ON THE CHOICE OF A BASIS FOR A LINEAR-IN-THE-PARAMETERS MODEL

Tomás Oliveira e Silva

Departamento de Electrónica e Telecomunicações / IEETA
Universidade de Aveiro, 3810-193 Aveiro, Portugal
e-mail: tos@det.ua.pt

ABSTRACT

The unit pulse response of a linear-in-the-parameters model lies in a given finite-dimensional subspace. This paper studies the problem of selecting a good basis for this subspace, in the case where it is known that the power spectral density of the input signal of the model is bounded, below and above, by two known power spectral densities. We attempt to make the condition number of the correlation matrix of the internal signals of the model as small as possible for the worst possible power spectral density of the input signal.

1 Introduction

This paper is concerned with linear, stable, linear-in-the-parameters models, whose general form is depicted in Fig. 1. In this figure, q is the advance operator, $G_k(z)$, $k = 1, \ldots, n$, are stable transfer functions, and the signals $x_k(t)$ are the internal signals of the model. Many adaptive filters can also be put in this form, in which case the weights w_k of the model become time-variant.

Referring to Fig. 1, the transfer function of the linear-in-the-parameters model is given by

$$G(z; \boldsymbol{w}) = \sum_{k=1}^{n} w_k \, G_k(z) = \boldsymbol{w}^{H} \boldsymbol{G}(z),$$

where the superscript ^{H} denotes complex conjugation followed by vector or matrix transposition, and where lower case letters in boldface represent column vectors. In many applications one chooses the weights that make

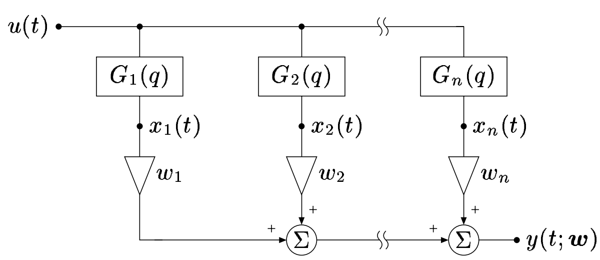

Figure 1: A general "linear in the parameters" model.

the model's output signal $y(t; \boldsymbol{w})$ as close as possible, in a least squares sense, to a given desired signal $d(t)$.

It is clear that the transfer functions $G(z; \boldsymbol{w})$ lie in the finite-dimensional subspace spanned by the transfer functions $G_1(z), \ldots, G_n(z)$, which we will assume to be linearly independent. Anyone attempting to use a model of this kind has to address two problems:

1. select an appropriate subspace, and
2. select a basis for the chosen subspace.

The first problem is quite difficult, so one usually uses a FIR model, in which case $G_k = z^{1-k}$. Fortunately, there exist some theoretical results that indicate that FIR models are the best choice if the only known *a priori* knowledge about the system we are trying to model is that it has some degree of asymptotic stability [1, 2]. More precisely, an FIR will minimize the worst modeling error if it is only known that the poles of the system we are trying to model lie inside a circle of radius $R < 1$, and that the energy of its unit pulse response is not larger than a given bound. There exist also classes of systems for which the best model is a Laguerre model [3], or a more general generalized orthonormal basis function model [4, 5, 6].

Once the subspace has been selected, one has to face the second problem: among the infinite number of basis (when $n > 1$) of the subspace, select one with "good properties". For example, one may be interested in using a basis than makes possible a cascade implementation of the model (e.g., a transversal filter), instead of the parallel implementation depicted in Fig. 1. In this paper we will study another possibility: we will attempt to choose a basis for which the condition number of the correlation matrix of the internal signals $x_k(t)$ of the model is as small as possible, given that the power spectral density of the input signal belongs to a given class of power spectral densities. From a practical point of view this is highly desired, not only because the model will have good numerical properties [7], but also because adaptive algorithms of the LMS family will converge faster to the optimal weights (if and when they are used to adjust the model's weights) [8].

The rest of the paper is organized as follows: in section 2 we present an sub-optimal solution to the problem posed in the previous paragraph; in section 3 we illustrate the main results of the paper with an example; and in section 4 we present some final remarks.

2 The problem and a sub-optimal solution

We will only analyze the case in which the power spectral density of the input signal, denoted by $\Phi_u(e^{i\omega})$, is known to satisfy the condition

$$\underline{\Phi}_u(e^{i\omega}) \leqslant \Phi_u(e^{i\omega}) \leqslant \bar{\Phi}_u(e^{i\omega}), \qquad \forall\omega,$$

where $\underline{\Phi}_u(e^{i\omega})$ and $\bar{\Phi}_u(e^{i\omega})$ are given absolutely continuous power spectral densities.[1]

The elements of the correlation matrix, $\boldsymbol{R}$, of the internal signals of the model are the inner products between the signals $x_k(t)$, Using Parseval's theorem, these inner products can be computed easily in the frequency domain, yielding

$$\boldsymbol{R} = \frac{1}{2\pi}\int_{-\pi}^{+\pi} \boldsymbol{G}(e^{i\omega})\boldsymbol{G}^{H}(e^{i\omega})\,\Phi_u(e^{i\omega})\,d\omega.$$

It is well known that one has [9]

$$\lambda_{\max}(\boldsymbol{R}) = \max_{\boldsymbol{w}^H\boldsymbol{w}=1} \boldsymbol{w}^H\boldsymbol{R}\boldsymbol{w},$$

and that

$$\lambda_{\min}(\boldsymbol{R}) = \min_{\boldsymbol{w}^H\boldsymbol{w}=1} \boldsymbol{w}^H\boldsymbol{R}\boldsymbol{w}.$$

In our case, trivial algebraic manipulations yield

$$\boldsymbol{w}^H\boldsymbol{R}\boldsymbol{w} = \int_{-\infty}^{+\infty} \left|\sum_{i=1}^{n} w_i^* G_i(i\omega)\right|^2 \Phi_u(i\omega)\,d\omega.$$

This leads to the tight bounds

$$\boldsymbol{w}^H\underline{\boldsymbol{R}}\boldsymbol{w} \leqslant \boldsymbol{w}^H\boldsymbol{R}\boldsymbol{w} \leqslant \boldsymbol{w}^H\bar{\boldsymbol{R}}\boldsymbol{w}$$

(the matrices $\underline{\boldsymbol{R}}$ and $\bar{\boldsymbol{R}}$ are defined in the obvious way, replacing $\Phi_u(e^{i\omega})$ by $\underline{\Phi}_u(e^{i\omega})$ and $\bar{\Phi}_u(e^{i\omega})$, respectively). Elementary considerations allow us to conclude that

$$\lambda_{\min}(\underline{\boldsymbol{R}}) \leqslant \lambda_{\min}(\boldsymbol{R}) \leqslant \lambda_{\max}(\boldsymbol{R}) \leqslant \lambda_{\max}(\bar{\boldsymbol{R}}).$$

It follows that the square of the condition number of $\boldsymbol{R}$ satisfies the bound

$$\frac{\lambda_{\max}(\boldsymbol{R})}{\lambda_{\min}(\boldsymbol{R})} \leqslant \frac{\lambda_{\max}(\bar{\boldsymbol{R}})}{\lambda_{\min}(\underline{\boldsymbol{R}})}.$$

Instead of attempting to minimize the left-hand side of this inequality, which is difficult due to its dependence on the power spectral density, we will minimize its right-hand side, which is easy. Since this bound is not tight, our solution will be sub-optimal.

[1] In [7] we treated only the case $\bar{\Phi}_u(e^{i\omega}) = \alpha\underline{\Phi}_u(e^{i\omega})$, with $\alpha > 1$.

If the elements of the vector $\boldsymbol{H}(z)$ are a known basis of our subspace, then any basis is given by

$$\boldsymbol{G}(z) = \boldsymbol{T}\boldsymbol{H}(z),$$

where $\boldsymbol{T}$ is any non-singular $n \times n$ matrix. Let

$$\boldsymbol{S} = \frac{1}{2\pi}\int_{-\pi}^{+\pi} \boldsymbol{H}(e^{i\omega})\boldsymbol{H}^{H}(e^{i\omega})\,\Phi_u(e^{i\omega})\,d\omega$$

be the correlation matrix for the known basis, with $\underline{\boldsymbol{S}}$ and $\bar{\boldsymbol{S}}$ defined in a similar way. Clearly, one has

$$\boldsymbol{R} = \boldsymbol{T}\boldsymbol{S}\boldsymbol{T}^{H}.$$

Using this parameterization for $\boldsymbol{G}(z)$ the square of the condition number of $\boldsymbol{R}$ becomes a function of $\boldsymbol{T}$, viz.,

$$\kappa(\boldsymbol{T}) = \frac{\lambda_{\max}(\boldsymbol{T}\boldsymbol{S}\boldsymbol{T}^{H})}{\lambda_{\min}(\boldsymbol{T}\boldsymbol{S}\boldsymbol{T}^{H})},$$

which has an upper bound that is also a function of $\boldsymbol{T}$, viz.,

$$\bar{\kappa}(\boldsymbol{T}) = \frac{\lambda_{\max}(\boldsymbol{T}\bar{\boldsymbol{S}}\boldsymbol{T}^{H})}{\lambda_{\min}(\boldsymbol{T}\underline{\boldsymbol{S}}\boldsymbol{T}^{H})}.$$

Without loss of generality we may assume that $\underline{\boldsymbol{S}} = \boldsymbol{I}$ (the identity matrix); for this to happen the fixed basis must the orthonormal with respect to the power spectral density $\underline{\Phi}(e^{i\omega})$.[2] Using this assumption and replacing $\boldsymbol{T}$ by its singular value decomposition, $\boldsymbol{T} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{H}$, yields

$$\bar{\kappa}(\boldsymbol{\Sigma},\boldsymbol{V}) = \frac{\lambda_{\max}(\boldsymbol{\Sigma}\boldsymbol{V}^{H}\bar{\boldsymbol{S}}\boldsymbol{V}\boldsymbol{\Sigma})}{\lambda_{\min}(\Sigma^2)}.$$

(Since $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary matrices, $\bar{\kappa}(\boldsymbol{T})$ does not depend on $\boldsymbol{U}$.) Because $\bar{\kappa}(\boldsymbol{\Sigma},\boldsymbol{V})$ is homogeneous in $\boldsymbol{\Sigma}$, and because $\boldsymbol{T}$ must be non-singular, we will force the smallest singular value of $\boldsymbol{\Sigma}$ to be one. This yields

$$\bar{\kappa}(\boldsymbol{\Sigma},\boldsymbol{V}) = \max_{\boldsymbol{w}^H\boldsymbol{w}=1} \boldsymbol{w}^H\boldsymbol{\Sigma}\boldsymbol{V}^{H}\bar{\boldsymbol{S}}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{w}, \qquad \sigma_{\min} = 1.$$

Since $\boldsymbol{\Sigma} \geqslant \boldsymbol{I}$, it follows that

$$(\boldsymbol{\Sigma}\boldsymbol{w})^{H}(\boldsymbol{\Sigma}\boldsymbol{w}) \geqslant 1.$$

Thus,

$$\max_{\boldsymbol{w}^H\boldsymbol{w}=1} \boldsymbol{w}^H\boldsymbol{\Sigma}\boldsymbol{V}^{H}\bar{\boldsymbol{S}}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{w} \geqslant \max_{\boldsymbol{w}^H\boldsymbol{w}=1} \boldsymbol{w}^H\boldsymbol{V}^{H}\bar{\boldsymbol{S}}\boldsymbol{V}\boldsymbol{w}.$$

Since $\boldsymbol{V}$ is an unitary matrix we finally have

$$\bar{\kappa}(\boldsymbol{\Sigma},\boldsymbol{V}) \geqslant \lambda_{\max}(\bar{\boldsymbol{S}}).$$

When $\boldsymbol{\Sigma} = \boldsymbol{I}$ equality holds in the previous inequality irrespective of the value of $\boldsymbol{V}$.[3] Thus

$$\min_{\text{non-singular } \boldsymbol{T}} \bar{\kappa}(\boldsymbol{T}) = \lambda_{\max}(\bar{\boldsymbol{S}}).$$

[2] A Gram-Schmidt orthonormalization can be used to replace any basis by an orthonormal basis.

[3] With this choice of $\boldsymbol{\Sigma}$ we may use $\boldsymbol{T}$ to diagonalize $\bar{\boldsymbol{S}}$, keeping $\underline{\boldsymbol{S}} = \boldsymbol{I}$.

Our previous equation shows that we have found a solution to the sub-optimal problem we posed in this section. Many more solutions exist, with $\boldsymbol{\Sigma} \geqslant \boldsymbol{I}$. Unfortunately, we do not have space to describe them here. We will only mention that the extra degrees of freedom can be used, e.g., to make the condition numbers of $\underline{\boldsymbol{S}}$ and $\bar{\boldsymbol{S}}$ equal or to minimize the condition number for a nominal power spectral density.

To summarize, we start with an arbitrary basis $\boldsymbol{H}(z)$, replace it with an orthonormal basis with respect to $\underline{\Phi}(e^{\mathrm{i}\omega})$. The resulting basis with minimize $\bar{\kappa}(\boldsymbol{T})$. Optionally, we may compute $\bar{\boldsymbol{S}}$ and diagonalize it.[4] The value of $\lambda_{\max}(\bar{\boldsymbol{S}})$ will give an upper bound of the worst condition number we may encounter. The first orthonormalization may also be done with respect to $\bar{\Phi}(e^{\mathrm{i}\omega})$, in which case $1/\lambda_{\min}(\underline{\boldsymbol{S}})$ will give an upper bound of the worst condition number we may encounter.

2.1 The continuous-time case

Since the standard bilinear transformation

$$z = \frac{a+s}{a-s}, \qquad \frac{dz}{z} = \frac{\sqrt{2a}}{a+s}\frac{\sqrt{2a}}{a-s}\,ds$$

induces an isomorphism between the s and z planes [11], all our results can be converted to the continuous-time domain simply by transforming signals using the rule

$$U(z) \longleftrightarrow \frac{\sqrt{2a}}{s+a}\,U\Big(\frac{a+s}{a-s}\Big)$$

and systems and power spectral densities using the rule

$$F(z) \longleftrightarrow F\Big(\frac{a+s}{a-s}\Big).$$

3 An example

In this example we will show that the upper bound determined in the previous section can be reasonably tight, and to show that apparently reasonable bases can give rise to condition numbers much larger than the best possible value. In order to allow a comparison of the results presented here with those presented in [7], we will specify some initial bases of our subspace, and the upper and lower bounds of the power spectral density, using Laplace transforms. To perform all necessary computations, they were transformed into Z-transforms using the ideas of subsection 2.1 using $a = 1$ (any positive value of a will produce the same results).

On our example the bounds for the input power spectral density are

$$\underline{\Phi}(s) = \left|\frac{10}{s+10}\right|^2_{s=\mathrm{i}\omega} = \frac{100}{\omega^2+100}$$

and

$$\bar{\Phi}(s) = \left|\frac{30}{s+20}\right|^2_{s=\mathrm{i}\omega} = \frac{900}{\omega^2+400}.$$

4 The previous two steps can be done at the same time using a generalized eigenvalue decomposition [10].

We will analyze four different bases for our chosen subspace. The first three are as follows:

$$\text{B1:} \quad \begin{cases} G_1(s) = \frac{\sqrt{2}}{s+1} \\ G_2(s) = \frac{\sqrt{4}}{s+2} \\ G_3(s) = \frac{\sqrt{6}}{s+3} \\ G_4(s) = \frac{\sqrt{8}}{s+4} \end{cases}$$

$$\text{B2:} \quad \begin{cases} G_1(s) = \frac{\sqrt{2}}{s+1} \\ G_2(s) = \frac{\sqrt{12}}{(s+1)(s+2)} \\ G_3(s) = \frac{\sqrt{120}}{(s+1)(s+2)(s+3)} \\ G_4(s) = \frac{\sqrt{2016}}{(s+1)(s+2)(s+3)(s+4)} \end{cases}$$

and

$$\text{B3:} \quad \begin{cases} G_1(s) = \frac{\sqrt{2}}{s+1} \\ G_2(s) = \frac{\sqrt{4}(s-1)}{(s+1)(s+2)} \\ G_3(s) = \frac{\sqrt{6}(s-1)(s-2)}{(s+1)(s+2)(s+3)} \\ G_4(s) = \frac{\sqrt{8}(s-1)(s-2)(s-3)}{(s+1)(s+2)(s+3)(s+4)} \end{cases}.$$

It can be easily verified that all basis functions presented above have unit energy. The impulse responses $g_k(t)$ of the basis B3 are orthonormal in the standard sense [12, 13]. A fourth "optimal" basis, B4, was computed, following the recommendations of section 2, and a fifth "symmetric" basis, B5, was also computed.

In Table 1 we present the values of $\lambda_{\max}(\boldsymbol{R})/\lambda_{\min}(\boldsymbol{R})$, which is the square of the condition number of $\boldsymbol{R}$, for the different bases, and for three power spectral densities: $\underline{\Phi}$, $\bar{\Phi}$, and for the worst power spectral density of the form

$$\Phi(\mathrm{i}\omega) = \begin{cases} \underline{\Phi}(\mathrm{i}\omega), & \text{if } 0 < |\omega| < \omega_o \\ \bar{\Phi}(\mathrm{i}\omega), & \text{if } |\omega| \geqslant \omega_o \end{cases}$$

or of the form

$$\Phi(\mathrm{i}\omega) = \begin{cases} \bar{\Phi}(\mathrm{i}\omega), & \text{if } 0 < |\omega| < \omega_o \\ \underline{\Phi}(\mathrm{i}\omega), & \text{if } |\omega| \geqslant \omega_o. \end{cases}$$

Basis	$\underline{\Phi}$	$\bar{\Phi}$	worst
B1	35025	32232	43930
B2	5418	4115	10141
B3	2.296	1.574	4.906
B4	1.000	1.461	3.218
B5	1.209	1.209	2.669

Table 1: Square of the condition numbers of $\boldsymbol{R}$ for the bases described in the text.

Note the very bad performance of the bases B1 and B2, and the good performance of the basis B3, all when compared with the "optimal" bases B4 and B5. In particular, the basis B1, which is some times used in "parallel" versions of IIR adaptive filters [14], is very bad. Using a cascade of low-pass sections, as in the basis B2, is also a bad idea. Using a cascade of all-pass sections, tapped by low-pass sections,[5] as in the basis B3, produces in this case very good results.

The good performance of the basis B3 can be explained by the fact that this basis is orthonormal with respect to a flat (unit) power spectral density, and by the fact that $\Psi(\mathrm{i}\omega)$ is approximately flat in the pass-bands of $G_k(\mathrm{i}\omega)$.

Finally, the worst performance of the basis B4 (3.218) should be compared with the minimum value of $\bar{\kappa}(\boldsymbol{T})$, which in this case is 3.302. This easy to compute bound is therefore reasonably tight.

4 Final remarks

The previous example demonstrates that the basis functions of a linear-in-the-parameters model or adaptive filter has to be chosen with care. Some bases that are easy to implement are quite bad in terms of the conditioning of the correlation matrix of its internal signals. A filter composed by orthogonal transfer functions in the classical sense performs quite well when the power spectral density of the input signal is reasonably flat in the pass band of the filters. In particular, an FIR filter, being composed by orthonormal basis functions (in the classical sense), usually performs well when the power spectral density is reasonably flat in the entire frequency axis.[6]

References

[1] A. Pinkus, *n-Widths in Approximation Theory*, vol. 7 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (A Series of Modern Surveys in Mathematics)*, Springer-Verlag, Berlin / New York, 1985.

[2] P. M. Mäkilä and J. R. Partington, "Robust approximate modelling of stable linear systems", *International Journal of Control*, vol. 58, no. 3, pp. 665–683, Sept. 1993.

[3] B. Wahlberg and P. M. Mäkilä, "On approximation of stable linear dynamical systems using Laguerre and Kautz functions", *Automatica*, vol. 32, no. 5, pp. 693–708, May 1996.

[4] P. S. C. Heuberger, P. M. J. Van den Hof, and O. H. Bosgra, "A generalized orthonormal basis for linear dynamical systems", *IEEE Transactions on Automatic Control*, vol. 40, no. 3, pp. 451–465, Mar. 1995.

[5] T. Oliveira e Silva, "A n-width result for the generalized orthonormal basis function model", in *Preprints of the 13th World Congress of the International Federation of Automatic Control*, San Francisco, USA, 1996, vol. I, pp. 375–380.

[6] J. Bokor, P. Heuberger, B. Ninness, T. Oliveira e Silva, P. Van den Hof, and B. Wahlberg, "Modelling and identification with orthogonal basis functions", Workshop notes of a pre-conference tutorial presented at the 14th IFAC World Congress, Beijing, China., July 1999.

[7] T. Oliveira e Silva, "On the design of linear-in-the-parameters adaptive filters", in *5th International Conference on Electronics, Circuits and Systems*, Lisbon, Portugal, 1998, vol. 1, pp. 385–388.

[8] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall Information and System Sciences Series. Prentice-Hall, Inc., Englewood Cliffs, N.J. 07632, second edition, 1991.

[9] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[10] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore and London, second edition, 1989.

[11] K. Steiglitz, "The equivalence of digital and analog signal processing", *Information and Control*, vol. 8, no. 5, pp. 455–467, Oct. 1965.

[12] W. H. Kautz, "Transient synthesis in the time domain", *IRE Transactions on Circuit Theory*, vol. CT-1, no. 3, pp. 29–39, Sept. 1954.

[13] Y. W. Lee, *Statistical Theory of Communication*, John Wiley & Sons, New York, 1960.

[14] J. J. Shynk, "Adaptive IIR filtering", *IEEE ASSP Magazine*, vol. 6, no. 2, pp. 4–21, Apr. 1989.

[15] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices", *IEEE Transactions on Information Theory*, vol. IT-18, no. 6, pp. 725–730, Nov. 1972.

[5] This can be implemented efficiently using balanced state space realizations of all-pass filters.

[6] From well known asymptotic results concerning the eigenvalues of Toeplitz matrices [15] we have, for FIR filters,

$$\frac{\lambda_{\max}(\boldsymbol{R})}{\lambda_{\min}(\boldsymbol{R})} \leqslant \frac{\max_\omega \Phi(e^{\mathrm{i}\omega})}{\min_\omega \Phi(e^{\mathrm{i}\omega})}.$$

This result is also valid for filters based on orthonormal transfer functions, such as Laguerre and Kautz filters.