Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems – MTNS 2010 • 5–9 July, 2010 • Budapest, Hungary

High resolution analysis via sparsity-inducing techniques: spectral lines in colored noise

Lipeng Ning, Tryphon T. Georgiou and Allen Tannenbaum

***Abstract*— The impact of sparsity-inducing techniques in signal analysis has been recognized for over ten years now and has been the key to a growing literature on the subject– commonly referred to as *compressive sensing*. The purpose of the present work is to explore such sparsity-inducing techniques in the context of system identification. More specifically we consider the problem of separating sinusoids in colored noise while at the same time identifying the dynamics that generate the wide-bandwidth noise-component. Our formalism relies on modeling the data as a superposition of a few unknown sinusoidal signals together with the output of an auto-regressive filter which is driven by white noise. Naturally, since neither the underlying dynamics nor any possible sinusoids present are known, the problem is ill-posed. We seek a sparse selection of sinusoids which together with the auto-regressive component can account for the data-set and, to this end, we propose a suitable modification of sparsity-inducing functionals (a la LASSO/Basis pursuit/etc.) which can generate admissible solutions-their sparsity being determined by tuning parameters.**

I. INTRODUCTION

The significance of sparsity, besides its mathematical convenience and elegance, stems from Occam's Razor and can often be justified on physical grounds. Indeed, the notion of sparsity has in recent years created an alternative paradigm to the more traditional minimal-dimension paradigm which has dominated system theory for the past fifty years. The search for the sparsest solution to an under-determined set of equalities/inequalities is in general daunting, and the emergence of this new paradigm can be traced to the relatively recent discovery that sparse solutions can be effectively computed via suitable ℓ_1-optimization problems, e.g., see [3], [5], [6] and the references therein. Applications abound, from image analysis to coding and signal decomposition. Our interest in the present work is a particular system identification problem which can be dealt with using similar tools.

The problem of separating signals from their linear mixtures is ubiquitous. It arises in radar, speech, imaging, and a host of other applications. In fact, some of the early work in signal analysis going back to the middle part of the twentieth century focused exclusively on separating sinusoids in white noise. This is dealt with by a number of classical techniques which have been continuously evolving (MUSIC, ESPRIT, etc., [18], [12], [13], [14]) as well as by recent ones using the tools of compressive sensing [8], [9], [10], [11], [7]. However, a more typical situation is when the noise is not white and the process is shaped by unknown dynamics. In such cases the performance of traditional techniques is often severely degraded. Thus, our interest is in formulating this particular problem in the context of system identification and utilize sparsity-inducing functionals to identify both a sparse selection of sinusoids from a known "dictionary" together with dynamics that account for time-correlations in the residual signal. Earlier work in system identification using compressive sensing techniques [4], [1], [17] focuses mostly on identifying sparse parameter vectors.

Supported in part by the National Science Foundation and by the Air Force Office of Scientific Research.

Lipeng Ning and Tryphon Georgiou are with the Department of Electrical & Computer Eng., Univ. of Minnesota, Minneapolis, MN 55455, USA.

Allen Tannenbaum is with the School of Electrical and Computer Engineering, Georgia Tech, Atlanta, GA 30332, USA.

II. SYSTEM DYNAMICS AND LINE SPECTRA

Consider a time series $\{y_k : k \in \mathbb{Z}\}$ which consists of a small number of pure sinusoids with additive colored noise. We model the colored noise as the output of an autoregressive (AR) filter with white input. Pure sinusoids can be equally well modeled as added to the input or the output of the filter—the only difference being a relative scaling of their amplitude/phase. The former has the advantage of leading to a convex formulation—thus, we consider the setting in Figure 1. Here, x_k is the periodic component and consists of sinusoids of unknown frequency, w_k represents white noise, and the two together form the input to the filter.

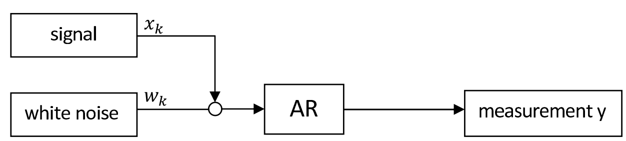

Fig. 1. Model for sinusoids in colored noise.

We seek an rth-order AR-model based on a finite observation record of $\{y_k, k \in \mathbb{Z}\}$. The periodic component is modeled as a sparse linear mixture of columns taken from a suitable "dictionary" matrix B. In our case this matrix is $n \times 2N$ and of the following form:

$$B := [B_{\rm sine}, B_{\rm cosine}],$$

$B_{\rm sine}(\ell, m) = \sin(\pi \ell m/N)$, $B_{\rm cosine}(\ell, m) = \cos(\pi \ell m/N)$, for $\ell = 0, 1, \ldots, n-1$ and $m = 1, 2, \ldots, N$. Thus, the periodic component is taken in the form $B\mathbf{v}$. The entries of $\mathbf{v}$ are labeled according to whether they correspond to sines or cosines:

$$\mathbf{v} := [v_{\sin,1}, ..., v_{\sin,N}, v_{\cos,1}, ..., v_{\cos,N}]^T.$$

Denote by

$$\mathbf{a} := [a_1, a_2, ..., a_r]^T$$

ISBN 978-963-311-370-7

the parameters of the AR-filter. Our modeling equation becomes

$$y_m = \sum_{k=1}^{r} a_k y_{m-k} + x_m + w_m,\ m \in \mathbb{Z}, \tag{1a}$$

or, over the available data-window,

$$\mathbf{y} = H\mathbf{a} + B\mathbf{v} + \mathbf{w} \tag{1b}$$

where $\mathbf{w}$ is the noise-vector,

$$\mathbf{y} := [y_1, y_2, ..., y_n]^T, \text{ and}$$

$$H := \begin{pmatrix} y_0 & y_{-1} & \cdots & y_{-r+1} \\ y_1 & y_0 & \cdots & y_{-r+2} \\ \vdots & & & \vdots \\ y_{n-1} & y_{n-2} & \cdots & y_{n-r} \end{pmatrix}.$$

For a given bound $\delta > 0$ on the energy of the noise, we are interested in determining

$$(\mathbf{a}, \mathbf{v}) = \operatorname{argmin}\{\|\mathbf{v}\|_0 \mid \|\mathbf{y} - H\mathbf{a} - B\mathbf{v}\|_2 \le \delta\}. \tag{2}$$

As usual, $\|\cdot\|_0$ denotes the number of nonzero entries whereas $\|\cdot\|_p$ denotes the p-th norm of a vector ($p \in \{1, 2, \infty\}$). Evidently, (2) is computationally intractable for large-size records. Thus, following the example of compressive sensing, we consider minimization of $\|\mathbf{v}\|_1$ instead. The ℓ_1-norm is in fact a convex relaxation of $\|\cdot\|_0$ and several powerful theorems provide conditions for the solution of the corresponding ℓ_1-minimization to be the sparsest possible. In the sequel we follow precisely such a program. Thus, we consider the following relaxation:

$$(\mathbf{a}, \mathbf{v}) = \operatorname{argmin}\{\|\mathbf{v}\|_1 \mid \|\mathbf{y} - H\mathbf{a} - B\mathbf{v}\|_2 \le \delta\}. \tag{3}$$

It is standard that the minimizer in (3) is also

$$(\mathbf{a}, \mathbf{v}) = \operatorname{argmin}\left\{\|\mathbf{v}\|_1 + \sigma\|\mathbf{y} - H\mathbf{a} - B\mathbf{v}\|_2^2\right\} \tag{4}$$

for a suitable value for $\sigma \ge 0$ which depends on δ. Typically, neither δ nor σ are known and can be thought of as tuning parameters that influence the sparsity of the solution.

It is obvious that as σ becomes large enough, the optimal pair $(\mathbf{a}, \mathbf{v})$ tends to the "least-squares" solution for which $\mathbf{v}$ is generally not sparse. At the other end, when $\sigma = 0^+$, we obtain the trivial solution where the entries of $\mathbf{v}$ are all zero. It turns out that for intermediate values there is a *phase*-like transition in the optimal value for $\|v\|_1$ and $\|v\|_0$ when these change from being zero. This is shown in Figures 2-3. While prior information on the level of noise may be helpful in deciding the sparsity of $\mathbf{v}$, robust plateaus are typical before transition into a higher value for $\|\mathbf{v}\|_0$ and can be used for guidance in choosing the proper weight. The exact value of σ_{critical} where the transition from $\|\mathbf{v}\|_0 = 0$ to $\|\mathbf{v}\|_0 = 1$ can be computed from the data as follows. Define

$$\begin{aligned} \Pi_H &:= H(H^T H)^{-1} H^T, \\ \Pi_H^\perp &:= I - \Pi_H, \\ B_{H^\perp} &:= \Pi_H^\perp B, \text{ and} \\ \mathbf{y}_{H^\perp} &:= \Pi_H^\perp \mathbf{y}. \end{aligned}$$

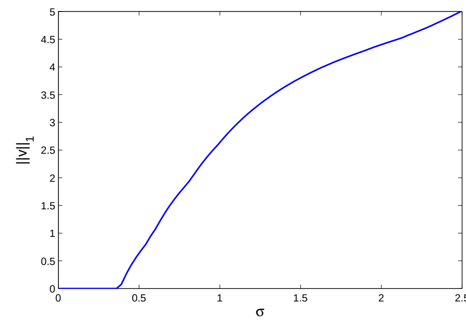

Fig. 2. $\|\mathbf{v}\|_1$ vs. σ

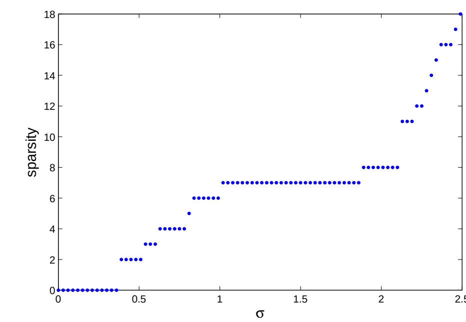

Fig. 3. $\|\mathbf{v}\|_0$ vs. σ

Proposition 1: Let $(\mathbf{a}, \mathbf{v})$ be the minimizer of (4) and

$$\sigma_{\text{critical}} := \frac{1}{2}\|B_{H^\perp}^T \mathbf{y}_{H^\perp}\|_\infty^{-1}.$$

If $\sigma \le \sigma_{\text{critical}}$, then $\|\mathbf{v}\|_0 = 0$ and $\mathbf{a} = (H^T H)^{-1} H^T \mathbf{y}$. If $\sigma > \sigma_{\text{critical}}$, then $\|\mathbf{v}\|_0 \ne 0$.

Proof: Problem (4) can be recast as

$$(\mathbf{a}, \mathbf{v}) = \operatorname{argmin}\left\{h\|\mathbf{v}\|_1 + \frac{1}{2}\|\mathbf{y} - H\mathbf{a} - B\mathbf{v}\|_2^2\right\} \tag{5}$$

for $h = 1/2\sigma$. The quadratic term can be written as

$$\|\mathbf{y} - H\mathbf{a} - B\mathbf{v}\|_2^2 = \|\Pi_H(\mathbf{y} - H\mathbf{a} - B\mathbf{v})\|_2^2 + \|\Pi_H^\perp(\mathbf{y} - B\mathbf{v})\|_2^2.$$

Since $\mathbf{a}$ is unconstrained, the optimal value of a satisfies

$$\Pi_H(\mathbf{y} - H\mathbf{a} - B\mathbf{v}) = 0.$$

Hence $\mathbf{a} = (H^T H)^{-1} H^T(\mathbf{y} - B\mathbf{v})$. Substituting the value of $\mathbf{a}$ above we have that

$$\mathbf{v} = \operatorname{argmin}\left\{h\|\mathbf{v}\|_1 + \frac{1}{2}\|\mathbf{y}_{H^\perp} - B_{H^\perp}\mathbf{v}\|_2^2\right\}. \tag{6}$$

Following [10],

$$\begin{aligned} \mathbf{v} = {} & \operatorname{argmin}\{\|B_{H^\perp}\mathbf{v}\|_2^2 \mid |B_{H^\perp}(\mathbf{y}_{H^\perp} - B_{H^\perp}\mathbf{v})|_i \le h, \\ & \text{for } 0 \le i \le 2N\}. \end{aligned} \tag{7}$$

Thus, if $h \ge \|B_{H^\perp}\mathbf{y}_{H^\perp}\|_\infty$, then $\mathbf{v} = 0$. □

In the next section we attempt to quantify the reliability and accuracy of estimates and motivate weighted optimization by appealing to certain facts related to maximum likelihood estimation.

III. Weighted optimization and maximum likelihood

Suppose that the frequencies of all sinusoidal components of $\{y_k,\ k \in \mathbb{Z}\}$ are known exactly and that $B_{\mathcal{S}}$ is a (full column-rank) matrix formed out of the columns of B corresponding to these known frequencies. It turns out, see [15], that the maximum likelihood estimation of the parameters a, v in the model coincide with the solution of

$$(\mathbf{a}_{\mathrm{ML}}, \mathbf{v}_{\mathrm{ML}}) = \operatorname{argmin} \|\mathbf{y} - H\mathbf{a}_{\mathrm{ML}} - B_{\mathcal{S}}\mathbf{v}_{\mathrm{ML}}\|_2^2. \tag{8}$$

This problem can be replicated in our earlier formalism via weighted optimization. More specifically, consider

$$(\mathbf{a}, \mathbf{v}) = \operatorname{argmin} \left\{ \|W\mathbf{v}\|_1 + \frac{1}{2}\|\mathbf{y} - H\mathbf{a} - B\mathbf{v}\|_2^2 \right\} \tag{9}$$

with $W = \operatorname{diag}(w_1, w_2, \ldots, w_{2N})$. Clearly, if $w_i = 0$ for any index corresponding to the columns of $B_{\mathcal{S}}$ and large otherwise, the solution of these two problems coincide.

Proposition 2: Let $\mathbf{y}$, H, B be as above, $\mathcal{S}$ be the index set of the non-zero entries of $\mathbf{v}_{\mathrm{ML}}$, and let $\mathcal{S}^c$ be its complement. Define matrices $B_{\mathcal{S}}$ and $B_{\mathcal{S}^c}$ containing the corresponding columns of B and define

$$\Pi_{[H\ B_{\mathcal{S}}]} := [H\ B_{\mathcal{S}}]([H\ B_{\mathcal{S}}]^T[H\ B_{\mathcal{S}}])^{-1}[H\ B_{\mathcal{S}}]^T$$

and $\Pi^{\perp}_{[H\ B_{\mathcal{S}}]} := I - \Pi_{[H\ B_{\mathcal{S}}]}$. If

$$\begin{aligned} w_i &= 0 && \text{for } i \in \mathcal{S} \\ w_i &\geq |(B_{\mathcal{S}^c})^T \Pi^{\perp}_{[H\ B_{\mathcal{S}}]}\mathbf{y}|_i && \text{for } i \in \mathcal{S}^c, \end{aligned}$$

then the solutions to (9) and (8) coincide.

Proof: The quadratic term in (9) is

$$\begin{aligned} \|\mathbf{y} - H\mathbf{a} - B\mathbf{v}\|_2^2 \quad = \quad & \|\Pi_{[HB_{\mathcal{S}}]}(\mathbf{y} - H\mathbf{a} - B_{\mathcal{S}}\mathbf{v}_{\mathcal{S}} \\ & -B_{\mathcal{S}^c}\mathbf{v}_{\mathcal{S}^c})\|_2^2 + \|\Pi^{\perp}_{[HB_{\mathcal{S}}]}(\mathbf{y} - B_{\mathcal{S}^c}\mathbf{v}_{cSc})\|_2^2. \end{aligned}$$

Since $\mathbf{a}, \mathbf{v}_{\mathcal{S}}$ are unconstrained, the first term on the right hand side must be zero for the minimizer. Hence,

$$\Pi_{[HB_{\mathcal{S}}]}(\mathbf{y} - H\mathbf{a} - B_{\mathcal{S}}\mathbf{v}_{\mathcal{S}} - B_{\mathcal{S}^c}\mathbf{v}_{\mathcal{S}^c}) = 0 \tag{10}$$

and

$$\mathbf{v}_{\mathcal{S}} = \operatorname{argmin} \left\{ \|W_{\mathcal{S}^c}\mathbf{v}_{\mathcal{S}^c}\|_1 + \frac{1}{2}\|\Pi^{\perp}_{[H\ B_{\mathcal{S}}]}(\mathbf{y} - B_{\mathcal{S}^c}\mathbf{v}_{\mathcal{S}^c}\|_2^2) \right\}.$$

As in the proof in Proposition 1, if

$$w_i \geq |(B_{\mathcal{S}^c})^T \Pi^{\perp}_{[HB_{\mathcal{S}}]}\mathbf{y}|_i \text{ for } i \in \mathcal{S}^c,$$

then $v_{\mathcal{S}^c} = 0$. Finally, (10) is equivalent to

$$\Pi_{[HB_{\mathcal{S}}]}(\mathbf{y} - H\mathbf{a} - B_{\mathcal{S}}v_{\mathcal{S}}) = 0.$$

which implies that $(\mathbf{a}, \mathbf{v}_{\mathcal{S}}) = (\mathbf{a}_{\mathrm{ML}}, \mathbf{v}_{\mathrm{ML}})$ in (8). □

Obviously, in practice, $\mathcal{S}$ is unknown. Thus, below, we suggest how this information may be sought in the data. Denote

$$Q = \operatorname{diag}\{q_1, q_2, \ldots, q_n\} \tag{11}$$

where q_i^2, $i = 1, 2, \ldots, n$ are the diagonal entries of the product $B^T_{H^\perp}B_{H^\perp}$ and $M = \max_{i \neq j} |(Q^{-1}B^T_{H^\perp}B_{H^\perp}Q^{-1})_{i,j}|$. Thus, M is the maximal coherence between the entries of $B_{H^\perp}$.

As will be elaborated in [16], the values $q_i, i = 1, 2, \ldots, 2N$ relate to the spectrum of the input signal at different frequencies and, in fact, the entries of $\mathbf{q} := [\frac{1}{\sqrt{q_j^2 + q_{j+N}^2}},\ j = 1, 2, \ldots, N]$ can be taken as a "pseudo spectrum" (same shape but not necessarily the same integral).

Proposition 3: Let $\mathbf{y}$, H, Q, B be as above, assume that B is square, and assume that there exists a pair $(\mathbf{v}_0, \mathbf{a}_0)$ satisfying

$$\|\mathbf{y} - H\mathbf{a}_0 - B\mathbf{v}_0\|_2 \leq \delta_0$$

for some $\delta_0 > 0$. Consider

$$\tilde{\mathbf{v}} = \operatorname{argmin}\{\|Q\mathbf{v}\|_1 \mid \|\mathbf{y}_{H^\perp} - B_{H^\perp}\mathbf{v}\|_2 \leq \delta\} \tag{12}$$

for $\delta \geq \delta_0$. If the sparsity $S := \|\mathbf{v}\|_0 < \frac{M+1}{4M}$, then

$$\|Q(\tilde{\mathbf{v}} - \mathbf{v}_0)\|_2^2 \leq \frac{(\delta_0 + \delta)^2}{1 + M - 4MS}.$$

For a proof see [16]. The above suggests that if W is chosen proportional to Q, the error between $\mathbf{v}$ in (9) and $\mathbf{v}_0$ is small. In some situations, the exact recovery of the set $\mathcal{S}$ can also be guaranteed [16]. We may also consider how close $\mathbf{a}$ in (9) is to $\mathbf{a}_{\mathrm{ML}}$. Since $\mathbf{a} = (H^TH)^{-1}H^T(\mathbf{y} - B\mathbf{v})$,

$$\begin{aligned} \mathbf{a} &= (H^TH)^{-1}H^T(y - B_{\mathcal{S}}\mathbf{v}_{\mathrm{ML}} + B_{\mathcal{S}}\mathbf{v}_{\mathrm{ML}} - B\mathbf{v}) \\ &= \mathbf{a}_{\mathrm{ML}} + (H^TH)^{-1}H^T(B_{\mathcal{S}}\mathbf{v}_{\mathrm{ML}} - B\mathbf{v}). \end{aligned}$$

Denote by $\mathbf{b}_i$ the ith column of B. It can be seen from the above that $\mathbf{a} - \mathbf{a}_{\mathrm{ML}}$ is insensitive to any difference between $B_{\mathcal{S}}\mathbf{v}_{\mathrm{ML}} - B\mathbf{v}$ corresponding to indices/frequencies where $(H^TH)^{-1}H^T\mathbf{b}_i$ is small. Analyzing this further suggests that this quantity is small at the stop-band of the AR-filter. Hence, error in identifying sinusoidal components within the stop band of the filter will not affect significantly the estimation of the filter dynamics.

The above propositions suggests that an effective *off-the-shelf* choice for W is Q. In the next section, we follow up with an alternative approach which updates W iteratively based on changes in the selection vector $\mathbf{v}$.

IV. Iterative re-weighting

The idea of the following steps originate in [2]. In light of the earlier discussion we begin by letting $W = Q$. We then solve (9) and update W in a way that promotes selection of sinusoids at frequencies where the *signal-to-noise ratio* (SNR) is large. The steps are as follows:

i) Chose $W^{[1]} = Q$ with Q as in (11).

ii) Update $W^{[k]} = \operatorname{diag}\{w_1^{[k]}, \ldots, w_{2N}^{[k]}\}$ using

$$\begin{aligned} w_i^{[k+1]} &= \frac{K}{|v^{[k]}_{\sin,i}| + |v^{[k]}_{\cos,i}| + \tau}, \\ w_{i+N}^{[k+1]} &= w_i^{[k+1]},\ i = 1, ..., N, \end{aligned}$$

for a choice of $\tau, K > 0$. (Here, τ is a "small" constant to prevent singularity of the above expression when the entries of $\mathbf{v}$ are zero whereas K may be adjusted according to the SNR.)

iiii) Terminate when $\|v^{[k+1]} - v^{[k]}\|_1$ is sufficiently small.

For a terminal value for the vector v,

$$\frac{|v_{\sin,i}| + |v_{\cos,i}|}{|v_{\sin,i}| + |v_{\cos,i}| + \tau} = \begin{cases} 1, & |v_{\sin,i}| + |v_{\cos,i}| \gg \tau; \\ 0, & |v_{\sin,i}| + |v_{\cos,i}| \approx 0. \end{cases}$$

Hence, $\|Wv\|_1 \approx (K \times \#\text{signals})$. On the other hand, $|v_{\sin,i}| + |v_{\cos,i}|$ relates to the "local" SNR. Thus, it is reasonable to assign K a value according to the smallest anticipated/detected amplitude of the spectral lines as those are reflected in $|v_{\sin,i}| + |v_{\cos,i}|$. A similar rationale, albeit in a different setting, was suggested in [15].

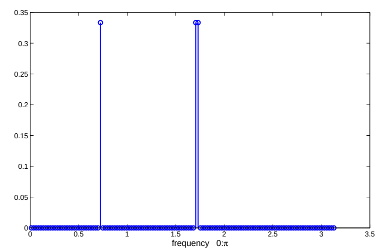

Fig. 4. True line spectra (normalized)

V. Illustrative examples

We highlight the performance of the scheme in the previous section with an example. Consider $B = [B_{\text{sine}},\, B_{\text{cosine}}]$ as before and of size 128×256. We generate data using an AR-filter with transfer function

$$f(z) = \frac{1}{1 - 1.8z^{-1} + 1.3z^{-2} - 0.4z^{-3}}$$

and unit variance white noise at the input along with three sinusoidal components at the following set of frequencies

$$\{30\pi/128,\, 70\pi/128,\, 71\pi/128\}.$$

The corresponding spectral lines and amplitudes are shown in Figure 4. The periodogram is shown in Figure 5 underscoring

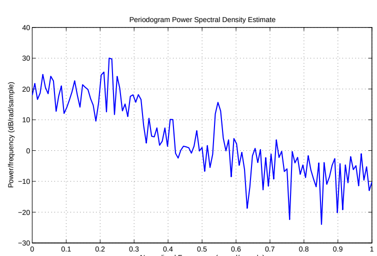

Fig. 5. Periodogram of y

the relatively poor SNR in the generated data. The resolution of the periodogram is insufficient to separate the sinusoids as these are closer than the Fourier limit. Further the presence of colored noise degrades the effectiveness of other traditional techniques [16]. It is interesting to compare with the "pseudo spectrum" $\mathbf{q}$. The power at the location of the sinusoidal components stands out.

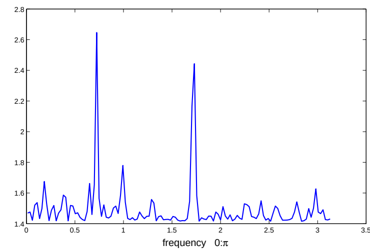

Fig. 6. "Pseudo spectrum" $\mathbf{q}$

We used a third order filter to model the noise-color and the methodology of Section IV to obtain the spectral lines shown in Figure 7. The estimated AR-filter parameters give

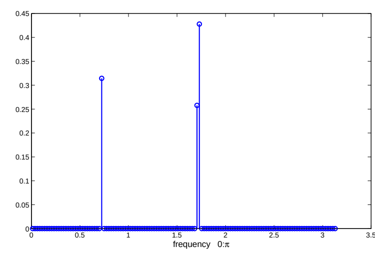

Fig. 7. Estimated line spectra (normalized)

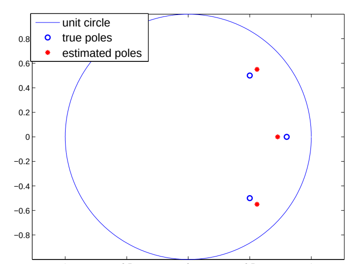

Fig. 8. Poles of true and estimated filters.

the transfer function

$$\hat{f}(z) = \frac{1}{1 - 1.8436z^{-1} + 1.4263z^{-2} - 0.4465z^{-3}}$$

as an approximate system model. Comparison of the poles of the true and estimated filters is shown in Figure 8. Both, the estimated system dynamics and the location of the spectral lines are in excellent agreement with the simulation parameters.

VI. Concluding remarks

We have cast the problem of separating sinusoids signals in colored noise into a compressive sensing setting. Performance guarantees have been obtained under suitable assumptions on the "dictionary" matrix of possible sinusoidal components and insight into possible choices for weighting/tuning parameters is provided. An iterative re-weighting method is discussed. We have found this to be quite effective in practice. Further analysis, both theoretical as well as experimental is needed as the tools of compressive sensing seems especially suited for this type of application.

REFERENCES

[1] M.S. Asif, W. Mantzel and J. Romberg, "Channel protection: random coding meets sparse channels," preprint, August 2009. http://arxiv.org/abs/0908.4265

[2] E. Candès, M. Wakin and S. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *Journal of Fourier Analysis and Applications*, **14(5-6)**: 877-905, December 2008.

[3] E.J. Candès, J. Romberg and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, **52(2)**: 489-509, February 2006.

[4] S.F. Cotter and B.D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Transactions on Communication*, **50(3)**: 374-377, March 2002.

[5] D. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proceedings of the National Academy of Sciences of the United States of America*, **100(5)**: 2197-2202, March 2003.

[6] D.L. Donoho, M. Elad and V.N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Transactions on Information Theory*, **52(1)**: 6-18, January 2006.

[7] M.F. Duarte and R.G. Baraniuk, "Spectral Compressive Sensing," preprint 2010. http://dsp.rice.edu/cs

[8] J.J. Fuchs, "Extension of the Pisarenko method to sparse linear arrays," *IEEE Transactions on Signal Processing*, **45(10)**: 2413-2421, 1997.

[9] J.J. Fuchs, "Recovery of exact sparse representations in the presence of bounded noise," *IEEE Transactions on Information Theory*, **51(10)**: 3601-3608, October 2005.

[10] J.J. Fuchs, "On sparse representation in arbitrary redundant bases," *IEEE Transactions on Information Theory*, **50(6)**: 1341-1344, June 2004.

[11] J.J. Fuchs, "On the use of sparse representation in the identification of line spectra," *17th World Congress IFAC*, pp. 10225-10229, 2008.

[12] T.T. Georgiou, "Signal Estimation via selective harmonic amplification: MUSIC redux," *IEEE Transactions on Signal Processing*, **48(3)**: 780-790, March 2000.

[13] T.T. Georgiou, "Spectral estimation by selective harmonic amplification," *IEEE Transactions on Automatic Control*, **46(1)**: 29-42, January 2001.

[14] T.T. Georgiou, "The Carathéodory-Fejér-Pisarenko decomposition and its multivariable counterpart," *IEEE Transactions on Automatic Control*, **52(2):** 212-228, February 2007.

[15] S.M. Kay, V. Nagesha, "Maximum likelihood estimation of signals in autoregressive noise," *IEEE Transactions on Signal Processing*, **42(1)**: 88-101, January 1994.

[16] L. Ning, T.T. Georgiou and A. Tannenbaum, "Separation of system dynamics and line spectra via sparse representation," preprint.

[17] G. Mileounis, B. Babadi, N. Kalouptsidis and V. Tarokh, "An adaptive greedy algorithm with application to sparse NARMA identification," *IEEE ICASSP 2010*.

[18] P. Stoica and R. Moses, *Introduction to Spectral Analysis*, Prentice Hall, 1997.