CHANNEL PHASE ESTIMATE IN TIME VARIANT SIMO SYSTEMS

Andrea Monti Guarnieri and Stefano Tebaldini

Dipartimento di Elettronica e Informazione, Politecnico di Milano, Italy

ABSTRACT

This paper introduces a novel ML based approach to channel identification for time variant SIMO (Single Input Multiple Output) systems fed by a stochastic process. We focus on the particular case where the unknowns are represented by the channels phases, that find applications in RADAR interferometry. Starting from the rigorous formulation of the ML estimator, we derive an approximation that makes use of mixers and FIR filters only. The computational efficiency and the robustness versus model errors of the resulting estimator make it suitable for its implementation is an adaptive framework. An application in topography reconstruction from real SAR (Synthetic Aperture Radar) data is presented.

1. INTRODUCTION

The present work approaches the problem of estimating the phase of a number of constant-envelope signals modulating a single realization of a stochastic process, according to the block diagram in Fig. 1.

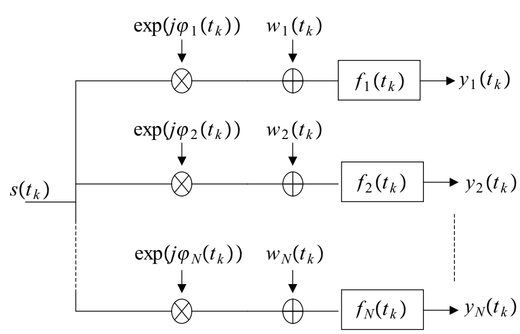

Fig. 1. Time variant SIMO system

The meaning of the quantities in Fig. 1 is the following:

- n: is the channel index $(n = 1...N)$
- $s(t_k)$: is the source, characterized as a realization of a white normal circular process $s(t_k) \sim N\left(0, \sigma_s^2\right)$, with σ_s^2 known.
- $y_n(t_k)$: is the observed data in each channel
- $\varphi_n(t_k)$: is the phase of the $n-th$ channel modulation. It represents the unknown of the problem.
- $f_n(t_k)$: is the $n-th$ channel filter. It is supposed to be a known LTI FIR filter
- $w_n(t_k)$: represents the superposed noise on the $n-th$ channel. It is assumed to be a white normal circular process: $w_n(t_k) \sim N\left(0, \sigma_w^2\right)$, with σ_w^2 known.

All the signals are supposed to be properly sampled (see section 2) at time instants t_k.

Such a model could apply to SAR interferometry (InSAR), where the estimate of the phases is required to infer information about the Earth's topography [1]. Nevertheless, the problem is posed in the most general terms.

We first observe that only $N-1$ phases may be estimated out of N outputs. This is a consequence of the invariance of the source statistics with respect to a multiplication for a complex exponential $\exp\left(j\varphi_n(t)\right)$. For this reason in the following we will assume $\varphi_1(t) \triangleq 0 \forall t$.

In an adaptive framework, we require an estimate to be accurate and local, therefore we assume a parametric model (a 1^{st} order polynomial is suite in most cases) for the unknown phases. In the following we will account for this by defining the unknown as $\varphi_n(t) = \varphi_n(t; \mathbf{c})$, where $\mathbf{c}$ is the vector of parameters to be estimated.

2. MODEL DISCRETIZATION AND OUTPUT STATISTICS

In most applications the model is time continuos, so that the real unknowns of the problem are the functions $\varphi_n(t)$, rather than their sampled versions. It may be shown that a proper discrete-time model is obtained by sampling the continuous-time outputs, $y_n(t)$, at a rate [2]:

$$f_s \geq B_y + B_\varphi, \tag{1}$$

B_y and B_φ being respectively the maximum (bilateral) bandwidth of the outputs $y_n(t)$, and of the modulating terms $\exp\left(j\varphi_n(t)\right)$. Under this condition it is possible to derive the statistical properties of the output samples directly from the discrete time model in Fig. 1.

We may express the model in Fig. 1 in a compact matricial formulation. Considering M samples out of each output and defining $D = L + M - 1$, where L is the length of the channel filters, we get:

$$\mathbf{y}_n = \mathbf{F}_n \mathbf{\Phi}_n \mathbf{s} + \mathbf{F}_n \mathbf{w}_n \tag{2}$$

where $\mathbf{y}_n$ is an $M \times 1$ column vector, $\mathbf{F}_n$ is the $M \times D$ matrix which implements the convolution of a sequence for the channel filter $f_n(t_k)$, $\mathbf{\Phi}_n$ is a diagonal $D \times D$ matrix such that

$$\mathbf{\Phi}_n = diag\left\{\exp(j\varphi_n(t_k))\right\} \quad k = 0, 1, ..., D-1 \tag{3}$$

$\mathbf{s}$ is a $D \times 1$ column vector, and so is the noise vector $\mathbf{w}_n$.

Under condition (1) it may be shown that the output vectors $\mathbf{y}_n$ are multivariate zero-mean normal circular processes with covariance [2]:

$$E\left[\mathbf{y}_n \mathbf{y}_m^H\right] = \left(\sigma_s^2 + \delta_{n-m} \cdot \sigma_w^2\right) \cdot \mathbf{F}_n \mathbf{\Phi}_n \mathbf{\Phi}_m^H \mathbf{F}_m^H \tag{4}$$

where δ_n is the unit sample sequence.

3. ML PHASE ESTIMATE

The ML estimate of the parameters vector $\mathbf{c}$ is given by:

$$\widehat{\mathbf{c}} = \arg\max_{\mathbf{c}} \left\{ L\left(\mathbf{y}_1, \mathbf{y}_2 ..., \mathbf{y}_N; \mathbf{c}\right)\right\} \tag{5}$$

where the likelihood function $L()$ follows directly from the expression of the joint pdf of the outputs.

$$L = p\left(\mathbf{y}_1, \mathbf{y}_2 ..., \mathbf{y}_N | \mathbf{c}\right) = \frac{\exp\left(-\mathbf{y}^H \mathbf{C}^{-1} \mathbf{y}\right)}{\pi^N \left|\mathbf{C}\right|} \tag{6}$$

where $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1^H & \mathbf{y}_2^H & ... & ... & \mathbf{y}_N^H \end{bmatrix}^H$ and $\mathbf{C}$ is the covariance matrix: $\mathbf{C} = E\left[\mathbf{y}\mathbf{y}^H\right]$.

The ML estimate achieves very good accuracy [3], provided that the data fit the model, as it will be shown later. Problems may arise with model errors, in dependence on the conditioning number of the covariance matrix. This aspect becomes more critical as the number of output channels increases, as a consequence of the strong correlations among the channels which occur for many combinations in the parameters space.

The impact of model errors is mitigated by adapting the estimate by means of a mobile window sliding over the data, so that only a simplified knowledge about the data statistic is required. However, the adaptive implementation raises severe computational problems, since the ML estimator (5) requires a matrix inversion for every value of the parameters characterizing the pdf of data ($\mathbf{c}$, σ_s^2, σ_w^2, and the matrices coefficients).

4. SIMPLIFIED ML ESTIMATE

The log-likelihood function of two channels L_{12} may be rewritten as follows, by means of the Bayes theorem:

$$\log\left(L_{12}\right) \equiv -\left\|\mathbf{y}_1 - \widehat{\mathbf{y}}_{1|2}\right\|^2_{\mathbf{C}^{-1}_{1|2}} - \log\left(\left|\mathbf{C}_{1|2}\right|\right) \tag{7}$$

where $\widehat{\mathbf{y}}_{1|2}$ is the optimum estimate (in the MMSE sense) of $\mathbf{y}_1$ from $\mathbf{y}_2$ and $\mathbf{C}_{1|2}$ is the covariance matrix of $\mathbf{y}_1$ conditioned to $\mathbf{y}_2$. The symbol $\equiv$ is used here with the meaning of equality but for a term independent from the unknowns $\mathbf{c}$.

The expression of $\widehat{\mathbf{y}}_{1|2}$ corresponds to the Best Linear Unbiased Estimator (BLUE) [4]:

$$\widehat{\mathbf{y}}_{1|2} = \gamma \mathbf{F}_1 \mathbf{\Phi}_1 \mathbf{\Phi}_2^H \mathbf{F}_2^\dagger \mathbf{y}_2 = \gamma \widetilde{\mathbf{y}}_{1|2} \tag{8}$$

where $\mathbf{F}_2^\dagger = \mathbf{F}_2^H \left(\mathbf{F}_2 \mathbf{F}_2^H\right)^{-1}$ is the pseudoinverse of $\mathbf{F}_2$ and

$$\gamma = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_w^2} = \frac{SNR}{1 + SNR} \tag{9}$$

So far, by equation (7) we may interpret the log-likelihood function as the L2 norm of the innovation $\mathbf{e}_{12} = \mathbf{y}_1 - \widehat{\mathbf{y}}_{1|2}$ defined by $\mathbf{C}^{-1}_{1|2}$.

$$\begin{aligned} \mathbf{C}_{1|2} &= E\left[\mathbf{e}_{12}\mathbf{e}_{12}^H | \mathbf{y}_2\right] = E\left[\mathbf{y}_1 \mathbf{y}_1^H\right] - E\left[\widehat{\mathbf{y}}_{1|2}\widehat{\mathbf{y}}_{1|2}^H\right] \\ &= \mathbf{F}_1\mathbf{F}_1^H - \gamma^2 \mathbf{F}_1 \mathbf{\Phi}_1 \mathbf{\Phi}_2^H \mathbf{F}_2^\dagger \mathbf{F}_2 \mathbf{\Phi}_2 \mathbf{\Phi}_1^H \mathbf{F}_1^H \end{aligned} \tag{10}$$

4.1. Reduction of costs

Most of the computational costs in equation (7) come from the kernel $\mathbf{C}^{-1}_{1|2}$, which is equivalent to M filters of M taps each, where M is the window length. In this section, we derive an approximation suitable to describe the whitening kernel $\mathbf{C}^{-1}_{1|2}$ by means of a single LTI filter.

Let us suppose that the estimate window is sufficiently long with respect to the channel filters $f_n\left(t_k\right)$. Under this hypothesis most of the output samples may be considered as the result of a circular, rather than linear, convolution - i.e.: we are neglecting the border effects -. As a consequence, we have that the channel matrices $\mathbf{F}_n$ may be considered as circulant, and thus diagonalized by the DFT (Discrete Fourier Transform) matrix $\mathbf{W}$:

$$\mathbf{F}_n \simeq \mathbf{W}\mathbf{\Lambda}_n\mathbf{W}^H \tag{11}$$

where the elements of the diagonal matrix $\mathbf{\Lambda}_n$ represent the DFT of the channel filter $f_n(t_k)$. Under hypothesis (11) the expression of the covariance matrix results

$$\begin{aligned} &\mathbf{C}_{1|2} = E\left[\mathbf{e}_{12}\mathbf{e}_{12}^H\right] \simeq \\ &\mathbf{W}\left(\mathbf{\Lambda}_1\mathbf{\Lambda}_1^H - \gamma^2 \mathbf{\Lambda}_1 \mathbf{W}^H \mathbf{\Phi}_1 \mathbf{\Phi}_2^H \mathbf{F}_2^\dagger \mathbf{F}_2 \mathbf{\Phi}_2 \mathbf{\Phi}_1^H \mathbf{W} \mathbf{\Lambda}_1^H\right)\mathbf{W}^H \end{aligned} \tag{12}$$

The bracketed kernel is diagonal only if the phase difference $\varphi_1(t_k) - \varphi_2(t_k)$ is linear, as happens for a 1^{st} order polynomial model for the phases (i.e. a frequency shift), but it shows a high diagonal predominance for any phase matrix. Therefore, the eigenvalues of $\mathbf{C}_{1|2}$ may be approximated by the elements on the kernel diagonal, referred to as $|\lambda_e(k)|^2$. This approximation is equivalent to considering the innovation as a stationary process, and thus $|\lambda_e(k)|^2$ represent its power spectrum. So far, equation (7) becomes:

$$\begin{aligned} &\log\left(L_{12}\right) \simeq \\ &\sum_k \left|Y_1(k) - \gamma \widetilde{Y}_{1|2}(k)\right|^2 |\lambda_e(k)|^{-2} + \sum_k \log\left(|\lambda_e(k)|^2\right) \end{aligned} \tag{13}$$

where $Y_1(k)$, $\widetilde{Y}_{1|2}(k)$ are the Fourier transform of $\mathbf{y}_1$, $\widetilde{\mathbf{y}}_{1|2}$.

Equation (13) may be implemented in time domain by means of a circular convolution, even though no cost reduction is obtained with respect to (7), since the circular filter is still equivalent to M LTI filters. However, since we're neglecting the border effects, the circular filter may be replaced by a *single* LTI filter, thus achieving a cost reduction factor of M. As an alternative, equation (13) may be implemented directly in the frequency domain. The time and frequency domain implementations are equivalent, since both of them are based on the norms of signals which are approximated at their borders. The same concepts apply to the block $\mathbf{F}_2^\dagger$ in equation (8).

The efficient time-domain estimator is shown in Fig. 2.a.

As a further advantage, this estimator may keep into account non-stationary features, such as changes in the SNR or even in the channel filters, simply by modifying the value of γ or the filter coefficients depending on the position of the estimate window.

4.2. Model relaxation

Under the hypothesis of low SNR it may be shown that the quantity [2]:

$$E\left[\sum_k |Y_1(k)|^2 |\lambda_e(k)|^{-2}\right] + \sum_k \log|\lambda_e(k)|^2 \tag{14}$$

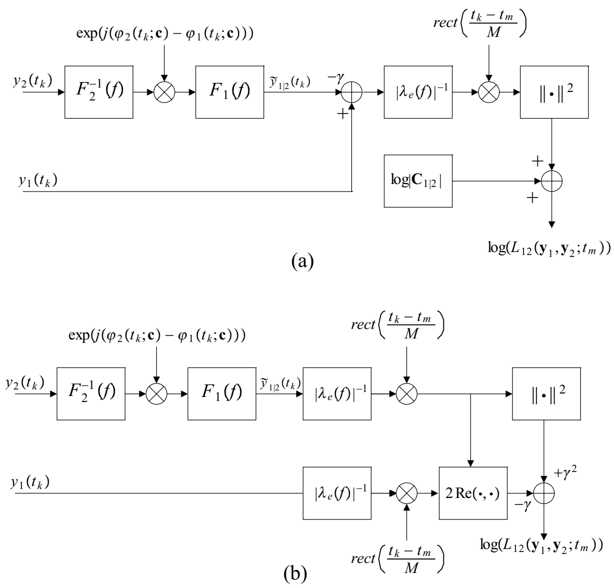

Fig. 2. Time domain implementation: the filters are implemented as LTI FIR and referred to by means of their Fourier transform. The output of both (a) and (b) is the log-likelihood computed at time t_m. (a) Simplified ML estimator (13); (b) simplified ML estimator under low SNR approximation (15)

doesn't depend on the phase parameters so that, in mean, the log-likelihood may be expressed as:

$$\log\left(L_{12}\right) \simeq \sum_k \left|\lambda_e(k)\right|^{-2} \left\{\gamma^2 \left|Y_{1|2}(k)\right|^2 - 2\,\mathrm{Re}\,\gamma Y_1(k) \widetilde{Y}^H(k)\right\} \tag{15}$$

According to the discussion in the last section, equation (15) may be implemented through the scheme in Fig. 2.b.

The low SNR hypothesis find its justification if we take into account the errors coming from an inaccurate knowledge of the real data generation mechanism. These model errors may be characterized quite sensibly as gaussian, so that their final effect is to raise the total amount of noise [5]. Hence the estimator (15) not only implies a significative cost reduction with respect to the exact ML implementation (5), but also allows to relax the hypotheses on which the model is based. Therefore, it does not require the exact knowledge about the source statistics, the parametrization of the phase matrices, and the channel filters.

4.3. Extension to multichannel

An extension of the concepts exposed to the case of an arbitrary number of channels is possible, but not straightforward. However, we experienced that, if the phases are represented by a 1^{st} order polynomial model, and the channel filters $f_n(t)$ are bandpass, the ML estimate (5) can be well approximated by optimizing the sum of the log-likelihood functions of every couple of data.

$$\widehat{\mathbf{c}} = \arg\max_{\mathbf{c}} \left\{ \sum_{m=1}^{N} \sum_{n=m+1}^{N} \log\left(L_{nm}\left(\mathbf{y}_n, \mathbf{y}_m; \mathbf{c}\right)\right) \right\} \tag{16}$$

This couple-based approach to the multichannel case is similar to the one in [6]. From a statistical point of view, this approach is optimum if the couples may be considered as uncorrelated. It can be shown that this method achieves the same performance as the optimum a-posteriori combination of the single estimates (considered as uncorrelated with each other), but without any need to know the exact variance of every single estimate [2].

5. EXPERIMENTAL RESULTS

This section shows the results achieved by running different estimators over a two output synthetic dataset. We refer to 2D data, described by the coordinates (t_k, x_h). The data are characterized as white in the x direction, so that the likelihood over a 2D window is simply given by the products of the single likelihoods computed along direction t.

First, we consider a data generated according to the model in Fig. 1. Let the phases be described by a first order polynomial model:

$$\begin{aligned} \varphi_1\left(t_k, x_h; f_0, \psi_0\right) &= 0 \\ \varphi_2\left(t_k, x_h; f_0, \psi_0\right) &= 2\pi f_0 t_k + \psi_0 \end{aligned} \tag{17}$$

Fig. 3 shows the RMSE of the estimate of the frequency f_0. In this case the model is perfectly known, and thus the best result is achieved by the ML estimator. The simplified estimator, implemented both through (13) and (15), achieves a good performance and gets closer to the ML as the window length increases. The Cramer Rao bound was computed numerically, according to (6). In order to provide a comparison with a commonly used frequency estimator, we plotted the curve relative to the estimates obtained by maximizing the periodogram of the Hermitian product of the two outputs.

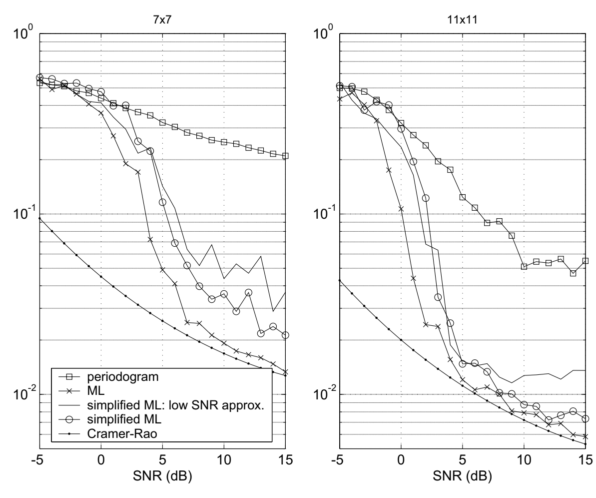

Fig. 3. Estimate of the frequency shift f_0 : Root Mean Square Error versus Signal to Noise Ratio. RMSE is normalized to the signal bandwidth B. The simplified ML estimators (13) and (15) are compared to the ML estimator, the periodogram, and the Cramer-Rao bound. The estimates are performed over a 7x7 (left) and 11x11 (right) estimate window. The true value of f_0 (normalized to the sampling frequency) is 0.2. The normalized bandwidth of the signals (B) is 0.4

The effect of model errors is taken into account in Fig. 4. The data simulate two SAR images correspondent to a *real* topography, so that the phase model (17) must be considered only as a first order approximation of the real phase. This introduces a model error, which causes stability problem to the ML estimator defined by (5).

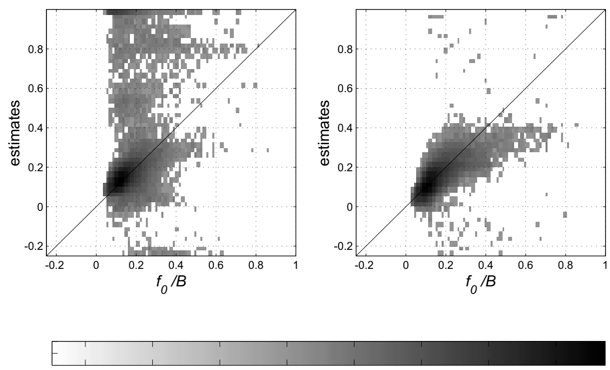

Fig. 4. Scatter plots: the colorbar indicates the logarithm of the number of points correspondent to a certain couple of values $(f_0/B, \widehat{f}_0/B)$. (left) ML estimator (5); (right) simplified ML estimator under low SNR approximation (15). Figure shows how the proposed method manages to avoid the stability problems experienced by the ML estimator.

6. AN EXAMPLE OF APPLICATION TO INSAR

Let us now show an example of usage of the simplified ML estimator applied to the estimate of topographic features by means of multibaseline SAR interferometry. The problem is really 2D, but can it be well approached by iterating two separable 1D estimates. As an example, let t_k, x_k be the coordinates, spanning the 2D image space, the first estimate is iterated in direction t_k by assuming the following 1^{st} order polynomial model for the $n - th$ channel phases:

$$\varphi_n \left(t_k, x_h; \nu_t, \psi_2 ... \psi_N\right) = 2\pi b_n \nu_t t_k + \psi_n \tag{18}$$

where b_n is a known parameter referred to as "normal baseline" and v_t the unknown parameter to be retrieved (usually referred to as “wavenumber shift" [1]), directly related to the topographic slopes (along t_k). The phase offsets ψ_n are nuisance parameters, and thus their estimate is not shown. The estimator is implemented through formula (16), computing the single log-likelihoods according to (15), so as to keep into account eventual model mismatches.

As an example, a dataset of $N = 6$ SAR images acquired form ENVISAT satellite close to Los Angeles has been processed, to derive the local estimate of slope-dependent parameter ν_t.

This estimate in shown in Fig. 5, compared to the same quantity computed using an external DEM (Digital Elevation Model) generated by SRTM (Shuttle Radar Topography Mission) in year 2000. The results are satisfying, as the estimates map manages to exploit the superior resolution of the ENVISAT data with respect to the SRTM DEM. Most errors are localized in correspondence of high topographic slopes, which represent a very critical case for every SAR system, or in correspondence of points where the single input model is not suitable due to scenery changes.

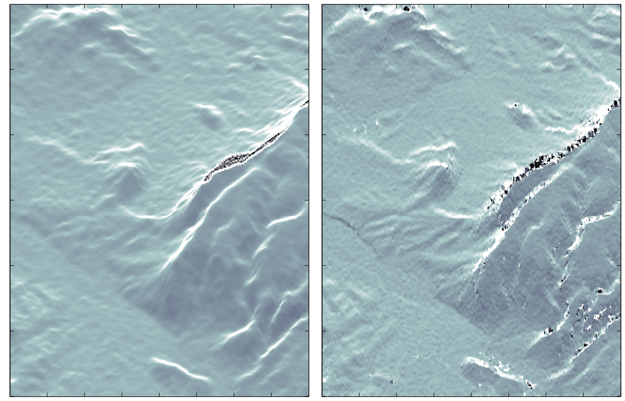

Fig. 5. (left) Map of ν_t computed from the SRTM DEM; (right) map of ν_t estimated from ENVISAT data by means of (15). The area is about $6 \times 4\ Km^2$, correspondent to 600×800 samples. The estimate window size is 7×9. The normalized bandwidth of data in the t_k direction is 0.4.

7. CONCLUSIONS

In this work we faced the problem of estimating the *modulations* applied to the same process from two or more filtered outputs. In the two channels case, we derived a ML-based solution suitable whenever the model of data is known only approximately and the computational costs represent a critical issue. We partially discussed the problem of the multichannel estimate (N>2), and proposed a solution which is simple, stable and revealed itself as effective in the applications.

Further studies are foreseen for optimizing the multichannel case, in order to design an estimator more adherent to the joint statistics of data. In a framework where the number of channel is high, carrying out this objective would also result in a significative cost reduction, as the estimator would automatically drop all the combinations non-informative about the phase estimate.

8. REFERENCES

[1] Richard Bamler and Philipp Hartl, “Synthetic aperture radar interferometry,” *Inverse Problems*, vol. 14, pp. R1–R54, 1998.

[2] Stefano Tebaldini, “Interferometria SAR multi-baseline,” M.S. thesis, Politecnico di Milano, July 2005.

[3] A. Pauciullo G. Fornaro, A. Monti Guarnieri and S. Tebaldini, “Joint multi-baseline SAR interferometry,” *EURASIP JASP*, vol. 20, pp. 3194–3205, Dec. 2005.

[4] Gianfranco Fornaro and Andrea Monti Guarnieri, “Minimum mean square error space-varying filtering of interferometric SAR data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 1, pp. 11–21, July 2002.

[5] A Tarantola, *Inverse Problems Theory*, Elsevier, New York, 1987.

[6] G Xu, H Liu, L Tong, and T Kailath, “A least-squares approach to blind channel identification,” *IEEE Transactions on Signal Processing*, vol. 43, pp. 2982–2993, Dec. 1995.