5.1 Adaptive IIR Filter Realization

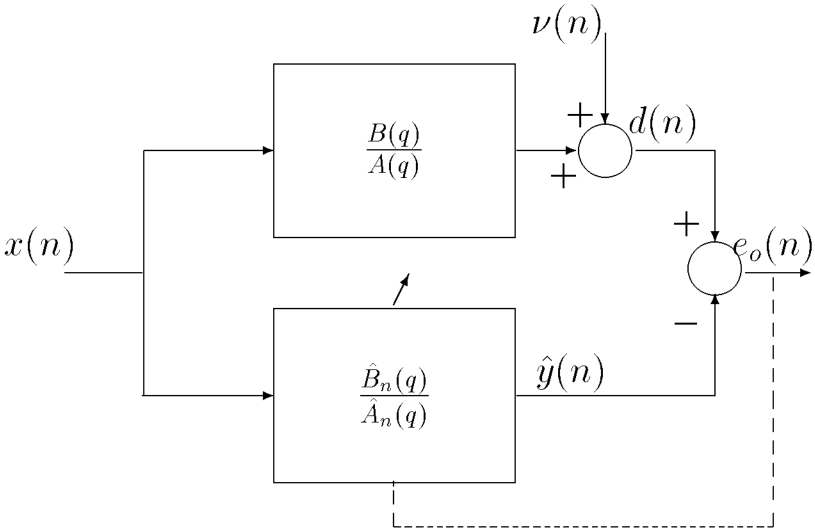

Figure 27: System identification configuration

The plant output is given by

$$d(n) = \frac{B(q)}{A(q)}x(n) + \nu(n)$$

where $\nu(n)$ is a zero mean measurement noise with bounded variance and uncorrelated with $x(n)$.

The IIR adaptive filter is

$$\hat{y}(n) = \frac{\hat{B}_n(q)}{\hat{A}_n(q)}x(n)$$

where

$$\hat{B}_n(q) = \sum_{k=0}^{M} \hat{b}_k(n)q^{-k} \qquad \hat{A}_n(q) = 1 + \sum_{k=1}^{N} \hat{a}_i(n)q^{-k}$$

where is assumed, without loss of generality that $M = N$.

The output error $e_o(n)$, defined by $e_o(n) = d(n) - \hat{y}(n)$, is minimized based in the following function

$$\xi_{OE} = (1/2)E\{e_o^2(n)\} = E\left\{\left[\left(\frac{B(q)}{A(q)} - \frac{\hat{B}_n(q)}{\hat{A}_n(q)}\right)x(n)\right]^2\right\} + E\{\nu^2(n)\}$$

An approach to solve this minimization problem is

$$\boldsymbol{\theta}(n+1) = \boldsymbol{\theta}(n) + \mu\boldsymbol{\nabla}(\xi_{OE})(n)e_o(n) \tag{48}$$

where

$$\boldsymbol{\theta}(n) = [\hat{a}_1(n), ..., \hat{a}_N, \hat{b}_0(n), ..., b_N(n)]^T$$

and

$$\boldsymbol{\nabla}(\xi_{OE})(n) = [\frac{\partial\xi_{OE}}{\partial\hat{a}_1}...\frac{\partial\xi_{OE}}{\partial\hat{a}_N}\frac{\partial\xi_{OE}}{\partial\hat{b}_0}...\frac{\partial\xi_{OE}}{\partial\hat{b}_N}]^T$$

where $\frac{\partial\xi_{OE}}{\partial\hat{a}_k}$ and $\frac{\partial\xi_{OE}}{\partial\hat{b}_j}$ are suitable recursive estimates.

To achieve faster convergence at the cost of additional complexity, a Gauss-Newton version algorithm can be contemplated

$$\boldsymbol{\theta}(n+1) = \boldsymbol{\theta}(n) + \mu\boldsymbol{P}(n+1)\boldsymbol{\nabla}(\xi_{OE})(n)e_o(n)$$

where

$$\boldsymbol{P}(n+1) = \left(\frac{1}{1-\mu}\right)\left(\boldsymbol{P}(n) - \frac{\boldsymbol{P}(n)\boldsymbol{\nabla}(\xi_{OE})(n)\boldsymbol{\nabla}^T(\xi_{OE})(n)\boldsymbol{P}(n)}{\frac{1-\mu}{\mu} + \boldsymbol{\nabla}^T(\xi_{OE})(n)\boldsymbol{P}(n)\boldsymbol{\nabla}^T(\xi_{OE})(n)}\right)$$

To obtain a recursive realization

$$\hat{y}(n) = -\sum_{k=1}^{N} \hat{a}_k(n)\hat{y}(n-k) + \sum_{k=0}^{N} \hat{b}_k(n)x(n-k)$$

In algorithm (48) the coefficients $\boldsymbol{\theta}(n)$ can be adapted as follows

$$(1/2)\nabla_{\boldsymbol{\theta}} E\{e_o^2(n)\} \approx e_o(n)\nabla_{\boldsymbol{\theta}} e_o(n) = -e_o(n)\nabla_{\boldsymbol{\theta}}\hat{y}(n)$$

where
$\nabla_{\boldsymbol{\theta}}\hat{y}(n) = [\frac{\partial\hat{y}(n)}{\partial\hat{a}_1(n)}, ..., \frac{\partial\hat{y}(n)}{\partial\hat{a}_N(n)}, \frac{\partial\hat{y}(n)}{\partial\hat{b}_0(n)}, ..., \frac{\partial\hat{y}(n)}{\partial\hat{b}_N(n)}]^T$
Then

$$\begin{aligned}\frac{\partial\hat{y}(n)}{\partial\hat{a}_k(n)} &= -\hat{y}(n-k) - \sum_{m=1}^{N} \hat{a}_m(n)\frac{\partial\hat{y}(n-m)}{\partial\hat{a}_k(n)} \\ \frac{\partial\hat{y}(n)}{\partial\hat{b}_k(n)} &= x(n-k) - \sum_{m=1}^{N} \hat{a}_m(n)\frac{\partial\hat{y}(n-m)}{\partial\hat{b}_k(n)}\end{aligned}$$

Using the slow convergence factor approximation

$$\boldsymbol{\theta}(n) \approx \boldsymbol{\theta}(n-1) \approx ... \approx \boldsymbol{\theta}(n-N) \tag{49}$$

Then

$$\begin{aligned}\frac{\partial\hat{y}(n)}{\partial\hat{a}_k(n)} &\approx -\hat{y}(n-k) - \sum_{m=1}^{N} \hat{a}_m(n)\frac{\partial\hat{y}(n-m)}{\partial\hat{a}_k(n-m)} \\ &= -\frac{1}{\hat{A}_n(q)}\hat{y}(n-k) \\ \frac{\partial\hat{y}(n)}{\partial\hat{b}_j(n)} &\approx x(n-j) - \sum_{m=1}^{N} \hat{a}_m(n)\frac{\partial\hat{y}(n-m)}{\partial\hat{b}_j(n-m)} \\ &= \frac{1}{\hat{A}_n(q)}x(n-j)\end{aligned}$$

for $1 \le k \le N$ and $0 \le j \le N$. By using (49) and defining

$$\begin{aligned}\hat{y}_f(n-1) &= \frac{\partial\hat{y}(n)}{\partial\hat{a}_1(n)} \\ x_f(n) &= \frac{\partial\hat{y}(n)}{\partial\hat{b}_0(n)}\end{aligned}$$

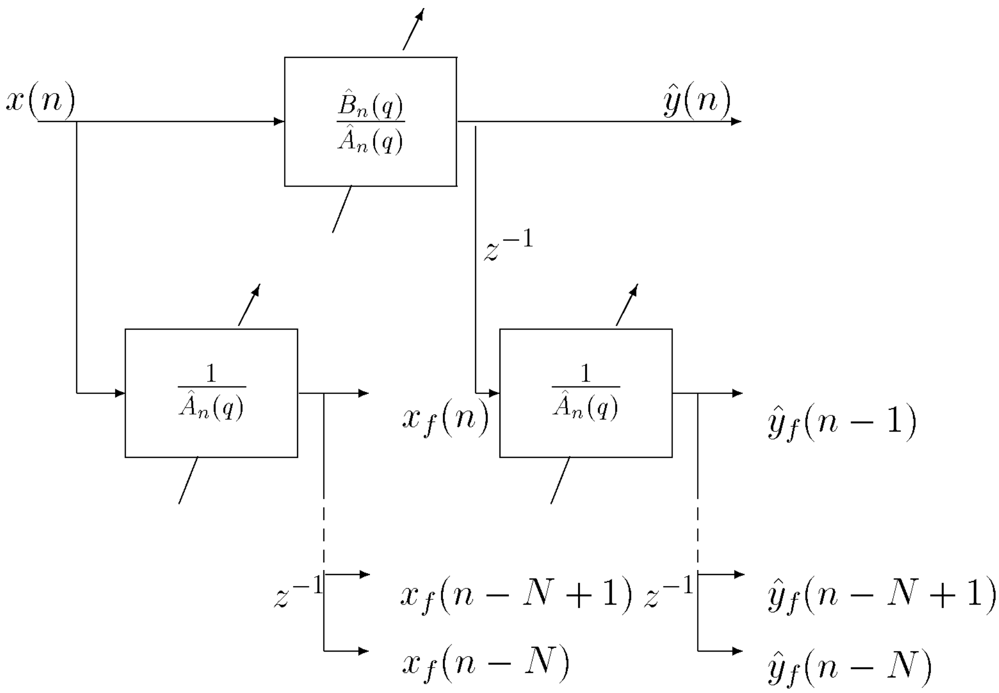

Figure 28: IIR adaptive filter realization wit hte **Recursive Gradient algorithm**

it is possible to find that

$$
\begin{aligned}
-\frac{\partial \hat{y}(n)}{\partial \hat{a}_k(n)} &= \hat{y}_f(n-k) \quad k = 2, ..., N \\
\frac{\partial \hat{y}(n)}{\partial \hat{b}_k(n)} &= x_f(n-k) \quad k = 1, ..., N
\end{aligned}
$$

With $\boldsymbol{\psi}(n) = [\hat{y}_f(n-1), ..., \hat{y}_f(n-N+1), x_f(n), ..., x_f(n-N+1)]^T$, the final form of the gradient version of the algorithm is

$$
\boldsymbol{\theta}(n+1) = \boldsymbol{\theta}(n) + \mu \boldsymbol{\psi}(n) e_o(n) \tag{50}
$$

5.2 Stationary points

- Objective: use of the $\mathcal{H}_2$ decomposition theorem to characterize properties of the algorithm.
- Witn $x(n)$ white noise and $N = M$, the stationary points related to the $\hat{b}_j$ coefficients

$$E\{\begin{bmatrix} \frac{1}{\overline{A}(q)}x(n) \\ \frac{1}{\overline{A}(q)}x(n-1) \\ \vdots \\ \frac{1}{\overline{A}(q)}x(n-N) \end{bmatrix} \left(\frac{B(q)}{A(q)} - \frac{\overline{B}(q)}{\overline{A}(q)}\right) x(n)\} \;=\; 0$$

and using the $\mathcal{H}_2$-inner product notation (with $x(n)$ unit white noise),

$$\left\langle \begin{bmatrix} \frac{1}{\overline{A}(z)} \\ \frac{1}{\overline{A}(z)} \\ \vdots \\ \frac{1}{\overline{A}(z)} \end{bmatrix}, \left(H(z) - \hat{H}(z)\right) \right\rangle \;=\; 0$$

- To use the Decomposition Theorem, we search for a function that

$$\left\langle \mathcal{C}_a(z), \left(H(z) - \hat{H}(z)\right) \right\rangle \;=\; 0$$

but this is satisfied for the $N+1$-controllability matrix

$$\mathcal{C}_a(z) \;=\; \begin{bmatrix} \frac{1}{\overline{A}(z)} \\ \frac{z^{-1}}{\overline{A}(z)} \\ \vdots \\ \frac{z^{-N}}{\overline{A}(z)} \end{bmatrix} = z(z\boldsymbol{I}_{N+1} - \boldsymbol{A}_a)^{-1}\boldsymbol{b}_a$$

where

$$\boldsymbol{A}_a = \begin{bmatrix} -a_1 & -a_2 & \cdots & -a_N & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$
$$\boldsymbol{b}_a = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

- The associated all-pass function is

$$\begin{aligned} V_a(z) &= \frac{det(\boldsymbol{I} - z\boldsymbol{A}_a)}{det(z\boldsymbol{I} - \boldsymbol{A}_a)} \\ &= z^{-1}\frac{a_N + a_{N-1}z^{-1} + ... + a_1 z^{-N-1} + z^{-N}}{1 + a_1 z^{-1} + ... + a_{N-1}z^{N-1} + a_N z^{-N}} = z^{-1}V(z) \end{aligned}$$

- By using the decomposition theorem, the equation of the stationary points with respect to the $\hat{b}_i$ coefficients is satisfied if and only if

$$H(z) - \hat{H}(z) = V(z)\sum_{k=1}^{\infty} g_k z^{-k}$$

where $g(z)$ is strictly causal.

- To optimize the MSOE as a function of $\hat{a}_i$ consider

$$E\left\{\begin{bmatrix} \frac{\overline{B}(q)}{\overline{A}^2(q)}x(n-1) \\ \frac{\overline{B}(q)}{\overline{A}^2(q)}x(n-2) \\ \vdots \\ \frac{\overline{B}(q)}{\overline{A}^2(q)}x(n-N) \end{bmatrix}\left(\frac{B(q)}{A(q)} - \frac{\overline{B}(q)}{\overline{A}(q)}\right)x(n)\right\} = 0$$

that in terms of the $\mathcal{H}_2$-inner product notation can be rewritten as

$$\left\langle \begin{bmatrix} \frac{z^{-1}}{\overline{A}(z)} \\ \frac{z^{-2}}{\overline{A}(z)} \\ \vdots \\ \frac{z^{-N}}{\overline{A}(z)} \end{bmatrix} \hat{H}(z), \left(H(z) - \hat{H}(z)\right) \right\rangle = 0$$

note that this is a necessary but not sufficient condition for the MSOE to be minimized with respect to the $\hat{a}_k$ coefficients.

- Since that the optimization with respect to b_i requires that $H(z) - \hat{H}(z) = V(z)g(z)$, then

$$\begin{aligned} 0 &= \left\langle \begin{bmatrix} \frac{z^{-1}}{\overline{A}(z)} \\ \frac{z^{-2}}{\overline{A}(z)} \\ \vdots \\ \frac{z^{-N}}{\overline{A}(z)} \end{bmatrix} \hat{H}(z), V(z)g(z) \right\rangle \\ &= \left\langle \begin{bmatrix} \frac{1}{\overline{A}(z)} \\ \frac{z^{-1}}{\overline{A}(z)} \\ \vdots \\ \frac{z^{-(N+1)}}{\overline{A}(z)} \end{bmatrix}, [V(z)\hat{H}(z^{-1})][zg(z)] \right\rangle \\ &= \left\langle \mathcal{C}(z), [V(z)\hat{H}(z^{-1})][zg(z)] \right\rangle \end{aligned}$$

- Some remarks:
  - Noting that $\hat{H}(z) = d + \boldsymbol{c}\mathcal{C}(z)$, then by the decomposition theorem, the inner product

$$\begin{aligned} < V(z^{-1})\hat{H}(z), z^{-k} > &= < \hat{H}(z), z^{-k}V(z) > \\ &= d < 1, z^{-k}V(z) > + \boldsymbol{c} < \mathcal{C}(z), z^{-k}V(z) >= 0 \end{aligned}$$

then $V(z^{-1})\hat{H}(z)$ is anticausal. Then $V(z)\hat{H}(z^{-1})$ is a causal function given by

$$V(z)\hat{H}(z^{-1}) = \frac{z^{-N}A(z^{-1})}{A(z)} \frac{B(z^{-1})}{A(z^{-1})} = \frac{z^{-N}B(z^{-1})}{A(z)}$$

- Since $g(z)$ is strictly causal, the function $zg(z)$ is causal.
- Then the product $[V(z)\hat{H}(z^{-1})][zg(z)]$ appears as a causal function and by the decomposition theorem, it must be causally divisible by $V(z)$.
- If $\hat{H}(z)$ has degree N, this implies that $[zg(z)]$ is causally divisible by $V(z)$ (i.e., the zeros of both functions coincide).
- Then $[zg(z)] = V(z)q(z)$ or $[zg(z)] = V(z)z^{-1}q(z)$, with $q(z) \in \mathcal{H}_2$

Theorem (Walsh): If degree of $\hat{H}(z)$ is N, then $\hat{H}(z)$ is a stationary point of $\|H(z) - \hat{H}(z)\|^2$ if and only if

$$H(z) - \hat{H}(z) = z^{-1}[V(z)]^2 q(z)$$

for some $q(z) \in \mathcal{H}_2$, where $V(z)$ is the all-pass function whose poles coincide with those of $\hat{H}(z)$.

- The main interpretation of this result is as an interpolation condition, i.e., for example, if $z_1, .., z_N$ are the poles of $\hat{H}(z)$ then

$$\begin{aligned} H(z_k^{-1}) &= \hat{H}(z_k^{-1}) && k = 1, ..., N, \\ \left.\frac{\partial H(z)}{\partial z}\right|_{z=z_k^{-1}} &= \left.\frac{\partial \hat{H}(z)}{\partial z}\right|_{z=z_k^{-1}} && k = 1, ..., N, \end{aligned}$$

5.2.1 Properties

- Using the *identifiability* of the proposed model, defining $n^* = min(N - n_a, M - n_b)$, then
  - if $n^* \geq 0$ the problem is of sufficient order,
  - else if $n^* < 0$ the system identification problem is of insufficient order.
- All results obtained by the decomposition theorem can be used in the insufficient order case, mainly the interpolation results.
- Is useful to consider the MSOE as

$$\xi_o = E\{e_o^2(n)\} = F_0 - 2\boldsymbol{F_1}(\boldsymbol{a})^T\boldsymbol{b} + \boldsymbol{b}^T\boldsymbol{F_2}(\boldsymbol{a})\boldsymbol{b}$$

  where $F_0 = E\{y^2(n)\}$, $\boldsymbol{F_1}(\boldsymbol{a}) = E\{[\frac{B(q)}{A(q)}x(n)][\frac{x(n-i)}{A(q)}]\}$ for $0 \leq i \leq n_b$, $\boldsymbol{F_2}(\boldsymbol{a}) = E\{[\frac{x(n-i)}{A(q)}][\frac{x(n-j)}{A(q)}]\}$ for $0 \leq i \leq n_b$, $1 \leq j \leq n_a$.
  where it is possible to verify the quadratic relation with the numerator coefficients. By minimizing with respect to $\boldsymbol{b}$, the reduced MSOE is given by

$$\xi_o^r = E\{e_o^2(n)\} = F_0 - \boldsymbol{F_1}(\boldsymbol{a})^T\boldsymbol{F_2}(\boldsymbol{a})\boldsymbol{F_1}(\boldsymbol{a})$$

- An approach to the analysis of the stationary points introduce the concept of *degenerate points*, i.e., such stationary points where $\hat{B}_n(q) = 0$.
- It can be shown that for many cases of insufficient order, i.e., $n^* \geq 0$, the existence of degenerate points implies the existence of saddle points and, as a consequence, determines the multimodality of the MSOE.

Also, in general (for sufficient and insufficient order cases)

- A stationary point that introduce a pole-zero cancellation is in fact a saddle point (not a minimum of the MSOE).

- *Normality condition of the MSOE*: Let $s_k = \min_{deg\ \hat{H}(z) \leq k} \|H(z) - \hat{H}(z)\|^2$ corresponding to the global minimum. Then $s_{k+1} < s_k$ (note that this is not a general property of every minimization method).

- Suppose the cost function admits k stationary points (including pole-zero cancellations), i.e., $\hat{H}_1(z), ..., \hat{H}_k(z)$. Let ϵ_i, $i = 1, ..., k$, the number of the negative eigenvalues of the Hessian matrix at the i-th stationary point. The *index* of stationary points is defined as $\sum_{i=1}^{k}(-1)^{\epsilon_i}$ where the sum is over all the stationary points. In particular, a local minimum will contribute a term $+1$ $(\epsilon_i = 0)$ to the above sum.

  Theorem: The index always equals one: $\sum_i (-1)^{\epsilon_i} = 1$

  Remarks:

  - If two or more stationary points exist, they can not all be minima.
  - If al candidate stationary points are expected to be minimum points, the cost function must have a sole stationary point, yielding a global minimum.
  - If two or more distinct minimum points occur, then saddle points must also be present.

5.2.2 ODE associated

The ODE associated to the gradient version algorithm (50), can be written

$$\frac{\partial \boldsymbol{\vartheta}(t)}{\partial t} = E\{\boldsymbol{\psi}(n)e_o(n)\}$$

and for the Gauss-Newton version

$$\begin{aligned}\frac{\partial \boldsymbol{\vartheta}(t)}{\partial t} &= \boldsymbol{\varrho}^{-1}(t)\mathbf{G}_1\left(\boldsymbol{\theta}_o - \boldsymbol{\vartheta}(t)\right) \\ \frac{\partial \boldsymbol{\varrho}(t)}{\partial t} &= \mathbf{G}_2 - \boldsymbol{\varrho}(t)\end{aligned}$$

where

$$\mathbf{G}_1 = E\{\ \boldsymbol{\psi}(n)\hat{\boldsymbol{\varphi}}^T(n)\}$$

with $\hat{\boldsymbol{\varphi}}(n) = [\hat{y}(n-1)..., \hat{y}(n-N), x(n)..., x(n-M)]^T$, and

$$\mathbf{G}_2 = E\{\ \boldsymbol{\psi}(n)\boldsymbol{\psi}^T(n)\}$$

In particular, to verify the (local) convergence of the ODE associated, a suitable Liapunov function can be easily found (by definition)

$$V(\boldsymbol{\vartheta}(t)) = \frac{1}{2}E\{e_o^2(n)\}$$

such that

$$\begin{aligned}(dV/dt) &= -\left[E\{\boldsymbol{\psi}(n)e_o(n)\}\right]^T \frac{d\,\boldsymbol{\vartheta}(t)}{d\,t} \\ &= -\left[E\{\boldsymbol{\psi}(n)e_o(n)\}\right]^T \left[E\{\boldsymbol{\psi}(n)e_o(n)\}\right] \\ &\leq 0\end{aligned}$$

Because of the existence of local minima, only local convergence of the previous algorithm can be guaranteed.

5.3 Alternative realizations

The particular choice of a realization of the IIR adaptive filter has influence in, between other issues:

- computational complexity (for stability check).
- convergence speed (exponential stability must be guaranteed).
- MSOE surface shape (mapping between direct-form parameter space and other form of adaptive filter parameters).

For the MSOE problem, however, the local minima problem can not be avoided.
Other possible realizations are

- Cascade: manifolds, high complexity of the gradient computation, low convergence speed, not guaranteed exponential stability.
- Parallel: manifolds, the lower complexity of the gradient, low convergence speed, not guaranteed exponential stability.
- Frequency domain: manifolds, high complexity, suitable convergence speed, not guaranteed exponential stability.
- Orthonormal: modeling different poles, low complexity, suitable convergence speed, not guaranteed exponential stability.
- Lattice: no manifolds, higher complexity than orthonormal, suitable convergence speed, guaranteed exponential stability.

In this characterization is not included aspects related to convergence analysis of the MSOE.

5.3.1 Parallel and cascade realizations

Is obtained from the partial fraction expansion of the transfer function of the adaptive filter at iteration n fixed, resulting in a sum of $N/2$ sections, as

$$H_k(q) = \frac{\hat{B}_n^k(q)}{\hat{A}_n^k(q)} = \frac{\hat{b}_{0k}(n) + \hat{b}_{1k}(n)q^{-1}}{1 - \hat{a}_{1k}(n)q^{-1} - \hat{a}_{2k}(n)q^{-2}}$$

where $k = 0, ..., N/2 - 1$.

- Stability check: $|\hat{a}_{2k}(n)| < 1$ and $|\hat{a}_{1k}(n)| < 1 - \hat{a}_{2k}$, for $k = 0, ..., (\frac{N}{2} - 1)$.
- MSOE surface with multiple global minima ($!N$), divided by **manifolds** when $a_{1k} = a_{1j}$ and $a_{2k} = a_{2j}$ for $k, j = 1, ..., N/2$. This leads in general to slow convergence speed.
- Ill conditioned in a Gauss-Newton algorithm (because the Hessian matrix will be close to singular over the manifold regions).
- Computational complexity comparable to (slightly lower than) the direct-form realization.

The cascade realization

- Analogous to the parallel realization, as a product of $N/2$ second-order sections.
- Low convergence speed by similar considerations on manifolds.
- High complexity in the gradient computation.

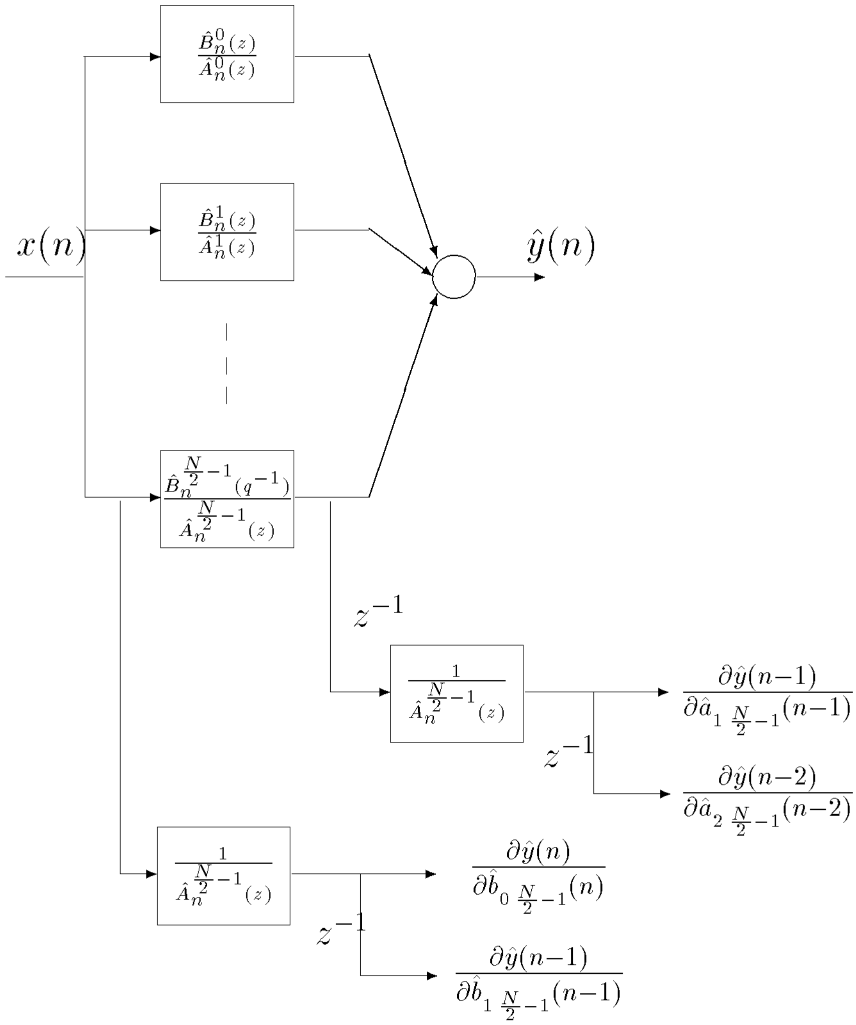

Figure 29: Parallel realization of the adaptive IIR filter with a detail of the last section coefficients.

5.3.2 Frequency Domain Parallel realization

- Include a Discrete Fourier Transform to uncorrelate the inputs to the complex first-order sections.
- A variant is a Frequency domain parallel realization using second order sections and the Discrete Cosine Transform or any other real transform.
- Reduce ill conditioning in Gauss-Newton algorithms, the manifold problem remain.
- Introduce additional complexity with the transform.
- Some simplifications can be considered, but lead in general to suboptimal MSOE performance.

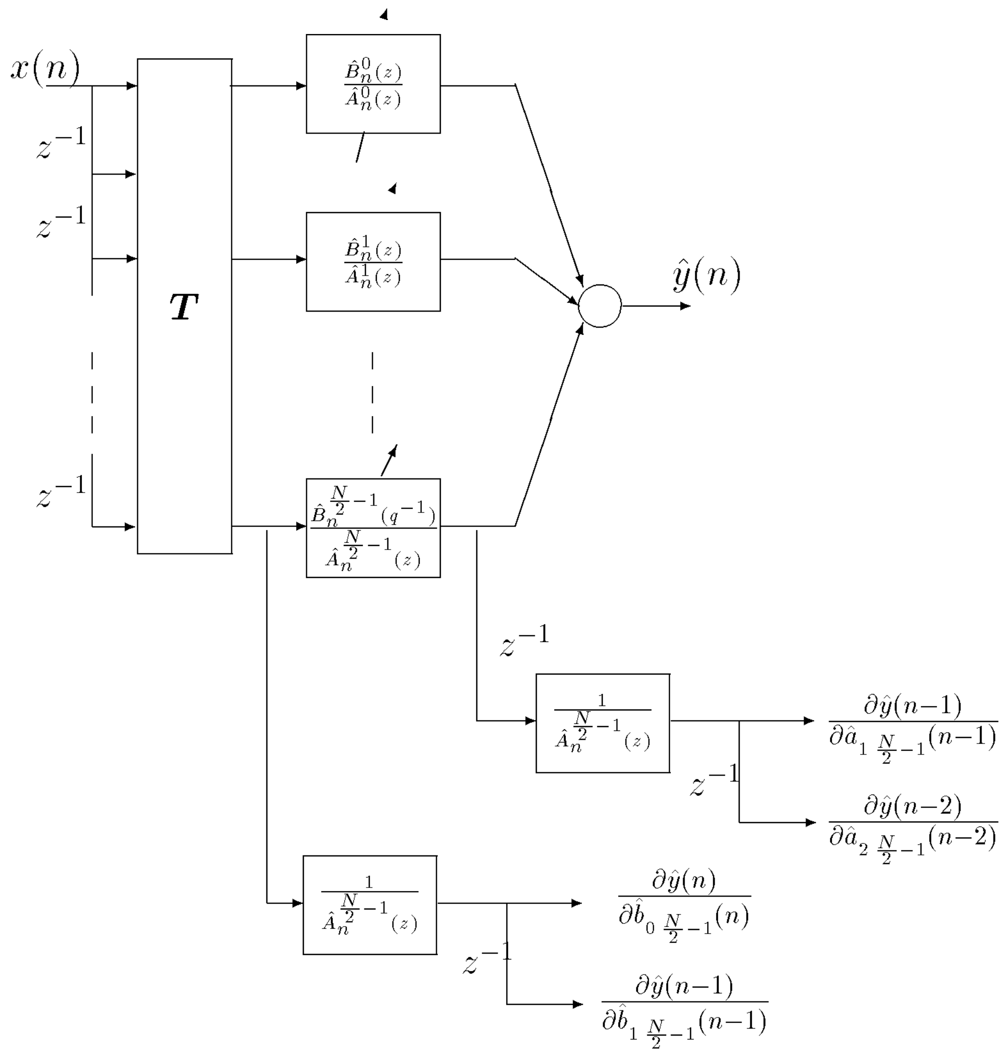

Figure 30: Frequency domain adaptive IIR filter with a detail of the last section coefficients.

5.3.3 Lattice realization

- Onto mapping with the direct form realization.
- High complexity for gradient computation (a new lattice for each gradient component), but $O(N)$ lattice realizations exists.
- The stability check for the *one multiplier* form (see the next figure) is $|\kappa_k(n)| < 1$.
- For the normalized form lattice realization, the following generic block

$$\boldsymbol{u}_k(n) = \begin{bmatrix} \frac{1}{1-\kappa_k(n)} & -\frac{\kappa_k(n)}{1-\kappa_k(n)} \\ -\frac{\kappa_k(n)}{1-\kappa_k(n)} & \frac{1}{1-\kappa_k(n)} \end{bmatrix} \boldsymbol{u}_{k-1}(n)$$

need to be replaced by

$$\boldsymbol{u}_k(n) = \begin{bmatrix} \cos\theta_k(n) & -\sin\theta_k(n) \\ \sin\theta_k(n) & \cos\theta_k(n) \end{bmatrix} \boldsymbol{u}_{k-1}(n)$$

where $\boldsymbol{u}_{k-1}(n)$ and $\boldsymbol{u}_{k-1}(n)$ are the output and input to section k, respectively.

- The normalized form lattice realization (described in the previous chapter) guarantee exponential stability if $\theta_k(n) < \pi/2$.

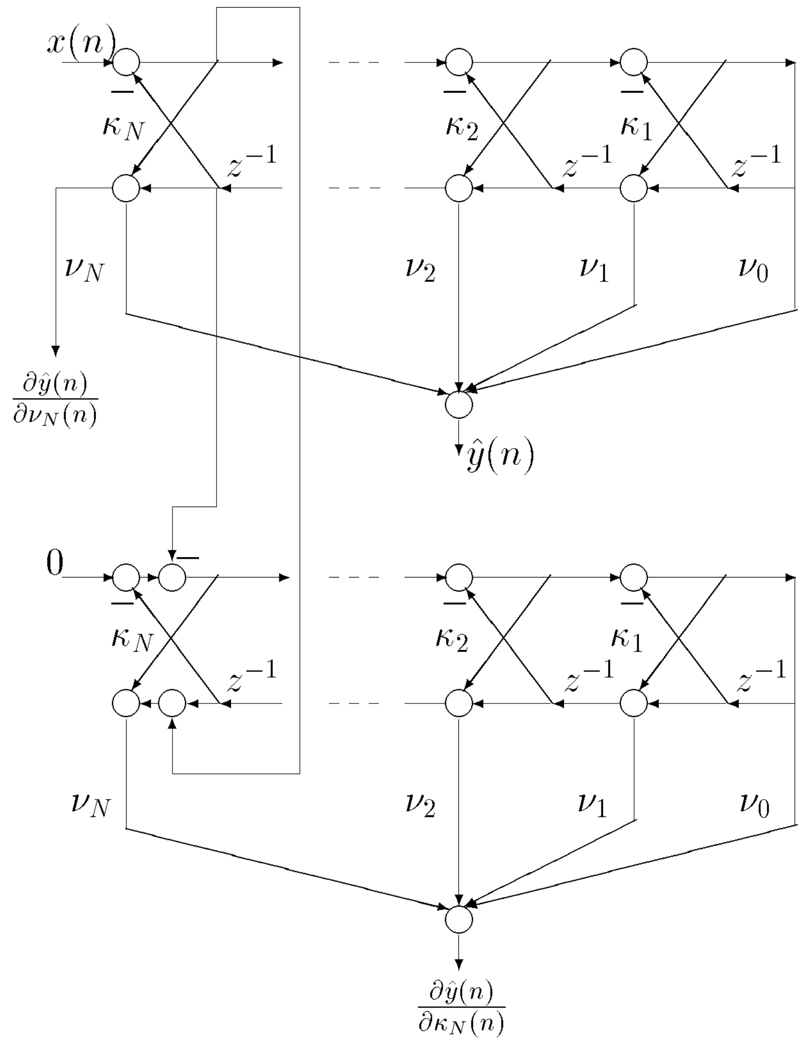

Figure 31: Lattice realization (with multiplier form) with the detail of gradient computation of $\kappa_N(n)$ and $\nu_N(n)$.

5.3.4 Orthogonal realization

Consider the orthonormal realization

$$y(n) = \left[\sum_{k=1}^{N/2}\{\nu_{2k-1}F'_k + \nu_{2k}F_k(z)\}\right] u(n) + \nu_0(n)u(n) \tag{51}$$

where

$$\begin{aligned}
F_k(z) &= \frac{N_k(z)}{D_k(z)}\prod_{i=1}^{k-1}\frac{\overline{D}_i(z)}{D_i(z)} \\
F'_k(z) &= \frac{N'_k(z)}{D_k(z)}\prod_{i=1}^{k-1}\frac{\overline{D}_i(z)}{D_i(z)} \\
D_k(z) &= 1 + \beta_{1_k}z^{-1} + \beta_{2_k}z^{-2} \\
\overline{D}_k(z) &= z^{-2}D_k(z^{-1}) \\
N_k(z) &= \alpha_{1_k} + \alpha_{2_k}z^{-1} \\
N'_k(z) &= \alpha'_{1_k} + \alpha'_{2_k}z^{-1}
\end{aligned}$$

$$\alpha_{1_k} = -\tfrac{1}{2}\sqrt{c}\left(\sqrt{a}+\sqrt{b}\right) \qquad \alpha'_{1_k} = \alpha_{2_k}$$
$$\alpha_{2_k} = \tfrac{1}{2}\sqrt{c}\left(\sqrt{a}-\sqrt{b}\right) \qquad \alpha'_{2_k} = \alpha_{1_k}$$

with: $a = 1 - \beta_{1_k} + \beta_{2_k}$, $b = 1 + \beta_{1_k} + \beta_{2_k}$ and $c = 1 - \beta_{2_k}$. Note that the definitions of α_{1_k}, α_{2_k}, α'_{1_k} and α'_{2_k} are required in order to maintain the basis functions $F_k(z)$ orthonormal, i.e.,

$$\|F_k(z)\| = 1 \qquad \langle F_k(z), F_l(z)\rangle = \delta_{k,l}$$

Then $e(n)$ that can be expressed as

$$\begin{aligned}
e(n) &= d(n) - y(n) = d(n) - \hat{H}(z)u(n) \\
&= d(n) - \left(\sum_{k=1}^{N} w_k(n) + \nu_0(n)u(n)\right)
\end{aligned}$$

where $d(n)$ is the reference signal and $w_k(n) = \nu_{2k-1}y'_k(n) + \nu_{2k}y_k(n) = \nu_{2k-1}F'_k(z)u(n) + \nu_{2k}F_k(z)u(n)$.

Using an stochastic gradient algorithm

$$\theta(n+1) = \theta(n) - \mu\nabla(n) \tag{52}$$

where

$$\begin{aligned}
\theta(n) = {} & [\nu_0(n)\,\nu_1(n)...\nu_{N-1}(n)\nu_N(n) \\
& \beta_{1_1}(n)\,\beta_{2_1}(n)...\beta_{1_{N/2}}(n)\,\beta_{2_{N/2}}(n)]^T
\end{aligned}$$

and $\nabla(n)$

$$\nabla(n) = -2e(n)[\frac{\partial y(n)}{\partial \nu_0} \cdots \frac{\partial y(n)}{\partial \nu_{N-1}} \frac{\partial y(n)}{\partial \nu_N} \frac{\partial y(n)}{\partial \beta_{1_1}} \frac{\partial y(n)}{\partial \beta_{2_1}} \cdots \frac{\partial y(n)}{\partial \beta_{1_{N/2}}} \frac{\partial y(n)}{\partial \beta_{2_{N/2}}}]^T \tag{53}$$

Then

$$\frac{\partial y(n)}{\partial \nu_k} = \begin{cases} F_k(n)u(n) & for\ k\ even \\ F'_k(n)u(n) & for\ k\ odd \end{cases} \tag{54}$$

$$\frac{\partial y(n)}{\partial \beta_{1_j}} = \sum_{k=1}^{N} \left(\nu_{2k-1} \frac{\partial y'_k(n)}{\partial \beta_{1_j}} + \nu_{2k} \frac{\partial y_k(n)}{\partial \beta_{1_j}} \right) \tag{55}$$

$$\frac{\partial y(n)}{\partial \beta_{2_j}} = \sum_{k=1}^{N} \left(\nu_{2k-1} \frac{\partial y'_k(n)}{\partial \beta_{2_j}} + \nu_{2k} \frac{\partial y_k(n)}{\partial \beta_{2_j}} \right) \tag{56}$$

$k = 0, 1...N$ $(F_0(z) = 1)$, $j = 1, ..., N/2$ where

$$\frac{\partial y_k(n)}{\partial \beta_{1_j}} = \begin{cases} 0 & if\ j > k \\ \left[F_k^a(z) - F_k(z) \frac{z^{-1}}{D_k(z)} \right] u(n) & if\ j = k \\ \left[F_k^j(z) \frac{z^{-1}}{D_j(z)} - F_k(z) \frac{z^{-1}}{D_j(z)} \right] u(n) & if\ j < k \end{cases}$$

$$\frac{\partial y_k(n)}{\partial \beta_{2_j}} = \begin{cases} 0 & if\ j > k \\ \left[F_k^b(z) - F_k(z) \frac{z^{-2}}{D_k(z)} \right] u(n) & if\ j = k \\ \left[F_k^j(z) \frac{1}{D_j(z)} - F_k(z) \frac{z^{-2}}{D_j(z)} \right] u(n) & if\ j < k \end{cases}$$

and

$$F_k^a(z) = \frac{\partial N_k(z)}{\partial \beta_{1_j}} \left(\prod_{i=1}^{k} \frac{\overline{D}_{i-1}(z)}{D_i(z)} \right)$$

$$F_k^b(z) = \frac{\partial N_k(z)}{\partial \beta_{2_j}} \left(\prod_{i=1}^{k} \frac{\overline{D}_{i-1}(z)}{D_i(z)} \right)$$

$$F_k^j(z) = N_k(z) \left(\prod_{i=1, i \neq j}^{k} \frac{\overline{D}_{i-1}(z)}{D_i(z)} \right)$$

with $\frac{\partial N_k(z)}{\partial \beta_{1_j}} = \frac{\partial \alpha_{1_k}}{\partial \beta_{1_j}} + \frac{\partial \alpha_{2_k}}{\partial \beta_{1_j}} z^{-1}$ and $\frac{\partial N_k(z)}{\partial \beta_{2_j}} = \frac{\partial \alpha_{1_k}}{\partial \beta_{2_j}} + \frac{\partial \alpha_{2_k}}{\partial \beta_{2_j}} z^{-1}$. Also

$$\frac{\partial \alpha_{1_k}}{\partial \beta_{1_j}} = \frac{1}{2} \frac{\alpha_{2_k}}{\sqrt{a\,b}} \qquad \frac{\partial \alpha_{2_k}}{\partial \beta_{1_j}} = \frac{1}{2} \left(\frac{\alpha_{1_k}}{c} - \frac{\alpha_{1_k}}{\sqrt{a\,b}} \right)$$

$$\frac{\partial \alpha_{1_k}}{\partial \beta_{2_j}} = \frac{1}{2} \frac{\alpha_{1_k}}{\sqrt{a\,b}} \qquad \frac{\partial \alpha_{2_k}}{\partial \beta_{1_j}} = -\frac{1}{2} \left(\frac{\alpha_{2_k}}{c} - \frac{\alpha_{2_k}}{\sqrt{a\,b}} \right)$$

Similar equations can be obtained for $\frac{\partial y'_k(n)}{\partial \beta_{1_j}}$, $\frac{\partial y'_k(n)}{\partial \beta_{2_j}}$, $F_k^{a\prime}(z)$, $F_k^{b\prime}(z)$, $\frac{\partial N'_k(z)}{\partial \beta_{1_j}}$, $\frac{\partial N'_k(z)}{\partial \beta_{2_j}}$ that corresponds to the odd numbered ν_k coefficients.
After some reordering, equations (55) and (56) can be rewritten as

$$\begin{aligned}
\frac{\partial y(n)}{\partial \beta_{1_j}} &= \left(\nu_{2j-1}F_j^{a\prime}(z)u(n) + \nu_{2j}F_j^{a}(z)u(n)\right) - \frac{z^{-1}}{D_j(z)}y(n) \\
&\quad + \frac{z^{-1}}{D_j(z)}\left[\sum_{k=1,k\neq j}^{N/2} w_k(n)\right] \qquad (57) \\
\frac{\partial y(n)}{\partial \beta_{2_j}} &= \left(\nu_{2j-1}F_j^{b\prime}(z)u(n) + \nu_{2j}F_j^{b}(z)u(n)\right) - \frac{z^{-2}}{D_j(z)}y(n) \\
&\quad + \frac{1}{D_j(z)}\left[\sum_{k=1,k\neq j}^{N/2} w_k(n)\right] \qquad (58)
\end{aligned}$$

- Although these equations are not suitable for direct implementation, they are useful for the analysis of possible simplifications in the updating algorithm.

- The third term of the right hand side of both equation (57) and equation (58) represents the main factor of increase in computational complexity, if the exact gradient so obtained is implemented.

- Assuming that $u(n)$ is white noise, the stationary points for MSOE minimization using this algorithm can be written as

$$\langle F_i(z), (H(z) - \hat{H}(z))\rangle = 0 \qquad (59)$$

$$\langle F'_i(z), (H(z) - \hat{H}(z))\rangle = 0 \qquad (60)$$

$$\begin{aligned}
&\langle\left(\nu_{2j-1}F_j^{a\prime}(z)u(n) + \nu_{2j}F_j^{a}(z)u(n)\right) - \frac{z^{-1}}{D_j(z)}\hat{H}(z) \\
&\quad + \frac{z^{-1}}{D_j(z)}\left[\sum_{k=1,k\neq j}^{N/2} w_k(n)\right], (H(z) - \hat{H}(z))\rangle = 0 \qquad (61) \\
&\langle\left(\nu_{2j-1}F_j^{b\prime}(z)u(n) + \nu_{2j}F_j^{b}(z)u(n)\right) - \frac{z^{-2}}{D_j(z)}\hat{H}(z) \\
&\quad + \frac{1}{D_j(z)}\left[\sum_{k=1,k\neq j}^{N/2} w_k(n)\right], (H(z) - \hat{H}(z))\rangle = 0 \qquad (62)
\end{aligned}$$

  for $i, j = 1, ..., N/2$.

- The mapping between the coefficients of the direct form realization and the coefficients of the orthogonal realization is not unique.

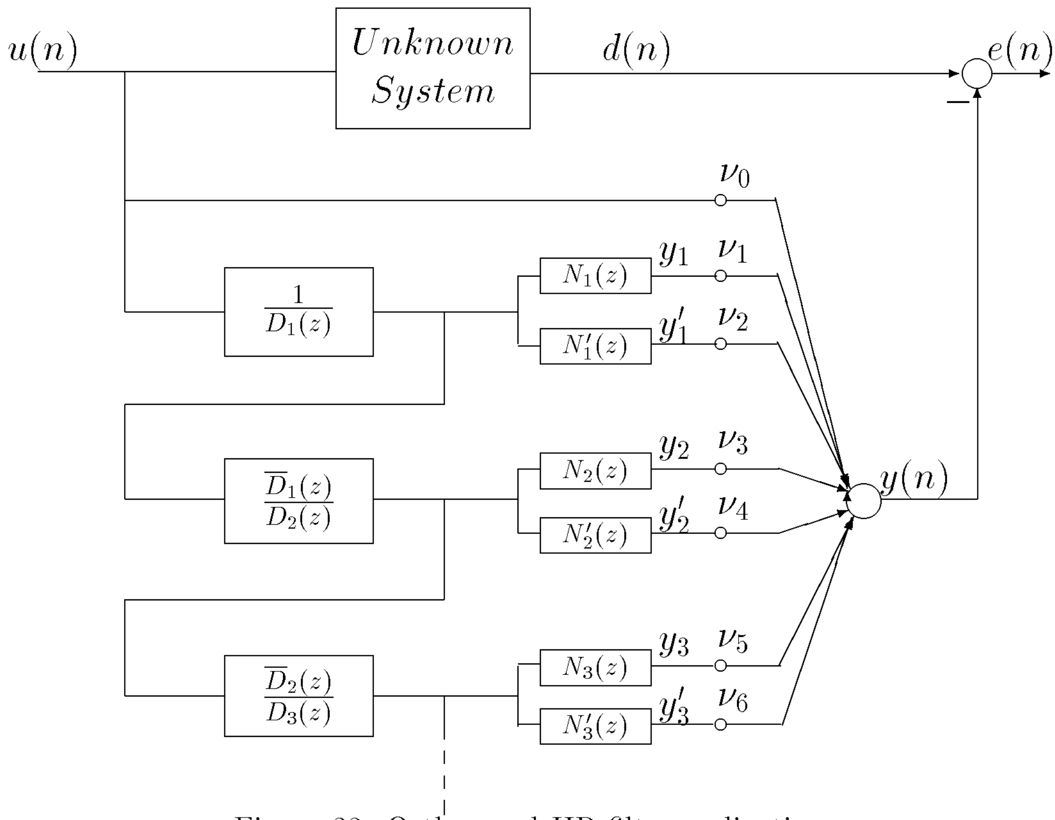

Figure 32: Orthogonal IIR filter realization

5.4 Simplified gradient orthonormal realizations

5.4.1 Lattice form algorithm

Consider first the lattice realization

$$\hat{H}(z) = \sum_{k=0}^{N} \nu_k F_k(z)$$

where $F_k(z) = \frac{\overline{D}_k(z)}{D(z)}$ and $\overline{D}_k(z)$ is the k-order Szego polynomial.

The *ideal* update formula,

$$\begin{aligned}
\theta_k(n+1) - \theta_k(n) &= -\frac{\mu}{2}\frac{\partial E\{e^2(n)\}}{\partial \theta_k} \\
&= \mu E\left\{e(n)\frac{\partial \hat{y}(n)}{\partial \theta_k}\right\} \\
&= \mu \left\langle \frac{\partial \hat{H}(z)}{\partial \theta_k}, \mathcal{S}_x(z)[H(z) - \hat{H}(z)] \right\rangle \\
&= \mu \left\langle \frac{\partial \hat{H}(z)}{\partial \theta_k}, f(z) \right\rangle
\end{aligned}$$

Because only a causal solution is interesting

$$\begin{aligned}
\left\langle \frac{\partial \hat{H}(z)}{\partial \theta_k}, \mathcal{S}_x(z)[H(z) - \hat{H}(z)] \right\rangle &= \left\langle \frac{\partial \hat{H}(z)}{\partial \theta_k}, f(z) \right\rangle \\
&= \left\langle \frac{\partial \hat{H}(z)}{\partial \theta_k}, f_0 + [f(z)]_+ \right\rangle
\end{aligned}$$

A similar conclusion can be obtained for the numerator coefficients ν_k, but in this case

$$\left\langle \frac{\partial \hat{H}(z)}{\partial \nu_k}, \mathcal{S}_x(z)[H(z) - \hat{H}(z)] \right\rangle = \langle F_k(z), f_0 + [f(z)]_+ \rangle$$

A stationary point with respect to ν_k then requires, using the decomposition theorem, that

$$f_0 + [f(z)]_+ = z^{-1}V(z)g(z)$$

where $g(z) \in \mathcal{H}_2$ and $V(z) = \frac{z^{-N}D_N(z^{-1}}{D_N(z)}$. This in particular implies that $f_0 = 0$.

In order to use this expression with the denominator coefficients consider first that

$$\begin{aligned}
\delta_k \hat{H}(z) = \frac{\partial \hat{H}(z)}{\partial \theta_k} &= \sum_{l=0}^{N} \nu_l \delta_k F_l(z) \\
&= \sum_{l=0}^{N} \nu_l \frac{D_N(z)\delta_k \overline{D}_l(z) - \overline{D}_l(z)\delta_k D_N(z)}{D_N^2(z)} \\
&= \sum_{l=0}^{N} \nu_l \frac{\delta_k \overline{D}_l(z)}{D_N(z)} - \hat{H}(z)\frac{\delta_k D_N(z)}{D_N(z)}
\end{aligned}$$

then

$$\begin{aligned}
\left\langle \delta_k \hat{H}(z), z^{-1}V(z)g(z) \right\rangle &= \left\langle \sum_{l=0}^{N} \nu_l \frac{\delta_k \overline{D}_l(z)}{D_N(z)}, z^{-1}V(z)g(z) \right\rangle \\
&\quad - \left\langle \hat{H}(z)\frac{\delta_k D_N(z)}{D_N(z)}, z^{-1}V(z)g(z) \right\rangle
\end{aligned}$$

$\delta_k \overline{D}_l(z)$ is a polynomial of order not exceeding l, exists c_i such that

$$\frac{\delta_k \overline{D}_l(z)}{D_N(z)} = \sum_{i=0}^{l} c_i \frac{\overline{D}_i(z)}{D_N(z)} = \sum_{i=0}^{l} c_i F_i(z)$$

Thus, $\{\frac{\delta_k \overline{D}_l(z)}{D_N(z)}\}_{l=0}^{N}$ can be expressed as linear combinations of $F_0(z), ..., F_N(z)$. But these functions are orthogonal to $z^{-1}V(z)g(z)$. Then the first term of the right hand of previous equation vanish.

Then this suggest the following update formula

$$\theta_k(n+1) - \theta_k(n) \;=\; -\mu e(n)\frac{\delta_k D_N(z)}{D_N(z)}\hat{H}(z)u(n)$$

Because the term $\frac{\delta_k D_N(z)}{D_N(z)}$ still can be simplified, further approximations can be introduced.

Consider the recursive portion of the lattice filter described by

$$\begin{bmatrix} \boldsymbol{x}(n+1) \\ w(n) \end{bmatrix} \;=\; \boldsymbol{Q}(\theta)\begin{bmatrix} \boldsymbol{x}(n) \\ u(n) \end{bmatrix}$$

or in the frequency domain

$$\begin{bmatrix} \frac{\overline{D}_0(z)}{D_N(z)} \\ \frac{\overline{D}_1(z)}{D_N(z)} \\ \vdots \\ \frac{\overline{D}_N(z)}{D_N(z)} \end{bmatrix} \;=\; \boldsymbol{Q}(\theta)\begin{bmatrix} z^{-1}\frac{\overline{D}_0(z)}{D_N(z)} \\ \vdots \\ z^{-1}\frac{\overline{D}_{N-1}(z)}{D_N(z)} \\ 1 \end{bmatrix} \tag{63}$$

Multiplying by $D_N(z)$ and considering a row-wise partition of the orthogonal Hessemberg $\boldsymbol{Q}^T(\theta) = [\boldsymbol{q}_1^T, ..., \boldsymbol{q}_{N+1}^T]^T$ matrix, that verifies

$$\boldsymbol{Q}^T(\theta) \;=\; \begin{bmatrix} \delta_1 \boldsymbol{q}_{N+1}^T(\theta)/\gamma_1 \\ \vdots \\ \delta_N \boldsymbol{q}_{N+1}^T(\theta)/\gamma_N \\ \boldsymbol{q}_{N+1}^T(\theta) \end{bmatrix}$$

where $\gamma_k = \prod_{l=k+1}^{N} \cos\theta_l$, $\gamma_N = 1$.

Then using the bottom row of the equation (63) we can write

$$D_N(z) \;=\; \boldsymbol{q}_{N+1}^T(\theta)\begin{bmatrix} \overline{D}_0(z) \\ \overline{D}_1(z) \\ \vdots \\ \overline{D}_N(z) \end{bmatrix}$$

Then

$$\delta_k D_N(z) = \delta_k \boldsymbol{q}_{N+1}^T(\theta) \begin{bmatrix} \overline{D}_0(z) \\ \overline{D}_1(z) \\ \vdots \\ \overline{D}_N(z) \end{bmatrix} + \boldsymbol{q}_{N+1}^T(\theta) \begin{bmatrix} \delta_k \overline{D}_0(z) \\ \delta_k \overline{D}_1(z) \\ \vdots \\ \delta_k \overline{D}_N(z) \end{bmatrix}$$

neglecting the second term and applying the special structure of the $\boldsymbol{q}_k^T(\theta)$,

$$\begin{aligned} \delta_k D_N(z) &= \delta_k \boldsymbol{q}_{N+1}^T(\theta) \begin{bmatrix} \overline{D}_0(z) \\ \overline{D}_1(z) \\ \vdots \\ \overline{D}_N(z) \end{bmatrix} \\ &= \gamma_k \boldsymbol{q}_k^T(\theta) \begin{bmatrix} \overline{D}_0(z) \\ \overline{D}_1(z) \\ \vdots \\ \overline{D}_N(z) \end{bmatrix} = \gamma_k z^{-1} \overline{D}_{k-1}(z) \end{aligned}$$

This suggest the use of the following approximation

$$\frac{\delta D_N(z)}{D_N(z)} \cong \gamma_k z^{-1} \frac{\overline{D}_{k-1}(z)}{D_N(z)} = \gamma_k z^{-1} F_k(z)$$

that leads to a **simplified partial gradient** lattice algorithm

$$\begin{bmatrix} \nu_0(n+1) \\ \vdots \\ \nu_N(n+1) \\ \theta_1(n+1) \\ \vdots \\ \theta_N(n+1) \end{bmatrix} = \begin{bmatrix} \nu_0(n) \\ \vdots \\ \nu_N(n) \\ \theta_1(n) \\ \vdots \\ \theta_N(n) \end{bmatrix} + \mu e(n) \begin{bmatrix} F_0(z) \\ \vdots \\ F_N(z) \\ -\gamma_1 z^{-1} \hat{H}(z) F_0(z) \\ \vdots \\ -\gamma_N z^{-1} \hat{H}(z) F_{N-1}(z) \end{bmatrix} u(n)$$

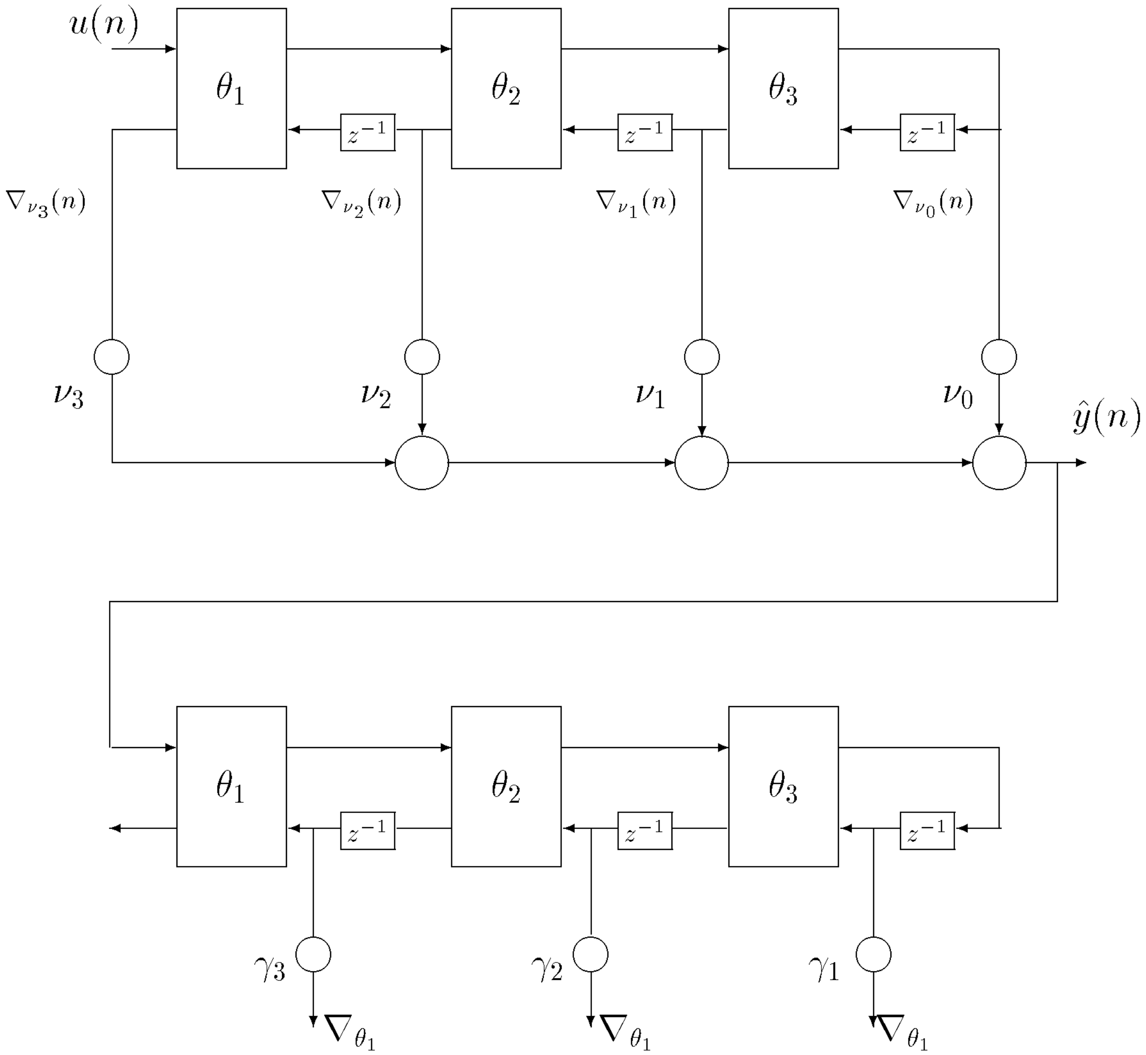

Figure 33: The simplified partial gradient algorithm for a third order recursive lattice filter.

5.4.2 Orthogonal form algorithm

- A simplified *partial gradient* algorithm that makes explicit use of the properties related to the orthogonal IIR filter structure with direct-form second-order sections is discussed.
- Consider the following

  Theorem: The third term of the inner product of equation (61) is equal to zero, i.e.

$$\langle \frac{z^{-1}}{D_j(z)} \left[\sum_{k=1,k\neq j}^{N/2} w_k(n)\right], (H(z) - \hat{H}(z))\rangle = 0 \qquad (64)$$

  for $i, j = 1, ..., N/2$.

- Outline of the proof: The cascade of second-order orthogonal realizations is represented by equation (51). Then, using the decomposition theorem with the stationary points related to β_{1_k} coefficients, we can write

$$\begin{aligned} & \langle \frac{z^{-1}}{D_j(z)} \left[\sum_{k=1,k\neq j}^{N/2} w_k(n)\right], (H(z) - \hat{H}(z))\rangle \\ = \quad & \langle \frac{z^{-1}}{D_j(z)} \left[\sum_{k=1,k\neq j}^{N/2} w_k(n)\right], V(z)g(z)\rangle \end{aligned}$$

  where $V(z) = V_{N/2}(z) = \prod_{k=1}^{N/2} \frac{\overline{D}_k(z)}{D_k(z)}$ and $g(z) \in \mathcal{H}_2$. Then

$$\begin{aligned} & \langle \left[\sum_{k=1,k\neq j}^{N/2} w_k(n)\right], \frac{z}{D_j(z^{-1})} \left[\prod_{k=1}^{N/2} \frac{z^{-2}D_k(z^{-1})}{D_k(z)}\right] g(z)\rangle \\ & = \langle \left[\sum_{k=1,k\neq j}^{N/2} w_k(n)\right], V_{N/2-1}(z)g'(z)\rangle = 0 \end{aligned}$$

  where $g'(z) = \frac{z^{-1}}{D_j(z)}g(z) \in \mathcal{H}_2$. The last equation vanishes because the second equality is a linear combination of the basis functions now $V_{N/2-1}(z)$.

- Similar results can be obtained with the stationary points related to the coefficients β_{2_k}.

Then, the *simplified partial gradient algorithm* can be written as

$$\theta(n+1) = \theta(n) - \mu\nabla'(n) \qquad (65)$$

where

$$\nabla'(n) = -2e(n)[\frac{\partial y(n)}{\partial \nu_0} \cdots \frac{\partial y(n)}{\partial \nu_{N-1}} \frac{\partial y(n)}{\partial \nu_N}$$
$$\frac{\partial y(n)}{\partial \beta_{1_1}} \frac{\partial y(n)}{\partial \beta_{2_1}} \cdots \frac{\partial y(n)}{\partial \beta_{1_{N/2}}} \frac{\partial y(n)}{\partial \beta_{2_{N/2}}}]^T \quad (66)$$

and

$$\frac{\partial y(n)}{\partial \nu_k} = \begin{cases} F_k(n)u(n) & for\ k\ even \\ F_k'(n)u(n) & for\ k\ odd \end{cases} \quad (67)$$

$$\frac{\partial y(n)}{\partial \beta_{1_j}} = \left(\nu_{2j-1} F_j^{a'}(z)u(n) + \nu_{2j} F_j^a(z)u(n)\right) - \frac{z^{-1}}{D_j(z)} y(n) \quad (68)$$

$$\frac{\partial y(n)}{\partial \beta_{2_j}} = \left(\nu_{2j-1} F_j^{b'}(z)u(n) + \nu_{2j} F_j^b(z)u(n)\right) - \frac{z^{-2}}{D_j(z)} y(n) \quad (69)$$

- The result is an efficient orthonormal algorithm with computational complexity similar to the direct form (or lattice) realization.
- Note that, the performance of the proposed algorithm is expected to be close to the full algorithm of equation (52) except for the iterations before the stationary behavior is reached.
- In these iterations the results obtained through Theorem 1 can not be applied. However, the following result is available

  Lemma: The stationary points of the partial gradient algorithm coincide with the stationary points related to the minimization of $E[e^2(n)]$.

5.5 Examples

- 1. $H(z) = \frac{0.05 - 0.4\, z^{-1}}{1 - 1.1314\, z^{-1} + 0.25\, z^{-2}}$
- 2. $H(z) = \frac{1 - 0.25\, z^{-1}}{1 - 0.1\, z^{-1} - 0.42\, z^{-2}}$

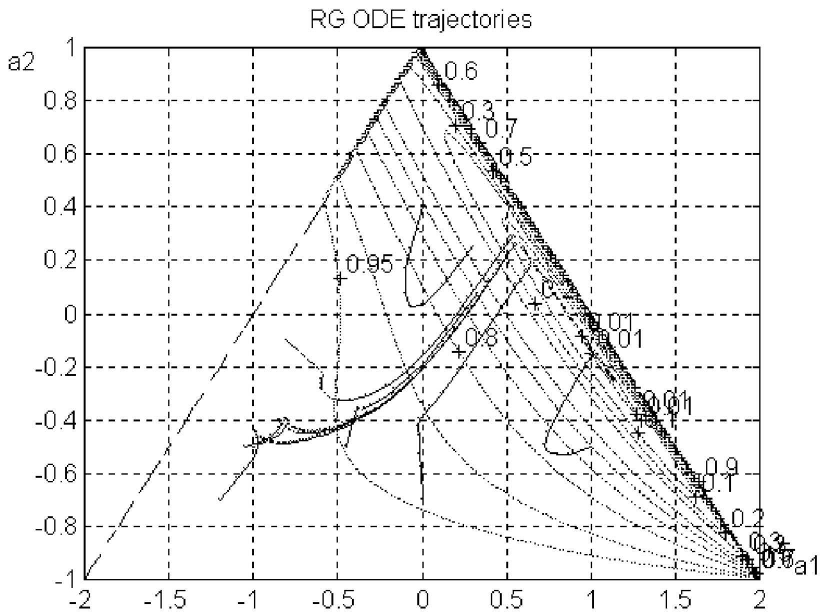

Figure 34: ODE trajectories, Recursive Gradient algorithm, sufficient order. Example 1.

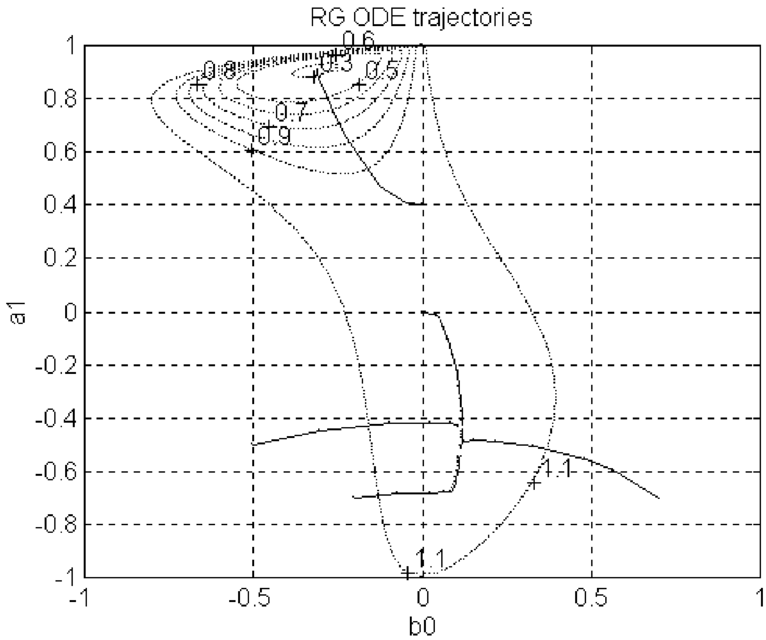

Figure 35: ODE trajectories, Recursive Gradient algorithm, insufficient order. Example 1.

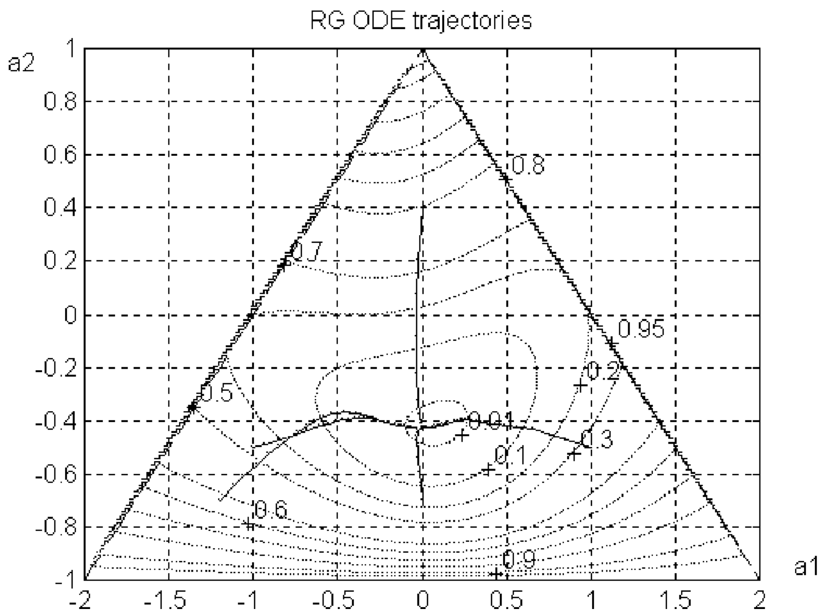

Figure 36: ODE trajectories, Recursive Gradient algorithm, sufficient order Example 2.

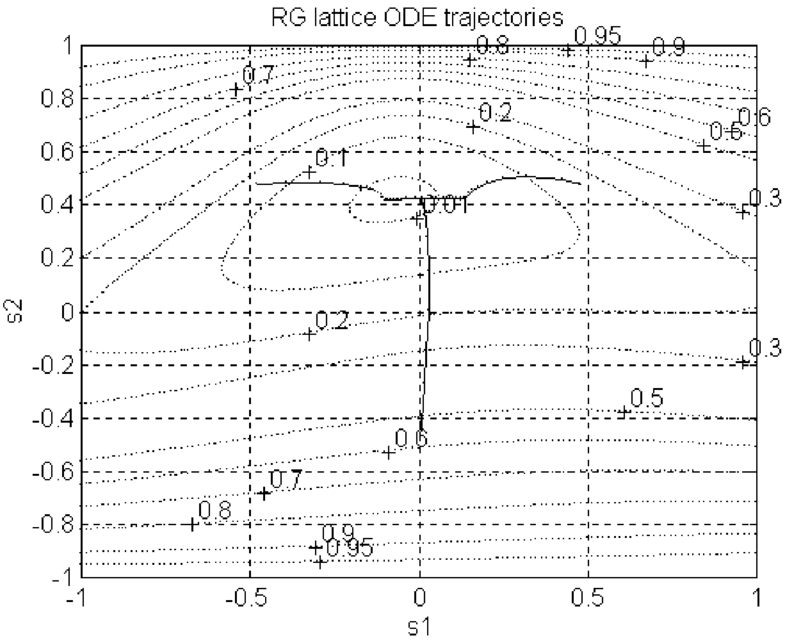

Figure 37: ODE trajectories, Partial Gradient Lattice algorithm, sufficient order. Example 2.

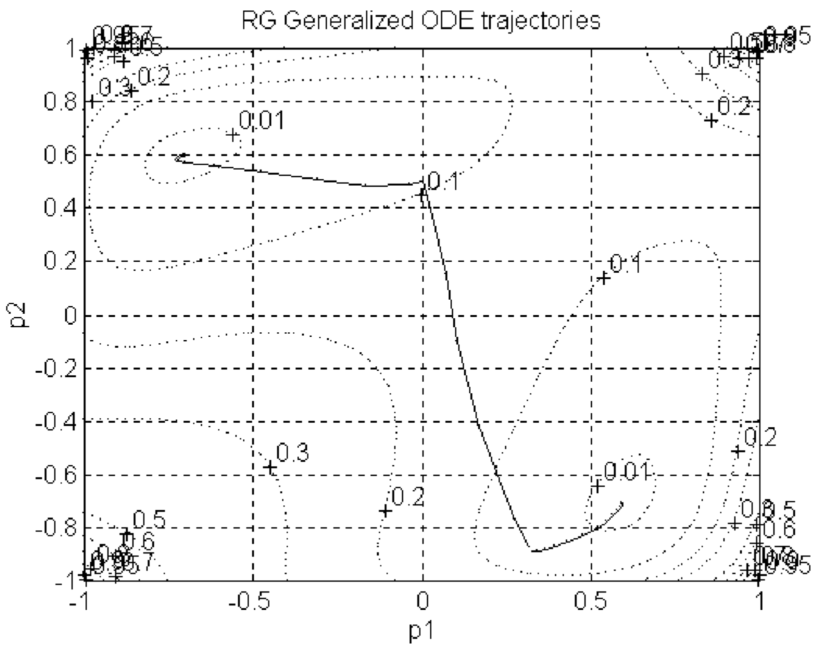

Figure 38: ODE trajectories, Partial Gradient Orthonormal algorithm, sufficient order. Example 2.

6 The equation error perspective

An IIR extension of the FIR adaptive filter

- Stationary points (bias) and ODE associated (stability conditions).
- Unitary norm variant.
- Instrumental variables.
- Bias reduction: BRLE.