Generalized quadratic minimization and a signal processing application

A. Gorokhov

Laboratoire des Signaux & Systèmes
Supélec Plateau du Moulon
91192 Gif-sur-Yvette Cédex FRANCE
e-mail: gorokhov@lss.supelec.fr

P. Stoica

Systems and Control Group
Uppsala University
Box 27, SE–751 03 Uppsala SWEDEN
e-mail: ps@syscon.uu.se

ABSTRACT

This paper deals with the problem of quadratic minimization subject to linear equality constraints. Contrary to the standard formulation, we assume the most general case of a possibly singular quadratic form. As we explain, the existing formal solution to this problem has several drawbacks. Our new approach is free from most of these drawbacks. It has a simple physical interpretation and is relatively easy to implement. Practical importance of this result lies in its numerous applications : filter design, spectral analysis, direction finding and blind deconvolution of multiple FIR channels. Here we focus on the blind deconvolution application.

General context

The linearly constrained quadratic optimization problem has been usually addressed in the case of strictly positive quadratic forms. The particular case of a *regular* quadratic form leads to a simple solution. However, in some applications the quadratic form to be minimized may be singular. A formal solution to such a generalized linearly constrained quadratic minimization problem is also available [1, 2]. The authors of [3] discuss the application of this result in signal processing with particular emphasis to the filter design problem. However, the existing solution in the general *singular* case suffers from a number of important drawbacks. As explained in the following section, this solution assumes knowledge of the rank of a certain matrix. Imprecise knowledge of that rank may lead to a hard failure. Additionally, the implementation of the existing solution is computationally burdensome in certain applications.

In this paper, we describe the entire set of solutions to the stated problem and prove that the known result of [1] yields the particular *minimum norm* solution. We also develop a new form of the latter solution. Contrary to the existing solution, the new method does not require any rank information and is also easy to implement. In addition, the new method has a clear interpretation : it may be obtained as a regularized version of the classical solution to the regular case. We prove that this method converges to the existing solution in the singular case as the regularization factor vanishes. Based on this property, we recommend the use of a very small regularization factor : this is in effect a main feature of our approach which distinguishes it from the standard regularization method.

Background

We consider the following optimization problem :

$$\min_{F\in\mathcal{F}} F^H R F, \quad \mathcal{F} = \{F \in \mathbb{C}^{M\times m} \ : \ F^H\Phi = C\}. \quad (1)$$

Here the superscript $(^H)$ denotes the conjugate transpose, R is an $M \times M$ Hermitian matrix, Φ and C are $M\times n$ and $m\times n$ matrices respectively with $m < M$ and $n < M$. In words, one needs to find an $M\times m$ minimizer of the quadratic form $F^H R F$ subject to a set of n linear constraints $F^H\Phi = C$. Recall that minimizing $F^H R F$ w.r.t. $F \in \mathcal{F}$ means finding some $F' \in \mathcal{F}$ such that $(F^H R F - F'^H R F')$ is a non-negative definite $m \times m$ matrix for all $F \in \mathcal{F}$. We assume that R is non-negative definite, *i.e.*, $R \geq 0$, and that the matrix Φ is full-rank. The existence of solutions to (1) is guaranteed by the fact that $R \geq 0$, which ensures that $F^H R F \geq 0$, and the full-rank property of Φ which ensures that $\mathcal{F}$ is not empty or equivalently, that $F^H\Phi = C$ holds for some F.

In some applications, R may be assumed to be strictly positive definite and well-conditioned. In such regular cases, problem (1) has a unique solution derived in [4] :

$$F_o = R^{-1}\Phi\left(\Phi^H R^{-1}\Phi\right)^{-1} C^H,$$
$$F_o^H R F_o = \left(\Phi^H R^{-1}\Phi\right)^{-1}. \quad (2)$$

In other applications, however, the matrix R is ill-conditioned or even exactly singular. For this general case, a particular minimizer of (1) presented in [1, 2] is given by :

$$F_P = [R + \Phi P\Phi^H]^{\#}\Phi\left(\Phi^H[R + \Phi P\Phi^H]^{\#}\Phi\right)^{-1} C^H,$$
$$F_P^H R F_P = C\left[\left(\Phi^H[R + \Phi\,\Phi^H]^{\#}\Phi\right)^{-1} - \mathrm{I}_n\right] C^H, \quad (3)$$

where P is an arbitrary $n \times n$ positive definite matrix, I_n is $n \times n$ identity matrix and the superscript $(^{\#})$ denotes the Moore-Penrose pseudo-inverse. The essential difference from (2) consists of replacing the inverse of R by the pseudo-inverse of $(R + \Phi P \Phi^H)$. The computation of this pseudo-inverse requires the knowledge of rank($R + \Phi P \Phi^H$) which is often unavailable. Moreover, the reliable computation of the pseudo-inverse in (3) from the available (empirical) quantities is complicated whenever the matrices R and Φ are ill-conditioned. An example of such a situation is the filter design problem addressed in [3]. Clearly rank estimation errors may lead to a hard failure of (3). Hence the direct implementation of this solution is not recommended in many realistic environments.

Another shortcoming of (3) comes from the necessity to calculate the pseudo-inverse of a singular $M \times M$ matrix instead of a simple inverse as in (2). The burden of this operation grows with increasing M. In high-resolution spectral analysis, for example as applied to large aperture radars, the use of (3) may pose serious complexity problems.

The aforementioned drawbacks of (3) motivated us to study the problem (1) in details, with the aim to find simpler solutions. In the following section we describe the whole set of solutions to (1) in the most general case. The main practical outcome of this work is a computationally efficient alternative to (3) which is free of the above mentioned drawbacks.

Main results

We first indicate that the most general condition ensuring the existence of solutions to (1) is given by:

$$R \;\geq\; 0, \qquad \text{rank}(\ \Phi\ ) = \text{rank}(\ [\Phi^T,\, C^T]^T\ ). \qquad (4)$$

Note that rank(Φ) $\leq$ rank($[\Phi^T,\, C^T]^T$) $\leq n$ so that the full-rank property of Φ, *i.e.*, rank(Φ) $= n$, is a sufficient condition for the right-hand equality in (4). Next, we establish the set of solutions to our problem.

Theorem 1 *Assume that (4) is verified. Then the entire set of solutions to (1) is given by* $\{F_V\}$ *:*

$$\begin{aligned} F_V &= [R + \Phi\Phi^H]^{\#}\Phi\left(\Phi^H[R+\Phi\Phi^H]^{\#}\Phi\right)^{\#}C^H + V,\\ F_V^H R F_V &= C\left[\left(\Phi^H[R+\Phi\Phi^H]^{\#}\Phi\right)^{\#} - \Phi^{\#}\Phi\right]C^H \end{aligned}$$

where V *is any* $M \times m$ *matrix such that* span{ V } $\subset$ null{ $R + \Phi\Phi^H$ }.

This statement means, in particular, that the solution F_P, which corresponds to F_V with $V = 0$, is invariant to the choice of P, *i.e.*, it always coincides with

$$F_\star = [\,R + \Phi\,\Phi^H\,]^{\#}\,\Phi\,\left(\Phi^H[\,R + \Phi\,\Phi^H\,]^{\#}\,\Phi\right)^{\#}C^H. \quad (5)$$

It is worth mentioning that $F_\star$ is a *minimum norm* solution to (1) since it is a particular case of F_V with $V = 0$.

From a practical standpoint, the direct calculation of $F_\star$ inherits all the drawbacks of (3) discussed in the previous section. An efficient alternative implementation of (5) is possible due to the following result.

Theorem 2 *Define the family* $\{F_\lambda\}$ *with scalar* $\lambda > 0$ *according to*

$$F_\lambda \triangleq [R + \lambda \mathrm{I}_M]^{-1}\Phi\left(\Phi^H[R+\lambda\,\mathrm{I}_M]^{-1}\Phi + \lambda \mathrm{I}_n\right)^{-1}C^H.$$

Then F_λ *and* $F_\lambda^H R F_\lambda$ *are continuous functions of* λ *on* $(0, \infty)$ *and*

$$\begin{aligned} &\lim_{\lambda\to 0} F_\lambda = F_\star, & (6)\\ &\lim_{\lambda\to 0} F_\lambda^H R F_\lambda = C\left[\left(\Phi^H[R+\Phi\Phi^H]^{\#}\Phi\right)^{\#} - \Phi^{\#}\Phi\right]C^H. \end{aligned}$$

In words, one can arbitrarily approach the solution to (1) in the most general case by setting λ sufficiently small. Moreover, F_λ may be regarded as a 'slightly regularized' version of the classical result (2). Indeed, a positive definite matrix R in (2) is replaced by $(R+\lambda\,\mathrm{I}_M)$ in the case of a possibly singular R.

Application to blind multichannel deconvolution

Later in the paper, we focus on the blind deconvolution of multiple FIR channels. We first recall the principle of the *ad-hoc* linear prediction method [5]. Based on the results of the previous section, we derive a new estimator which yields deterministic deconvolution if the model order is known and remains consistent when the order is over-estimated.

Consider the standard setup of blind identification of multiple FIR filters :

$$x(t) = \sum_{\tau=0}^{L} \mathbf{h}(\tau)\,s(t-\tau) + e(t), \qquad t \in \mathbb{Z}. \qquad (7)$$

In the absence of noise $\{e\,(t)\}_{t\in\mathbb{Z}}$, $\{x\,(t)\}_{t\in\mathbb{Z}}$ stands for the q-variate series observed at the output of a linear system described by the $q \times 1$ impulse response $\{\mathbf{h}(\tau)\}_{\tau=0}^{L}$ and excited by the scalar stationary series $\{s\,(t)\}_{t\in\mathbb{Z}}$. Our task is the identification of system parameters $\mathbf{h}(0), \ldots, \mathbf{h}(L)$, up to a non-essential scale factor, from a number T of consecutive samples of $\{x\,(t)\}_{t\in\mathbb{Z}}$. An important condition of this setup is that the order L of the system is unknown. A model order L' is then chosen so that $L' \geq L$. Define $h(z) = \sum_{\tau=0}^{L}\mathbf{h}(\tau)\,z^{-\tau}$, the transfer function of the system. To ensure the global identifiability of $h(z)$ from the second order statistics of $\{x\,(t)\}_{t\in\mathbb{Z}}$, we assume that $h(z) \neq 0$ for any $z \in \mathbb{C}$, and that $\{s\,(t)\}_{t\in\mathbb{Z}}$ is a zero-mean unit-variance white noise. Under these hypotheses, the authors of [5] showed that in the absence of noise, $\{x\,(t)\}_{t\in\mathbb{Z}}$ is also a finite-order AR series, *i.e.*, there exists N and $q \times q$ matrices $\mathbf{A}(1), \ldots, \mathbf{A}(N)$ such that

$$x(t) \;+\; \sum_{\tau=1}^{N}\mathbf{A}(\tau)\,x(t-\tau) \;=\; \mathbf{h}(0)\,s(t), \qquad t \in \mathbb{Z}, \;(8)$$

and $\{\mathbf{h}(0)\,s\,(t)\}_{t\in\mathbb{Z}}$ is the innovation of $\{x\,(t)\}_{t\in\mathbb{Z}}$. An important fact is that $\mathbf{A} \triangleq [\mathbf{A}(1), \ldots, \mathbf{A}(N)]$ satisfying (8) is not unique. Additionally, note that the innovation series has a singular covariance matrix : $D = \mathbb{E}\left\{\mathbf{h}(0)\, s(t)\, s(t)^H\, \mathbf{h}(0)^H\right\} = \mathbf{h}(0)\,\mathbf{h}(0)^H$. In fact, $\mathbf{A}$ and D satisfy the generalized Yule-Walker equation :

$$[\mathrm{I}_q,\, \mathbf{A}]\, \mathbf{R}_x = [D,\, 0_{q\times qN}], \quad \mathbf{R}_x = \mathbb{E}\{\underline{x}(t)\, \underline{x}(t)^H\} \quad (9)$$

where $\underline{x}(t) = [x(t)^T, \ldots, x(t-N)^T]^T$. Hence consistent estimates $\hat{\mathbf{A}}$ and $\hat{D}$ of a particular $\mathbf{A}$ (usually minimum norm solution to (9)) and D may be obtained from the empirical covariance matrix $\hat{\mathbf{R}}_x = (T-N)^{-1}\sum_{t=N+1}^{T} \underline{x}(t)\, \underline{x}(t)^H$, see [5, 6]. An estimate of $\mathbf{h}(0)$ up to a scaling factor can be calculated using the fact that $D = \mathbf{h}(0)\,\mathbf{h}(0)^H$. Let $\hat{u}$ be the $q\times 1$ eigenvector corresponding to the dominant eigenvalue of $\hat{D}$ such that $\|\hat{u}\|^2$ is equal to this dominant eigenvalue. Since $\mathrm{rank}(\ D\ ) = 1$, we should have $\hat{u}\, \hat{u}^H \approx \hat{D}$. It was shown (see [6]), that in the noiseless case

$$\hat{u} = \hat{\alpha}\, \mathbf{h}(0), \qquad \text{and} \qquad \hat{D} = |\hat{\alpha}|^2\, D \qquad (10)$$

holds for any $T > N + L$, where $|\hat{\alpha}| \xrightarrow{\mathrm{P}} 1$ as $T \to \infty$, while the argument of $\hat{\alpha}$ is undefined.

As shown in [5], the $\hat{\mathbf{A}}$ and $\hat{u}$ yield an estimate of the parameters $\mathbf{h}(0), \ldots, \mathbf{h}(L)$ up to a multiplicative scalar. The main drawback of the *ad-hoc* approach proposed in [5] is a poor estimation accuracy even in the absence of noise unless T is large.

A statistical approach to estimating $\mathbf{h}(0), \ldots, \mathbf{h}(L)$, derived in this paper, consists of fitting a suitable function of these parameters to the estimates $\hat{\mathbf{A}}$ and $\hat{u}$. Using (7) rewritten in a compact *operator* form as $x(t) = [h(z)]\, s(t)$, and a similar operator form for its autoregressive counterpart (8), we find :

$$[A(z)]\, x(t) = \mathbf{h}(0)\, s(t) \quad \Rightarrow \quad A(z)\, h(z) = \mathbf{h}(0),$$

where $A(z) = \mathrm{I}_q + \sum_{\tau=1}^{N} \mathbf{A}(\tau)\, z^{-\tau}$. The last equation implies, for any order $L' \geq L$, that :

$$\mathcal{T}_{L'}(A)\, \mathbf{h} \,=\, [\mathbf{h}(0)^T, 0_{1\times q(N+L')}], \qquad (11)$$

where $\mathcal{T}_{L'}(A)$ is a block Toeplitz matrix with $(N+L'+1)$ vertical and $(L'+1)$ horizontal blocks, the first block column $[\mathrm{I}_q, \mathbf{A}(1)^T, \ldots, \mathbf{A}(N)^T,\ 0\ldots 0]$ and the first block row $[\mathrm{I}_q, 0\ldots 0]$ and $\mathbf{h} = [\mathbf{h}(0)^T, \ldots, \mathbf{h}(L)^T, 0_{1\times q(L'-L)}]$.

Our model fitting approach consists of minimizing the residual of this equation w.r.t. the set of parameters in h after replacing $\mathcal{T}_{L'}(A)$ and $[\mathbf{h}(0)^T, 0_{1\times q(N+L')}]^T$ by $\mathcal{T}_{L'}(\hat{A})$ and $[\hat{u}^T, 0_{1\times q(N+L')}]^T$, respectively. More exactly, we consider the minimization problem

$$\hat{\mathbf{h}}_W = \arg\min_{\vartheta} \|\, \mathcal{T}_{L'}(\hat{A})\, \vartheta - [\hat{u}^T, 0_{1\times q(N+L')}]^T\, \|_W, \quad (12)$$

where ϑ is a $q(L'+1)\times 1$ dummy vector of parameters, W is a $q(N+L'+1)\times q(N+L'+1)$ non-negative definite weighting matrix and $\|v\|_W^2 \triangleq v^H\, W\, v$. Assuming that the inverse matrix below exists, the solution to (12) is

$$\hat{\mathbf{h}}_W = \left(\mathcal{T}_{L'}(\hat{A})^H\, W\, \mathcal{T}_{L'}(\hat{A})\right)^{-1} \times$$
$$\times\ \mathcal{T}_{L'}(\hat{A})^H\ W\ [\hat{u}^T, 0_{1\times q(N+L')}]^T. \qquad (13)$$

The arbitrary scaling factor $\hat{\alpha}$ in the estimate $\hat{u}$ also appears in $\hat{\mathbf{h}}_W$. To take this non-significant indeterminacy into account, we define $\hat{\vartheta}_W$ such that $\hat{\mathbf{h}}_W = \hat{\alpha}\, \hat{\vartheta}_W$.

Our goal is to find the *optimal* W that minimizes the variance of the estimation error $(\hat{\vartheta}_W - \mathbf{h})$. To this end, we show that the minimization problem (12) is asymptotically equivalent to the problem (1). Let $\vartriangle\!\hat{\mathbf{A}} \triangleq \hat{\mathbf{A}} - \mathbf{A}$. Using the fact that $\vartriangle\!\hat{\mathbf{A}} \xrightarrow{\mathrm{P}} 0$, one can show that

$$\hat{\vartheta}_W \doteq \arg\min_{\vartheta}\, (\,\Phi\,\vartheta - \chi\,)^H\ W\ (\,\Phi\,\vartheta - \chi\,) \qquad (14)$$
$$\Phi = \mathcal{T}_{L'}(A),\ \ \chi = [\tilde{\mathbf{h}}(0)^T, 0_{1\times q(N+L')}]^T - \mathcal{T}_{L'}(\vartriangle\!\hat{A})\, \mathbf{h},$$

where $(\doteq)$ denotes an asymptotic equivalence (after neglecting the term $\mathcal{T}_{L'}(\vartriangle\!\hat{A})\,(\vartheta - \mathbf{h})$) and $\tilde{\mathbf{h}}(0) = \hat{\alpha}^{-1}\hat{u}$. Like in (12)-(13), the closed-form solution to (14) yields

$$\hat{\vartheta}_W \doteq \left(\Phi^H\, W\, \Phi\,\right)^{-1} \Phi^H\, W\, \chi. \qquad (15)$$

Denote $F_W = W\,\Phi\,(\,\Phi^H\, W\,\Phi\,)^{-1}$, then $\hat{\vartheta}_W = F_W^H\, \chi$. By (11) and (14), $\chi \xrightarrow{\mathrm{P}} [\mathbf{h}(0)^T, 0_{1\times q(N+L')}]^T = \Phi\,\mathbf{h}$. First of all, the class of *consistency* of (12) must be ensured which implies that $\hat{\vartheta}_W \xrightarrow{\mathrm{P}} \mathbf{h}$, or equivalently, $F_W^H\, \chi \xrightarrow{\mathrm{P}} \mathbf{h}$ for all $\mathbf{h}$. This latter yields $F_W^H\, \Phi\,\mathbf{h} = \mathbf{h}$, *i.e.*, the set of linear constraints $F_W^H\, \Phi = \mathrm{I}$. The covariance matrix of $\hat{\vartheta}_W$ may be written as $F_W^H\ \mathbb{E}\{(\chi - \Phi\,\mathbf{h})(\chi - \Phi\,\mathbf{h})^H\}\ F_W$. Define the normalized residual error $\xi_T = \sqrt{T}\,(\Phi\,\mathbf{h} - \chi)$ and its asymptotic covariance matrix $R = \lim_{T\to\infty} \mathbb{E}\{\,\xi_T\, \xi_T{}^H\}$. Now, the minimum (asymptotic) variance estimate $\hat{\vartheta}_W = F_W^H\, \chi$ is given by F_W which minimizes $F_W^H\, R\, F_W$ subject to $F_W^H\, \Phi = \mathrm{I}$, *i.e.*, the problem (1). Since $F_W = W\,\Phi\,(\,\Phi^H\, W\,\Phi\,)^{-1}$, the solutions $F_\star$ and F_λ (see (5) and theorem 2), imply two optimal weightings :

$$W_\star = [R + \mathcal{T}_{L'}(A)\, \mathcal{T}_{L'}(A)^H]^{\#}; \quad W_\lambda = [R + \lambda\, \mathrm{I}_q]^{-1} \quad (16)$$

with $\lambda \to 0$. The term $\lambda\, \mathrm{I}_n$ in F_λ is neglected since Φ is a full-rank matrix which guarantees that $\lim_{\lambda\to 0} F_\lambda$ remains unchanged after removing $\lambda\, \mathrm{I}_n$.

We next study the optimal estimator *in the absence of noise.* According to (14), $\xi_T = \sqrt{T}(\,\mathcal{T}_{L'}(\vartriangle\!\hat{A})\ \mathbf{h} + [(\mathbf{h}(0) - \tilde{\mathbf{h}}(0))^T, 0_{1\times q(N+L')}]^T\,)$. Due to (10), we have in the absence of noise $\mathbf{h}(0) = \tilde{\mathbf{h}}(0)$. Hence $R = \lim_{T\to\infty} T\ \mathbb{E}\{\,\mathcal{T}_{L'}(\vartriangle\!\hat{A})\, \mathbf{h}\, \mathbf{h}^H\, \mathcal{T}_{L'}(\vartriangle\!\hat{A})^H\}$. As shown in [6], the matrix R has the following expression :

$$R = \begin{bmatrix} 0_q & 0 & 0 \\ 0 & \mathrm{I}_{N+L}\otimes D & 0 \\ 0 & 0 & 0_{q(L'-L)} \end{bmatrix}, \qquad (17)$$

where $(\otimes)$ denotes the Kronecker product of matrices. Note that this $q(L+1)\times q(L+1)$ matrix is rank-deficient. Indeed, $\mathrm{rank}(\ D\ ) = 1$ implies that $\mathrm{rank}(\ R\ ) = N+1$. Hence the classical solution (2) designed for positive definite R is not applicable in this case. The estimate $\hat{R}$ of R is calculated assuming L' as the true order, *i.e.*, as if $L = L'$. Clearly this estimate is not consistent unless the true order is known. In the latter case, $\hat{R}$

coincides with R up to a constant, even for finite T (recall that we are in the noiseless case). Indeed, (10) yields $\hat{R} = |\hat{\alpha}|^2 R$. Consequently, the estimate of W_λ yields the same result as the true weighting : $\hat{W}_\lambda = [\hat{R} + \lambda \, \mathrm{I}_M]^{-1}$ $= |\hat{\alpha}|^{-2} [R + \hat{\lambda} \, \mathrm{I}_M]^{-1}$, $\hat{\lambda} = \lambda \, / \, |\hat{\alpha}|^2$. Unlike $\hat{W}_\lambda$, the use of $\hat{W}_\star = [\hat{R} + \mathcal{T}_{L'}(\hat{A}) \, \mathcal{T}_{L'}(\hat{A})^H]^\#$ is not equivalent to the use of $W_\star$, for finite T, because of the errors in the estimate $\hat{\mathbf{A}}$. Also recall that the implementation of $\hat{W}_\star$ is computationally expensive and requires a reliable rank estimation prior to pseudo-inversion. However the most important argument for the use of $\hat{W}_\lambda$ is that the weighting W_λ (and therefore $\hat{W}_\lambda$) yields the *perfect reconstruction* of $\mathbf{h}$ up to a multiplicative scalar when $L' = L$. This property holds under the technical condition $L' \leq N + 2\,L$, see [6]. The latter condition also ensures the consistency of $\hat{\mathbf{h}}_{W_\lambda}$ when $L' > L$. We thus recommend the use of (13) with $\hat{W}_\lambda$, $\lambda \to 0$.

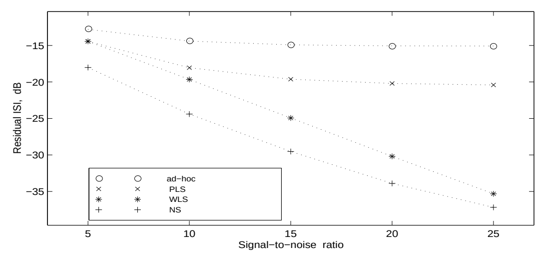
Obviously the optimality $\hat{W}_\lambda$ does not hold in *the presence of noise*. A study of the class (13) of estimators in the case of the model (7) as well as the derivation of the optimal weighting W has been undertaken in [6]. Based on a rather complicated optimal solution, the following sub-optimal weighting was derived :

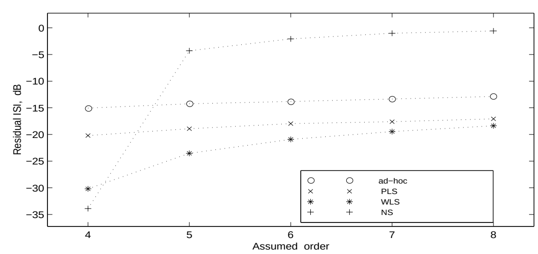
$$\hat{W}_\sigma = [\hat{R} + \hat{\sigma}^2 \, \mathrm{I}_M]^{-1}, \tag{18}$$

where $\hat{\sigma}^2$ is an estimate of the noise power. This weighting is a natural generalization of $\hat{W}_\lambda$ since $\hat{W}_\sigma$ reduces to $\hat{W}_\lambda$ as $\sigma \to 0$. In the case of known order ($L' = L$), $\hat{\mathbf{h}}_\sigma$ converges to a scaled version of $\mathbf{h}$ as $\sigma^2 \to 0$, even for finite T. This result follows from our discussion in the noise-free case.

Numerical study

Our objective is to justify the use of the sub-optimal Weighted Least Squares (WLS) estimator given by (13) and (18) as an alternative to the *ad-hoc* method presented in [5], the *plain* Least Squares (PLS) estimator given by (13) with $W = \mathrm{I}_M$ and the noise subspace based identification technique (NS) described in [7]. The simulations were performed by mimicking a digital communication scenario. The observation signal $\{y\,(t)\}_{t \in \mathbb{Z}}$ is generated as the output of $q = 4$ identical equispaced antennas with $h(z)$ describing the propagation channel between the emitter and the receiving array. The input signal $\{s\,(t)\}_{t \in \mathbb{Z}}$ has a QAM-4 digital modulation (transmitted at rate 500 symbols/s and shaped by the raised-cosine filter with rolloff 1/2). The received signal is sampled at the baud rate. The average signal-to-noise ratio (SNR) per antenna $SNR = \frac{1}{q\,\sigma^2} \sum_\tau \|\mathbf{h}(\tau)\|^2$ is set to 20 dB, unless otherwise stated. On the basis of $T = 300$ consecutive samples of $\{y\,(t)\}_{t \in \mathbb{Z}}$, we use the *ad-hoc* method, PLS, WLS and NS to estimate $h(z)$. The quality of estimation is measured as follows. We calculate the linear system $\hat{\gamma}(z) = \hat{h}(z)^\# \, h(z)$ which verifies $\hat{\gamma}(z) \to \hat{\alpha}^{-1}$ since $\hat{h}(z) \to \hat{\alpha}\, h(z)$ as $T \to \infty$. Writing the Laurent series expansion $\hat{\gamma}(z) = \sum_\tau \underline{\hat{\gamma}}(\tau)\, z^{-\tau}$, we observe that the coefficients $\underline{\hat{\gamma}}(\tau)$, $\tau \neq 0$ vanish as $T \to \infty$. Note that these coefficients specify the residual *inter-symbol interference* (ISI) if the input signal is estimated via the *minimum norm* left inversion of $h(z)$. We will study the quantity $ISI = \sum_{\tau \neq 0} |\underline{\hat{\gamma}}(\tau)|^2$ which is not is not affected by the indeterminacy introduced by $\hat{\alpha}$. In the following figures, the symbols ('$\cdots o \cdots$'), ('$\cdots x \cdots$'), ('$\cdots * \cdots$') and ('$\cdots + \cdots$') show the residual ISI, averaged over 100 independent Monte-Carlo trials, for the *ad-hoc* solution, PLS, WLS and NS, respectively.





The above figures show the residual ISI versus the average SNR value ($L' = 4$) and the model order L'. Unlike the NS method, which lacks consistency for large L', the WLS shows an acceptable estimation accuracy which is better than the accuracy of PLS.

References

[1] T. Söderström and P. Stoica, *System Identification*. Prentice-Hall International (UK) Ltd., 1989.

[2] C.R. Rao and S.K. Mitra, *Generalized Inverse of Matrices and its Applications*. John Wiley & Sons, Inc., 1971.

[3] P. Stoica, A. Jakobsson, J. Li, "Matrix optimization result with dsp applications," in *Digital Signal Processing - a review journal*, vol. 6, pp. 195–201, 1996.

[4] O.L. Frost III, "An algorithm for linear constrained adaptive array processing," *Proc. IEEE*, vol. 60, pp. 926–935, Aug. 1972.

[5] K. Abed-Meraïm, E. Moulines and P. Loubaton, "Prediction error methods for second-order blind identification," *IEEE Tr. on Sig. Proc.*, vol. 45, pp. 694–705, Mar. 1997.

[6] A. Gorokhov, *Blind separation of convolutive mixtures : second order methods*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, 1997.

[7] E. Moulines, P. Duhamel, J.F. Cardoso and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Tr. on Sig. Proc.*, vol. 43, pp. 516–526, Feb. 1995.