ADAPTIVE PARAMETER ESTIMATION USING INTERIOR POINT OPTIMIZATION TECHNIQUES: CONVERGENCE ANALYSIS

Kaywan H. Afkhamie and Zhi-Quan Luo

Communications Research Laboratory
Department of Electrical and Computer Engineering
McMaster University
Hamilton, Ontario, Canada, L8S 4K1

ABSTRACT

Interior Point Optimization techniques have recently emerged as a new tool for developing parameter estimation algorithms [1, 2]. These algorithms aim to take advantage of the fast convergence properties of interior point methods, to yield, in particular, fast transient performance. In this paper we develop a simple *analytic center* based algorithm, which updates estimates with a constant number of computation (independent of number of samples). The convergence analysis shows that the asymptotic performance of this algorithm matches that of the well-known least squares filter (provided some mild conditions are satisfied). Some numerical simulations are provided to demonstrate the fast transient performance of the interior point algorithm.

1. INTRODUCTION

Consider the following system:

$$y_i = \mathbf{x}_i^T \mathbf{w}_* + v_i \qquad i = 1, 2, \ldots$$

where $y_i \in \Re$ is the output, $\mathbf{x}_i \in \Re^M$ is the input vector, $\mathbf{w}_* \in \Re^M$ is the unknown parameter vector and $v_i \in \Re$ is additive measurement noise. The parameter estimation problem is to identify the parameter vector $\mathbf{w}_*$ from input/output pairs $\{\mathbf{x}_i, y_i\}$. The same task is also referred to as adaptive system identification and has applications in adaptive filtering [3].

The error measure used most commonly to evaluate the performance of a particular parameter estimate $\mathbf{w}$ is the mean-squared error:

$$\mathcal{F}_n(\mathbf{w}) = \sum_{i=1}^{n} \lambda^{n-i} [y_i - \mathbf{x}_i^T \mathbf{w}]^2$$

where λ is the forgetting factor typically included to allow the estimation algorithm to adapt to slow changes in the parameter vector $\mathbf{w}$. It will be helpful to rewrite $\mathcal{F}_n(\mathbf{w})$ in vector notation,

$$\mathcal{F}_n(\mathbf{w}) = \mathbf{y}_n^T \Lambda \mathbf{y}_n - 2\mathbf{w}_n^T \mathbf{p}_{xy}(n) + \mathbf{w}_n^T \mathbf{R}_{xx}(n) \mathbf{w}_n$$

where

$$\begin{aligned} \mathbf{y}_n &= [y_1, y_2, \ldots, y_n]^T, \\ \Lambda &= diag(\sqrt{\lambda^{n-1}}, \ldots, \sqrt{\lambda}, 1), \\ \mathbf{p}_{xy}(n) &= \sum_{i=1}^{n} \lambda^{n-i} \mathbf{x}_i y_i, \\ \mathbf{R}_{xx}(n) &= \sum_{i=1}^{n} \lambda^{n-i} \mathbf{x}_i \mathbf{x}_i^T. \end{aligned}$$

$\mathbf{R}_{xx}(n)$ and $\mathbf{p}_{xy}(n)$ are the input signal auto-correlation matrix and the cross-correlation vector of desired and input signal, respectively (both are exponentially weighted). The least-squares solution is obtained by solving $\nabla \mathcal{F}_n(\mathbf{w}) = 0$, which yields

$$\hat{\mathbf{w}}_{LS}(n) = \mathbf{R}_{xx}^{-1}(n) \mathbf{p}_{xy}(n). \tag{1.1}$$

$\hat{\mathbf{w}}_{LS}(n)$ sometimes suffers from poor transient performance due to ill-conditioning of $\mathbf{R}_{xx}(n)$ (caused, for example, from a lack of observation data).

In Section 2 we develop an adaptive estimation algorithm using concepts from interior point optimization. In contrast to other algorithms based on similar ideas [2] our algorithm has a computational complexity that does not increase with the number of samples. In Section 3 we show that the asymptotic performance of our algorithm matches that of the least-squares estimator. Finally, we include some numerical simulations to demonstrate the potential of our algorithm, especially with regard to its transient performance.

2. ALGORITHM IPM

In the optimization theory literature interior point algorithms are often applied to the convex feasibility problem: to find

a feasible point y in a convex set Γ starting from an initial bounded search region $\Omega_0 \supset \Gamma$. At each iteration the following steps are performed

1. Check feasibility of current estimate y_{n-1}.
2. If infeasible, add a new convex cut to obtain Ω_n (note that we maintain $\Gamma \subset \Omega_n \subset \Omega_{n-1}$).
3. Update y_{n-1} to the new "center" y_n of Ω_n.

Since the sequence of sets Ω_n are constantly shrinking, eventually the center y_n will satisfy $y_n \in \Gamma$. We shall give details on how to modify the feasibility set Ω_n and on how to determine its "center" as we apply the interior point algorithm to the parameter estimation problem.

To apply the generic interior point method described above to our parameter estimation problem, we first reformulate the problem in the context of convex feasibility. At each iteration n we look for a filter $\mathbf{w}_n$ at the center of

$$\Omega_n = \{\mathbf{w} \in \Re^M \mid \mathcal{F}_n(\mathbf{w}) \leq \tau_n^2, \|\mathbf{w}\|^2 \leq R^2\}.$$

Note that τ_n is typically selected in such a way that our initial point $\mathbf{w}_{n-1}$ remains in the interior of the search region. But since $\mathbf{w}_{n-1}$ might still be close to the boundary we try to re-center deeper into the interior of Ω_n. To update to a new estimate $\mathbf{w}_n$ we define the *logarithmic barrier function* of Ω_n as

$$\phi_n(\mathbf{w}) = -\log(\tau_n^2 - \mathcal{F}_n(\mathbf{w})) - \log(R^2 - \|\mathbf{w}\|^2).$$

The function $\phi_n(\mathbf{w})$ is convex and approaches infinity on the boundary of Ω_n. The (unique) global minimizer of $\phi_n(\mathbf{w})$ is called the *analytic center* of Ω_n. The gradient and Hessian of $\phi_n(\mathbf{w})$ are given by

$$\nabla\phi_n(\mathbf{w}) = \frac{\nabla\mathcal{F}_n}{s_1} + \frac{2\mathbf{w}}{s_2}, \tag{2.2}$$

$$\nabla^2\phi_n(\mathbf{w}) = \frac{(\nabla\mathcal{F}_n)^T\nabla\mathcal{F}_n}{s_1^2} + \frac{\nabla^2\mathcal{F}_n}{s_1} + \frac{4\mathbf{w}^T\mathbf{w}}{s_2^2} + \frac{2\mathbf{I}}{s_2} \tag{2.3}$$

where s_1 and s_2 are slack variables defined by $s_1 := \tau_n^2 - \mathcal{F}_n(\mathbf{w})$, and $s_2 := R^2 - \|\mathbf{w}\|^2$. The gradient and Hessian of $\mathcal{F}_n(\mathbf{w})$ are given by

$$\begin{aligned} \nabla\mathcal{F}_n &= -2\mathbf{p}_{xd}(n) + 2\mathbf{R}_{xx}(n)\mathbf{w}, \\ \nabla^2\mathcal{F}_n &= 2\mathbf{R}_{xx}(n). \end{aligned}$$

We summarize the steps of Algorithm IPM1 as follows:

Step 1: Initialization. Let $\beta > 0$, $R > 0$ be given. Set

$$\mathbf{w}_0 = \mathbf{0}, \quad \mathbf{p}_{xy}(0) = \mathbf{0}$$
$$\mathbf{R}_{xx}(0) = \mathbf{0}, \quad \nabla\mathcal{F}_0(\mathbf{0}) = \mathbf{0}$$

Step 2: Updating. Acquire new data $\mathbf{x}_n, y_n$. Then recursively update

$$\begin{aligned} \mathbf{p}_{xy}(n) &= \lambda\mathbf{p}_{xy}(n-1) + \mathbf{x}_n y_n, \\ \mathbf{R}_{xx}(n) &= \lambda\mathbf{R}_{xx}(n-1) + \mathbf{x}_n\mathbf{x}_n^T. \end{aligned}$$

Then

- update $\nabla\mathcal{F}_n(\mathbf{w}_{n-1}) = -2\mathbf{p}_{xd}(n) + 2\mathbf{R}_{xx}(n)\mathbf{w}_{n-1}$
- let

$$r_n = \max\left\{\frac{1}{n}, \sqrt{\nabla\mathcal{F}_n^T(\mathbf{w}_{n-1})(\nabla^2\phi_{n-1}(\mathbf{w}_{n-1})^{-1}\nabla\mathcal{F}_n(\mathbf{w}_{n-1})}\right\}$$

$$\begin{aligned} s_1(n) &= \tau_n - \mathcal{F}_n(\mathbf{w}_{n-1}) \\ &= \beta r_n \end{aligned}$$

- $\nabla\phi_n(\mathbf{w}_{n-1}), \nabla^2\phi_n(\mathbf{w}_{n-1})$ using Eqs. (2.2), (2.3)

Step 3: Centering. For $n > 0$, perform Newton iterations starting from $\mathbf{w}_{n-1}$:

$$\mathbf{w} := \mathbf{w} - (\nabla^2\phi_n(\mathbf{w}))^{-1}\nabla\phi_n(\mathbf{w}) \tag{2.4}$$

until ϕ_n is (approximately) minimized. □

Some remarks are in order. The choice of constant parameters β and R will be discussed more in Section 3. In Step 2 the update of the slack variable implies the following update of the threshold τ_n,

$$\tau_n = \mathcal{F}_n(\mathbf{w}_{n-1}) + \beta r_n$$

We have chosen to update $s_1(n)$ directly, to save ourselves the computation of $\mathcal{F}_n(\mathbf{w}_{n-1})$. If more than one Newton iterations are needed in Step 3, we have to re-evaluate $\nabla\phi_n(\mathbf{w})$, $\nabla^2\phi_n(\mathbf{w})$ at the new iterate. While we can show theoretically, that a constant number of Newton iterations are sufficient to find an approximate analytic center of ϕ_n, in practice taking just one Newton iteration ensures adequate convergence. This results in substantial computational savings since the inversion of the Hessian in (2.4) is computationally the most expensive step in the algorithm. Since $\mathcal{F}_n$, $\nabla\mathcal{F}_n$ and $\nabla^2\mathcal{F}_n$ can all be calculated recursively, the overall complexity per iteration of Algorithm IPM1 is $O(M^3)$, independent of n.

3. CONVERGENCE ANALYSIS

We now analyze the asymptotic convergence of the IPM1 algorithm (its steady state behaviour). Two conditions will be required in the analysis.

Condition 1 Weak Persistent Excitation.
There exist $n_0 > 0, \sigma > 0$ *such that for all* $n > n_0$

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T \geq \sigma \mathbf{I}.$$

Condition 1 will be used in the proof of Lemma 3 to upper bound the Hessian inverse $[\nabla^2 \phi_n(\mathbf{w}_{n-1})]^{-1}$. Below we outline the overall analysis by providing several lemmas that are required for the final convergence theorem. Due to the limited space, proofs will not be included in this short paper.

Lemma 1 *Assume the least-squares solution exists and is bounded,* $\|\hat{\mathbf{w}}_{LS}(n)\| \leq t$. *Let* $R > t$, *then*

$$\mathbf{w}_n = \arg \min_{\mathbf{w} \in \Omega_n} \phi_n(\mathbf{w}),$$

the analytic center of Ω_n *is also bounded, and*

$$\|\mathbf{w}_n\| \leq \|\hat{\mathbf{w}}_{LS}(n)\| \leq t.$$

Lemma 2 *In algorithm IPM,* $\mathbf{w}_{n-1}$ *remains feasible with respect to the updated feasible set* Ω_n.

Lemma 1 is intuitively satisfied since in the Interior Point method we minimize both the mean-squared error as well as the norm of the parameter estimate $\mathbf{w}$. The proof of Lemma 2 is straight forward. One only needs to show that the slacks s_1, s_2 remain positive at $\mathbf{w}_{n-1}$.

Lemma 3 *Suppose Condition 1 holds and the noise* v_i *is i.i.d, then*

$$\langle \nabla \phi_n(\mathbf{w}_{n-1}), [\nabla^2 \phi_n(\mathbf{w}_{n-1})]^{-1} \nabla \phi_n(\mathbf{w}_{n-1}) \rangle^{\frac{1}{2}}$$

$$\leq \frac{1}{\beta} + \frac{\sqrt{2}Rt}{R^2 - t^2}.$$

Lemma 3 is important because it implies that $\mathbf{w}_{n-1}$ is an approximate analytic center of Ω_n. Thus, only a constant number of Newton iterations are required in the re-centering step of IPM 1. Lemma 3 is also instrumental in the proof of Theorem 1.

Theorem 1 *Suppose Condition 1 is satisfied and the measurement noise is i.i.d., then*

$$E(\|\mathbf{w}_* - \mathbf{w}_n\|^2) = O\left(\frac{1}{n}\right),$$

where w_* *is the true weight vector.*

Theorem 1 states that the analytic center $\mathbf{w}_n$ will minimize the mean-squared error cost function. Since it does so at the same rate as the least-squares estimator $\hat{\mathbf{w}}_{LS}(n)$ we can say that their asymptotic performance is equivalent.

4. SIMULATIONS

In this section we present results from two different simulation experiments in which we compare the estimation accuracy of Algorithm IPM to that of a typical least-squares algorithm, RLS. First, the potential difference in transient performance of the algorithms is demonstrated, and in a second experiment we raise some new issues by considering an abruptly changing channel.

For the first experiment, the input vector sequence is generated from a white Gaussian process, the noise sequence is also white Gaussian at a SNR of $30dB$, and the unknown parameter vector is given by

$$\mathbf{w}_* = [0.1 \ -0.21 \ 0.35 \ 0.4 \ -0.6 \ 0.3 \ 0.8 \ -0.95 \ 0.5 \ -0.7]^T.$$

Figure 1 shows the error in the parameter vector

$$\|\hat{\mathbf{w}} - \mathbf{w}_*\|^2$$

for RLS and IPM, averaged over 100 independent Monte Carlo trials. It is verified that after an initial transient phase both RLS and IPM show the same asymptotic behaviour. Until approximately iteration 20, however, IPM has a lower estimation error. The norm constraint in Ω_n does not allow the parameter estimate to fluctuate wildly with incoming data. Parameter settings for the two algorithms are as follows:

IPM: $\beta = 2, R = 100, \mathbf{w}_0 = \mathbf{0}, \lambda = 0.99$
RLS: $\mathbf{R}_{xx}(0) = 10^{-6}\mathbf{I}, \mathbf{w}_0 = \mathbf{0}, \lambda = 0.99$

The transient convergence of a parameter estimation algorithm becomes even more significant when the parameter vector to be identified changes in time. In our second experiment the parameter vector starts off at the same values as before and at iteration 100 it changes to

$$\mathbf{w}_* = [0.02 \ -0.04 \ 0.07 \ 0.08 \ -0.12 \ 0.06 \ 0.16 \ -0.19 \ 0.1 \ -0.14]^T.$$

It is well know that the "plain" RLS algorithm performs poorly after an abrupt change, because at the time of the change the state of $\mathbf{R}_{xx}^{-1}$ corresponds to a poor initialization [4]. The problem is addressed by using a "sliding-window" approach in which all data outside the current window is erased from memory (*e.g.*, [5]). It is interesting to note that implementing a sliding-window version of IPM is straightforward. We simply compute the cross-correlation vector and auto-correlation matrix over a limited time window. For the second experiment SNR is increased to $80dB$ and the length of the sliding window is 20. Otherwise all parameters remain the same. Figure 2 shows the estimation error of the sliding-window versions of RLS and IPM. It can be observed that IPM sustains very fast convergence until it reaches the steady-state error level. The sliding-window RLS on the other hand drops out of the "fast-convergence" mode sooner and thus takes much longer to come to a steady state.

5. CONCLUSIONS

We have presented a novel Interior Point Optimization approach to parameter estimation, and shown its asymptotic convergence to match that of the least-squares solution. Our algorithm compares well to an alternative approach based on similar ideas [2], as it does not require a bounded noise assumption, and its computational complexity does not increase with the number of observations. The main practical advantage of the new algorithm lies in its fast transient performance. Simulation examples were shown to demonstrate the potential of the new method in this respect.

6. REFERENCES

[1] K. Afkhamie, Z.-Q. Luo, and K. M. Wong, "Adaptive System Identification using Interior Point Optimization," in *Proceedings of the 9^{th} IEEE Workshop on Statistical Signal and Array Processing*, (Portland, Oregon), Sep 1998.

[2] E. Bai, M. Fu, R. Tempo, and Y. Ye, "Analytic Center Approach to Parameter Estimation: Convergence Analysis," in *Proceedings of ICASSP 1998*, (Seattle, WA), May 1998.

[3] S. Haykin, *Adaptive Filter Theory*. NJ: Prentice Hall, third ed., 1998.

[4] G. V. Moustakides, "Study of the transient phase of the forgetting factor RLS," *IEEE Transactions on Signal Processing*, vol. 45, pp. 2468–2476, October 1997.

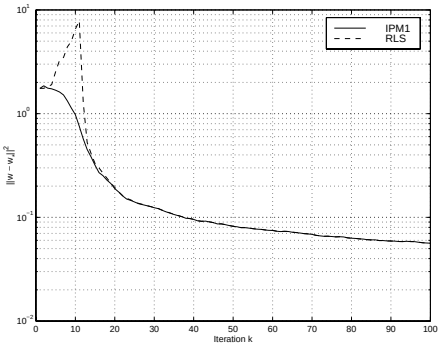

Figure 1: Example 1: Transient performance in parameter estimation

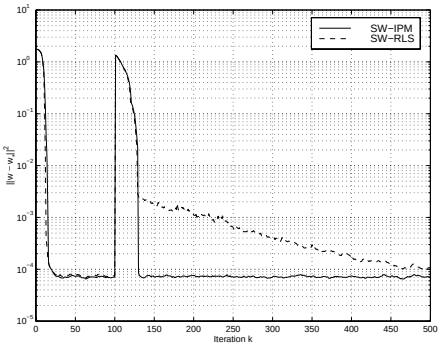

Figure 2: Example 2: Changing Channel characteristics at Iteration 100

[5] H. Liu and Z. He, "A sliding-exponential window RLS adaptive filtering algorithm: Properties and Applications," *Signal Processing*, vol. 45, pp. 357–368, 1995.