ON THE CONSISTENCY OF ℓ_1 -NORM BASED AR PARAMETERS ESTIMATION IN A SPARSE MULTIPATH ENVIRONMENT

Arie Yeredor

School of Electrical Engineering, Tel-Aviv University P.O.Box 39040, Tel-Aviv 69978, Israel arie@eng.tau.ac.il

ABSTRACT

When an autoregressive (AR) process is observed through a sparse multipath environment, its AR parameters may be estimated by searching for a symmetric Finite Impulse Response (FIR) filter, which, when convolved with the observed signal's autocorrelation sequence, yields the sparsest output. The zeros of that filter would then correspond to the poles of the AR process. When the ℓ_0 -norm of the output is used as a measure of its sparsity, consistency of the resulting estimate (under some simple conditions) is readily obtained. However, due to problematic aspects of ℓ_0 norm minimization, it is often more convenient to resort to ℓ_1 -norm minimization. A question of major interest in this context is whether (and if so, under what conditions) consistency of the resulting estimate is maintained. By analyzing the perturbations of the ℓ_1 -norm about the desired solution, we derive (and illustrate) specific conditions for consistency. We show that when the multipath reflections are sufficiently sparse, consistency is guaranteed for a very wide range of AR parameters and reflection gains.

Index Terms— multipath, sparsity, ℓ_1 minimization, deconvolution, consistency.

1. INTRODUCTION

Exploitation of sparsity-related properties in various signalprocessing contexts has gained considerable attention in recent years. The most natural measure of sparsity of a discrete time-series (or vector) is its ℓ_0 -norm, which counts the number of nonzero elements. Unfortunately, however, the ℓ_0 -norm is rather difficult to work with, due to its severe discontinuity and non-convexity. An appealing alternative to the ℓ_0 -norm is the ℓ_1 -norm (sum of absolute values), which on one hand enjoys continuity and convexity properties, but on the other hand might be somewhat less truthful to the notion of sparsity. Therefore, while minimization of the ℓ_0 -norm (in various contexts) can often be shown to yield a desired solution, it is not always clear whether substitution thereof by the more convenient ℓ_1 -norm minimization is still guaranteed to yield the same desired solution.

It is therefore of interest to try to identify (in the respective contexts) conditions under which the ℓ_1 alternative can be used without compromising the ability to attain the sparsest solution. For example, in the context of over-complete dictionaries (closely related to compressed sensing), Donoho *et al.* (e.g.,[1, 2]) have shown that under certain conditions the ℓ_1 minimization can yield the same desired results as ℓ_0 minimization.

In this paper we consider the spectral estimation of an all-poles autoregressive (AR) process observed through a sparse multipath environment. A possible strategy for estimating the AR parameters of the signal of interest (SOI) can be to find a Finite Impulse Response (FIR) filter (of proper length), which, when convolved with the estimated correlation sequence of the observed signal, yields the sparsest output sequence. Then, the zeros of this FIR filter can serve as estimates of the SOI's poles, whereas the resulting sparse signal (output of the convolution) is associated with the (sparse) autocorrelation sequence of the multipath reflections profile. Such an approach was recently proposed in [3] (under tha acronym SPARE - SParsity-based AR Estimation), where it was also shown that under some nearly trivial conditions, minimization of the ℓ_0 -norm of that convolution can yield a consistent estimate (namely, recovers the exact poles of the SOI if the exact autocorrelation of the observed signal is used). However, as mentioned above, ℓ_0 norm minimization is nearly unfeasible in practice, and it is therefore substituted (in [3]) with ℓ_1 -norm minimization. Although the ℓ_1 minimization was shown (in simulations) to perform well (outperforming competing estimators), conditions for its consistency remained an open question.

To explore conditions for consistency of the ℓ_1 -norm minimization in this problem, we shall analyze the ℓ_1 -norm of the convolution result in the vicinity of the desired solution. We shall show that small-errors in one of the estimated poles inflict two different error-terms on the ℓ_1 -norm, and that the balance between these two terms implies conditions for the consistency of the resulting estimate.

In the following section we shall specify the problem formulation and briefly outline the SPARE approach of [3]. In Section 3, which contains our main result, we shall analyze the relation between errors in the estimated poles and perturbations in the ℓ_1 -norm, and identify (and illustrate) the conditions for consistency.

2. PROBLEM FORMULATION

Let s[n] denote the SOI, an AR process of known order P,

$$s[n] = -\sum_{k=1}^{P} a_k s[n-k] + w[n] \quad , \quad \forall n$$
 (1)

where w[n] is a zero-mean, unit-variance white process ("driving-noise"), and the parameters $a_1, a_2, ..., a_P$ are the unknown AR parameters, such that the polynomial $A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_P z^{-P}$ has all its P roots inside the unit-circle (in the Z-plane).

Assume now that s[n] is received through some sparse, isolated-multipaths environment. The received signal x[n]can be modeled as

$$x[n] = \sum_{m=0}^{M} g_m s[n - \tau_m]$$
(2)

where M denotes the number of multipath reflections (in addition to the direct, main path), $\{g_m\}_{m=0}^M$ denote the unknown path-loss coefficients and $\{\tau_m\}_{m=0}^M$ denote the respective unknown propagation delays¹. Note that this model assumes that all multipath delays are integer multiples of the sampling period (hence all τ_m are integers). This is certainly a simplifying assumption. However, even if the true delays are fractional, this model can still hold, but then each single reflection would give rise to several (actually an infinite number of) delays and "interpolation" coefficients, and the sparsity of the reflections profile would be somewhat weakened. And yet, it was demonstrated in [3] that SPARE works well even in the presence of fractional delays.

It is desired to estimate the AR coefficients a_1, \ldots, a_P (alternatively, the poles of A(z)) from the observed signal $x[n], n = 1, \ldots, N$.

Now let $R_{ss}[\ell] \stackrel{\Delta}{=} E[s[n+\ell]s[n]]$ denote the autocorrelation of s[n]. The Z-transform of $R_{ss}[\ell]$ is the spectrum

$$S_{ss}(z) \stackrel{\triangle}{=} \frac{1}{A(z)A(1/z)}.$$
(3)

Let $H(z) \stackrel{\triangle}{=} A(z)A(1/z)$ denote a symmetric Finite Impulse Response (FIR) filter (of length 2P + 1), whose impulse response is denoted $h[\ell]$. Obviously, the convolution

of $R_{ss}[\ell]$ with $h[\ell]$ results in an impulse (Kronecker's delta function, $\delta[\ell]$). Now consider the autocorrelation $R_{xx}[\ell]$ of the observed signal x[n]. Likewise, the convolution of the same $h[\ell]$ with $R_{xx}[\ell]$ would result in some symmetric sequence $\beta[\ell]$, whose Z-transform is given by $B(z)B^*(1/z^*)$, where B(z) is the Z-transform of the set of coefficients $\{b_\ell\}_{\ell=0}^{T_M}$, related to the multipath coefficients and delays via

$$b_{\ell} = \begin{cases} g_m & \text{if } \exists m \mid \tau_m = \ell \\ 0 & \text{otherwise.} \end{cases}$$
(4)

In other words, the result of the convolution of $h[\ell]$ with the (true) autocorrelation of the observed signal x[n] is the autocorrelation of the multipath channel's impulseresponse, which, under the (often realistic) assumption of a sparse multipath profile, is supposed to be sparse as well.

It was therefore proposed in [3] to search for the symmetric FIR filter $\hat{h}[\ell]$ of length 2P + 1, whose convolution with the (estimated) autocorrelation $\hat{R}_{xx}[\ell]$ of the observed signal yields the sparsest result. Under asymptotic conditions (infinite observation time, weak ergodicity of the SOI), $\hat{R}_{xx}[\ell]$ can be assumed to coincide with the true $R_{xx}[\ell]$. Then, it was shown in [3] that if the ℓ_0 -norm of the convolution result is used as a measure of its sparsity, then under certain conditions on the minimum delay-difference between the multipath reflections, the zeros of the SOI (hence a consistent estimate² thereof can be extracted from the minimizing filter $\hat{h}[\ell]$).

However, as mentioned earlier, the ℓ_0 -norm is very difficult to minimize, due to its severe discontinuity and nonconvexity. Thus, the SPARE algorithm minimizes the ℓ_1 norm of the convolution, rather than its ℓ_0 -norm. To evade a trivial solution, a linear scaling constraint on the coefficients of $\hat{h}[\ell]$ is incorporated. The resulting linearly-constrained convex minimization problem admits a convenient, unique solution (e.g., using [4] or standard linear programming), which was shown to perform well in simulations. And yet, it remained unclear whether (or when) this alternative measure of sparsity indeed provides a consistent estimate.

In the following section we shall analyze the sensitivity of the ℓ_1 -norm to deviations from the true solution, and establish conditions for consistency of the resulting estimate. However, we shall substitute the linear scaling constraint with a somewhat different (nonlinear) constraint. Thus, strictly speaking, the consistency proof will not apply directly to the original version of the SPARE algorithm in [3], but would still provide insight as to conditions under which an ℓ_1 -based objective function can (or cannot) yield a consistent estimate in the context of our problem.

¹Without loss of generality we can assume that $0 = \tau_0 < \tau_1 < \cdots < \tau_M.$

²To be precise, in order to properly guarantee consistency, a thresholdmodified variant of the ℓ_0 -norm has to be used, such that the threshold value vanishes asymptotically with the observation length N.

3. CONDITIONS FOR CONSISTENCY OF THE ℓ_1 -BASED ESTIMATE

As mentioned above, our scaling constraint (on the symmetric FIR filter $\hat{h}[\ell]$) in this analysis will not be linear. We constrain the Z-transform $\hat{H}(z)$ of $\hat{h}[\ell]$ to take the form

$$\hat{H}(z) = \hat{D}(z)\hat{D}(1/z)$$
 , $\hat{D}(z) = \prod_{k=1}^{P} (1 - \hat{d}_k z^{-1})$ (5)

meaning that the multiplicative components $\hat{D}(z)$ and $\hat{D}(1/z)$ of $\hat{H}(z)$ are both *monic* polynomials. Note that this is still merely a scaling constraint, since any symmetric FIR filter (of length 2P + 1) can be represented in such a form up to some scaling.

For simplicity, let us assume first that the SOI is an AR(1) process, namely that $A(z) = 1 - az^{-1}$. Under the above scaling constraint, $\hat{D}(z)$ would simply be given by $\hat{D}(z) = 1 - \hat{d}z^{-1}$ (with some \hat{d}), so that $\hat{H}(z) = -\hat{d}z + (1 + \hat{d}^2) - \hat{d}z^{-1}$. Now, observe the following relation, which can be easily verified:

$$C_s(z) \stackrel{\triangle}{=} \frac{(1 - \hat{d}z^{-1})(1 - \hat{d}z)}{(1 - az^{-1})(1 - az)} = U + \frac{V}{(1 - az^{-1})(1 - az)},$$
(6)

where $U = \hat{d}/a$ and $V = (\hat{d} - a)(\hat{d} - \frac{1}{a})$. Obviously, $C_s(z)$ is the Z-transform of the convolution between the FIR filter $\hat{h}[\ell]$ and the SOI's true correlation $R_{ss}[\ell]$. Assume now, that \hat{d} is in the vicinity of the true solution, namely $\hat{d} = a + \epsilon$, where ϵ is infinitely small. Substituting into U and V and ignoring the term which is quadratic in ϵ , we obtain that

$$C_s(z) \approx 1 + \frac{\epsilon}{a} - \frac{\epsilon}{a} \cdot \frac{1 - a^2}{(1 - az^{-1})(1 - az)}.$$
 (7)

Recognizing that the inverse Z-transform of the term $(1 - a^2)/((1 - az^{-1})(1 - az))$ is the sequence $a^{|\ell|}$, we conclude that the inverse Z-transform of $C_s(z)$ is given by

$$c_s[\ell] \approx \begin{cases} 1 & \ell = 0\\ -\frac{\epsilon}{a}a^{|\ell|} & \ell \neq 0 \end{cases}$$
(8)

(and, of course, as expected, $c_s[\ell] = \delta[\ell]$ when $\epsilon = 0$). In the presence of multipath reflections, the observations' autocorrelation $R_{xx}[\ell]$ will differ from $R_{ss}[\ell]$, and its convolution with $\hat{h}[\ell]$ will not be given by $c_s[\ell]$. The key question here is whether or not this convolution will still obtain its minimal ℓ_1 -norm when $\hat{d} = a$ (namely, when $\epsilon = 0$).

Again, for simplicity of the exposition we temporarily employ a simplifying assumption of a *single* multipath component with arbitrary gain g and (integer, positive) delay τ . Consequently, we have

$$R_{xx}[\ell] = (1+g^2) \cdot R_{ss}[\ell] + g \cdot R_{ss}[\ell-\tau] + g \cdot R_{ss}[\ell+\tau].$$
(9)

The convolution between $\hat{h}[\ell]$ and $R_{xx}[\ell]$, denoted $c_x[\ell]$, is therefore similarly given by

$$c_x[\ell] = (1+g^2) \cdot c_s[\ell] + g \cdot c_s[\ell-\tau] + g \cdot c_s[\ell+\tau].$$
(10)

Combining (8) and (10), we obtain

$$c_{x}[\ell] \approx \begin{cases} 1 + g^{2} - \frac{\epsilon}{a} \cdot 2ga^{\tau} & \ell = 0\\ g - \frac{\epsilon}{a} \cdot ((1 + g^{2})a^{\tau} + ga^{2\tau}) & \ell = \pm \tau\\ -\frac{\epsilon}{a} \cdot ((1 + g^{2})a^{|\ell|} + ga^{|\ell - \tau|} + ga^{|\ell + \tau|}) & \text{o.w.} \end{cases}$$
(11)

We now wish to determine whether or not the ℓ_1 -norm of $c_s[\ell]$ obtains a minimum (with respect to ϵ) at $\epsilon = 0$. To this end, we denote the ℓ_1 -norm as $f(\epsilon)$, and observe that $f(\epsilon)$ is composed of three terms:

$$f(\epsilon) \approx \alpha \cdot |\epsilon| + \beta \cdot \epsilon + \gamma,$$
 (12)

where α is obtained by summing the absolute values of $c_x[\ell]$ over all $\ell \neq 0, \pm \tau$,

$$\alpha = \frac{2}{|a|} \sum_{\substack{\ell=1\\\ell\neq\tau}}^{\infty} |(1+g^2)a^{|\ell|} + ga^{|\ell-\tau|} + ga^{|\ell+\tau|}|, \quad (13)$$

and β and γ are obtained from the absolute values of the terms $c_x[0]$ and $c_x[\pm \tau]$:

$$\beta = -2ga^{\tau - 1} - \operatorname{sign}(g)a^{\tau - 1}(1 + g^2 - ga^{\tau}), \quad (14)$$

$$\gamma = 1 + g^2 + 2|g|. \tag{15}$$

Having obtained the general form of $f(\epsilon)$ (recall that the approximation becomes exact as ϵ tends to zero), we observe that $f(\epsilon)$ would take a (possibly local) minimum value (of γ) at $\epsilon = 0$ if and only if $|\beta| < \alpha$. In Figs.1a,b we show typical shapes of $f(\epsilon)$ vs. ϵ , compared to the approximation (12). Also, the ℓ_2 -norm of $c_s[\ell]$ is shown for reference (illustrating that the consistency is a particular property of the ℓ_1 -norm). In Fig.1a we show a case in which $|\beta| < \alpha$, such that $f(\epsilon)$ indeed obtains a minimum at $\epsilon = 0$, whereas in Fig.1b we show a case in which $|\beta| > \alpha$, such that $\epsilon = 0$ is no longer a minimum. Note that in both cases the ℓ_2 -norm does not take a minimum at $\epsilon = 0$, meaning that minimization thereof would not yield a consistent estimate.

Given the expressions³ (13), (14) for α and β (resp.), it is possible to determine (as a function of a, g and τ) whether or not the ℓ_1 -norm minimization (with the specified scaling constraint) would yield a consistent estimate in a certain setup. In fact, it is possible to identify "consistencyregions" in the a - g plane for each τ , as we illustrate in Fig.2. As could be expected, when the multipath reflection

³The expression (13) for α can be further simplified by obtaining an explicit expression for the infinite part of the summation. We omit the explicit expressions due to lack of space.



Figure 1: The ℓ_1 -norm and its approximation (12) vs. the deviation ϵ of \hat{d} from a. Fig.1a (left) depicts the case $a = 0.6, g = 0.4, \tau = 3$, such that $|\beta| < \alpha$, and the (consistent) minimum is obtained at $\epsilon = 0$. Fig.1b (right) depicts a case in which $|\beta| > \alpha$, causing inconsistency. The ℓ_2 -norm is shown for reference.

is very close ($\tau = 1, 2$), the sparsity of the multipath profile is "questionable", and the resulting estimate is not consistent in a significant part of the a - g plane. However, as the reflection gets farther away, the regions of inconsistency become significantly smaller and practically negligible.

We now turn to relax the simplifying assumptions made earlier. We first address the assumption of a single multipath: Obviously, the same derivation can be extended to the case of several multipath reflections. The only modification would be the addition of "exception" cases in the expression (10) for $c_x[\ell]$. Then, the expressions for α , β and γ would change accordingly, as β and γ would encompass all of the "exception" cases, and α would be left with the summation over all other values of ℓ . While the expressions would become more cumbersome, they would all still depend directly on a, on all g-s and on all τ -s, so it would still be easy to predict, in any given scenario, whether or not the ℓ_1 -based estimation would be consistent, simply based on the relation between $|\beta|$ and α .

Regarding the assumption of an AR(1) SOI: assume first that A(z) has several poles, but that all poles are real-valued. Consider the ℓ_1 -norm of the convolution between $R_{xx}[\ell]$ and $\hat{h}[\ell]$, as $\hat{h}[\ell]$ is perturbed about its desired value $h[\ell]$. Since $h[\ell]$ is constrained by (5), we may equivalently consider perturbation of any one of the zeros of $\hat{D}(z)$, say \hat{d}_k (for some k < P). Obviously, after such perturbation, the convolution result will be exactly the same as if the SOI were an AR(1) process with parameter a_k , and $\hat{h}[\ell]$ were a length-3 (2·1+1) FIR with $\hat{d} = a_k + \epsilon_k$ (ϵ_k denoting the perturbation in \hat{d}_k). Thus, all of our results above would hold, simply by substituting a with a_k . If the consistency condition $|\beta| < \alpha$ is satisfied for each $a_k, k = 1, \ldots, P$, then the true solution is guaranteed to be a minimum of the ℓ_1 norm - at least with respect to each d_k separately (note that in general this does not necessarily mean that it is minimum



Figure 2: Regions of inconsistency in the a - g plane for four values of τ . The colored regions indicate values of a and g for which $\beta | > \alpha$, rendering the ℓ_1 -norm inconsistent. As τ increases (and the multipath profile becomes more truly sparse), these regions become significantly smaller and consistency prevails nearly everywhere. The '+'-s indicate the points corresponding to Figs.1a,b

with respect to all d_k -s together). When A(z) has complexvalued poles (in conjugate pairs), the expressions (13), (14) cannot be readily used, since each poles-pair has to be perturbed together (the FIR must be real-valued). Thus, the expression for $c_s[\ell]$ becomes more complicated (depending on a complex-valued ϵ), and is omitted due to lack of space. Nevertheless, similar conditions for consistency in such cases can be obtained along the same guidelines.

4. CONCLUSION

We addressed the consistency of a sparsity-based approach for AR parameters estimation in a sparse multipath environment, which is based on ℓ_1 -norm minimization of the convolution between the observed correlation and a scaleconstrained FIR filter. Analyzing the sensitivity of the ℓ_1 norm to perturbations in the filter, we identified a condition for its minimum to be obtained at zero perturbation (namely, at the desired solution). Thus, asymptotically, as the observation length tends to infinity, such that the estimated correlation approaches the true correlation, the minimizing solution is obtained at the true value (provided that our consistency conditions are satisfied).

5. REFERENCES

- D.L. Donoho and M. Elad, "Optimally sparse representation from overcomplete dictionaries via l₁-norm minimization," *Proc. Nat. Acad. Sci. USA* 100 (5), pp. 21972202, March 2002.
- [2] Y. Tsaig and D.L. Donoho, "Breakdown of equivalence between the minimal *l*₁-norm solution and the sparsest solution," *Signal Processing* 86(3), pp.533-548, March 2006.
- [3] A. Yeredor, "Multipath mitigation in spectral estimation using l₁ minimization," *Proceedings of EUSIPCO 2008*, 2008.
- [4] R.H. Bartels, A.R. Conn and J.W. Sinclair, "Minimization techniques for piecewise differentiable functions: The ℓ₁ solution to an overdetermined linear system," *SIAM Journal on Numerical Analysis*, 15(2), pp.224-241, Apr. 1978.