GENERALIZED LINEAR-IN-PARAMETER MODELS

Tuomas Paatero
Helsinki University of Technology
Laboratory of Acoustics and Audio Signal Processing
P. O. Box 3000, 02015 Espoo
tuomas.paatero@hut.fi

Abstract

This paper presents the generalized linear-in-parameter model concept as a means to adduce the somewhat overlooked origins of familiar parameter estimation methods, to deduce alternative signal and system representations, and as an introduction to rational orthonormal filter structures.

I. Introduction

In digital signal processing (DSP), a FIR (finite impulse response) or moving average (MA) model is often preferred over other models because of its unconditional stability and efficient parameter estimation methods. This praxis-driven choice between MA and ARMA (autoregressive MA) models may overlook the possibility of using other well-behaving linear-in-parameter models than MA. Even a belief that the non-recursiveness is the key to the good qualities is possible. However, there exists an infinite variety of recursive linear-in-parameter model structures with similar principal properties to the MA model and with almost all the same parameter estimation methods.

By defining a generalized linear-in-parameter model (GLM) based on linearly independent modelling signals it is possible to deduce a more general "signal theory" for the development of alternative signal representations and model structures. GLM provides a direct view into the approximation problem and a natural introduction to general orthonormal structures, but it also suggests the possibility of using non-orthogonal models. The GLM construction is made in a Hilbert space although all the needed formulas and definitions could be made explicitly. Besides being elegant, the Hilbert space framework makes it possible to present theory and methods without the need for distinguishing between stochastic and deterministic signals, or between time and transfer domain representations.

II. Generalized linear model structure

An efficient function space description is based on the fact that causal and stable (CS) signals and CS linear time-invariant (LTI) systems belong to Hilbert spaces of causal finite-energy signals and LTI systems. In the discrete time-domain, $\ell^2(\mathbb{N})$ is the space of causal and finite-energy complex signals, or stationary zero-mean and finite variance complex signals, and the corresponding transfer-domain space is denoted by $H^2(\mathbb{E})$. The following definition is based on the Hilbert space projection theorem.

A. *Definition of GLM*

A model of the form $\hat{H}(z) = \sum_{i=0}^{N} w_i G_i(z)$ is a generalized linear model (GLM) for the system $y = H[x]$, $x, y \in \ell^2(\mathbb{N})$, $H \in H^2(\mathbb{E})$, if the modeling signals $x_i = G_i[x]$, $i = 0, \ldots, N$, are linearly independent $\ell^2(\mathbb{N})$ signals. In the least-square (LS) or minimum mean-square-error (MMSE) sense, the optimal model parameters are obtained from the normal equations

$$\sum_{i=0}^{N} w_i (x_i, x_j) = (y, x_j), \; 0 \le j \le N, \quad (1)$$

where $(\cdot,\cdot)$ denotes the inner product. The matrix form of equation (1) is $\mathbf{Rw} = \mathbf{p}$.

B. *Properties of the correlation matrix*

The solution to the equations (1) is $\mathbf{w} = \mathbf{R}^{-1}\mathbf{p}$, but as in the FIR case, indirect methods are preferred to avoid matrix inversion. Many of those methods rely on the fact that the correlation matrix $\mathbf{R}$ is always hermitian and positive semi-definite. In fact, only methods relying on Toepliz form or (impulse-input) orthogonality of the correlation matrix are excluded in general. On the other hand, it is interesting to see into which form the GLM structure is forced with these assumptions. For a transversal structure, the Toepliz or block-Toepliz form of $\mathbf{R}$ corresponds to identical all-pass (AP) sections enabling the use of Levinson-Durbin-algorithm and a connection to lattice structures. The orthogonality assumption forces the GLM into a transversal AP structure with output tap "normalization" related purely to the subsequent AP section. It can be proven that orthogonal GLM structures have exactly the same input dependent upper and lower bounds for the eigenvalues of $\mathbf{R}$ than in the FIR case. These bounds are independent of the model order, and for a rational GLM, of the choice of poles.

C. *Existence of concrete GLM structures*

If the model is presumed to be rational, then the GLM assumption can be seen as a restriction on the rational fraction decomposition defining the model structure. For a rational GLM, the linear independency is guaranteed e.g. for transversal structures of (non-constant) CSLTI rational AP blocks with possible output tap weighting filters of the same order.

III. Rational orthonormal GLM structures

In the signal processing context, rational orthonormal model (ROM) structures, based on orthonormalization of continious-time complex exponentials, were first introduced by Kautz, Huggins and Young [5, 14]. Much earlier, Lee and Wiener proposed synthesis networks based on some orthonormal structures but in a more restricted form [8]. The discrete-time version can be attributed to Broome [2]. Apparently not known to the aforementioned, rational orthonormal function expansions were deduced in the 1920's to prove equivalency between rational approximations and interpolations, and the least-square problem [13]. More recent publications, few and scattered, can be found in the fields of system identification [11, 4, 9] and FIR-to-IIR filter conversion [3, 1].

A. *Kautz functions*

Deducable in many ways, the lowest order rational orthonormal functions are of the form [13, 9]

$$G_i(z) = \alpha_i \frac{\sqrt{1 - z_i z_i^*}}{z^{-1} - z_i^*} \prod_{j=0}^{i} \frac{z^{-1} - z_j^*}{1 - z_j z^{-1}}, \quad (2)$$

$i = 0, 1, \ldots,$ defined by any sequence of points $(z_i)_{i=0}^{\infty}$ in the unit disk and with a free rotation parameter α_i of modulus one. Functions (2) form an orthonormal set in $H^2(\mathbb{E})$, and this orthonormal set is complete, or a base, with a moderate restriction on the poles [13]. The corresponding time functions $g_i(n) \in \ell^2(\mathbb{N})$, $i \in \mathbb{N}$, impulse responses or inverse z-transforms of (2), have the same properties in $\ell^2(\mathbb{N})$. With functions (2), the parallel GLM structure degenerates into a trasversal Kautz filter (Fig. 1).

A basis representation of a signal $h(n) \in \ell^2(\mathbb{N})$ or $H(z) \in H^2(\mathbb{E})$ is obtained as a Fourier series expansion of $h(n)$ or $H(z)$ with respect to (2) and the Fourier partial sums correspond to LS or MMSE sense optimal approximations $\hat{h}(n)$ or $\hat{H}(z)$. Evaluation of the Fourier coefficients, $c_i = (h, g_i) = (H, G_i)$, can be implemented by feeding the signal $h(-n)$ to the Kautz filter and reading the tap outputs $x_i(n) = G_i[h(-n)]$ at $n = 0$: $c_i = x_i(0)$. Input-output-data identification with the Kautz model is based on (1): the evaluation (or approximation) of the inner products and the solving of the normal equations, or

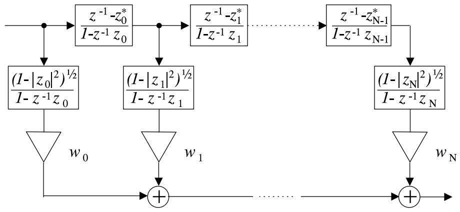

Figure 1: *The Kautz filter. For $z_i = 0$ in (2) it degenerates to a FIR filter and for $z_i = a, -1 < a < 1$,, it is a Laguerre filter, where the tap filters are replaced by a prefilter. For higher order identical sections, Laguerre form is restored with vectorized filter sections, internal signals and model parameters. The general real Kautz filter has a transversal structure of 2nd order blocks with dual tap outputs [2].*

combined techniques. Signal representation and approximation are special cases of identification: $\mathbf{R} = \mathbf{I}$ in (1) corresponding to impulse or (appropriate) white noise input.

A Kautz filter produces real modelling signals only in the case of real poles. From a sequence of real or complex conjugate poles it is possible to form real orthonormal functions with a slightly modified structure. Some of the deductions are made directly on the real function assumption [2]. Moreover, the state-space approach to structures with identical blocks is based on balanced realizations of real rational AP functions [4].

B. Methods for the choosing of poles

Even in the one-pole Laguerre case, it is impossible to optimize the pole position analytically. A relation between optimal model parameters and error energy surface stationary points with respect to the pole positions may be used to select a global minima [9]. A classification of systems may be utilized to associate systems and basis functions [12].

Orthonormal expansions, obtained directly in the ROM approximation situation or implicitly by orthogonalizing the GLM modeling signals, distribute the signal energy orthogonally in the parameters. This means that the orthonormal expansion coefficients are independent of approximation or model order selection and that signal, model and modelling error energies are related in the simple form $(x, x) = \sum_{i=0}^{N} |c_i|^2 + (e, e)$, which can be utilized in pole location and model order optimization. Many of the parameter estimation methods can be seen as orthogonalization operations on the modelling signals or on the correlation matrix.

The concept of a complementary signal [14] may be used to directly optimize the AP structure for a given finite impulse response $h(n)$: the output $a(n)$ of the AP chain with input $h(-n)$ is a decomposition of the signal energy where the approximation error energy is a finite sum of the form $E = \sum_{n=-m}^{0} a(n)$. A network structure for parallel calculation of the partial derivatives of E with respect to the poles and a gradient search may be used to optimize the AP structure [3]. The same principle has been implicitly utilized in a proposed recursive AP structure generation [1] with better guarantee for optimality.

The contrast between the well-defined fixed model parameter estimation task and the complicated and non-optimal model structure optimization makes it tempting to use sophisticated guesses and iterative search in model selection. In some situations, the tuning of a set of separate resonances and the corresponding time-constants could be an appropriate approach. In general, a priori knowledge of the system and indirect means, such as AR or ARMA modelling, can be incorporated. For a chosen set of poles there is still the choice of pole ordering to be considered.

The choosing of poles is intimately related to model order selection. A practical way to restrict the stuctures is to use identical AP sections because then the pole and model order selection problems are essentially separated. Identical AP sections also provide a "change of variable" interpretation and a related transformation generalization: if $A(z)$ is the (appropriate) mth order AP section generating the basis, then the system $y(n) = H[x(n)]$ can be transformed to the Hambo domain, $\tilde{\mathbf{y}}(w) = \tilde{\mathbf{H}}(w)\tilde{\mathbf{x}}(z)$, where $\tilde{\mathbf{x}}, \tilde{\mathbf{y}} \in H_2^m(\mathbb{E})$ and $\tilde{\mathbf{H}} \in H_2^{m\times m}(\mathbb{E})$ are the Hambo transforms of the signals and the system [4]. $\tilde{\mathbf{H}}(w)$ is obtained by a change of variable, $z^{-1} = \mathbf{N}(w)$, in $H(z)$ where $\mathbf{N} \in H_2^{m\times m}$ is the state-space realization "inverse" of $A(z)$. Hambo transform is a generalization of Fourier and Laguerre transforms and it can be used to break up the identification problem into m Laguerre-type problems which makes the basis optimization easier.

IV. Proposed and potential applications

A. Filter design

A rational GLM structure is a fixed pole IIR filter, unconditionally stable, with almost all the same parameter estimation methods as the FIR filter. Especially direct AP structure optimization schemes are appealing. On the other hand, manual tuning of pole positions and time-constants could be appropriate e.g. when intelligent and non-technical criteria are desired in modelling. An attempt on musical instrument body modellin will be done on this basis.

B. Adaptive filtering

Methods like least-mean-square (LMS) and recursive least-square (RLS) algorithms, and their variants, are well-suited for GLM because of their "lightness", even in LTI modelling. To benefit from complicating things with the GLM, one has to incorporate some invariant features of the system to the basis, or use an adaption mechanism for the poles. For example, in modelling a shifting strong dominant resonance, a GLM defined by a complex conjugate pole pair and tuned with a single-pole model could be appropriate.

Simulations on the LMS algorithm reveal that GLM is particularly competitive in situations where FIR filters require a huge filter order and IIR filters have stability problems. As before, the GLM with identical blocks has its counterparts in adaptive lattice algorithms. Convergence considerations for the FIR model are also valid for orthonormal GLM structures.

C. System identification

There is a long tradition in the Laguerre domain SI based on Lagurre signal transformation. Recently, there has been some interest in other GLM structures as well [11, 4, 9]. A structure with identical blocks is often preferred to enable efficient state-space representations and easy model reduction. Methods for basis function identification and the question of windowing will be considered in this context later by the author.

D. Linear prediction

Warped linear prediction (WLP) is a GLM application, deduced somewhat independently of the ROM tradition, where delay elements in the LP analysis and synthesis filters are replaced with identical first order AP filters [10, 6]. The pre-filter implied by the Laguerre filter is often omitted to accomplish desired residue signal spectral shaping. WLP has been well studied but, could more complex structures be of use? One attempt in this direction has been made [7]. A more thorough view to windowing and orthogonality could also benefit WLP.

References

[1] Brandenstein, H. and Unbehauen, R. 1998. Least-Squares Approximation of FIR by IIR Digital Filters. *IEEE Transactions on Signal Processing.* Vol. 46. No. 1. 21–30.

[2] Broome, P. W. 1965. Discrete Orthonormal Sequences. *Journal of the Association for Computing Mashinery.* Vol. 12. No. 2. 151–168.

[3] Friedman, D. H. 1981. On Approximating an FIR Filter Using Discrete

Orthonormal Exponentials. *IEEE Transactions on Acoustics, Speech and Signal Processing.* Vol. ASSP-29. No. 4. 923–831.

[4] Heuberger, P. S. C., Van den Hof, P. M. J. and Bosgra, O. H. 1995. A Generalized Orthonormal Basis for Linear Dynamical Systems. *IEEE Translations on Automatic Control.* Vol 40. No 3. 451–465.

[5] Huggins, W. H. 1956. Signal Theory. *IRE Transactions on Circuit Theory.* Vol. CT–3. 210–216.

[6] Laine, U. K., Karjalainen, M. and Altosaar, T. 1994. Warped Linear Prediction in Speech and Audio Processing. *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Adelaide, Australia, 1994. III–349–352.

[7] Laine, U. K. 1995. Generalized Linear Prediction Based on Analytic Signals. *Proc. ICASSP-95*, Detroit, MI. 1701–1704.

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

[9] Oliveira e Silva, T. 1995. Rational Ortonormal Functions on the Unit Circle and the Imaginary Axis, with Applications in System Identification. **URL:***ftp://inesca.inesca.pt/pub/tos /English/rof.ps.gz*

[10] Strube, H. W. 1980. Linear Prediction on a Warped Frequency Scale. *J. Acoust. Soc. Am. Vol. 68. No 4. 1071–1076.*

[11] Wahlberg, B. 1994. System Identification Using Kautz Models. *IEEE Translations on Automatic Control.* Vol. 39. No. 6. 1276–1282.

[12] Wahlberg, B. and Mäkilä, P. 1996. On Approximation of Stable Linear Dynamical Systems Using Laguerre and Kautz Functions. *Automatica.* Vol. 32. No. 5. 693–708.

[13] Walsh, J. L. 1969. *Interpolation and Approximation by Rational Functions in the Complex Domain.* American Mathematical Society, Providence, Rhode Island. First ed. 1935. 405 s.

[14] Young, T. Y. and Huggins, W. H. 1962. 'Complementary' Signals and Orthogonalized Exponentials. *IRE Transactions on Circuit Theory.* Vol. CT–9. 362–370.