Estimation of large data sets on the basis of sparse sampling

Anatoli Torokhti, Phil Howlett
CIAM, School of Inf. Techn. & Math. Sci.
University of South Australia, SA 5095
Email: anatoli.torokhti@unisa.edu.au

Hamid Laga
PBRC, School of Inf. Techn. & Math. Sci.
University of South Australia, SA 5095
Email: hamid.laga@unisa.edu.au

***Abstract*—We propose a new technique which allows us to estimate any random signal from a large set of noisy observed data on the basis of samples of only a few reference signals.**

I. INTRODUCTION

A. Motivation

In many applications associated with difficult environments, *a priori* information on signals of interest can be obtained only at *a few* given times $\{t_j\}_1^p \subset T = [a,b] \subset \mathbb{R}$ where $a = t_1 < t_2 < \cdots < t_{p-1} < t_p = b$ whereas it is required to estimate the signals at *any time* $t \in T$. Typical examples are devices deployed in the stratosphere, underground or underwater. The choice of points t_j might be beyond our control (e.g. in geophysics and defence tascs). For any $t \in T$, the signal is a stochastic vector. We consider large sets of such signals where each signal is associated with a particular $t \in T$. The observations are noisy and also *large*. Thus, all we can exploit is noisy observations and a *sparse* information on reference signals given by samples of the signal set at times $\{t_j\}_1^p$.

B. Formalization of the problem

To formalize the problem, we denote by Ω the set of all experimental outcomes[1], by $\mathcal{K}_x = \{\boldsymbol{x}_\omega \mid \omega \in \Omega\}$ a set of reference stochastic signals and by $\mathcal{K}_y = \{\boldsymbol{y}_\omega \mid \omega \in \Omega\}$ a set of observed signals[2]. Note that, theoretically, $\mathcal{K}_x$ and $\mathcal{K}_y$ are infinite signal sets. In practice, however, sets $\mathcal{K}_x$ and $\mathcal{K}_y$ are finite and large, each with, say, N signals. To each random outcome $\omega \in \Omega$ we associate a unique signal pair $(\boldsymbol{x}_\omega, \boldsymbol{y}_\omega)$ where $\boldsymbol{x}_\omega : T \to \mathcal{C}^{0,1}(T, \mathbb{R}^m)$ and $\boldsymbol{y}_\omega : T \to \mathcal{C}^{0,1}(T, \mathbb{R}^n)$[3]. Write

$$\mathcal{P} = \mathcal{K}_x \times \mathcal{K}_y = \{(\boldsymbol{x}_\omega, \boldsymbol{y}_\omega) \mid \omega \in \Omega\} \quad (1)$$

for the set of all such signal pairs. For each $\omega \in \Omega$, the components $\boldsymbol{x}_\omega = \boldsymbol{x}_\omega(t), \boldsymbol{y}_\omega = \boldsymbol{y}_\omega(t)$ are Lipschitz continuous vector-valued functions on T [1].

We wish to construct an estimator $F^{(p-1)}$ that estimates each reference signal $\boldsymbol{x}_\omega(t)$ in $\mathcal{P}$ from related observed input $\boldsymbol{y}_\omega(t)$ under the restriction that *a priori* information on only a *few* reference signals, $\boldsymbol{x}_\omega(t_1), \ldots, \boldsymbol{x}_\omega(t_p)$, is available where $p \ll N$.

In more detail, this restriction implies the following. Let us denote by $\mathcal{K}_x^{(p)}$ a set of p signals $\boldsymbol{x}_\omega(t_1), \ldots, \boldsymbol{x}_\omega(t_p)$ for which *a priori* information is available. A set of associated observed signals $\boldsymbol{y}_\omega(t_1), \ldots, \boldsymbol{y}_\omega(t_p)$ is denoted by $\mathcal{K}_y^{(p)}$. Then for all $\boldsymbol{y}_\omega(t)$ that do not belong to $\mathcal{K}_y^{(p)}$, $\boldsymbol{y}_\omega(t) \notin \mathcal{K}_y^{(p)}$, estimator $F^{(p-1)}$ is said to be the *blind* estimator [2], [3], [4], [5] since no information on $\boldsymbol{x}_\omega(t) \notin \mathcal{K}_x^{(p)}$ is available. If $\boldsymbol{y}_\omega(t) \in \mathcal{K}_y^{(p)}$ then $F^{(p-1)}$ becomes a *nonblind* estimator since information on $\boldsymbol{x}_\omega(t) \in \mathcal{K}_x^{(p)}$ is available. Thus, depending on $\boldsymbol{y}_\omega(t)$, estimator $F^{(p-1)}$ is classified differently. Therefore, such a procedure of estimating reference signals in $\mathcal{K}_x$ is here called the *almost blind* estimation.

C. Differences from known techniques

We would like to note that the *almost blind* estimation is different from known methods such as nonblind [6]–[18], semiblind and blind techniques [2]–[5], [19]–[22][4]. Indeed, at each particular time $t \in T$, the input of the *almost blind* estimator $F^{(p-1)}$ developed below in this paper, is a random vector $\boldsymbol{y}_\omega(t)$. Thus, for different $t \in T$, the input is a different random vector $\boldsymbol{y}_\omega(t)$ but we

[1]We write $\{\Omega, \Sigma, \mu\}$ for a probability space where $\Sigma \subset \Omega$ is a sigma-algebra of measurable sets known as the event space and μ is a non-negative probability measure with $\mu(\Omega) = 1$.

[2]In an intuitive way, $\mathbf{y}$ can be regarded as a noise-corrupted version of $\mathbf{x}$. For example, $\mathbf{y}$ can be interpreted as $\mathbf{y} = \mathbf{x} + \mathbf{n}$ where $\mathbf{n}$ is white noise. We do not restrict ourselves to this simplest version of $\mathbf{y}$ and assume that the dependence of $\mathbf{y}$ on $\mathbf{x}$ and $\mathbf{n}$ is arbitrary.

[3]The space $\mathcal{C}^{0,1}(T, \mathbb{R}^p)$ is the set of vector-valued Hölder continuous functions $\boldsymbol{f}$ of order 1 with $\boldsymbol{f}(t) \in \mathbb{R}^p$ and $\|\boldsymbol{f}(s) - \boldsymbol{f}(t)\| \leq K|s-t|$. See [1], p. 96.

[4]The literature on these subjects is very abundant. Here, we listed only some related references.

wish to keep *the same estimator* $F^{(p-1)}$ for any $t \in T$, i.e. for any observed signal $\boldsymbol{y}_\omega(t)$ in the set $\mathcal{K}_y$.

By known techniques in [2]–[16] and [19]–[22], an estimator (here, we choose the united term 'estimator' to denote an equalizer or a system) is specifically designed for *each* particular input–output pair represented by random vectors. That is, for different inputs (observed signals) $\boldsymbol{y}_\omega(t)$, known techniques require different specified estimators and the number of estimators should be equal to a number of processed signals. In the case of *large signal sets*, such approaches become inconvenient because the number of signals N can be very large as it is supposed in this paper. For example, in problems related to DNA analysis, $N = \mathcal{O}(10^4)$. Therefore, the inconvenient (burdened, difficult) restriction of using *a priori* information on only p reference signals, with $p \ll N$, is quite significant. At the same time, beside difficulties that this restriction imposes on the estimation procedure, we use it in a way that allows us to avoid the hard work associated with known techniques applied to large signal sets. To the best of our knowledge, the exception is the methodology in [17], [18] where for estimation of a set of signals, the single estimator is constructed. The estimation techniques in [17], [18] exploit information in the form of a vector obtained, in particular, from averaging over signals in $\mathcal{K}_x^{(p)}$.

Further, the semiblind techniques are not applicable to the considered problem because they require a knowledge of some 'parts' of each reference signal in $\mathcal{K}_x$ (e.g., see [3], [5], [19]) but it is not the case here. Although the blind techniques allow us to avoid this restriction, it is known that they have difficulties related to accuracy and computational load. In the problem under consideration, the advantage is a knowledge of some (small) part of the set of reference signals. It is natural to use this advantage in the estimator structure and we will do it in Section II.

Nonblind estimators [6]–[16] are not applicable here because they require *a priori* information on each reference signal in $\mathcal{K}_x$ (e.g., a knowledge of covariance matrix $E[\boldsymbol{x}_\omega \boldsymbol{y}_\omega^T]$ where E is the expectation operator). In particular, it is known that there are significant advantages in adaptive or recursive estimators (e.g., associated with Kalman filtering approach) and it may well be possible to embed our estimator into such an environment—but that is not our particular concern here. Further, we note that much of the literature on piecewise linear estimators [23]–[26] seems to be *not directly relevant* to the estimator proposed here. In the first instance papers such as [23]–[26] are mostly concerned with the theoretical problems of approximation by piecewise linear functions on multi-dimensional domains which is *not the case here*.

Also, unlike many known techniques, we consider the case of observations corrupted by an arbitrary noise (not by an additive noise only) and design the estimator in terms of the Moore-Penrose pseudo-inverse matrix [27]. Therefore it is always well defined.

II. The main results

In this section we outline the rationale for the proposed estimator and state the main results.

A. Some preliminaries

The proposed estimator $F^{(p-1)}$ is adaptive to a sparse set $\mathcal{K}_x^{(p)}$.

The conceptual device behind the proposed estimator is a linear interpolation of an optimal incremental estimation applied to random signal pairs $(\boldsymbol{x}_\omega(t_j), \boldsymbol{y}_\omega(t_j))$ and $(\boldsymbol{x}_\omega(t_{j+1}), \boldsymbol{y}_\omega(t_{j+1}))$, for $j = 1, \ldots, p-1$, interpreted an extension of the least squares linear (LSL) estimator (see, for example, [6], [11], [16]).
Although this idea may seem reasonable the detailed justification of the new estimator is not straightforward and requires careful analysis. We shall do this by establishing an upper bound for the associated error and by showing that this upper bound is directly related to the expected error for an incremental application of the optimal LSL estimator. In Section II-B below, we will show that such an estimator is possible under quite unrestrictive assumptions.

Since the estimator proposed below is based on an extension of the LSL estimator it is convenient to sketch known related results here. Consider a *single* random signal pair $(\boldsymbol{x}(\omega), \boldsymbol{y}(\omega))$ where $\boldsymbol{x} \in L^2(\Omega, \mathbb{R}^m)$ and $\boldsymbol{y} \in L^2(\Omega, \mathbb{R}^n)$ with zero mean $(E[\boldsymbol{x}], E[\boldsymbol{y}]) = (\mathbf{0}, \mathbf{0})$, where $\mathbf{0}$ is the zero vector. Note that here, $\boldsymbol{x}$ and $\boldsymbol{y}$ do not depend on t as above. The estimate $\widehat{\boldsymbol{x}}$ of the reference vector $\boldsymbol{x}$ by the optimal least squares linear estimator is given by

$$\widehat{\boldsymbol{x}}(\omega) = E_{\boldsymbol{xy}} E_{\boldsymbol{yy}}^{\dagger}\, \boldsymbol{y}(\omega) \tag{2}$$

where $E_{\boldsymbol{xy}} = E[\boldsymbol{x}\boldsymbol{y}^T]$ and $E_{\boldsymbol{yy}} = E[\boldsymbol{y}\boldsymbol{y}^T]$ are known covariance matrices and $E_{\boldsymbol{yy}}^{\dagger}$ is the Moore-Penrose pseudo-inverse of $E_{\boldsymbol{yy}}$. By the LSL estimator, matrices $E_{\boldsymbol{xy}}$ and $E_{\boldsymbol{yy}}^{\dagger}$ should be specified for each signal pair $(\boldsymbol{x}(\omega), \boldsymbol{y}(\omega))$.

Further, for a justification of our estimator, we need some more notation as follows. It is convenient to denote $\boldsymbol{x}(t,\omega) = \boldsymbol{x}_\omega(t)$ and $\boldsymbol{y}(t,\omega) = \boldsymbol{y}_\omega(t)$ so that $\boldsymbol{x}(t,\omega) \in \mathbb{R}^m$ and $\boldsymbol{y}(t,\omega) \in \mathbb{R}^n$.

B. The piecewise LSL interpolation estimator

For each signal pair (or vector function pair) in the set $\mathcal{P}$, $(\boldsymbol{x}(t,\omega), \boldsymbol{y}(t,\omega))$, we assume that $(E[\boldsymbol{x}(t,\cdot)], E[\boldsymbol{y}(t,\cdot)]) = (\boldsymbol{0},\boldsymbol{0})$. To begin the estimation process we need to find an initial estimate $\widehat{\boldsymbol{x}}(t_1,\omega)$. It is assumed this can be found by some known method. Further, let us consider a signal estimation procedure at $t_2,\cdots,t_p$. We use an inductive argument to define an incremental estimation procedure. Consider a typical interval $[t_j, t_{j+1}]$ and define incremental random vectors

$$\boldsymbol{v}_j(\omega) = \boldsymbol{x}(t_{j+1},\omega) - \boldsymbol{x}(t_j,\omega) \in \mathbb{R}^m, \quad (3)$$

$$\boldsymbol{w}_j(\omega) = \boldsymbol{y}(t_{j+1},\omega) - \boldsymbol{y}(t_j,\omega) \in \mathbb{R}^n \quad (4)$$

and construct the optimal linear estimate

$$\widehat{\boldsymbol{v}}_j(\omega) = E_{\boldsymbol{v}_j\boldsymbol{w}_j} E^\dagger_{\boldsymbol{w}_j\boldsymbol{w}_j}\, \boldsymbol{w}_j(\omega) \quad (5)$$

of the increment $\boldsymbol{v}_j(\omega)$ for each $j = 1,\ldots,p-1$. We will write

$$B_j = E_{\boldsymbol{v}_j\boldsymbol{w}_j} E^\dagger_{\boldsymbol{w}_j\boldsymbol{w}_j} \in \mathbb{R}^{m\times n}. \quad (6)$$

Define the estimate at point t_{j+1} by setting $\widehat{\boldsymbol{x}}(t_{j+1},\omega) = \widehat{\boldsymbol{x}}(t_j,\omega) + \widehat{\boldsymbol{v}}_j(\omega)$. Thus we have

$$\begin{aligned}\widehat{\boldsymbol{x}}(t_{j+1},\omega) &= \widehat{\boldsymbol{x}}(t_j,\omega) + B_j[\boldsymbol{y}(t_{j+1},\omega) - \boldsymbol{y}(t_j,\omega)] \\ &= \boldsymbol{\epsilon}_j(\omega) + B_j\boldsymbol{y}(t_{j+1},\omega) \quad (7)\end{aligned}$$

where we write

$$\boldsymbol{\epsilon}_j(\omega) = \widehat{\boldsymbol{x}}(t_j,\omega) - B_j\boldsymbol{y}(t_j,\omega). \quad (8)$$

Note that this definition can be rewritten more suggestively as

$$\widehat{\boldsymbol{x}}(t_j,\omega) = \boldsymbol{\epsilon}_j(\omega) + B_j\boldsymbol{y}(t_j,\omega) \quad (9)$$

for each $j = 1,\ldots,p-1$.

The formula (7) shows that on each subinterval $[t_j, t_{j+1}]$ the estimate of the reference signal at t_{j+1} is obtained from the initial estimate at t_j by adding the optimal LSL estimate of the increment.

Now, consider a signal estimation at any $t \in [a,b]$. The first step is simply to extend the formulæ (7) and (9) to all $t \in [t_j, t_{j+1}]$ by defining

$$\widehat{\boldsymbol{x}}(t,\omega) = \boldsymbol{\epsilon}_j(\omega) + B_j\boldsymbol{y}(t,\omega). \quad (10)$$

Thus the incremental estimation across each subinterval is extended to every point within the subinterval. Because of determining estimate $\widehat{\boldsymbol{x}}(t_{j+1},\omega)$ in the form (5)–(7) we interpret this procedure as the *LSL piecewise interpolation*.

The incremental estimations are collected together in the following way. For each $j = 1,2,\ldots,p-1$, write

$$F_j[\boldsymbol{y}(t,\omega)] = \boldsymbol{\epsilon}_j(\omega) + B_j\boldsymbol{y}(t,\omega) \quad (11)$$

for all $t \in [t_j, t_{j+1}]$ and hence define the *piecewise LSL interpolation estimator* by setting

$$F^{(p-1)}[\boldsymbol{y}(t,\omega)] = \sum_{j=1}^{p-1} F_j[\boldsymbol{y}(t,\omega)][u(t-t_j) - u(t-t_{j+1})] \quad (12)$$

for all $t \in [a,b]$ where $u(t) = \begin{cases} 1 & \text{for } t > 0 \\ 0 & \text{otherwise.}\end{cases}$ is the unit step function. Thus we can now use the estimate

$$\widehat{\boldsymbol{x}}(t,\omega) = F^{(p-1)}[\boldsymbol{y}(t,\omega)] \quad (13)$$

for all $(t,\omega) \in T\times\Omega$. The idea of a piecewise LSL interpolation estimator on T seems intuitively reasonable for signals with a well defined gradient over T.

We note that by (6)-(13), the estimator $F^{(p-1)}$ is adaptive to a variation of signals in $\mathcal{K}_x^{(p)}$. A change of signals in $\mathcal{K}_x^{(p)}$ implies a change of determinations of sub-estimators B_j in (6) and keep the same structure of the $F^{(p-1)}$.

C. Justification of the LSL interpolation estimator

We wish to justify the proposed estimator by establishing an upper bound for the associated error.

To explain the technical details we introduce some further terminology.

Let us denote $\|\boldsymbol{x}(t,\cdot)\|^2_\Omega = \int_\Omega \|\boldsymbol{x}(t,\omega)\|^2 d\mu(\omega)$. Assume that for all $t \in T$, we have

$$\|\boldsymbol{x}(t,\cdot)\|^2_\Omega < \infty \quad \text{and} \quad \|\boldsymbol{y}(t,\cdot)\|^2_\Omega < \infty, \quad (14)$$

where $\|\boldsymbol{x}(t,\omega)\|$ and $\|\boldsymbol{y}(t,\omega)\|$ are the Euclidean norms for $\boldsymbol{x}(t,\omega)$ and $\boldsymbol{y}(t,\omega)$ for each $(t,\omega) \in T\times\Omega$, respectively. Thus we will say that the signals are square integrable in ω and write $\boldsymbol{x}(t,\cdot) \in L^2(\Omega)$ and $\boldsymbol{y}(t,\cdot) \in L^2(\Omega)$ for each fixed $t \in T$.

For each $t \in T$, let $\mathcal{F} = \{\boldsymbol{f} : T\times\Omega \to \mathbb{R}^m \mid \boldsymbol{f}(t,\cdot) \in L^2(\Omega,\mathbb{R}^m)\}$ and define

$$\begin{aligned}\|\boldsymbol{f}\|_{T,\Omega} &= \frac{1}{b-a}\int_{T\times\Omega} \|\boldsymbol{f}(t,\omega)\|\; dt\, d\mu(\omega) \\ &= \frac{1}{b-a}\int_T E[\|\boldsymbol{f}(t,\cdot)\|]\, dt\end{aligned}$$

for each $\boldsymbol{f} \in \mathcal{F}$ where $\|\boldsymbol{f}(t,\omega)\|$ is the Euclidean norm of $\boldsymbol{f}(t,\omega)$ on $\mathbb{R}^m$ for all $(t,\omega) \in \mathbb{R}^m$. Suppose that for all $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$ there exist constants $\gamma_j, \delta_j > 0$ such that

$$\|\boldsymbol{x}(s,\omega) - \boldsymbol{x}(t,\omega)\| \leq \gamma_j |s-t|, \qquad (15)$$

$$\|\boldsymbol{y}(s,\omega) - \boldsymbol{y}(t,\omega)\| \leq \delta_j |s-t| \qquad (16)$$

for all $(s,\omega),(t,\omega) \in [t_j, t_{j+1}] \times \Omega$, i.e. we suppose that the Lipschitz constants in (15) are independent of ω.

The error bound for the piecewise LSL interpolation estimator is established in Theorem 1 below.

Theorem 1: If condition (15) is satisfied then the error $\epsilon_p = \|\boldsymbol{x} - F^{(p-1)}[\boldsymbol{y}]\|_{T,\Omega}$ associated with the piecewise LSL interpolation estimator satisfies the inequality

$$\epsilon_p \leq \max_{j=1,\ldots,p-1} \{(\gamma_j + \|B_j\|_2 \delta_j)|t_{j+1} - t_j| \qquad (17)$$

$$+ \left[\|E^{1/2}_{\boldsymbol{v}_j,\boldsymbol{v}_j}\|^2_F - \|E_{\boldsymbol{v}_j\boldsymbol{w}_j}(E^{1/2}_{\boldsymbol{w}_j\boldsymbol{w}_j})^\dagger\|^2_F\right]^{1/2}\} \qquad (18)$$

where $\|B_j\|_2$ denotes the 2-norm given by the square root of the largest eigenvalue of $B_j^T B_j$ and $\|\cdot\|$ denotes the Frobenius norm.

III. Conclusion

The piecewise least squares linear (LSL) interpolation estimator was developed to estimate a large set of random signals of interest from the set of observed data. The distinctive feature is that *a priori* information can be obtained on only a *few* reference signals in the form of samples. Since no information of the major part of the set of reference signals is known, such a procedure is called *almost blind* estimation.

The proposed estimator mitigates to some extent the difficulties associated with existing estimation approaches such as the necessity to know information (in the form of a sample, for instance) on *each* random reference signal; invertibility of the matrices used to define the estimators; and demanding computational work.

References

[1] E. Zeidler, *Applied Functional Analysis, Applications to Mathematical Physics*, Applied Mathematical Sciences 108, Springer, 1997.

[2] Y. Hua, Fast maximum likelihood for blind identification of multiple FIR channels, *IEEE Trans. on Signal Processing,* 44, No. 3, pp. 661-672, 1996.

[3] K. Georgoulakis and S. Theodoridis, Blind and semi-blind equalization using hidden Markov models and clustering techniques, *Sifnal Processing,* 80, Issue 9, pp. 1795-1805, 2000.

[4] V. Zarzoso and P. Comon, Blind and Semi-Blind Equalization Based on the Constant Power Criterion, *IEEE Trans. on Signal Processing,* 53, 11, pp. 4363–4375, 2005.

[5] C.–Y. Chi, C.–H. Chen, C.–C. Feng, C.–Y. Chen, *Blind Equalization and System Identification*, Springer, 2006.

[6] Y. Hua, M. Nikpour, and P. Stoica, Optimal Reduced-Rank estimation and filtering, *IEEE Trans. on Signal Processing,* vol. 49, pp. 457-469, 2001.

[7] Y. Hua and W. Q. Liu, Generalized Karhunen-Loève transform, *IEEE Signal Processing Letters*, vol. 5, pp. 141-143, 1998.

[8] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, ISBN 0-13-048434-2, 2002.

[9] J. Chen, J. Benesty, Y. Huang, and S. Doclo, New Insights Into the Noise Reduction Wiener Filter, *IEEE Trans. on Audio, Speech, and Language Processing,* **14**, No. 4, 1218 - 1234, 2006.

[10] M. Spurbeck and P. Schreier, Causal Wiener filter banks for periodically correlated time series, *Signal Processing,* 87, 6, pp. 1179-1187, 2007.

[11] J. S. Goldstein, I. Reed, and L. L. Scharf, A Multistage Representation of the Wiener Filter Based on Orthogonal Projections, *IEEE Trans. on Information Theory,* vol. 44, pp. 2943-2959, 1998.

[12] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*, J. Wiley & Sons, 2001.

[13] A. P. Torokhti and P. G. Howlett, An Optimal Filter of the Second Order, *IEEE Trans. on Signal Processing*, **49**, No 5, 1044-1048, 2001.

[14] A. Torokhti and P. Howlett, Method of recurrent best estimators of second degree for optimal filtering of random signals, *Signal Processing*, **83**, 5, 1013-1024, 2003.

[15] A. Torokhti and P. Howlett, Optimal Transform Formed by a Combination of Nonlinear Operators: The Case of Data Dimensionality Reduction, *IEEE Trans. on Signal Processing*, **54**, No. 4, 1431-1444, 2006.

[16] A. Torokhti and P. Howlett, *Computational Methods for Modelling of Nonlinear Systems*, Mathematics in Science and Engineering, **212**, series editor C. K. Chui, elsevier, 2007.

[17] A. Torokhti and P. Howlett, Filtering and Compression for Infinite Sets of Stochastic Signals, *Signal Processing*, 89, pp. 291-304, 2009.

[18] A. Torokhti and J. Manton, Generic Weighted Filtering of Stochastic Signals, *IEEE Trans. on Signal Processing,* 57, issue 12, pp. 4675-4685, 2009.

[19] H. A. Cirpan and M. K. Tsatsanis, Stochastic Maximum Likelihood Methods for Semi-Blind Channel Estimation, *IEEE Signal Processing Letters,* 5, No. 1, pp. 21–24,1998.

[20] G. Kutz and D. Raphaeli, Maximum-Likelihood Semiblind Equalization of Doubly Selective Channels Using the EM Algorithm, *EURASIP Journal on Advances in Signal Processing*, Springer Open J., 2010.

[21] D. He and H. Leung, Semi-Blind Identification of ARMA Systems Using a Dynamic-Based Approach, *IEEE Trans. on Circuits and Systems-I*, 52, 1, pp. 179–190, 2005.

[22] J. Even and K. Sugimoto, An ICA approach to semi-blind identification of strictly proper systems based on interactor polynomial matrix, *Int. J. Robust Nonlinear Control*, 17, pp. 752768, 2007.

[23] S. Kang and L. Chua, A global representation of multidimensional piecewise-linear functions with linear partitions, *IEEE Trans. on Circuits and Systems,* 25 Issue:11, pp. 938 - 940, 1978.

[24] L.O. Chua and A.-C. Deng, Canonical piecewise-linear representation, *IEEE Trans. on Circuits and Systems,* **35** Issue:1, pp. 101 - 111, 1988.

[25] P. Julian, A. Desages, B. D'Amico, Orthonormal high-level canonical PWL functions with applications to model reduction, *IEEE Trans. on Circuits and Systems I: Fundamental Theory and Applications,* 47 Issue:5, pp. 702 - 712, 2000.

[26] J.E. Cousseau, J.L. Figueroa, S. Werner, T.I. Laakso, Efficient Nonlinear Wiener Model Identification Using a Complex-Valued Simplicial Canonical Piecewise Linear Filter, *IEEE Trans. on Signal Processing*, 55 5, pp. 1780 - 1792, 2007.

[27] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications,* John Wiley & Sons, New York, 1974.

[28] P. G. Howlett, P. Pudney, and X. Vu, Local energy minimization in optimal train control, *Automatica*, **45**(11), 2692–2698, 2009. DOI: 10.1016/j.automatica.2009.07.028.