Katholieke Universiteit Leuven

Departement Elektrotechniek ESAT-SISTA/TR 07-61

Pole optimization in discrete-time Kautz and Laguerre filters[1]

Toon van Waterschoot[2] [3], Moritz Diehl[2], and Marc Moonen[2]

May 2007

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.be* in the directory *pub/sista/vanwaterschoot/reports/07-61.pdf*

[2]K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group SCD(SISTA), Kasteelpark Arenberg 10, B-3001 Leuven, Belgium, Tel. +32 16 321927, Fax +32 16 321970, WWW: *http://www.esat.kuleuven.be/sista-cosic-docarch*. E-mail: *toon.vanwaterschoot@esat.kuleuven.be*.

[3]Toon van Waterschoot is a Research Assistant with the I.W.T. (Flemish Institute for Scientific and Technological Research in Industry). This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of K.U.Leuven Research Council: CoE EF/05/006 Optimization in Engineering (OPTEC), the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office IUAP P6/04 (DYSCO, "Dynamical systems, control, and optimization", 2007-2011), and the Concerted Research Action GOA-AMBioRICS. The scientific responsibility is assumed by its authors.

Pole Optimization in Discrete-Time Kautz and Laguerre Filters

Project report Numerical Optimization course Fall 2006*

Toon van Waterschoot†

May 27, 2007

1 Introduction

Discrete-time Kautz and Laguerre filters are gaining interest in system identification, parameter estimation, and signal processing, as alternatives to the traditional finite impulse response (FIR) and infinite impulse response (IIR) filters. The strength of the Kautz and Laguerre filters lies in the fact that they can model systems having a long impulse response with much fewer parameters than the traditional (FIR) filters, while having only one pole (in the Laguerre case) or one complex conjugate pole pair (in the Kautz case).

Consider the following input-output relation of a linear time-invariant (LTI) system,

$$y(t) = G(q)u(t) + e(t). \tag{1}$$

where $t \in \mathbb{Z}$ denotes a shorthand notation for the discrete time variable tT_s after sampling at sampling frequency $f_s = 1/T_s$, and q denotes the time shift operator, i.e., $q^{-k}u(t) = u(t-k)$. The input signal is assumed to be deterministic, while the noise term and hence the output signal are stochastic quantities. The LTI system, which is assumed causal and stable, can be characterized by its impulse response, which is possibly of infinite duration,

$$G(q) = g_0 + g_1 q^{-1} + g_2 q^{-2} + \ldots = \sum_{i=0}^{\infty} g_i q^{-i}. \tag{2}$$

*This report was also submitted as an internal report ESAT-SISTA TR 07-61, and is accompanied by a MATLAB software package.

†Katholieke Universiteit Leuven, ESAT-SCD, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium, Tel. +32 16 321927, Fax +32 16 321970, E-mail toon.vanwaterschoot@esat.kuleuven.be

If $G(q)$ is unknown, the system identification problem related to (1) consists of two challenges: finding a suitable model and estimating the model parameters. The two most widely used models for representing LTI systems are the FIR or all-zero model,

$$B(q) = b_0 + b_1 q^{-1} + \ldots + b_{L_B} q^{-L_B} \tag{3}$$

and the IIR or pole-zero model

$$\frac{B(q)}{A(q)} = \frac{b_0 + b_1 q^{-1} + \ldots + b_{L_B} q^{-L_B}}{1 + a_1 q^{-1} + \ldots + a_{L_A} q^{-L_A}}. \tag{4}$$

The FIR model is by far the most popular, certainly in signal processing applications, because when combined with a least-squares (LS) criterion to estimate the model parameters, it yields a quadratic and hence convex objective function. Its main disadvantage is that a large number of model parameters are needed when modelling systems with a slowly decaying impulse response, hence increasing the computational complexity and deteriorating the conditioning of the related parameter estimation problem. The IIR model, on the other hand, is better suited to model systems with a long impulse response in an efficient way. However, when estimating the IIR model parameters, one is usually confronted with a non-convex objective function and, moreover, the estimated system model is not always guaranteed to be stable.

A compromise approach, which combines the advantages of the FIR and IIR models, is based on the theory of rational basis functions (RBFs). If the time shift operators q^{-i} in the FIR model are replaced with discrete-time filters $\Psi_i(q, \boldsymbol{\eta})$, with $\{\Psi_i(q, \boldsymbol{\eta})\}_{i=0}^{n_B}$ a set of RBFs having a common free parameter vector $\boldsymbol{\eta}$, we obtain the so-called RBF model for the system $G(q)$,

$$B(q, \boldsymbol{\eta}) = b_0 \Psi_0(q, \boldsymbol{\eta}) + b_1 \Psi_1(q, \boldsymbol{\eta}) + \ldots + b_{n_B} \Psi_{n_B}(q, \boldsymbol{\eta}). \tag{5}$$

If the RBFs and the free parameter vector $\boldsymbol{\eta}$ are appropriately chosen, then the RBF model is capable of modelling the system equally well as the FIR model, but with much fewer parameters, i.e., $n_B << L_B$.

Two types of RBFs that are well-known for their good modelling capabilities are the Kautz and Laguerre RBFs, which are moreover orthonormal and form a complete set in the Hilbert space l^2. The discrete-time Kautz and Laguerre basis functions were originally formulated in [1]. The Kautz basis functions are parametrized by a set of complex conjugate pole pairs $\{\zeta_i, \bar{\zeta}_i\}$, which are often chosen to coincide in one pole pair $\{\zeta, \bar{\zeta}\}$, corresponding to the assumption that the system to be modelled has a dominating resonant mode [2]. The discrete-time Kautz basis functions are then defined as follows:

$$\Psi_{2i}(q, \boldsymbol{\eta}) = |1 + \zeta| \sqrt{\frac{1 - \zeta\bar{\zeta}}{2}} \frac{(z^{-1} - 1)}{(1 - \zeta z^{-1})(1 - \bar{\zeta} z^{-1})} \prod_{l=0}^{i-1} \frac{(z^{-1} - \zeta)(z^{-1} - \bar{\zeta})}{(1 - \zeta z^{-1})(1 - \bar{\zeta} z^{-1})}$$

$$\Psi_{2i+1}(q, \boldsymbol{\eta}) = |1 - \zeta| \sqrt{\frac{1 - \zeta\bar{\zeta}}{2}} \frac{(z^{-1} + 1)}{(1 - \zeta z^{-1})(1 - \bar{\zeta} z^{-1})} \prod_{l=0}^{i-1} \frac{(z^{-1} - \zeta)(z^{-1} - \bar{\zeta})}{(1 - \zeta z^{-1})(1 - \bar{\zeta} z^{-1})}$$

with $\zeta \triangleq \alpha + j\beta$ and $\boldsymbol{\eta} \triangleq [\alpha, \beta]^T$.

The discrete-time Laguerre basis functions [1] are parametrized by a set of real poles $\{\alpha_i\}$, which are usually chosen to coincide in one single real pole α, hence assuming that the system has a dominating slow mode [2]. The Laguerre basis functions are defined as

$$\Psi_i(q, \boldsymbol{\eta}) = \frac{\sqrt{1-\alpha^2}}{1-\alpha z^{-1}} \prod_{l=0}^{i-1} \frac{z^{-1}-\alpha}{1-\alpha z^{-1}} \tag{6}$$

where $\boldsymbol{\eta} \triangleq [\alpha]$.

2 The objective function

We will now derive the objective function for estimating the RBF pole parameters in $\boldsymbol{\eta}$ when a system identification problem is tackled using the RBF model in (5). The derivation is largely based on [2], so we will only give the main results here. We start by formulating an LS criterion to jointly estimate the RBF model parameter vector $\mathbf{b} \triangleq [b_0, b_1, \ldots, b_{n_B}]$ and the RBF pole parameter vector $\boldsymbol{\eta}$,

$$f(\mathbf{b}, \boldsymbol{\eta}) = \sum_{t=1}^{N} \varepsilon^2(t, \mathbf{b}, \boldsymbol{\eta}) \tag{7}$$

$$= \sum_{t=1}^{N} \big(y(t) - \boldsymbol{\phi}^T(t, \boldsymbol{\eta})\mathbf{b}\big)^2 \tag{8}$$

where the last equation is obtained by substituting the RBF model $B(q, \boldsymbol{\eta})$ for $G(q)$ in (1), with the regression vector defined as

$$\boldsymbol{\phi}(t, \boldsymbol{\eta}) \triangleq \big[\Psi_0(q, \boldsymbol{\eta})u(t) \quad \ldots \quad \Psi_{n_B}(q, \boldsymbol{\eta})u(t)\big]^T. \tag{9}$$

The LS criterion (7) can be written in matrix notation as follows:

$$f(\mathbf{b}, \boldsymbol{\eta}) = \boldsymbol{\varepsilon}^T(\mathbf{b}, \boldsymbol{\eta})\boldsymbol{\varepsilon}(\mathbf{b}, \boldsymbol{\eta}) \tag{10}$$

$$= \big(\mathbf{y} - \boldsymbol{\Phi}(\boldsymbol{\eta})\mathbf{b}\big)^T\big(\mathbf{y} - \boldsymbol{\Phi}(\boldsymbol{\eta})\mathbf{b}\big) \tag{11}$$

with

$$\boldsymbol{\varepsilon}(\mathbf{b}, \boldsymbol{\eta}) \triangleq \big[\varepsilon(1, \mathbf{b}, \boldsymbol{\eta}) \quad \ldots \quad \varepsilon(N, \mathbf{b}, \boldsymbol{\eta})\big]^T \tag{12}$$

$$\mathbf{y} \triangleq \big[y(1) \quad \ldots \quad y(N)\big]^T \tag{13}$$

$$\boldsymbol{\Phi}(\boldsymbol{\eta}) \triangleq \big[\boldsymbol{\phi}(1, \boldsymbol{\eta}) \quad \ldots \quad \boldsymbol{\phi}(N, \boldsymbol{\eta})\big]^T. \tag{14}$$

The LS criterion (11) is nonlinear in $\boldsymbol{\eta}$, but linear in $\mathbf{b}$, and can be shown to be a separable nonlinear LS criterion, see [2], [3]. This means that by substituting the linear LS estimate for $\mathbf{b}$ (which depends on $\boldsymbol{\eta}$),

$$\hat{\mathbf{b}}(\boldsymbol{\eta}) = \big(\boldsymbol{\Phi}^T(\boldsymbol{\eta})\boldsymbol{\Phi}(\boldsymbol{\eta})\big)^{-1}\boldsymbol{\Phi}^T(\boldsymbol{\eta})\mathbf{y} \tag{15}$$

$$= \boldsymbol{\Phi}^\dagger(\boldsymbol{\eta})\mathbf{y} \tag{16}$$

in (11), we obtain a nonlinear criterion which only depends on $\boldsymbol{\eta}$,

$$\boxed{\tilde{f}(\boldsymbol{\eta}) = \mathbf{y}^T \big(\mathbf{I} - \boldsymbol{\Phi}(\boldsymbol{\eta})\boldsymbol{\Phi}^\dagger(\boldsymbol{\eta})\big)^T \big(\mathbf{I} - \boldsymbol{\Phi}(\boldsymbol{\eta})\boldsymbol{\Phi}^\dagger(\boldsymbol{\eta})\big)\mathbf{y}} \tag{17}$$

and which is of much lower dimension than the original criterion in (11). Both criteria can be shown to be equivalent [2], [3], i.e.,

$$\arg\min_{\mathbf{b},\boldsymbol{\eta}} f(\mathbf{b}, \boldsymbol{\eta}) = \begin{bmatrix} \hat{\mathbf{b}}(\boldsymbol{\eta}) \\ \arg\min_{\boldsymbol{\eta}} \tilde{f}(\boldsymbol{\eta}) \end{bmatrix}. \tag{18}$$

The remainder of this report is devoted to minimizing the objective function in (17) w.r.t. the RBF pole parameters in $\boldsymbol{\eta}$.

3 The optimization algorithm

The RBF pole parameter vector $\boldsymbol{\eta}$ (which has dimension 2×1 in the Kautz case and 1×1 in the Laguerre case) can be estimated iteratively as follows (with k denoting the iteration number):

$$\boxed{\hat{\boldsymbol{\eta}}^{(k+1)} = \hat{\boldsymbol{\eta}}^{(k)} + t_k \mathbf{p}^{(k)}.} \tag{19}$$

The search direction $\mathbf{p}^{(k)}$ is calculated using either of the following methods:

1. steepest descent (SD):

$$\mathbf{p}^{(k)} = -\nabla \tilde{f}(\hat{\boldsymbol{\eta}}^{(k)}) \tag{20}$$

2. Gauss-Newton (GN):

$$\mathbf{p}^{(k)} = -\Big(\big(\nabla \tilde{\boldsymbol{\varepsilon}}(\boldsymbol{\eta})\big)^T \nabla \tilde{\boldsymbol{\varepsilon}}(\boldsymbol{\eta})\Big)^{-1} \big(\nabla \tilde{\boldsymbol{\varepsilon}}(\boldsymbol{\eta})\big)^T \tilde{\boldsymbol{\varepsilon}}(\boldsymbol{\eta}) \tag{21}$$

with $\tilde{\boldsymbol{\varepsilon}}(\boldsymbol{\eta}) \triangleq \big(\mathbf{I} - \boldsymbol{\Phi}(\boldsymbol{\eta})\boldsymbol{\Phi}^\dagger(\boldsymbol{\eta})\big)\mathbf{y}$

3. Quasi-Newton with BFGS update (BFGS):

$$\begin{cases} \mathbf{p}^{(k)} &= -\mathbf{B}_k^{-1} \nabla \tilde{f}(\hat{\boldsymbol{\eta}}^{(k)}) \\ \mathbf{B}_{k+1} &= \mathbf{B}_k - \frac{\mathbf{B}_k \mathbf{s}_k \mathbf{s}_k^T \mathbf{B}_k}{\mathbf{s}_k^T \mathbf{B}_k \mathbf{s}_k} + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} \end{cases} \tag{22}$$

with $\mathbf{y}_k \triangleq \nabla \tilde{f}(\boldsymbol{\eta}^{(k+1)}) - \nabla \tilde{f}(\boldsymbol{\eta}^{(k)})$ and $\mathbf{s}_k \triangleq \boldsymbol{\eta}^{(k+1)} - \boldsymbol{\eta}^{(k)}$, and using Powell's trick, with $\gamma_3 = 10^{-2}$,

$$\mathbf{y}_k^T \mathbf{s}_k < \gamma_3 \mathbf{s}_k^T \mathbf{B}_k \mathbf{s}_k \Rightarrow \text{replace } \mathbf{y}_k \text{in (22) with } \tilde{\mathbf{y}}_k = \lambda \mathbf{y}_k + (1-\lambda)\mathbf{B}_k \mathbf{s}_k \tag{23}$$

with

$$\lambda = (1 - \gamma_3) \frac{\mathbf{s}_k^T \mathbf{B}_k \mathbf{s}_k}{\mathbf{s}_k^T \mathbf{B}_k \mathbf{s}_k - \mathbf{y}_k^T \mathbf{s}_k}. \tag{24}$$

The step length t_k is calculated using backtracking, combined with a condition which should ensure both a sufficient decrease and a stable pole estimate [2]. From the Kautz and Laguerre filter expressions in Section 1, it can be seen that the stability condition is met by ensuring that the RBF poles are inside the unit circle in the complex plane, i.e.,

$$\|\hat{\boldsymbol{\eta}}^{(k)} + t_k \mathbf{p}^{(k)}\|_2 < 1. \tag{25}$$

We should stress that this approach is not ideal, since the algorithm can get stuck near the stability boundary. A better approach would be to take into account the stability condition when calculating the search direction $\mathbf{p}^{(k)}$, by relating the inequality constraint in (25) to the objective function, and performing a constrained optimization.

In [2], the sufficient decrease condition is expressed as

$$\tilde{f}(\hat{\boldsymbol{\eta}}^{(k)} + t_k \mathbf{p}^{(k)}) < \tilde{f}(\hat{\boldsymbol{\eta}}^{(k)}) \tag{26}$$

which is, however, not strong enough to ensure convergence to a local minimizer [4, Ch. 3]. We will use Armijo's sufficient decrease condition instead. Moreover, we add another two conditions, one to stop the iterative algorithm when the backtracking procedure reduces the step length below some minimal value, e.g., $t_{\min} = 10^{-3}$ (to avoid an infinite loop when the algorithm gets stuck near the stability boundary), and another one to continue the backtracking procedure when $\tilde{f}(\hat{\boldsymbol{\eta}}^{(k)} + t_k \mathbf{p}^{(k)}) = \text{NaN}$ (which may occur close to the stability boundary), regardless of the outcome of Armijo's condition. The entire backtracking procedure can be summarized as follows, starting at $t_k = t_{\max} = 1$:

$$\left(\begin{array}{c} \Big(\tilde{f}(\hat{\boldsymbol{\eta}}^{(k)} + t_k \mathbf{p}^{(k)}) > \tilde{f}(\hat{\boldsymbol{\eta}}^{(k)}) + \gamma t_k \big(\nabla \tilde{f}(\hat{\boldsymbol{\eta}}^{(k)})\big)^T \mathbf{p}^{(k)}\Big) \\ \vee \Big(\|\hat{\boldsymbol{\eta}}^{(k)} + t_k \mathbf{p}^{(k)}\|_2 \geq 1\Big) \\ \vee \Big(\tilde{f}(\hat{\boldsymbol{\eta}}^{(k)} + t_k \mathbf{p}^{(k)}) = \text{NaN}\Big) \end{array}\right) \wedge (t_k \geq t_{\min}) \Rightarrow t_k \leftarrow \beta t_k \tag{27}$$

with $\gamma = 0.1$ and $\beta = 0.9$.

Finally, the optimization algorithm stops when

$$\Big(\big|\big(\nabla \tilde{f}(\hat{\boldsymbol{\eta}}^{(k)})\big)^T \mathbf{p}^{(k)}\big| \leq \text{tol}\Big) \vee (k > k_{\max}) \vee (t_k < t_{\min}) \tag{28}$$

with $\text{tol} = 10^{-6}$ and $k_{\max} = 50$.

The gradient vector $\nabla \tilde{f}(\hat{\boldsymbol{\eta}}^{(k)})$ and the GN Hessian approximation $\big(\nabla \tilde{\boldsymbol{\varepsilon}}(\boldsymbol{\eta})\big)^T \nabla \tilde{\boldsymbol{\varepsilon}}(\boldsymbol{\eta})$ can be calculated as outlined in [2].

4 Computer simulations

4.1 Example 1: Kautz filters

We examine the effect of calculating the search direction $\mathbf{p}^{(k)}$ with three different methods (SG, GN, and BFGS), by considering the Kautz filter example

presented in [2, Sec. V-A-1],[5]. The system to be identified is described by

$$G(q) = \frac{0.0890 - 0.2199q^{-1} + 0.2866q^{-2} - 0.2199q^{-3} + 0.0890q^{-4}}{1 - 2.6918q^{-1} + 3.5992q^{-2} - 2.4466q^{-3} + 0.8288q^{-4}} \tag{29}$$

and the input signal is obtained by high-pass filtering a zero-mean unit-variance Gaussian white noise signal $w(t)$,

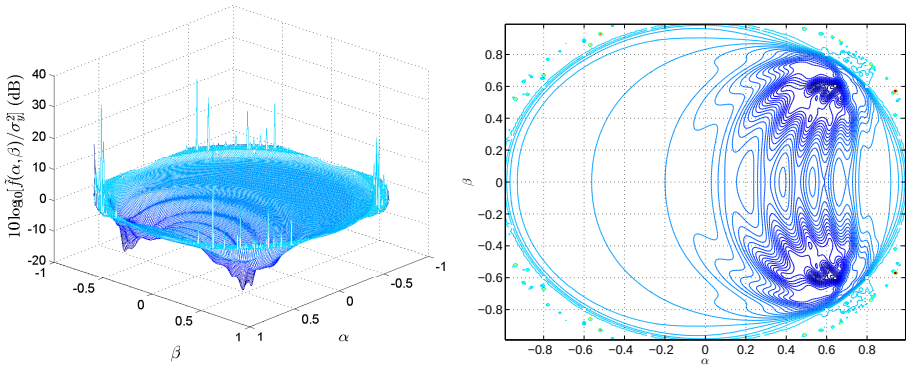
$$u(t) = \frac{1}{1 + 0.8q^{-1}} w(t). \tag{30}$$

The measurement noise $e(t)$ in (1) is assumed to be zero, the number of input and output samples is $N = 3000$, and the number of Kautz basis functions used is $n_B + 1 = 8$. The objective function $\tilde{f}(\boldsymbol{\eta})$, normalized with the output signal variance σ_y^2 is plotted on a logarithmic scale versus $\alpha \in (-1, 1)$ and $\beta \in (-1, 1)$ in Figs. 1(a) (surface plot) and 1(b) (contour plot). The objective function is symmetric around $\beta = 0$ (because the Kautz filters are real-valued) and has global minima at $\zeta \approx 0.6212 \pm j0.5790$. From the contour plot, it can be seen that there are many local minima.

In [2], the optimization algorithm with a GN search direction is applied using three different initial conditions, in the neighbourhood of the global minimum: $\hat{\boldsymbol{\eta}}^{(0)} = [0.69, 0.54]^T$ (case 1), $\hat{\boldsymbol{\eta}}^{(0)} = [0.55, 0.65]^T$ (case 2), and $\hat{\boldsymbol{\eta}}^{(0)} = [0.43, 0.35]^T$ (case 3). We also apply the SD and BFGS search directions to this problem, starting from the same initial conditions. The locus of the estimates $\hat{\boldsymbol{\eta}}^{(k)}$ is shown on a contour plot of the objective function, in the neighbourhood of the global minimum, for cases 1-3 in Figs. 2(a)-(c). Apparently, none of the algorithms succeeds in finding the global minimum, and they all get stuck near a local minimum. One remarkable observation is that the BFGS algorithm in case 1 produces a huge step from the neighbourhood of the global minimum in the first quadrant of the complex plane, to the neighbourhood of the global minimum in the fourth quadrant (see Fig. 2(d)).

The value of the objective function, evaluated at estimates $\hat{\boldsymbol{\eta}}^{(k)}$ obtained during subsequent iterations, is plotted versus the number of iterations in Figs. 3(a)-(c) for cases 1-3. These convergence plots show that, on average, the three different search directions yield a comparable estimation error, and that the GN algorithm converges slightly slower than the other ones. It should be noted that calculating the GN search direction is much more expensive from a computational point of view, as compared to the BFGS and SD search directions, the latter being the cheapest one to evaluate.

Finally, the truncated impulse responses of the system models $B(q, \hat{\boldsymbol{\eta}})$, evaluated at the final estimate of case 2, are compared to the truncated impulse response of the true system $G(q)$ in Figs. 3(d)-(f). The SD, GN, and BFGS estimated impulse responses are nearly identical, but are rather poor estimates of the true impulse response. To improve the fit, either the number of Kautz basis functions $n_B + 1$ should be increased, or the optimization algorithm should be started at an initial estimate which is even closer to the global minimum.



(a) (b)

Figure 1: Kautz filter example, objective function: (a) surface plot, (b) contour plot.

4.2 Example 2: Laguerre filters

In a second example, the optimization algorithm outlined in Section 3 is applied to a system identification problem using Laguerre filters. We adopt the example given in [2, Sec. V-A-2], [6], in which a system with transfer function

$$G(q) = \frac{1 - 3.8q^{-1} + 1.96z^{-2} - 0.297z^{-3}}{1 - 0.9z^{-1} - 0.19z^{-2} + 0.201z^{-3} - 0.027z^{-4}} \tag{31}$$

is identified using $N = 100$ samples of a unit impulse input signal. There is again no measurement noise, and the system model $B(q, \boldsymbol{\eta})$ is made up of $n_B + 1 = 8$ Laguerre basis functions. The objective function $\tilde{f}(\boldsymbol{\eta})$, normalized with σ_y^2, is plotted versus $\alpha \in (-1, 1)$ in Fig. 4(a). The objective function appears to have a convex shape, hence there is only one local minimum, equal to the global minimum.

In [2], this problem was tackled using an optimization algorithm with a GN search direction. We also apply the SD and BFGS search directions to this example. All algorithms start from $\alpha = 0$, which corresponds to the traditional FIR model. The locus of the estimates $\hat{\boldsymbol{\eta}}^{(k)} = \alpha^{(k)}$ is plotted in Figs. 4(b)-(d). It can be seen that all algorithms converge to the global minimum, but the number of iterations needed for convergence is very different. The convergence curves in Fig. 5(a) clearly show that the BFGS search direction is far superior over the other two, and the SD method even outperforms the GN method. Finally, the truncated impulse responses obtained with the final estimate of all three algorithms, are compared to the truncated true system impulse response in Figs. 5(b)-(d). The fit is relatively good for all three search directions.

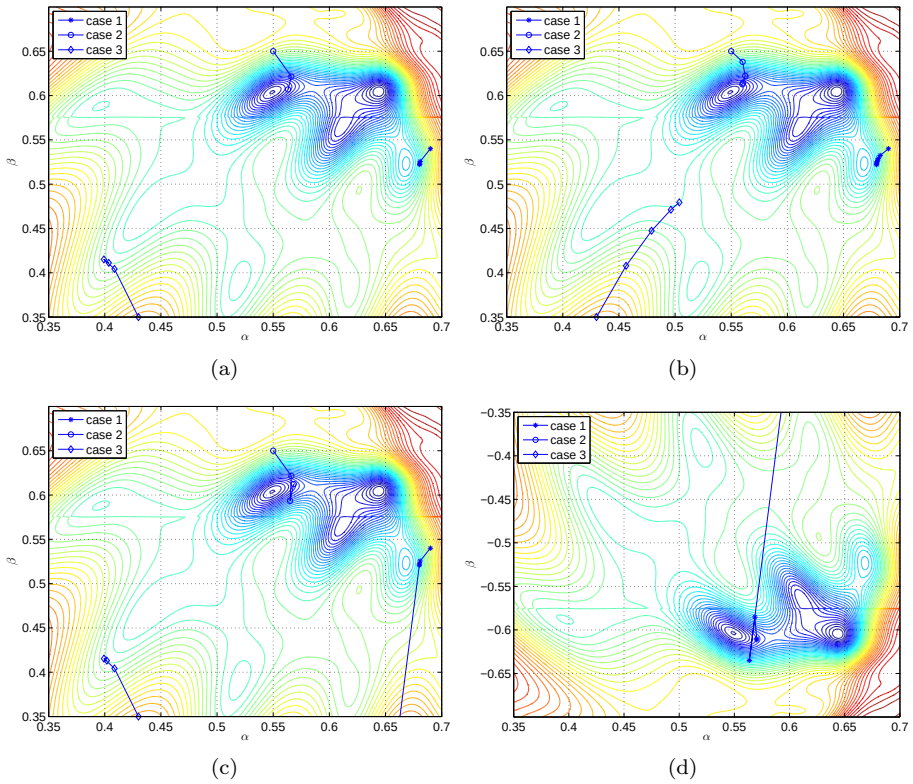


Figure 2: Kautz filter example, locus of $\hat{\boldsymbol{\eta}}^{(k)}$ for three different initial conditions: (a) SD search direction, (b) GN search direction, (c) BFGS search direction (first quadrant), (d) BFGS search direction (fourth quadrant).

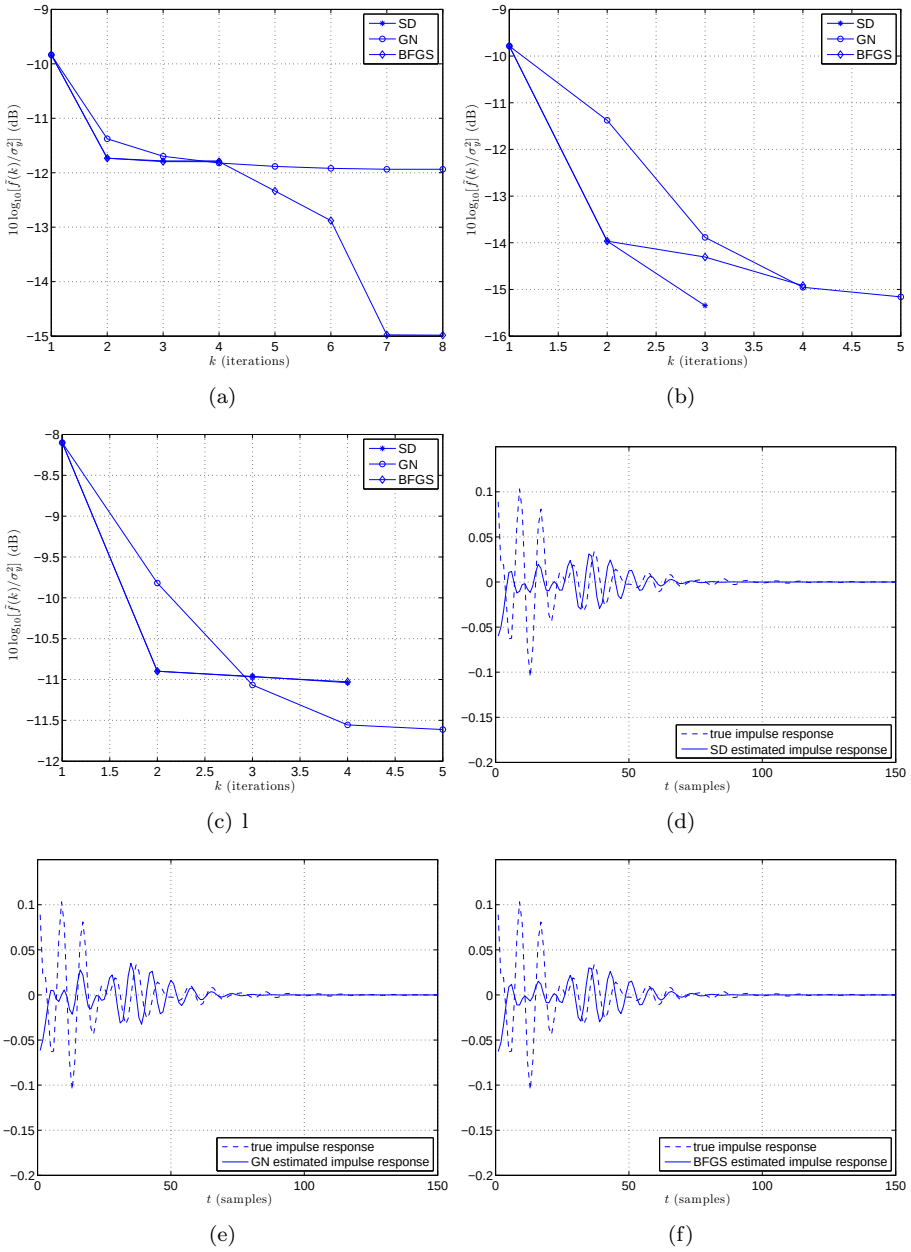


Figure 3: Kautz filter example, convergence curves for different initial conditions: (a) case 1, (b) case 2, (c) case 3; impulse response estimates obtained in case 2: (d) SD search direction, (e) GN search direction, (f) BFGS search direction.

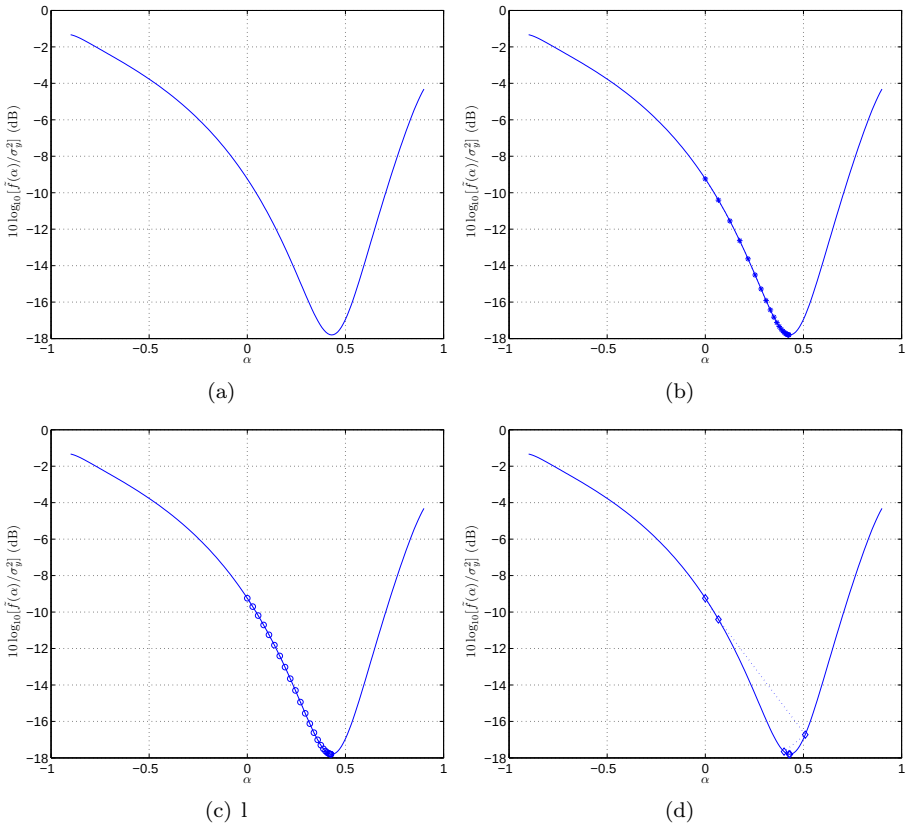


Figure 4: Laguerre filter example: (a) objective function; locus of $\hat{\boldsymbol{\eta}}^{(k)}$: (b) SD search direction, (c) GN search direction, (d) BFGS search direction.

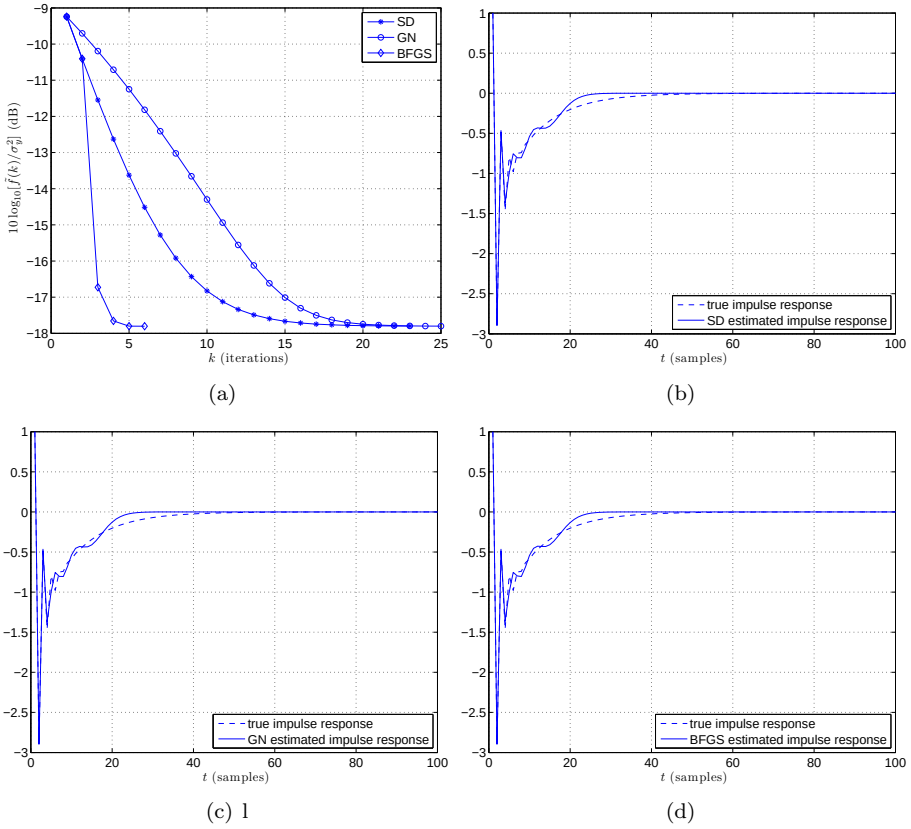


Figure 5: Laguerre filter example: (a) convergence curves; impulse response estimates: (b) SD search direction, (c) GN search direction, (d) BFGS search direction.

5 Conclusion

We have extended the work in [2], in which the application of a Gauss-Newton (GN) optimization algorithm to a system identification problem is considered by using discrete-time Kautz and Laguerre filters. As an alternative to the GN search direction, which is computationally expensive to calculate, we have also applied cheaper steepest descent (SD) and quasi-Newton (BFGS) optimization algorithms to the RBF system identification problem. Moreover, we have suggested to use a stronger sufficient decrease condition than the one used in [2]. Simulation results indicate that if the objective function is convex, the SD and BFGS algorithms outperform the GN algorithm in terms of convergence speed. If the objective function has a non-convex shape, the three algorithms have comparable performance, and a proper choice of the initial conditions is crucial for obtaining a useful estimate of the system impulse response.

References

[1] P. W. Broome, "Discrete orthonormal sequences," *J. Assoc. Comput. Machinery*, vol. 12, no. 2, pp. 151–168, Dec. 1965.

[2] L. S. H. Ngia, "Separable nonlinear least-squares methods for efficient off-line and on-line modeling of systems using Kautz and Laguerre filters," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 48, no. 6, pp. 562–579, June 2001.

[3] G. H. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM J. Numer. Anal.*, vol. 10, no. 2, pp. 413–432, Apr. 1973.

[4] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York, NY, USA: Springer, 2006.

[5] T. O. e Silva, "Optimality conditions for truncated Kautz networks with two periodically repeating complex conjugate poles," *IEEE Trans. Autom. Control*, vol. 40, no. 2, pp. 342–346, Feb. 1995.

[6] N. Tanguy, R. Morvan, P. Vilbe, and L. C. Calvez, "Improved method for optimum choice of free parameter in orthogonal approximations," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2576–2578, Sept. 1999.