Piqué Muntané, I. & Fernández-Getino García, M. J. (20-22 July 2022). *Parametric Approximation to Optimal Averaging in Superimposed Training Schemes under Realistic Time-Variant Channels* [proceedings]. In 13th IEEE/IET International Symposium on Communication Systems, Networks and Digital Signal Processing, Porto, Portugal.

DOI: 10.1109/CSNDSP54353.2022.9907996

Parametric Approximation to Optimal Averaging in Superimposed Training Schemes under Realistic Time-Variant Channels

Ignasi Piqué Muntané
Department of Signal Theory and Communications
Carlos III University of Madrid
Madrid, Spain
ignasip@tsc.uc3m.es

M. Julia Fernández-Getino García
Department of Signal Theory and Communications
Carlos III University of Madrid
Madrid, Spain
mjulia@tsc.uc3m.es

***Abstract*—Superimposed Training (ST) with orthogonal frequency division multiplexing (OFDM) scheme has become an attractive solution to meet the goals of the fifth generation (5G) of mobile communications, by improving the channel estimation performance, which is one of the main challenge in multiple input multiple output (MIMO) systems. This technique does not hinder the throughput, however, it introduces an intrinsic interference since the data and the reference symbols are sent together. In order to mitigate it, several studies propose a time averaging over several OFDM received symbols, where the optimal length of this averaging can be analytically computed by solving a transcendental equation. In this paper, this optimal averaging is approximated by a low complexity parametric approach based on a multiple linear regression model that inputs two parameters, the signal-to-noise ratio (SNR) and the relative speed between the transmitter and receiver, which effectively represents the variability of the channel in time. Results show that the approximated solutions give an error of 0.05% on average and 7% at most in terms of the provided mean square error (MSE) of the channel estimation.**

***Index Terms*—OFDM, Superimposed Training, Time-Variant Channel, Channel Estimation, Least Squares, Averaging, Multiple Linear Regression.**

I. Introduction

Orthogonal frequency division multiplexing (OFDM) is the selected scheme to meet the goals of the 5-th generation (5G) of wireless communications [1]. In conjunction with the millimeter wave (mmWave) technique and massive multiple input multiple output (MIMO) layout, a very robust and reliable transmission with high throughput can be achieved [2], [3]. The main problem with this deployment is the channel estimation stage, which is mandatory in order to effectively transmit data. The massive number of channel links to be estimated forces the system to set apart a lot of reference symbols, also known as pilots, which, at the end, hinder the capacity [4].

For this reason, superimposed training (ST) technique has become an appealing solution because the pilot symbols are added over the data symbols, and the transmission of both is performed at the same time-frequency resource [5]. The main benefit of this scheme is the improvement of the throughput, since there is not an exclusive allocation of resources for signalling. However, as a downside, an intrinsic interference due to the superimposed pilot existence cannot be avoided.

In order to mitigate this interference, a general approach to ST stipulates that before estimating the channel, a mandatory averaging of the received signal must be performed in the time domain [5]–[8]. Thus, the power of the noise and the interference can be significantly reduced, yielding a much better channel estimation.

Thanks to the channel correlation and its high resemblance over time, the length of the averaging can be optimized analytically, as [8] proved. Nonetheless, these optimal solutions have to be computed by solving a transcendental equation, which sometimes, it may imply an increase of the computational cost of the system. With that in mind, in this paper, a parametrical approach, based on a multiple linear regression model, aims to fit the optimum averaging values and provide approximate solutions.

The paper contributions are summarized as follows:

- The trend and intrinsic relationships between the optimum averaging solutions and the coherence number of symbols has been illustrated. Also, a multiple linear regression model has been proposed to approximate them in terms of two parameters, the signal-to-noise ratio (SNR), and the relative speed between the transmitter and the receiver.
- The boundary where valid input parameters can be employed, so the regression polynomial is fitted to the original values has been defined. Also, the cleansed data points from this region have been used to estimate the weights of the model.
- The performance and accuracy of the model has been addressed, and the parametric solutions have been compared in terms of the error between the mean square error (MSE) of the channel estimation at the approximated averaging values and the minimum state-of-the-art MSE.

This work has been supported by a fellowship from the Spanish Ministry of Science and Innovation, under grant PRE2018-084315, and, by the Spanish National Project IRENE-EARTH (PID2020-115323RB-C33 / AEI / 10.13039/501100011033).

The remainder of this paper is organized as follows. In Section II, it is defined the system model where a channel model with a realistic correlation in the time domain is introduced. In Sections III and IV, the optimum averaging of the MSE is addressed analytically and parametrically, respectively. Then, in Section V some numerical results verify the analysis and illustrate the accuracy of the multiple linear regression model. Finally, in Section VI, the conclusions are summarized.

Notation: x, $\mathbf{x}$ and $\mathbf{X}$ represent a scalar, a vector, and a matrix, respectively, where $\left[\mathbf{X}\right]_{(m,:)}$ are the elements from the m-th row; $(\cdot)^T$ and $(\cdot)^{-1}$ denote the transpose and the inverse operation of a matrix, respectively; $\otimes$ is the matrix Kronecker product and $||$ refers to the concatenation of two vectors; $\mathbb{E}\left\{\cdot\right\}$ is the mean, and $\lfloor\cdot\rceil$ is the nearest integer operation; $\mathbb{N}_0$, $\mathbb{R}_{>0}$ and $\mathbb{R}_{\geq 0}$, are the sets of non-negative integers, positive and non-negative real numbers, respectively; and, $(\cdot)^c$ is the complementary of a set. In addition, any reference to the logarithmic scale is defined with "log" in base 10.

II. System model

In this section, the transmitted and received signal of the ST scheme are described, and the realistic channel model is presented. This analysis is intended for a downlink (DL) transmission scheme of a system with K subcarriers.

A. Transmitter

To start with, the ST technique is uniformly implemented within all the resources of the OFDM grid. The transmitted signal x_m^k for the k-th subcarrier and m-th OFDM symbol is computed as the addition of the data symbols s_m^k over superimposed pilot symbols c_m^k,

$$x_m^k = s_m^k + c_m^k, \quad \begin{cases} \forall\, k \ \in [0, \cdots, K-1] \\ \forall\, m \ \in \mathbb{N}_0 \end{cases} , \tag{1}$$

where the power of the complete symbol is P. Then, the ratio between the power of data symbols (P_s) and superimposed pilots (P_c) is defined with the power allocation factor $\big(\beta \in [0,1]\big)$ as follows,

$$P = P_s + P_c \begin{cases} P_s = (1-\beta)P \\ P_c = \beta P \end{cases} . \tag{2}$$

Once the signal is generated, it enters the OFDM transmitter scheme where some guardbands are appended in order to avoid intercarrier interference (ICI), an inverse discrete Fourier transform (IDFT) converts the signal in the time domain, and the cyclic prefix is added in order to mitigate the intersymbol interference (ISI).

B. Receiver

An OFDM receiver scheme with perfect synchronization of the received signal is considered, in which the CP of each OFDM symbol is removed, a discrete Fourier transform (DFT) converts the signal back into the frequency domain and the previously appended guardbands are extracted, too. Then, the k-th subcarrier and m-th OFDM received symbol is obtained,

$$y_m^k = H_m^k x_m^k + w_m^k = H_m^k s_m^k + H_m^k c_m^k + w_m^k \ , \tag{3}$$

with H_m^k being the channel and w_m^k being the additive white Gaussian noise (AWGN) coefficients, both in the frequency domain. Also, these random variables behave as complex Gaussian processes that obey the statistics of $H_m^k \sim \mathcal{CN}\left(0, \frac{\sigma_h^2}{K}\right)$ and $w_m^k \sim \mathcal{CN}\left(0, \sigma_w^2\right)$, with σ_h^2 and σ_w^2 being the powers of the channel model and the AWGN, respectively. As it can be seen, the multiplicative nature of the OFDM scheme shows how the data symbols interfere over the pilots, thus, complicating the channel estimation process.

For this reason, it is common to perform an averaging of the received signal in order to reduce the noise power, and mitigate the data interference [5]–[8]. Since this operation is performed at each subcarrier, the averaging of N_t OFDM symbols ($N_t \in \mathbb{N}_0$) in the time domain for the k-th subcarrier can be expressed as follows,

$$\bar{y}^k = \sum_{m'=0}^{N_t-1} \frac{1}{N_t} y_{m'}^k \ . \tag{4}$$

Then, the least squares (LS) algorithm computes the channel estimation minimizing the following cost function [9],

$$J_{cost}(\widehat{H}_{\text{LS}}^k) = \left| \frac{c^*}{\beta P}\, \bar{y}^k - \widehat{H}_{\text{LS}}^k \right|^2 \quad \rightarrow \quad \widehat{H}_{\text{LS}}^k = \frac{c^*}{\beta P}\, \bar{y}^k \ , \tag{5}$$

where the previously averaged signal provides the channel estimation estimation $\widehat{H}_{\text{LS}}^k$ for all the averaged symbols of the k-th subcarrier. Moreover, the linear and logarithmic expression of SNR are defined, respectively, as follows,

$$SNR = \frac{P\sigma_h^2}{\sigma_w^2} \quad , \quad SNR^{dB} = 10\,\log\big(SNR\big) \ . \tag{6}$$

C. Channel model

The channel model employed in this analysis follows the realistic correlation profile in the time domain. Its general expression [8] between the m-th and m'-th OFDM symbol, where any correlation model (ρ_t) can be used, is simplified as,

$$\mathbb{E}\left\{ \left(H_m^k\right)^* H_{m'}^k \right\} = \frac{\sigma_h^2}{K} \rho_t\left(\gamma\, \Delta m\right) \ , \tag{7}$$
$$\Delta m = |m - m'| \ , \quad \gamma = 2\pi \frac{f_d}{\Delta f}\left(1 + \frac{L_{cp}}{K}\right)$$

where γ is a variable that takes into account the effect of the OFDM scheme, with Δf being the subcarrier spacing, L_{cp} being the number of samples of the CP, and f_d the Doppler frequency. Implicitly, the speed between the user equipment (UE) and the base station (BS), defined as v, in km/h, is involved in this correlation since the $f_d = \frac{v}{3.6\,c} f_c$, where c is the speed of light constant and f_c is the carrier frequency at which the signal is modulated.

$$MSE = \Psi_v(N_t) = \frac{\sigma_h^2}{K}\left(1 - \frac{1}{(N_t)^2}\sum_{m_1=0}^{N_t-1}\sum_{m_2=0}^{N_t-1}\rho_t\left(\gamma\,|m_1 - m_2|\right)\right) + \frac{1}{N_t\beta}\left(\frac{\sigma_h^2}{K}(1-\beta) + \frac{\sigma_w^2}{P}\right) \tag{9}$$

$$\psi_v(n_t) = \frac{\sigma_h^2}{K}\left(1 - \frac{2}{(\gamma\, n_t)^2}\Big(\cos(\gamma\, n_t) + \gamma\, n_t\, \mathrm{Si}(\gamma\, n_t) - 1\Big)\right) + \frac{1}{n_t\beta}\left(\frac{\sigma_h^2}{K}(1-\beta) + \frac{\sigma_w^2}{P}\right) \tag{10}$$

In the literature, an approximation of the coherence time (T_{coh}) [10], or equivalently the coherence number of symbols (N_c), is defined as,

$$T_{coh} \sim \frac{0.423}{f_d}\,, \quad N_c \sim \left\lfloor \frac{T_{coh}}{T_{sym}} \right\rceil = \left\lfloor \frac{\kappa}{v} \right\rceil \sim N_c(v)\,, \tag{8}$$
$$T_{sym} = \frac{\left(1 + \frac{L_{cp}}{K}\right)}{\Delta f}\;\;,\quad \kappa = \frac{0.423 \cdot 3.6\, c}{f_c}\frac{\Delta f}{\left(1 + \frac{L_{cp}}{K}\right)}$$

where N_c is computed after applying a unit conversion to T_{coh}, in which the factor is the symbol duration of the OFDM scheme (T_{sym}), and κ is a constant that gathers all the previous parameters. This rule-of-thumb computation determines the duration such that the channel coefficients remain fairly constant with a correlation, at least, higher than 70% [11]. It must be noted the dependency of N_c to v, henceforth, any reference to this value will be equivalent to $N_c(v)$.

III. Optimal Averaging Based on Solving a Transcendental Equation

In this section, the optimal length of OFDM symbols to be averaged is presented. This analytical solution, which is computed following a calculus approach, guarantees the minimum MSE of the channel estimation.

First of all, as the analysis in [8] showed, the MSE expression of the channel estimation is presented in (9). This MSE takes into account the correlation of the channel coefficients (ρ_t) since it is integrated inside the parenthesis. Additionally, the number of symbols to be averaged (N_t) is an important feature because it defines how much of the integration must be computed. Thus, if the averaging is performed within very correlated channel coefficients, the integration term is almost equal to N_t^2, and the overall MSE is reduced, as expected.

From this expression, the MSE can be computed with any correlation profile, e.g. $\rho_t(\gamma\Delta m) = \mathrm{sinc}(\gamma\Delta m)$, which is an approximation of the realistic correlation model from [12], that can simplify the MSE formula as (10) [8]. In this case, it is worth noting that the MSE has been extended into the continuous domain ($\Psi_v \to \psi_v$), as well as its argument ($N_t \to n_t \in \mathbb{R}_{>0}$). Then, following a calculus approach, analytical solutions of the optimum averaging can be obtained by solving the transcendental equation,

$$\mathrm{Si}\left(\gamma\, n_t^{opt}\right) - \gamma\, n_t^{opt}\,\mathrm{sinc}^2\left(\frac{\gamma}{2}\, n_t^{opt}\right) - \frac{\gamma\left(1 + \frac{K}{SNR} - \beta\right)}{2\,\beta} = 0\,. \tag{11}$$

Finally, the optimum averaging can be computed after applying the closest integer value to the continuous solution, $N_t^{opt} = \left\lfloor n_t^{opt} \right\rceil$, so the minimum MSE can be achieved.

IV. Proposed Parametric Approximation Based on a Multiple Linear Regression

As previous Section III presented, the optimum solutions N_t^{opt} are computed by finding the roots of a transcendental equation for a given γ and SNR parameters. Implicitly, since γ depends on v (7), the obtention of the optimum solutions implies that for each input pair of parameters (v, SNR), a root-finding algorithm, e.g. the Newton method, must be executed. In fact, its computational cost depends on the effectivity of the initial point introduced in the algorithm [13].

For this reason, it is proposed a pragmatic and low-complex approach based on approximating the optimum averaging solutions through a regression model that depends on the input arguments. As the results from [8] showed, the optimum averaging values (N_t^{opt}) follow an almost log-log linear relationship with two parameters: the SNR, and the coherence number of symbols (N_c). Thus, the implementation of these parametric curves will enhance the cost effectivity of computing approximated averaging solutions, and overall, it will improve the energy saving of the system.

It is well-known that regression models are widely employed in machine learning (ML), which try to fit curves to observed data points by describing a relationship between dependent and independent variables [14]. Although the multiple linear regression model is not the most popular model due to its multiple downsides [15], e.g. the high influence that far points may interfere in the regression, or the difficuties that polynomials show by not achieving rapid turns, it has been selected as the best approach. The main reason to employ this model is that in very predictible and well-conditioned regions, it can provide accurate approximations with very low computational complexity.

Then, for a set of given data points of size n, which are the points that the parametric approach aims to fit, the complete multiple linear regression model expression is [16],

$$y_i = \log(N_{t,i}^{opt}) = \alpha_0 + \sum_{j=1}^{p} \alpha_j\, x_{i,j} + \varepsilon_i \tag{12}$$
$$\forall\, i = 1, 2, ..., n\;, \quad \text{with } p < n$$

where, the given polynomial is dependent on p input variables, also known as explanatory variables. Therefore, for the i-th data point, $N_{t,i}^{opt}$ and y_i are its linear and logarithmic averaging values, respectively, ε_i is its error value, and, $x_{i,j}$ is its j-th explanatory variable, with α_j, as the associated j-th weight[(1)].

From this expression, the estimation of the weights is performed with an LS computation of the following matrix definition [16],

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \;\rightarrow\; \widehat{\boldsymbol{\alpha}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}\,, \tag{13}$$

where the previously indexed values are gathered as vectors,

$$\begin{aligned} \mathbf{y} &= \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^T & \boldsymbol{\varepsilon} &= \begin{bmatrix} \varepsilon_1 & \cdots & \varepsilon_n \end{bmatrix}^T \\ \boldsymbol{\alpha} &= \begin{bmatrix} \alpha_0 & \cdots & \alpha_p \end{bmatrix}^T & \widehat{\boldsymbol{\alpha}} &= \begin{bmatrix} \widehat{\alpha}_0 & \cdots & \widehat{\alpha}_p \end{bmatrix}^T \end{aligned}$$
$$\mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,p} \end{bmatrix}. \tag{14}$$

As previously mentioned, SNR^{dB} and N_c are going to be the basis of the independent explanatory variables of the model. Then, the complete input variables of the regression model can be constructed with the following general expression [17],

$$\mathbf{x}_i = [\mathbf{X}]_{(i,:)} = \left(\mathbf{x2}_i \otimes \mathbf{x1}_i\right)^T \tag{15}$$

where $\begin{cases} \mathbf{x1}_i = \begin{bmatrix} 1 & \log\left(N_c(v_i)\right) & \cdots & \log\left(N_c(v_i)\right)^{p'} \end{bmatrix}^T \\ \mathbf{x2}_i = \begin{bmatrix} 1 & SNR_i^{dB} & \cdots & \left(SNR_i^{dB}\right)^{p'} \end{bmatrix}^T \end{cases}$

with p' being the highest order of any parameter within the polynomial. Since there are cross-terms between components of $\mathbf{x1}_i$ and $\mathbf{x2}_i$, also known as interaction terms, the final order of the polynomial is $p = (p'+1)^2$. In order to reduce this order, these cross-terms can be neglected with the following input parameter definition,

$$\mathbf{x}_i = \left(\mathbf{x1}_i \,||\, \mathbf{x2'}_i\right)^T \tag{16}$$
$$\text{with } \mathbf{x2'}_i = \begin{bmatrix} SNR_i^{dB} & \cdots & \left(SNR_i^{dB}\right)^{p'} \end{bmatrix}^T.$$

which reduces the order of the polynomial to $p = 2\,p' + 1$.

At last, the final fitted averaging values $\left(\widehat{N}_{t,i}^{opt}\right)$ for the i-th data point are obtained after applying the operation of rounding the closest integer to the approximated solution $\left(\widehat{n}_{t,i}^{opt}\right)$, which is computed with the estimated weights,

$$\widehat{y}_i = \log(\widehat{n}_{t,i}^{opt}) = \widehat{\alpha}_0 + \sum_{j=1}^{p} \widehat{\alpha}_j\, x_{i,j}\;, \quad \widehat{N}_{t,i}^{opt} = \left\lfloor \widehat{n}_{t,i}^{opt} \right\rceil. \tag{17}$$

V. Numerical Results

In this section, the multiple linear regression model is implemented. The weights of the model ($\widehat{\boldsymbol{\alpha}}$) are estimated with a set of valid data points computed from a realistic system model. After that, the accuracy of the approximated solutions is addressed in terms of the error between their respective MSE and the analytical MSE.

[(1)]Even though, the variable notation coincides with the transmitted and received signal variables, from now on, any reference to x and y will refer to the regression model notation.

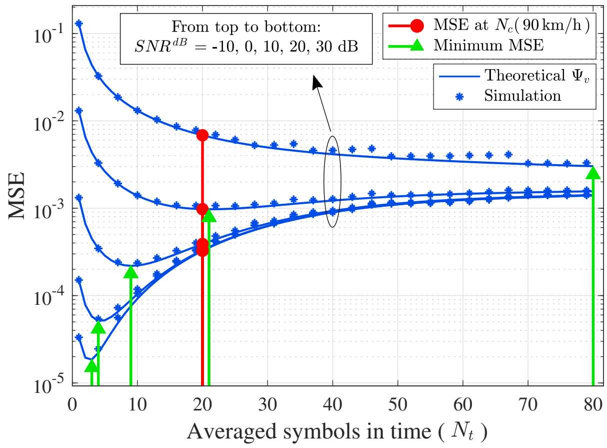

Fig. 1. Representation of the MSE curves for different SNR values and a relative speed between UE and BS of 90 km/h.

To start with, the system model is implemented in the mmWave regime with a $f_c = 28$ GHz. The subcarrier spacing is $\Delta f = 120$ kHz and the overhead of the cyclic prefix is approximately 7% (L_{cp}/K).

A. State of the Art Optimal MSE

First of all, as Fig. 1 shows, the MSE from expression (9) is plotted for a system model with relative speed of 90 km/h, and a range of SNR values from -10 to 30 dB. As it can be seen, the theoretical curves (blue lines) that employ the approximated correlation of $\rho_t\left(\gamma\Delta m\right) = \mathrm{sinc}\left(\gamma\Delta m\right)$, match the simulated curves (blue stars) computed with the realistic channel model from [12]. Additionally, the minimum MSE values (green triangles) are achieved at the optimal averagings, which have been computed by solving the transcendental equation (11).

From this picture, two kind of curves can be distinguished: in the first one, at high SNR values, a change of slope is shown, where the minimum solution can be obtained before the MSE rises up again until the error floor; and, in the second one, at low SNR values, a downward trend is maintained, which flattens asymptotically. In the first type of curve, in general, it is fulfilled $N_t^{opt} < N_c$, where the minimum MSE is enhanced almost one order of magnitude in contrast to the MSE at N_c. Whereas in the second type of curve, the best strategy is to average at $N_t^{opt} \rightarrow \infty$.

B. Selection of Valid Input Parameters Set

In Fig. 2, the optimum solutions from the transcendental equation are plotted against N_c. In this case, the averaging results are computed for a wider range of N_c values (bottom axis) or equivalently a range of speeds (top axis). Clearly, there exists a linear relationship in log-log scale between N_t^{opt} and $N_c(v)$, which is consistently sustained for all the range of SNR values (plotted with different symbols at specific values and a colour gradient from violet to red at intermediate values).

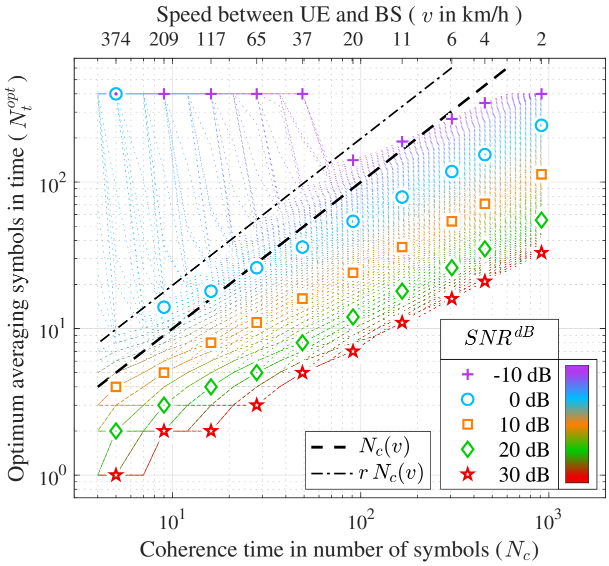

Fig. 2. Relation between optimal averaging solutions and the coherence number of symbols, for different SNR values.

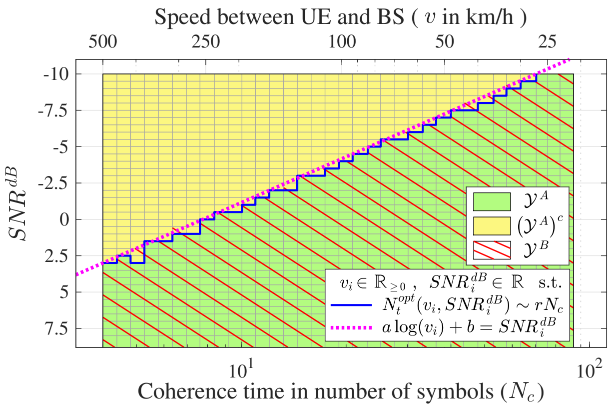

Fig. 3. Illustration of valid input parameters for the multiple linear regression model defined with $\mathcal{Y}^A$ and $\mathcal{Y}^B$.

As it can be seen, there are some outlier data points (top-left solutions) that must be excluded in the training of the weights in order to perform a proper regression. Thus, the solutions from the, previously mentioned, first kind of the MSE curves, i.e. the solutions that fulfill $N_t^{opt} < N_c$, and fall below N_c (black dashed line), are the valid data points that guarantee accurate approximations. Specifically, this region can be slightly extended (black dash-dotted line) with the ratio of a constant value, like $r \in \mathbb{R}$.

As Fig. 3 shows, these valid data points are addressed from a different point of view, where the input parameters, placed in a grid, are shaded depending on the nature of their respective data point. To start with, the set of valid parameters that provide $N_t^{opt} < rN_c$ (green area) are defined as,

$$\mathcal{Y}^A = \left\{ \begin{matrix} v_i \in \mathbb{R}_{\geq 0} \\ SNR_i^{dB} \in \mathbb{R} \end{matrix} \middle| N_t^{opt}(v_i, SNR_i^{dB}) < rN_c \right\}, \quad (18)$$

In contrast to this set, the rest of parameters which give outlier data points $\left((\mathcal{Y}^A)^c \right)$ are shaded in yellow. The main issue about this set definition is that in order to check if the input parameters provide a valid data point, the condition to be fulfilled employs N_t^{opt}, which in turn, is the solution pretended to be approximated.

Alternatively, as it can be seen in the figure, the boundary between $\mathcal{Y}^A$ and $(\mathcal{Y}^A)^c$ (blue line) can be approximated with another linear regression (dotted magenta line). The formal definition of this approximated set of cleansed input parameters (shaded area with red stripes) is,

$$\mathcal{Y}^B = \left\{ \begin{matrix} v_i \in \mathbb{R}_{\geq 0} \\ SNR_i^{dB} \in \mathbb{R} \end{matrix} \middle| a\log(v_i) + b - SNR_i^{dB} < 0 \right\} \quad (19)$$

where, $a \in \mathbb{R}$ and $b \in \mathbb{R}$ are the slope and the offset of the linear boundary, respectively, which have been computed employing two, already known, valid parameters from the green region,

$$a = \frac{SNR_2^{dB} - SNR_1^{dB}}{\log(v_2) - \log(v_1)} \ , \quad b = SNR_2^{dB} - a\log(v_2) \quad (20)$$

Obviously, the approximation of $\mathcal{Y}^A$ with $\mathcal{Y}^B$ is considered to be accurate since the shaded area with red stripes covers all the green region. Henceforth, in order to compute approximate solutions with the parametric approach, the input parameters must belong to $\mathcal{Y}^B$.

C. Validation of the Proposed Parametric Approach

Once the data set is cleansed and the valid input parameters are determined, the regression model can be implemented. In Fig. 4, it has been plotted the regressive curves for a range of SNR values (coloured lines from red to green). The model weights (α) are estimated when the explanatory variables contain interaction terms, as in (15), and the order of dependency is $p' = 4$. Similarly to Fig. 2, the boundaries of N_c with and without the threshold r (dashed-dotted and dashed lines, respectively), are plotted, too. After running some tests, r has been extended up to $r = 2$, which guarantees a well-conditioned region of valid data points (filled blue dots).

As the figure depicts, the regression model fits with accuracy the trend of the optimum averagings within the region of valid data points. However, in order to analyze the accuracy of the model, it has to be compared with the true optimum averagings. In fact, it has to be compared in terms of the final MSE, since it is the metric to be optimized. Then, the accuracy in percentage is computed for the maximum and the mean error between the MSE at the approximated averaging and the minimum MSE, as follows,

$$\epsilon^{Max} = \frac{\max(\epsilon)}{\Psi_v(N_t^{opt})} \cdot 100 \ , \quad \epsilon^{Avg} = \frac{\mathbb{E}\{\epsilon\}}{\Psi_v(N_t^{opt})} \cdot 100 \quad (21)$$
$$\text{with } \epsilon = \left| \Psi_v(N_t^{opt}) - \Psi_v(\widehat{N}_t^{opt}) \right| .$$

In Fig. 5, the values of ϵ^{Max} (blue left axis) and ϵ^{Avg} (red right axis) are plotted for regression models with different p' values, which, as previously mentioned, represent the highest

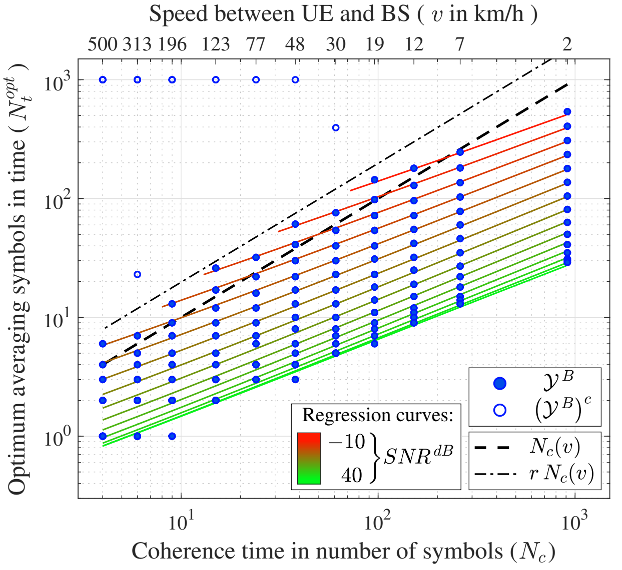

Fig. 4. Representation of the multiple linear regression model over cleansed data points. The input variables contain interaction terms following (15), with order $p' = 4$.

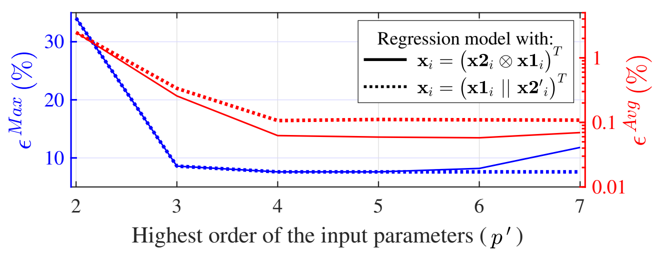

Fig. 5. Accuracy in terms of ϵ^{Max} (blue left axis) and ϵ^{Avg} (red right axis) for different order p'.

order any input parameter can take. The curves show the performance when the explanatory variables employ interaction terms (solid line), as in (15), or when they do not use them (dotted line), like in (16).

This figure shows that the minimum error is achieved when $p' = \{4, 5, 6\}$, with $\epsilon^{Max} = 7\%$, and $\epsilon^{Avg} = 0.05\%$ and 0.1% for a model with and without interaction terms, respectively. Otherwise, if $p' = 2$, the regression model is very inaccurate with $\epsilon^{Max} = 35\%$ and $\epsilon^{Avg} = 2.5\%$. Finally, the downside of the interaction terms appears when $p' = 7$, since the high order polynomial overfits the data points increasing the maximum error up to 12%.

VI. Conclusion

In this paper, the optimal averaging in time for an ST scheme has been computed from a parametric approach. The proposed low complexity approximation avoids the analytical solution, which must solve a transcendental equation. It has been defined the multiple linear regression model which requires, as input parameters, the speed between the UE and the BS, and the SNR of the system. Also, the region where these parameters provide robust approximations has been addressed.

At the end, the accuracy between the minimum MSE and the MSE computed from the proposed solutions shows a maximum error of 7%, and a mean error of 0.05%. With this parametric approach, an almost optimal ST with a robust performance can be easily deployed in practical systems where the computational cost is critical.

References

[1] A. Gupta and R. K. Jha, “A survey of 5G network: Architecture and emerging technologies,” *IEEE Access*, vol. 3, pp. 1206–1232, 2015.

[2] X. Wang, L. Kong, F. Kong, F. Qiu, M. Xia, S. Arnon, and G. Chen, “Millimeter wave communication: A comprehensive survey,” *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 1616–1653, 2018.

[3] E. G. Larsson, “Massive MIMO for 5G: Overview and the road ahead,” in *2017 51st Annual Conference on Information Sciences and Systems (CISS)*, 2017, pp. 1–1.

[4] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, “Superimposed pilots: An alternative pilot structure to mitigate pilot contamination in massive MIMO,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 3366–3370.

[5] W. Huang, C. Li, and H. Li, “On the power allocation and system capacity of OFDM systems using superimposed training schemes,” *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 1731–1740, 2009.

[6] J. C. Estrada-Jiménez and M. J. Fernández-Getino García, “Partial-data superimposed training with data precoding for OFDM systems,” *IEEE Transactions on Broadcasting*, vol. 65, no. 2, pp. 234–244, 2019.

[7] K. Chen-Hu, M. J. Fernández-Getino García, A. M. Tonello, and A. G. Armada, “Pilot pouring in superimposed training for channel estimation in CB-FMT,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 3366–3380, 2021.

[8] I. Piqué Muntané and M. J. Fernández-Getino García, “Optimum averaging of superimposed training schemes in OFDM under realistic time-variant channels,” *IEEE Access*, vol. 9, pp. 115 620–115 631, 2021.

[9] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. USA: Prentice-Hall, Inc., 1993.

[10] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. USA: Prentice Hall PTR, 2001.

[11] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, 1st ed. USA: Cambridge University Press, 2008.

[12] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems: Modeling, Methodology and Techniques*, 2nd ed. USA: Kluwer Academic Publishers, 2000.

[13] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3rd ed. USA: Cambridge University Press, 2007.

[14] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

[15] A. Gelman and G. Imbens, “Why high-order polynomials should not be used in regression discontinuity designs,” *Journal of Business & Economic Statistics*, vol. 37, no. 3, pp. 447–456, 2019. [Online]. Available: https://doi.org/10.1080/07350015.2017.1366909

[16] N. Draper and H. Smith, *Applied Regression Analysis*, 3rd ed. New York: Wiley, 1998. [Online]. Available: https://doi.org/10.1002/9781118625590

[17] D. Varsamis, N. Karampetakis, and P. Mastorocostas, “Transformations between two-variable polynomial bases with applications,” *Applied Mathematics and Information Sciences*, vol. 10, no. 4, pp. 1303–1311, 2016.