$, $\boldsymbol{H}'_i$, and $\boldsymbol{R}'_i$ in the measurement updates of the filters.

APPENDIX C

THE KHATRI-RAO MATRIX PRODUCT

Definition C.1: Let $\boldsymbol{A}$ be a $P \times N$ matrix and $\boldsymbol{B}$ be a $M \times N$ matrix. The Khatri-Rao product of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ is defined by

$$\boldsymbol{A} \odot \boldsymbol{B} = \begin{bmatrix} [\boldsymbol{A}]_{(:,1)} \otimes [\boldsymbol{B}]_{(:,1)} & \cdots & [\boldsymbol{A}]_{(:,N)} \otimes [\boldsymbol{B}]_{(:,N)} \end{bmatrix} \tag{C.1}$$

where $\otimes$ is the Kronecker matrix product, $[\cdot]_{(:,n)}$ selects the nth column of the matrix argument, and $A \odot B$ has dimensions $PM \times N$.

REFERENCES

[1] N. Gordon, D. Salmond, and C. Ewing, "Bayesian state estimation for tracking and guidance using the bootstrap filter," *Journal of Guidance, Control, and Dynamics*, vol. 18, pp. 1434–1443, 1995.

[2] M. D. Butala, F. Kamalabadi, R. A. Frazin, and Y. Chen, "Dynamic tomographic imaging of the solar corona," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, pp. 755–766, 2008.

[3] G. Evensen, "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *Journal of Geophysical Research*, vol. 99, pp. 10 143–10 162, 1994.

[4] J. Geweke, "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica*, vol. 57, pp. 1317–1339, 1989.

[5] N. P. Willis and Y. Bresler, "Optimal scan for time varying tomographic imaging I: Theoretical analysis and fundamental limitations," *IEEE Transactions on Image Processing*, vol. 4, pp. 642–653, 1995.

[6] N. P. Willis and Y. Bresler, "Optimal scan for time varying tomographic imaging II: Efficient design and experimental validation," *IEEE Transactions on Image Processing*, vol. 4, pp. 654–666, 1995.

[7] J. L. Duerk and O. P. Simonetti, "Review of MRI gradient waveform design methods with application in the study of motion," *Concepts in Magnetic Resonance*, vol. 5, pp. 105–122, 1993.

[8] T. Ernst, O. Speck, L. Itti, and L. Chang, "Simultaneous correction for interscan patient motion and geometric distortions in echoplanar imaging," *Magnetic Resonance in Medicine*, vol. 42, pp. 201–205, 1999.

[9] Z.-P. Liang and P. C. Lauterbur, *Principles of Magnetic Resonance Imaging*. New York, NY: Wiley-IEEE Press, 2000.

[10] S. Bonnet, A. Koenig, S. Roux, P. Hugonnard, R. Guillemaud, and P. Grangeat, "Dynamic X-ray computed tomography," *Proceedings of the IEEE*, vol. 91, pp. 1574–1587, 2003.

[11] H. C. van de Hulst, "The electron density of the solar corona," *Bulletin of the Astronomical Institutes of the Netherlands*, vol. 11, pp. 135–150, 1950.

[12] R. A. Frazin, "Tomography of the solar corona. I. A robust, regularized, positive estimation method," *Astrophysical Journal*, vol. 530, pp. 1026–1035, 2000.

[13] M. D. Butala, R. A. Frazin, and F. Kamalabadi, "Three-dimensional estimates of the coronal electron density at times of extreme solar activity," *Journal of Geophysical Research*, vol. 110, p. A09S09, 2005.

[14] A. K. George, M. D. Butala, R. A. Frazin, F. Kamalabadi, and Y. Bresler, "Time-resolved CT reconstruction using the ensemble Kalman filter," in *Proceedings of the IEEE International Symposium on Biomedical Imaging*, Paris, France, 2008, pp. 1489–1492.

[15] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian estimation," *IEE Proceedings F – Radar and Signal Processing*, vol. 140, pp. 107–113, 1993.

[16] J. S. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems," *Journal of the American Statistical Association*, vol. 93, pp. 1032–1044, 1998.

[17] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. New York, NY: Springer-Verlag, 2001.

[18] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, pp. 174–188, 2002.

[19] P. M. Djurić, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Míguez, "Particle filtering," *IEEE Signal Processing Magazine*, vol. 20, pp. 19–38, 2003.

[20] F. Daum and J. Huang, "Curse of dimensionality and particle filters," in *Proceedings of the IEEE Aerospace Conference*, vol. 4, Big Sky, MT, 2003, pp. 1979–1993.

[21] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, pp. 197–208, 2000.

[22] T. Bengtsson, P. Bickel, and B. Li, "Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems," *Probability and Statistics: Essays in Honor of David A. Freedman*, vol. 2, pp. 316–334, 2008.

[23] N. Vaswani, "Particle filtering for large dimensional state spaces with multimodal observation likelihoods," *IEEE Transactions on Signal Processing*, vol. 56, pp. 4583–4597, 2008.

[24] D. Watzenig, M. Brandner, and G. Steiner, "A particle filter approach for tomographic imaging based on different state-space representations," *Measurement Science and Technology*, vol. 18, pp. 30–40, 2007.

[25] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.

[26] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[27] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, NJ: Prentice-Hall, 2000.

[28] D. Baroudi, J. Kaipio, and E. Somersalo, "Dynamical electric wire tomography: A time series approach," *Inverse Problems*, vol. 14, pp. 799–813, 1998.

[29] M. Vahkonen, P. A. Karjalainen, and J. P. Kaipio, "A Kalman filter approach to track fast impedance changes in electrical impedance tomography," *IEEE Transactions on Biomedical Engineering*, vol. 45, pp. 486–493, 1998.

[30] A. Seppänen, M. Vauhkonen, P. J. Vauhkonen, E. Somersalo, and J. P. Kaipio, "State estimation with fluid dynamical evolution models in process tomography - an application to impedance tomography," *Inverse Problems*, vol. 17, pp. 467–483, 2001.

[31] V. Kolehmainen, S. Prince, S. R. Arridge, and J. P. Kaipio, "State-estimation approach to the nonstationary optical tomography problem," *Journal of the Optical Society of America*, vol. 20, pp. 876–888, 2003.

[32] R. A. Frazin, M. D. Butala, A. Kemball, and F. Kamalabadi, "Time-dependent reconstruction of nonstationary objects with tomographic or interferometric measurements," *Astrophysical Journal Letters*, vol. 635, pp. L197–L200, 2005.

[33] R. Daley, *Atmospheric Data Analysis*. New York, NY: Cambridge University Press, 1991.

[34] J. M. Lewis and J. C. Derber, "The use of adjoint equations to solve a variational adjustment problem with advective constraints," *Tellus*, vol. 37A, pp. 309–322, 1985.

[35] D. P. Rohler, "A covariance-subset Kalman filter and the image reconstruction problem," Ph.D. dissertation, Case Western Reserve University, Cleveland, OH, 1983.

[36] R. W. Schunk, L. Scherliess, J. J. Sojka, D. C. Thompson, D. N. Anderson, M. Codrescu, C. Minter, T. J. Fuller-Rowell, R. A. Heelis, M. Hairston, and B. M. Howe, "Global assimilation of ionospheric measurements (GAIM)," *Radio Science*, vol. 39, p. RS1S02, 2004.

[37] G. A. Hajj, B. D. Wilson, C. Wang, X. Pi, and I. G. Rosen, "Data assimilation of ground GPS total electron content into a physics-based ionosphere model by use of the Kalman filter," *Radio Science*, vol. 39, p. RS1S05, 2004.

[38] A. Asif and J. M. F. Moura, "Data assimilation in large time-varying multidimensional fields," *IEEE Transactions on Image Processing*, vol. 8, pp. 1593–1607, 1999.

[39] T. M. Chin, A. J. Mariano, and E. P. Chassignet, "Spatial regression and multiscale approximations for sequential data assimilation in ocean models," *Journal of Geophysical Research*, vol. 104, pp. 7991–8014, 1999.

[40] H. S. Oster and Y. Rudy, "The use of temporal information in the regularization of the inverse problem of electrocardiography," *IEEE Transactions on Biomedical Engineering*, vol. 39, pp. 65–75, 1992.

[41] D. H. Brooks, G. F. Ahmad, R. S. MacLeod, and G. M. Maratos, "Inverse electrocardiography by simultaneous imposition of multiple constraints," *IEEE Transactions on Biomedical Engineering*, vol. 46, pp. 3–18, 1999.

[42] U. Schmitt and A. K. Louis, "Efficient algorithms for the regularization of dynamic inverse problems: I. Theory," *Inverse Problems*, vol. 18, pp. 645–658, 2002.

[43] F. Greensite, "The temporal prior in bioelectromagnetic source imaging problems," *IEEE Transactions on Biomedical Engineering*, vol. 50, pp. 1152–1159, 2003.

[44] Y. Zhang, A. Ghodrati, and D. H. Brooks, "An analytical comparison of three spatio-temporal regularization methods for dynamic linear inverse problems in a common statistical framework," *Inverse Problems*, vol. 21, pp. 357–382, 2005.

[45] M. A. Cane, A. Kaplan, R. N. Miller, B. Tang, E. C. Hackert, and A. J. Busalacchi, "Mapping tropical Pacific sea level: Data assimilation via a reduced state Kalman filter," *Journal of Geophysical Research*, vol. 101, pp. 22 599–22 617, 1996.

[46] L. Scherliess, R. W. Schunk, J. J. Sojka, and D. C. Thompson, "Development of a physics-based reduced state Kalman filter for the ionosphere," *Radio Science*, vol. 39, p. RS1S04, 2004.

[47] P. W. Fieguth, D. Menemenlis, and I. Fukumori, "Mapping and pseudoinverse algorithms for ocean data assimilation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, pp. 43–51, 2003.

[48] G. Burgers, P. J. van Leeuwen, and G. Evensen, "Analysis scheme in the ensemble Kalman filter," *Monthly Weather Review*, vol. 126, pp. 1719–1724, 1998.

[49] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*. Berlin, Germany: Springer-Verlag, 2007.

[50] M. D. Butala, R. A. Frazin, Y. Chen, and F. Kamalabadi, "Tomographic imaging of dynamic objects with the ensemble Kalman filter," *IEEE Transactions on Image Processing*, vol. 18, pp. 1573–1587, 2009.

[51] P. L. Houtekamer and H. L. Mitchell, "A sequential ensemble Kalman filter for atmospheric data assimilation," *Monthly Weather Review*, vol. 129, pp. 123–137, 2001.

[52] T. M. Hamill, J. S. Whitaker, and C. Snyder, "Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter," *Monthly Weather Review*, vol. 129, pp. 2776–2790, 2001.

[53] J.-H. Park and A. Kaneko, "Assimilation of coastal acoustic tomography data into a barotropic ocean model," *Geophysical Research Letters*, vol. 27, pp. 3373–3376, 2000.

[54] M. D. Butala and F. Kamalabadi, "Optimal dynamic tomography for wide-sense stationary spatial random fields," in *Proceedings of the IEEE International Conference on Image Processing*, Cairo, Egypt, 2009, pp. 593–596.

[55] G. Christakos, *Random Field Models in Earth Sciences*. San Diego, CA: Academic Press, 1992.

[56] F. Natterer, *The Mathematics of Computerized Tomography*. New York, NY: Wiley, 1986.

[57] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction: Space-invariant tomographs," *IEEE Transactions on Image Processing*, vol. 5, pp. 1346–1358, 1996.

[58] M. Butala, R. Hewett, R. Frazin, and F. Kamalabadi, "Dynamic three-dimensional tomography of the solar corona," *Solar Physics*, vol. 262, pp. 495–509, 2010.

[59] O. Cohen, I. V. Sokolov, I. I. Roussev, C. N. Arge, W. B. Manchester, T. I. Gombosi, R. A. Frazin, H. Park, M. D. Butala, F. Kamalabadi, and M. Velli, "A semiempirical magnetohydrodynamical model of the solar wind," *Astrophysical Journal Letters*, vol. 654, pp. L163–L166, 2007.

[60] N. Barbey, F. Auchère, T. Rodet, and J.-C. Vial, "A time-evolving 3D method dedicated to the reconstruction of solar plumes and results using extreme ultraviolet data," *Solar Physics*, vol. 248, pp. 409–423, 2008.

[61] E. Vanmarcke, *Random Fields: Analysis and Synthesis*. Cambridge, MA: MIT Press, 1983.

[62] A. W. Naylor and G. R. Sell, *Linear Operator Theory in Engineering and Science*. New York, NY: Springer-Verlag, 1982.

[63] B. Hajek, "Notes for ECE 534: An exploration of random processes for engineers," 2009. [Online]. Available: http://www.ifp.illinois.edu/~hajek/Papers/randomprocesses.html

[64] R. M. Perry, "On reconstructing a function on the exterior of a disk from its radon transform," *Journal of Mathematical Analysis and Applications*, vol. 59, pp. 324–341, 1977.

[65] E. T. Quinto, "Singular value decomposition and inversion methods for the exterior Radon transform and a spherical transform," *Journal of Mathematical Analysis and Applications*, vol. 95, pp. 437–448, 1983.

[66] M. D. Altschuler, "Reconstruction of the global-scale three-dimensional solar corona," in *Image Reconstructions from Projections: Implementation and Applications*. Berlin, Germany: Springer-Verlag, 1979, pp. 105–145.

[67] T. J. Bogdan and B. C. Low, "The three-dimensional structure of magnetostatic atmospheres. II - Modeling the large-scale corona," *Astrophysical Journal*, vol. 306, pp. 271–283, 1986.

[68] T. Neukirch, "On self-consistent three-dimensional analytic solutions of the magnetohydrostatic equations." *Astronomy and Astrophysics*, vol. 301, pp. 628–639, 1995.

[69] R. A. Frazin, F. Kamalabadi, and M. A. Weber, "On the combination of differential emission measure analysis and rotational tomography for 3D solar EUV imaging," *Astrophysical Journal*, vol. 628, pp. 1070–1080, 2005.

[70] E. Landi, G. Del Zanna, P. R. Young, K. P. Dere, H. E. Mason, and M. Landini, "CHIANTI - an atomic database for emission lines - VII. New data for X-rays and other improvements," *Astrophysical Journal Supplement Series*, vol. 162, pp. 261–280, 2006.

[71] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. New York, NY: Cambridge University Press, 1991.

[72] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins University Press, 1996.

[73] A. H. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Signal Processing Magazine*, vol. 11, pp. 18–60, 1994.

[74] A.-I. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*. Washington, DC: Winston, 1977.

[75] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 2024–2036, 1989.

[76] R. Furrer and T. Bengtsson, "Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants," *Journal of Multivariate Analysis*, vol. 98, pp. 227–255, 2007.

[77] J. R. Guerci, "Theory and application of covariance matrix tapers for robust adaptive beamforming," *IEEE Transactions on Signal Processing*, vol. 47, pp. 977–985, 1999.

[78] P. J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *Annals of Statistics*, vol. 36, pp. 199–227, 2008.

[79] G. Gaspari and S. E. Cohn, "Construction of correlation functions in two and three dimensions," *Quarterly Journal of the Royal Meteorological Society*, vol. 125, pp. 723–757, 1999.

[80] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice Hall, 1993.

[81] A. K. Jain and S. Ansari, "Radon transform theory for random fields and optimum image reconstruction from noisy projections," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 9, 1984, pp. 495–498.

[82] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[83] A. E. Bryson and M. Frazier, "Smoothing for linear and nonlinear dynamic systems," Wright-Patterson Air Force Base, Dayton, OH, Technical Report TDR-63-119, 1963.

[84] P. J. van Leeuwen and G. Evensen, "Data assimilation and inverse methods in terms of a probabilistic formulation," *Monthly Weather Review*, vol. 124, pp. 2898–2913, 1996.

[85] G. Evensen and P. J. van Leeuwen, "An ensemble Kalman smoother for nonlinear dynamics," *Monthly Weather Review*, vol. 128, pp. 1852–1867, 2000.

[86] S. P. Khare, J. L. Anderson, T. J. Hoar, and D. Nychka, "An investigation into the application of an ensemble Kalman smoother to high-dimensional geophysical systems," *Tellus*, vol. 60A, pp. 97–112, 2008.

[87] J. R. Guerci, *Space-Time Adaptive Processing for Radar*. Norwood, MA: Artech House, 2003.

[88] S. U. Pillai, J. R. Guerci, and S. R. Pillai, "Efficient tapering methods for STAP," in *Proceedings of the IEEE Radar Conference*, Philadelphia, PA, 2004, pp. 408–413.

[89] C. R. Rao and M. B. Rao, *Matrix Algebra and Its Applications to Statistics and Econometrics*. Singapore: World Scientific Publishing Company, 1998.

[90] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. Philadelphia, PA: IEEE Press, 1988.

[91] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, pp. 215–223, 1979.

[92] A. C. Bovik, Ed., *Handbook of Image and Video Processing*, 2nd ed. Burlington, MA: Academic Press, 2005.

[93] K. A. Khalsa and J. A. Fessler, "Resolution properties in regularized dynamic MRI reconstruction," in *Proceedings of the IEEE International Symposium on Biomedical Imaging*, 2007, pp. 456–459.

[94] S. Y. Chun and J. A. Fessler, "Spatial resolution and noise properties of regularized motion-compensated image reconstruction," in *Proceedings of the IEEE International Symposium on Biomedical Imaging*, 2009, pp. 863–866.

[95] D. N. Baker, "What is space weather?" *Advances in Space Research*, vol. 22, pp. 7–16, 1998.

[96] M. J. Aschwanden, *Physics of the Solar Corona*. Berlin, Germany: Springer-Verlag, 2004.

[97] F. Natterer and F. Wübbeling, *Mathematical Methods in Image Reconstruction*. Philadelphia, PA: SIAM, 2001.

[98] M. D. Altschuler and R. M. Perry, "On determining the electron density distribution of the solar corona from K-coronameter data," *Solar Physics*, vol. 23, pp. 410–428, 1972.

[99] R. A. Frazin and P. Janzen, "Tomography of the solar corona. II. Robust, regularized, positive estimation of the three-dimensional electron density distributions from LASCO-C2 polarized white-light images," *Astrophysical Journal*, vol. 570, pp. 408–422, 2002.

[100] J. T. Mueller, H. Maldonado, and A. S. Driesman, "STEREO: The challenges," *Acta Astronautica*, vol. 53, pp. 729–738, 2003.

[101] H. Alfvén, "Existence of electromagnetic-hydrodynamic waves," *Nature*, vol. 150, pp. 405–406, 1942.

[102] P. A. Davidson, *An Introduction to Magnetohydrodynamics*. Cambridge, United Kingdom: Cambridge University Press, 2001.

[103] D. F. Elmore, J. T. Burkepile, A. J. Darnell, A. R. Lecinski, and A. L. Stanger, "Calibration of a ground-based solar coronal polarimeter," in *Proceedings of the Society of Photo-Optical Instrumentation Engineers*, vol. 4843, 2003, pp. 66–75.

[104] G. E. Brueckner, R. A. Howard, M. J. Koomen, C. M. Korendyke, D. J. Michels, J. D. Moses, D. G. Socker, K. P. Dere, P. L. Lamy, A. Llebaria, M. V. Bout, R. Schwenn, G. M. Simnett, D. K. Bedford, and C. J. Eyles, "The large angle spectroscopic coronagraph (LASCO)," *Solar Physics*, vol. 162, pp. 357–402, 1995.

[105] V. Domingo, B. Fleck, and A. I. Poland, "The SOHO mission: An overview," *Solar Physics*, vol. 162, pp. 1–37, 1995.

[106] R. A. Howard, J. D. Moses, A. Vourlidas, J. S. Newmark, D. G. Socker, S. P. Plunkett, C. M. Korendyke, J. W. Cook, A. Hurley, J. M. Davila, W. T. Thompson, O. C. St Cyr, E. Mentzell, K. Mehalick, J. R. Lemen, J. P. Wuelser, D. W. Duncan, T. D. Tarbell, C. J. Wolfson, A. Moore, R. A. Harrison, N. R. Waltham, J. Lang, C. J. Davis, C. J. Eyles, H. Mapson-Menard, G. M. Simnett, J. P. Halain, J. M. Defise, E. Mazy, P. Rochus, R. Mercier, M. F. Ravet, F. Delmotte, F. Auchere, J. P. Delaboudinière, V. Bothmer, W. Deutsch, D. Wang, N. Rich, S. Cooper, V. Stephens, G. Maahs, R. Baugh, D. McMullin, and T. Carter, "Sun Earth connection coronal and heliospheric investigation (SECCHI)," *Space Science Reviews*, vol. 136, pp. 67–115, 2008.

[107] J.-P. Delaboudinière, G. E. Artzner, J. Brunaud, A. H. Gabriel, J. F. Hochedez, F. Millier, X. Y. Song, B. Au, K. P. Dere, R. A. Howard, R. Kreplin, D. J. Michels, J. D. Moses, J. M. Defise, C. Jamar, P. Rochus, J. P. Chauvineau, J. P.

Marioge, R. C. Catura, J. R. Lemen, L. Shing, R. A. Stern, J. B. Gurman, W. M. Neupert, A. Maucherat, F. Clette, P. Cugnon, and E. L. van Dessel, "EIT: Extreme-ultraviolet imaging telescope for the SOHO mission," *Solar Physics*, vol. 162, pp. 291–312, 1995.

[108] M. A. Weber, E. E. Deluca, L. Golub, and A. L. Sette, "Temperature diagnostics with multichannel imaging telescopes," in *Proceedings IAU Symposium No. 223*, A. V. Stepanov, E. Benevolenskaya, and A. Kosovichev, Eds. Cambridge, MA: Cambridge University Press, 2004, pp. 321–328.

[109] L. Golub, E. DeLuca, G. Austin, J. Bookbinder, D. Caldwell, P. Cheimets, J. Cirtain, M. Cosmo, P. Reid, A. Sette, M. Weber, T. Sakao, R. Kano, K. Shibasaki, H. Hara, S. Tsuneta, K. Kumagai, T. Tamura, M. Shimojo, J. McCracken, J. Carpenter, H. Haight, R. Siler, E. Wright, J. Tucker, H. Rutledge, M. Barbera, G. Peres, and S. Varisco, "The X-ray telescope (XRT) for the Hinode mission," *Solar Physics*, vol. 243, pp. 63–86, 2007.

[110] T. Kosugi, K. Matsuzaki, T. Sakao, T. Shimizu, Y. Sone, S. Tachikawa, T. Hashimoto, K. Minesugi, A. Ohnishi, T. Yamada, S. Tsuneta, H. Hara, K. Ichimoto, Y. Suematsu, M. Shimojo, T. Watanabe, S. Shimada, J. M. Davis, L. D. Hill, J. K. Owens, A. M. Title, J. L. Culhane, L. K. Harra, G. A. Doschek, and L. Golub, "The Hinode (Solar-B) mission: An overview," *Solar Physics*, vol. 243, pp. 3–17, 2007.

[111] W. T. Thompson, J. M. Davila, R. R. Fisher, L. E. Orwig, J. E. Mentzell, S. E. Hetherington, R. J. Derro, R. E. Federline, D. C. Clark, P. T. Chen, J. L. Tveekrem, A. J. Martino, J. Novello, R. P. Wesenberg, O. C. St Cyr, N. L. Reginald, R. A. Howard, K. I. Mehalick, M. J. Hersh, M. D. Newman, D. L. Thomas, G. Card, and D. Elmore, "The COR1 inner coronagraph for STEREO-SECCHI," *Proceedings of the Society of Photo-Optical Instrumentation Engineers*, vol. 4853, pp. 1–11, 2003.

[112] M. Kramar, S. Jones, J. Davila, B. Inhester, and M. Mierla, "On the tomographic reconstruction of the 3D electron density for the solar corona from STEREO COR1 data," *Solar Physics*, vol. 259, pp. 109–121, 2009.

[113] W. T. Thompson and N. L. Reginald, "The radiometric and pointing calibration of SECCHI COR1 on STEREO," *Solar Physics*, vol. 250, pp. 443–454, 2008.

[114] G. Evensen, "The ensemble Kalman filter: Theoretical formulation and practical implementation," *Ocean Dynamics*, vol. 53, pp. 343–367, 2003.

[115] J. A. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography," *IEEE Transactions on Image Processing*, vol. 5, pp. 493–506, 1996.

[116] P. A. Franaszek and B. Liu, "On a class of linear time-varying filters," *IEEE Transactions on Information Theory*, vol. IT-13, pp. 477–481, 1967.

[117] P. A. Franaszek, "On linear systems which preserve wide sense stationarity," *SIAM Journal on Applied Mathematics*, vol. 15, pp. 1481–1484, 1967.

[118] S. Giordano and S. Mancuso, "Coronal rotation at solar minimum from UV observations," *Astrophysical Journal*, vol. 688, pp. 656–668, 2008.

[119] T. Wiegelmann and B. Inhester, "Magnetic field modeling and tomograph: First steps towards a consistence reconstruction of the solar corona," *Solar Physics*, vol. 214, pp. 287–312, 2003.

[120] S. F. Arnold, *Mathematical Statistics*. Englewood Cliffs, NJ: Prentice-Hall, 1990.

[121] M. D. Butala, J. Yun, Y. Chen, R. A. Frazin, and F. Kamalabadi, "Asymptotic convergence of the ensemble Kalman filter," in *Proceedings of the IEEE International Conference on Image Processing*, San Diego, CA, 2008, pp. 825–828.

[122] R. Durett, *Probability: Theory and Examples*, 2nd ed. Belmont, CA: Duxbury Press, 1995.

[123] C. J. Johns and J. Mandel, "A two-stage ensemble Kalman filter for smooth data assimilation," *Environmental and Ecological Statistics*, vol. 15, pp. 101–110, 2008.