ON THE STABILITY AND CONVERGENCE OF FEINTUCH'S ALGORITHM FOR ADAPTIVE IIR FILTERING

MARKUS RUPP AND ALI H. SAYED

Department of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106–9560

Abstract—Gradient-descent adaptive algorithms are among the most widely used in current practice, with many different variants that generally fit into two major groups: one group includes algorithms that are especially suited for FIR (or finite-impulse-response) modeling, while the other group includes algorithms that are tailored for IIR (or infinite-impulse-response) modeling. In the first group, the regression (or data) vectors do not depend on the unknown parameters, which leads to convenient linear models that often facilitate the analysis of the algorithms. In the second category, on the other hand, the regression vectors are dependent on the unknown parameters, thus giving rise to nonlinear functionals and to a richer structure that requires a more thorough analysis. This paper focuses on a widely used adaptive IIR algorithm, the so-called Feintuch's algorithm, and provides a study of its robustness, stability, and convergence properties in a deterministic framework.

I. INTRODUCTION

This paper focuses on a particular adaptive algorithm of gradient-descent type that is often used in IIR modeling, the so-called Feintuch's recursion [1] (see also the survey papers [2,3]). It is well-known that in the IIR context, certain nonlinear functionals arise that lead to error surfaces with possibly multiple local minima, and the main challenge then is to verify whether a given algorithm exhibits global convergence or not (i.e., relative to a global minimum). One of the contributions of the present work is to exploit an intrinsic feedback structure that is present in the nonlinearities in Feintuch's case, along with a so-called small-gain theorem (widely used in system theory), in order to provide convergence and stability analysis for Feintuch's algorithm. Connections with the so-called strictly positive real (SPR) condition that is often cited in the literature will also be clarified.

The following notational conventions will be useful to remember. We shall use small boldface letters to denote vectors and the symbol "$*$" for Hermitian conjugation (complex conjugation for scalars).

This material is based on work supported by the National Science Foundation under Award No. MIP-9409319. A. H. Sayed is an Assistant Professor in the Department of Electrical and Computer Engineering, UCSB, CA 93106. M. Rupp is a postdoctoral fellow. His work is also supported by a scholarship from DAAD (German Academic Exchange Service) as well as the scientific division of NATO (Wissenschaftsausschuß der NATO).

II. FEINTUCH'S ALGORITHM

Consider a linear time-invariant system that is described by a recursive (i.e., pole-zero or IIR) difference equation of the form

$$\begin{aligned} y(i) &= \sum_{k=1}^{M_a} a_k\, y(i-k) + \sum_{k=0}^{M_b-1} b_k\, x(i-k) \qquad (1)\\ &= \begin{bmatrix} \mathbf{y}_{i-1} & \mathbf{x}_i \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \equiv \mathbf{u}_i \mathbf{w}, \end{aligned}$$

where $\mathbf{a} = \text{col}\{a_1, \ldots, a_{M_a}\}$ and $\mathbf{b} = \text{col}\{b_0, \ldots, b_{M_b-1}\}$ are column vectors, while $\mathbf{x}_i = \begin{bmatrix} x(i) & \ldots & x(i-M_b+1) \end{bmatrix}$ and $\mathbf{y}_{i-1} = \begin{bmatrix} y(i-1) & \ldots & y(i-M_a) \end{bmatrix}$. The row vector $\mathbf{u}_i = \begin{bmatrix} \mathbf{y}_{i-1} & \mathbf{x}_i \end{bmatrix}$ is called the *data* vector. Also, $\mathbf{w}$ is a column vector that contains the parameters $\mathbf{a}$ and $\mathbf{b}$, and will be referred to as the *weight* vector. The following widely used compact notation, based on the shift operator q defined by $q^{-1}[x(k)] = x(k-1)$, will also be used in the sequel:

$$y(i) = \frac{B(q^{-1})}{1 - A(q^{-1})}[x(i)], \qquad (2)$$

where $A(q^{-1}) = \sum_{k=1}^{M_a} a_k q^{-k}$, $B(q^{-1}) = \sum_{k=0}^{M_b-1} b_k q^{-k}$, and the ratio $B/(1-A)$ is further assumed irreducible.

The problem of interest is the following: given noisy measurements $\{d(\cdot)\}$ of the output of the system, $\{y(\cdot)\}$, in response to a known input sequence, $\{x(\cdot)\}$, say

$$d(i) = y(i) + v(i) = \mathbf{u}_i \mathbf{w} + v(i), \qquad (3)$$

where $v(i)$ denotes the measurement noise, estimate the system parameters $\mathbf{a}$ and $\mathbf{b}$ (or $\mathbf{w}$) so as to meet a certain optimality criterion. In the literature, this criterion is often motivated in a stochastic setting[1] by defining the error surface (in the output-error formulation) in terms of the variance of the noise mismatch $\{v(\cdot)\}$,

$$J_o(\mathbf{w}) = E\left[\left|d(i) - \frac{B(q^{-1})}{1 - A(q^{-1})}[x(i)]\right|^2\right]. \qquad (4)$$

[1]The stochastic argument is used here as a motivation and for review purposes. We shall shortly drop all stochastic assumptions and approach the problem from a deterministic point of view.

0-7803-2431-5/95 $4.00 © 1995 IEEE

The $J_o(\mathbf{w})$ is clearly a nonlinear functional of the unknown parameters $\{a_k, b_k\}$ and it may thus exhibit several local minima. An approximate solution that seeks to minimize $J_0(\mathbf{w})$ over $\{a_k, b_k\}$, and which is based on instantaneous-gradient ideas, is Feintuch's recursion [1]. It takes the following form: start with initial guesses $\{\hat{y}(-1), \hat{y}(-2), \ldots, \hat{y}(-M_a)\}$ and compute successive estimates $\hat{y}(i)$, for $i \geq 0$, via the recursion:

$$\begin{aligned}\hat{y}(i) &= \sum_{k=1}^{M_a} a_k(i-1)\hat{y}(i-k) + \sum_{k=0}^{M_b-1} b_k(i-1)x(i-k), \\ &= \underbrace{\begin{bmatrix} \hat{\mathbf{y}}_{i-1} & \mathbf{x}_i \end{bmatrix}}_{\hat{\mathbf{u}}_i} \begin{bmatrix} \mathbf{a}_{i-1} \\ \mathbf{b}_{i-1} \end{bmatrix} \equiv \hat{\mathbf{u}}_i \mathbf{w}_{i-1},\end{aligned}$$

where $\{a_k(i-1), b_k(i-1)\}$ denote estimates of $\{a_k, b_k\}$ at time $(i-1)$, and $\hat{\mathbf{y}}_{i-1} = \begin{bmatrix} \hat{y}(i-1) & \ldots & \hat{y}(i-M_a) \end{bmatrix}$. The parameter estimates are, in turn, updated as follows:

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \hat{\mathbf{u}}_i^* [d(i) - \hat{\mathbf{u}}_i \mathbf{w}_{i-1}], \quad (5)$$

with initial guesses $\mathbf{w}_{-1}$ and $\hat{\mathbf{y}}_{-1}$, and where μ is a positive so-called step size.

Although claimed to be a stable algorithm in [1], experimental results in [4] showed that the algorithm could lead to incorrect solutions as well as exhibit an unstable behaviour. One of the contributions of this paper is to provide deterministic conditions on the data, and on the recursive part $(1 - A(q^{-1}))$, in order to guarantee convergence and l_2-stability of the algorithm in a sense precised in Theorem 2 below.

But some aspects of Feintuch's recursion, which will be very relevant in future discussions, should be stressed at this point. First, it is immediate to verify that the $d(i)$ can be re-expressed as (compare with (3) where $\mathbf{u}_i$ is here replaced by $\hat{\mathbf{u}}_i$)

$$d(i) = \hat{\mathbf{u}}_i \mathbf{w} + \hat{v}(i), \quad (6)$$

where the modified noise term $\hat{v}(i)$ is related to the original noise sequence $v(i)$ as follows:

$$\begin{aligned}\hat{v}(i) &= v(i) + A(q^{-1})[y(i) - \hat{y}(i)] \\ &= v(i) + \frac{A(q^{-1})}{1 - A(q^{-1})}[\hat{e}_a(i)], \quad (7)\end{aligned}$$

where $\hat{e}_a(i)$ is defined by $\hat{e}_a(i) = \hat{\mathbf{u}}_i(\mathbf{w} - \mathbf{w}_{i-1}) = \hat{\mathbf{u}}_i \tilde{\mathbf{w}}_{i-1}$.

III. Local Passivity Relations

It will be shown in the sequel that relation (7) leads to an interesting feedback structure whose stability analysis is facilitated. This is due to an existing contraction mapping that is characteristic of gradient recursions of the form (5)-(6), as established in [5] by following a simple Cauchy-Schwarz argument for vectors in an Euclidean space - see also the last section in [6]. The relevant facts are briefly outlined in the sequel. In particular, the next theorem provides a collection of so-called passivity relations that were shown in [5] to hold at every time instant i, for all possible $\hat{v}(i)$ and for all possible initial guesses $\mathbf{w}_{-1} \neq \mathbf{w}$.[2] In the statement of the theorem we use the symbols $\hat{e}_p(i) = \hat{\mathbf{u}}_i(\mathbf{w} - \mathbf{w}_i)$ and $\gamma(i) = [\mu^{-1} - \|\hat{\mathbf{u}}_i\|_2^2]$.

Theorem 1 (Energy Bounds) *The following local energy bounds always hold at each time instant i:*

$$\begin{aligned}\frac{\mu^{-1}\|\mathbf{w} - \mathbf{w}_i\|_2^2 + |\hat{e}_a(i)|^2}{\mu^{-1}\|\mathbf{w} - \mathbf{w}_{i-1}\|_2^2 + |\hat{v}(i)|^2} &\leq 1, \\ \frac{|\hat{e}_a(i)|^2 + |\hat{e}_p(i)|^2}{\mu^{-1}\|\mathbf{w} - \mathbf{w}_{i-1}\|_2^2 + |\hat{v}(i)|^2} &\leq 1, \\ \frac{\gamma(i)\|\mathbf{w} - \mathbf{w}_i\|_2^2 + |\hat{e}_p(i)|^2}{\gamma(i)\|\mathbf{w} - \mathbf{w}_{i-1}\|_2^2 + |\hat{v}(i)|^2} &\leq 1, \\ \frac{|\hat{e}_a(i)|^2 + |\hat{e}_a(i+1)|^2}{\mu^{-1}\|\mathbf{w} - \mathbf{w}_{i-1}\|_2^2 + |\hat{v}(i)|^2} &\leq 1.\end{aligned}$$

where it is assumed that $\mu\|\hat{\mathbf{u}}_i\|_2^2 \leq 1$ for the first three bounds, while $\mu \leq \min\left\{1/\|\hat{\mathbf{u}}_i\|_2^2, 1/\|\hat{\mathbf{u}}_{i+1}\|_2^2\right\}$ for the last bound.

We may add that the above bounds also hold for time-variant step-sizes $\mu(i)$ and that the analysis provided in this paper can be accordingly modified to accommodate this case [5]. Also, the above local bounds show, on a step-by-step basis, how the energies of the apriori and aposteriori residuals, $\hat{e}_a(i)$ and $\hat{e}_p(i)$, compare with the energies of the disturbances due to $\hat{v}(i)$ and to the weight estimation errors, $(\mathbf{w} - \mathbf{w}_{i-1})$ or $(\mathbf{w} - \mathbf{w}_i)$.

IV. A Contraction Mapping

Assume now that we run the gradient recursion (5) from time $i = 0$ up to time N and that $\mu\ \|\hat{\mathbf{u}}_i\|_2^2 \leq 1$ at each time instant i. It then follows that the first inequality in Theorem 1 holds for each $0 \leq i \leq N$,

$$|\hat{e}_a(i)|^2 \leq \mu^{-1}\|\mathbf{w} - \mathbf{w}_{i-1}\|_2^2 - \mu^{-1}\|\mathbf{w} - \mathbf{w}_i\|_2^2 + |\hat{v}(i)|^2 .$$

Summing over i we conclude that we must have (we now use the simplifying notation $\tilde{\mathbf{w}}_i = \mathbf{w} - \mathbf{w}_i$)

$$\frac{\mu^{-1}\|\tilde{\mathbf{w}}_N\|_2^2 + \sum_{i=0}^{N} |\hat{e}_a(i)|^2}{\mu^{-1}\|\tilde{\mathbf{w}}_{-1}\|_2^2 + \sum_{i=0}^{N} |\hat{v}(i)|^2} \leq 1. \quad (8)$$

The numerator of (8) is the sum of the energies of the (modified) apriori residuals $\hat{e}_a(i)$ over $0 \leq i \leq N$, and the energy of the final weight-error at time N. Likewise, the sum in the denominator consists of two terms: the energy of the modified noise signal over the same time interval and the energy of the weight error due to the initial guess. Consequently, (8) establishes a global energy bound over the interval of duration $(N + 1)$: it states that the (block lower triangular) matrix that maps the modified noise signals $\{\hat{v}(i)\}_{i=0}^{N}$ and the initial uncertainty $\mu^{-1/2}\tilde{\mathbf{w}}_{-1}$ to the

[2] The case $\mathbf{w}_{-1} = \mathbf{w}$ (the true weight vector) is excluded so as to avoid having a ratio with both zero numerator and denominator. However, here and in later places in the paper, we can avoid this technicality by simply working all through with differences rather than ratios. For example, the first ratio in Theorem 1 can be rewritten as $(\mu^{-1}\|\mathbf{w} - \mathbf{w}_i\|_2^2 + |\hat{e}_a(i)|^2) - (\mu^{-1}\|\mathbf{w} - \mathbf{w}_{i-1}\|_2^2 + |\hat{v}(i)|^2) \leq 0$.

modified apriori residuals $\{\hat{e}_a(i)\}_{i=0}^N$ and the final weight error $\mu^{-1/2}\tilde{\mathbf{w}}_N$ is always a contraction mapping – see Figure 1 further ahead. This means that the 2–induced norm of this mapping, denoted by $\mathcal{T}_N$, is always upper bounded by one ($\|\mathcal{T}_N\|_{2,ind} \leq 1$) – the 2–induced norm is also written as $\|\mathcal{T}_N\|_\infty$ due to connections with a frequency domain interpretation that often arises in the control context. The algorithm is thus said to be a *robust algorithm*. Alternatively, if we denote by $\Delta_N(\mathbf{w}_{-1}, \hat{v}(\cdot))$ the difference between the numerator and the denominator of (8), then we also conclude that we always have, for any $\mathbf{w}_{-1}$ and $\hat{v}(\cdot)$,

$$\Delta_N(\mathbf{w}_{-1}, \hat{v}(\cdot)) \leq 0. \tag{9}$$

We shall now expand on the significance of the global bounds (8) or (9).

V. Stability Analysis: Feedback Structure

The convergence (to a local or global minimum) analysis of Feintuch's algorithm is still an open problem and has been discussed rather controversially in literature (see, e.g., [4]). A standard statement, based on Popov's hyperstability theorem, is that convergence may follow if the recursive part of the transfer function satisfies a strict positive-real (SPR) condition (see, e.g., [7,8]), viz.,

$$\text{Real}\left\{\frac{1}{1-A(e^{j\omega})}\right\} - \frac{1}{2} > 0 \quad \text{for} \quad 0 \leq \omega \leq 2\pi. \tag{10}$$

This condition has been used successfully for the so-called HARF algorithm [10], which is a gradient-based version of Landau's scheme [7], but is only approximately true for other filter structures, including Feintuch's algorithm, mainly because the regression vectors $\hat{\mathbf{u}}_i$ are constructed differently. In Feintuch's case, predicted estimates $\hat{y}(\cdot)$ are used while in the HARF and Landau schemes filtered estimates are used, as well as a different choice for the step-size parameter. The discussion in this section exploits the contractive relation (8) (or (9)), along with the feedback structure implied by (7), in order to provide an exact condition (cf. (13) further ahead) that will guarantee the l_2–stability and convergence of Feintuch's scheme in the sense defined in Theorem 2.

Motivated by (7), we let $\mathcal{F}_N$ denote the $(N+1)\times(N+1)$ leading triangular operator that describes the action of $A/(1-A)$ over the first $(N+1)$ samples of $\{\hat{e}_a(\cdot)\}$ (in the absence of initial conditions). We can then represent the mapping from $\{\mu^{-1/2}\tilde{\mathbf{w}}_{-1}, v(i)\}_{i=0}^N$ to $\{\mu^{-1/2}\tilde{\mathbf{w}}_N, \hat{e}_a(i)\}_{i=0}^N$ as a feedback cascade with $\mathcal{F}_N$ in the feedback loop as in Figure 1. Note that this is now a mapping from the *original* disturbances $\{v(\cdot)\}$ rather than the modified disturbances $\{\hat{v}(\cdot)\}$. By exploiting the fact that $\mathcal{T}_N$ is a contraction (as shown by (8)), one can establish the following result.[3]

Theorem 2 (l_2–Stability and Convergence) *Consider Feintuch's algorithm (5) and assume $\mu\|\hat{\mathbf{u}}_i\|_2^2 \leq 1$ for every i. Assume further that the original plant, $B/(1-A)$, is stable. If the following condition is satisfied,*

[3] This can also be seen as an immediate consequence of a so-called small gain theorem in system analysis [11].

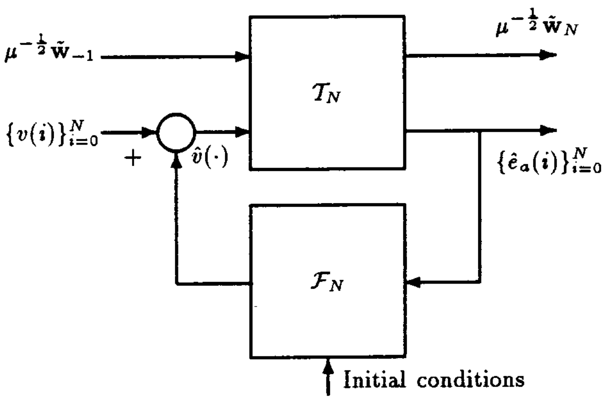

Figure 1: Feedback structure of Feintuch's algorithm.

$$\|\mathcal{T}_N\|_\infty \|\mathcal{F}_N\|_\infty < 1\,, \tag{11}$$

then the mapping from the disturbances $\{\mu^{-\frac{1}{2}}\tilde{\mathbf{w}}_{-1}, v(\cdot)\}$ to $\{\hat{e}_a(\cdot)\}$ is l_2–stable with finite gain in the following sense,

$$\sqrt{\sum_{i=0}^N |\hat{e}_a(i)|^2} \leq \tag{12}$$

$$\frac{\|\mathcal{T}_N\|_\infty}{1-\|\mathcal{T}_N\|_\infty\|\mathcal{F}_N\|_\infty}\left[\mu^{-\frac{1}{2}}\|\tilde{\mathbf{w}}_{-1}\|_2 + \sqrt{\sum_{i=0}^N |v(i)|^2} + \beta'_N\right],$$

where β'_N is a finite positive constant.

In the limit, if the noise sequence $\{v(\cdot)\}$ has finite energy, then the above bound, which also holds as $N\to\infty$, implies that convergence is guaranteed, i.e., $\lim_{i\to\infty}\hat{e}_a(i) = 0$.

Two remarks are due here concerning condition (11) and the convergence result. First, note that since it is already known that $\|\mathcal{T}_N\|_\infty \leq 1$, then a sufficient condition for (11) to hold is to simply require $\|\mathcal{F}_N\|_\infty < 1$ or

$$\max_\omega \left|\frac{A(e^{j\omega})}{1-A(e^{j\omega})}\right| < 1, \quad 0 \leq \omega \leq 2\pi. \tag{13}$$

It is easy to see that for (13) to hold, it is necessary that

$$0 < \text{Real}\left\{\frac{1}{1-A(e^{j\omega})}\right\} < 2.$$

That is, the real part cannot assume arbitrary positive values (compare with the SPR condition (10)).

As for the convergence result, recall that it follows from (7) that

$$y(i) - \hat{y}(i) = \frac{1}{1-A(q^{-1})}[\hat{e}_a(i)].$$

Consequently, if $\hat{e}_a(i)$ converges to zero then we also get $\lim_{i\to\infty}\hat{y}(i) = y(i)$. It also follows from the relation $\hat{e}_a(i) = (1-A(q^{-1}))[y(i)] - (1-A(q^{-1}))[\hat{y}(i)]$, that $\hat{y}(\cdot)$ cannot become arbitrarily unbounded as time progresses since otherwise the $\hat{e}_a(\cdot)$ may assume large values and the bound (12) in the theorem will then be violated. The boundedness of $\hat{y}(\cdot)$ is consistent and, in fact, a consequence of the assumption $\mu\|\hat{\mathbf{u}}_i\|_2^2 \leq 1$. In this regard, the convergence statement in the theorem should be interpreted to mean the following: if the $\hat{y}(\cdot)$ are assumed bounded then, under the conditions stated in the theorem, the sequence $\{\hat{y}(\cdot)\}$ is guaranteed to converge to the sequence $\{y(\cdot)\}$.

The boundedness assumption on $\{\hat{y}(\cdot)\}$ can be relaxed by allowing for a time-variant step-size $\mu(i)$ in (5) – see also [9]. In this case, the requirement

$$\mu \leq \inf_i \frac{1}{\|\hat{\mathbf{u}}_i\|_2^2} ,$$

is replaced by $\mu(i) \leq 1/(\|\hat{\mathbf{u}}_i\|_2^2)$, at every i. It can then be shown that a feedback structure similar to the one in Figure 1 still applies but with a time-variant mapping $\mathcal{F}_N$ in the feedback loop. The mapping will be time-variant since it will depend not only on $A/(1-A)$ but also on the time-variant step sizes $\{\mu(\cdot)\}$. A sufficient condition for stability would also require $\mathcal{F}_N$ to be contractive. These extensions will be addressed elsewhere.

Note also that our derivation is based on the assumption that the signals are deterministic, i.e. all kinds of signals are allowed. In a stochastic setting, the signals are often restricted to a certain class, for example white Gaussian random processes. For this case, it is sufficient to require that[4]

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{A(e^{j\omega})}{1-A(e^{j\omega})}\right|^2 d\omega < 1.$$

VI. Minimax Performance of Feintuch's Algorithm

The global property (8) is valid for any initial guess $\mathbf{w}_{-1}$ and for any noise sequence $\hat{v}(\cdot)$, as long as μ is properly bounded (and assuming a non-zero denominator). One might then wonder whether the bound in (8) is tight or not. That is, are there disturbances $\{\mathbf{w}_{-1}, \hat{v}(\cdot)\}$ for which the ratio can be made arbitrarily close to one? The answer is positive. To clarify this, we follow [5] and rewrite (5) into the alternative form $\mathbf{w}_i = \mathbf{w}_{i-1} + \mu\hat{\mathbf{u}}_i^* [\hat{e}_a(i) + \hat{v}(i)]$. We can now envision a noise sequence $\hat{v}(i)$[5] that satisfies $\hat{v}(i) = -\hat{e}_a(i)$, at each time instant i. In this case, Feintuch's recursion trivializes to $\mathbf{w}_i = \mathbf{w}_{i-1}$ for all i; thus leading to $\mathbf{w}_N = \mathbf{w}_{-1}$ and the ratio in (8) will be one. This means that the maximum value of the ratio in (8), over the unknowns $\{\mathbf{w}_{-1} \neq \mathbf{w}, \hat{v}(\cdot)\}$, is equal to one,

$$\max_{\{\mathbf{w}_{-1},\hat{v}(\cdot)\}} \left\{ \frac{\mu^{-1}\|\mathbf{w}-\mathbf{w}_N\|_2^2 + \sum_{i=0}^N |\hat{e}_a(i)|^2}{\mu^{-1}\|\mathbf{w}-\mathbf{w}_{-1}\|_2^2 + \sum_{i=0}^N |\hat{v}(i)|^2} \right\} = 1. \quad (14)$$

Alternatively, we also have

$$\max_{\{\mathbf{w}_{-1},\hat{v}(\cdot)\}} \{\Delta_N(\mathbf{w}_{-1}, \hat{v}(\cdot))\} = 0.$$

[4]Here, l_2–stability is defined in terms of a mapping between variances of stochastic quantities. More details on this topic will be provided elsewhere.

[5]Although $\hat{v}(i)$ depends on $\hat{e}_a(i)$, it is always possible to find a $v(i)$ such that $\hat{v}(i) = -\hat{e}_a(i)$, viz., $v(i) = \frac{-1}{(1-A(q^{-1}))}[\hat{e}_a(i)]$.

Another question of interest is the following: how does the gradient recursion (5) compare with other possible causal recursive algorithms.? So let $\mathcal{A}$ denote any such algorithm and assume we perform the following experiment on $\mathcal{A}$. We initialize it with $\mathbf{w}_{-1} = \mathbf{w}$ and define the noise sequence $\hat{v}(i)$ in terms of the resulting (successive) apriori estimation errors as follows: $\hat{v}(i) = -\hat{e}_a(i)$ for $0 \leq i \leq N$. Then it always holds that

$$\sum_{i=0}^N |\hat{v}(i)|^2 \leq \mu^{-1}\|\mathbf{w}-\mathbf{w}_N\|_2^2 + \sum_{i=0}^N |\hat{e}_a(i)|^2 ,$$

no matter what the resulting value of $\mathbf{w}_N$ is. Therefore, this particular choice of initial guess ($\mathbf{w}_{-1} = \mathbf{w}$) and noise sequence $\{\hat{v}(\cdot)\}$ will always result in a difference Δ_N that is nonnegative. This implies that for any causal algorithm it always holds that

$$\max_{\{\mathbf{w}_{-1},\hat{v}(\cdot)\}} \{\Delta_N(\mathbf{w}_{-1}, \hat{v}(\cdot))\} \geq 0.$$

For Feintuch's recursion (5) we were able to show that the maximum has to be exactly zero because the global property (9) already provided us with an inequality in the other direction. We can therefore state that among all causal algorithms, Feintuch's recursion (5) is one that solves the following optimization problem:

$$\min_{Algorithm} \left\{ \max_{\{\mathbf{w}_{-1},\hat{v}(\cdot)\}} \Delta_N(\mathbf{w}_{-1}, \hat{v}(\cdot)) \right\}, \quad (15)$$

and that the optimal value is equal to zero.

References

[1] P. L. Feintuch, "An adaptive recursive LMS filter," *Proc. IEEE*, vol. 64, no. 11, pp. 1622–1624, Nov. 1976.

[2] C. R. Johnson, "Adaptive IIR filtering, current results and open issues," *IEEE Trans. Inform. Theory*, vol. 30, no. 2, pp. 237–250, Mar. 1984.

[3] J. J. Shynk, "Adaptive IIR filtering," *IEEE ASSP Magazine*, vol. 6, no. 2, Apr. 1989.

[4] C. R. Johnson, Jr., and M.G. Larimore, "Comments on and additions to "An adaptive recursive LMS filter"," *Proc. IEEE*, vol. 65, no. 9, pp. 1399–1402, Sep. 1977.

[5] A. H. Sayed and M. Rupp, "On the robustness, convergence, and minimax performance of instantaneous-gradient adaptive filters," *Proc. Asilomar Conference on Signals, Systems, and Computers*, Oct. 1994.

[6] A. H. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Signal Processing Magazine*, vol. 11, no. 3, pp. 18–60, July 1994.

[7] I. D. Landau, "A feedback system approach to adaptive filtering," *IEEE Transactions on Information Theory*, vol. 30, no. 2, pp. 251-262, March 1984.

[8] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*, MIT Press, 1983.

[9] D. A. Schoenwald and P. Kokotovic, "Boundedness conjecture for an output error adaptive algorithm," *International Journal of Adaptive Control and Signal Processing*, vol. 4, pp. 27–47, 1990.

[10] F. Colin, N. Cowan, P. M. Grant et al., *Adaptive Filters*. Englewood Cliffs, NJ: Prentice–Hall, 1985.

[11] M. Vidyasagar. *Nonlinear Systems Analysis*. Prentice Hall, New Jersey, second edition, 1993.