ITERATIVE GRAM-SCHMIDT ORTHONORMALIZATION FOR EFFICIENT PARAMETER ESTIMATION

Hongya Ge

Dept. of Electrical and Computer Engineering New Jersey Institute of Technology, Newark, NJ 07102

ABSTRACT

We present an efficient method for estimating nonlinearly entered parameters of a linear signal model corrupted by additive noise. The method uses the Gram-Schmidt orthonormalization procedure in combination with a number of iterations to de-bias and re-balance the coupling between non-orthogonal signal components efficiently. Projection interpretation is provided as rationale of the proposed iterative algorithm. Computer simulations are conducted to show the effectiveness of the algorithm.

1. INTRODUCTION

Many parameter estimation problems can be reduced to the problem of estimating non-linearly entered parameters of a linear signal model corrupted by additive noise. Typical application examples are frequency estimation problem in time series analysis [1 - 4], time delay and direction of arrivals (DoA) estimation problem in array signal processing [6 - 7], and parametric modeling of empirical data [5]. Recent research efforts in this area are mainly focused on two aspects. One is to develop computationally efficient algorithms to estimate non-linear signal parameters, the other is to develop algorithms that provide very low SNR threshold [1 - 7].

This work presents a novel idea of efficiently estimating *non-linear* signal parameters using the Gram-Schmidt (G-S) orthonormalization procedure in combination with a few number of iterations. To demonstrate the main ideas of this work, we choose the problem of estimating two frequencies of sinusoids embedded in white Gaussian noise as our application example. Results of this work can also be applied to delay estimation and DoA estimation [5,7]. Using the projection notation and interpretation of section 4, the proposed algorithm can be generalized straightforwardly to more general case, where more than just two sinusoids are present in the data.

2. NOTATION AND PROBLEM FORMULATION

We assume that the available discrete-time data can be modeled as,

$$y[t] = a_1 e^{j 2 \pi f_1 t} + a_2 e^{j 2 \pi f_2 t} + n[t],$$

$$n = 0, 1, \dots, N - 1.$$
(1)

where a_1 and a_2 are complex amplitudes of the two sinusoids; t is the sample index; f_1 and f_2 are nonlinearly entered signal parameters; n[t] is complex white Gaussian noise with zero mean and variance $2\sigma^2$.

In this work, we assume that a_1 , a_2 , f_1 , f_2 , and σ^2 are all deterministic and unknown. We are mainly interested in estimating frequencies f_1 and f_2 based on N samples of the available data in (1). To facilitate the inner product formulation in the G-S procedure, we rewrite (1) in matrix notation,

$$\mathbf{y} = a_1 \mathbf{s}_1(f_1) + a_2 \mathbf{s}_2(f_2) + \mathbf{n} ,$$

$$= \underbrace{\left[\mathbf{s}_1(f_1) \mathbf{s}_2(f_2) \right]}_{\mathbf{S}(\boldsymbol{\theta})} \cdot \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \mathbf{n} , \qquad (2)$$

with $\mathbf{y} = \begin{bmatrix} y[0] & y[1] & \cdots & y[N-1] \end{bmatrix}^T$; $\boldsymbol{\theta} = \begin{bmatrix} f_1 & f_2 \end{bmatrix}^T$; $\mathbf{s}_i(f_i) = \begin{bmatrix} 1 & e^{j \ 2 \ \pi \ f_i} & \cdots & e^{j \ 2 \ \pi \ f_i \ (N-1)} \end{bmatrix}^T$, (i = 1, 2); and $\mathbf{n} = \begin{bmatrix} n[0] & n[1] & \cdots & n[N-1] \end{bmatrix}^T$.

Note that the matrix notation in (2) is a general model for linear signal in additive noise, which arises in many signal processing applications. Even though the nonlinearly entered parameters f_1 and f_2 are separately contained in signal terms $a_1 \mathbf{s}_1(f_1)$ and $a_2 \mathbf{s}_2(f_2)$, only the data \mathbf{y} that contain the combined signal corrupted by noise are available for estimation purpose. The optimum maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$ corre-

This work was supported in part by the CECOM, the Office of Sponsored Research, NJIT, and the Air Force Office of Scientific Research under contract AFOSR F49620-93-0026.

sponds to the global maximum of the compressed likelihood function (CLF) [1, 6],

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \left\{ L(\boldsymbol{\theta}) \right\},
= \arg \max_{\boldsymbol{\theta}} \left\{ \mathbf{y}^{H} \mathbf{P}_{\mathbf{S}(\boldsymbol{\theta})} \mathbf{y} \right\},$$
(3)

where *H* denotes complex conjugate transpose, and $\mathbf{P}_{\mathbf{S}(\boldsymbol{\theta})} = \mathbf{S}(\boldsymbol{\theta}) (\mathbf{S}^{H}(\boldsymbol{\theta}) \mathbf{S}(\boldsymbol{\theta}))^{-1} \mathbf{S}^{H}(\boldsymbol{\theta})$ is a projection matrix associated with subspace $\langle \mathbf{S}(\boldsymbol{\theta}) \rangle$. The CLF in (3) can be written as [1,6]

where $\beta(f_1, f_2) = \mathbf{s}_1^H(f_1) \, \mathbf{s}_2(f_2) = \sum_{t=0}^{N-1} e^{j \, 2 \, \pi \, (f_2 - f_1) \, t};$ $Y(f_i) = \sum_{t=0}^{N-1} y[t] \, e^{-j \, 2 \, \pi \, f_i \, t}$ is simply the DFT of data \mathbf{y}

evaluated at f_i (i = 1, 2).

In order to reduce the computations in (3), various computationally efficient algorithms, such as the K-T [1], the KiSS/IQML [2,3], the AP [4], and the FML [6], have been proposed. In this work, we proposed a G-S orthonormalization based iterative projection algorithm. The idea of using the G-S orthonormalization in combination with a few number of iterations is to de-bias and re-balance the coupling between the usually *non-orthogonal* signal components, so that the task of searching for the maximum of a multi-dimensional CLF can be simplified into a few iterative 1-D searches.

3. PROPOSED ESTIMATION ALGORITHM

The proposed algorithm provides efficient estimates for both frequencies by combining the G-S orthonormalization with a few iterations. The major steps of the algorithm are summarized as follows,

- Initial estimates (i=1) [8]: $\hat{f}_{1}^{(i)} = \arg \max_{f} \left\{ |Y(f)|^{2} \right\},$ $\hat{f}_{2}^{(i)} = \arg \max_{f} \left\{ W(f - \hat{f}_{1}^{(i)}) \cdot |R(f; \hat{f}_{1}^{(i)})|^{2} \right\},$ (5)
- Refined estimates (for $i = 2, 3, \cdots$):

$$\begin{split} \hat{f}_{1}^{(i)} &= \arg \max_{f} \left\{ W(f - \hat{f}_{2}^{(i-1)}) \cdot \left| R(f; \hat{f}_{2}^{(i-1)}) \right|^{2} \right\} \\ \hat{f}_{2}^{(i)} &= \arg \max_{f} \left\{ W(f - \hat{f}_{1}^{(i)}) \cdot \left| R(f; \hat{f}_{1}^{(i)}) \right|^{2} \right\}, \end{split}$$

$$(6)$$

where W(f) is a weighting function used to de-bias the effect of \hat{f}_1 on estimating f_2 in the initialization stage, and the effect of \hat{f}_2 on estimating f_1 in the refinement stage. It is an even function given by [8],

$$W(f) = \begin{cases} \frac{1}{1 - \frac{1}{N^2} \left| \sum_{n=0}^{N-1} e^{-j 2\pi f n} \right|^2}, & f \neq 0 \\ 0, & f = 0 \end{cases}$$
(7)

Y(f) and $R(f; \hat{f}_j^{(i)})$, (j = 1, 2) are the DFTs of the original data y[t] and the residual data r[t], respectively,

$$r[t] = y[t] - \frac{1}{N} \sum_{m=0}^{N-1} y[m] e^{-j 2\pi \hat{f}_{j}^{(i)}(m-t)}, \qquad (8)$$

$$R(f; \hat{f}_j^{(i)}) = Y(f) - \frac{1}{N} Y(\hat{f}_j^{(i)}) \cdot \beta(f, \hat{f}_j^{(i)}) .$$
(9)

The weighting function W(f) is found to be related to the $\beta(f_1, f_2)$ in (4) through,

$$W(f_1 - f_2) = \frac{1}{1 - \frac{1}{N^2} |\beta(f_1, f_2)|^2}.$$
 (10)

4. PROJECTION INTERPRATATION AND RATIONALE OF PROPOSED ALGORITHM

To understand the above proposed algorithm, we provide the projection interpretation of the algorithm in this section. The orthogonal projection matrix $\mathbf{P}_{\mathbf{S}(\boldsymbol{\theta})}$ in (3) can be decomposed as follows [7],

$$\mathbf{P}_{\mathbf{S}(\boldsymbol{\theta})} = \mathbf{P}_{\mathbf{S}_1} + \mathbf{P}_{\mathbf{P}_{\mathbf{S}_1}^{\perp} \mathbf{S}_2}, \qquad (11)$$

$$=\mathbf{P}_{\mathbf{S}_{2}}+\mathbf{P}_{\mathbf{P}_{\mathbf{S}_{2}}^{\perp}\mathbf{S}_{1}},$$
(12)

where $\mathbf{P}_{\mathbf{s}_{i}} = \mathbf{s}_{i} (\mathbf{s}_{i}^{H} \mathbf{s}_{i})^{-1} \mathbf{s}_{i}^{H}$, (i = 1, 2) is the projection metrix associated with subspace $\langle \mathbf{s}_{i} \rangle$. While $\mathbf{P}_{\mathbf{P}_{\mathbf{s}_{i}}^{\perp} \mathbf{s}_{j}}$, (i, j = 1, 2) is the projection matrix associated with subspace $\langle \mathbf{P}_{\mathbf{s}_{i}}^{\perp} \mathbf{s}_{j} \rangle$, which is the part of $\langle \mathbf{s}_{j} \rangle$ that is unaccounted for by the subspace $\langle \mathbf{s}_{i} \rangle$. In the above simplified notation, we used \mathbf{s}_{1} and \mathbf{s}_{2} to denote $\mathbf{s}_{1}(f_{1})$ and $\mathbf{s}_{2}(f_{2})$, respectively.

Using formulae (11) and (12), the CLF $L(f_1, f_2)$ in equations (3) and (4) can, therefore, be rewritten as,

$$\mathbf{y}^{H} \mathbf{P}_{\mathbf{S}(\boldsymbol{\theta})} \mathbf{y} = \mathbf{y}^{H} \mathbf{P}_{\mathbf{S}_{1}} \mathbf{y} + \mathbf{y}^{H} \mathbf{P}_{\mathbf{P}_{\mathbf{S}_{1}}^{\perp} \mathbf{S}_{2}} \mathbf{y}, \qquad (13)$$

$$= \mathbf{y}^{H} \mathbf{P}_{\mathbf{S}_{2}} \mathbf{y} + \mathbf{y}^{H} \mathbf{P}_{\mathbf{P}_{\mathbf{S}_{2}} \mathbf{S}_{1}} \mathbf{y} .$$
(14)

These two equations provide the rationale of the iterative G-S based estimation algorithm in formulae (5) and (6). To fully understand the observation, we concentrate only on the equation (13). The decomposed two terms in (13) can be precisely written as,

$$\mathbf{y}^{H} \mathbf{P}_{\mathbf{S}_{1}} \mathbf{y} = \frac{1}{N} |\mathbf{s}_{1}^{H} \mathbf{y}|^{2} = \frac{1}{N} |Y(f_{1})|^{2},$$
 (15)

with $L(\hat{f}_1, f) = \frac{|Y(f)|^2 - \frac{2}{N} \mathcal{R}e\{Y^*(\hat{f}_1) \beta(\hat{f}_1, f_2) Y(f)\}}{1 - \frac{1}{N^2} |\beta(\hat{f}_1, f_2)|^2},$ which is different from $W(f - \hat{f}_1) \cdot \left| R(f; \hat{f}_1) \right|^2$ in (17).

$$\mathbf{y}^{H} \mathbf{P}_{\mathbf{P}_{\mathbf{S}_{1}}^{\perp} \mathbf{S}_{2}} \mathbf{y} = \mathbf{y}^{H} \mathbf{P}_{\mathbf{S}_{1}}^{\perp} \mathbf{s}_{2} \left(\mathbf{s}_{2}^{H} \mathbf{P}_{\mathbf{S}_{1}}^{\perp} \mathbf{s}_{2}\right)^{-1} \mathbf{s}_{2}^{H} \mathbf{P}_{\mathbf{S}_{1}}^{\perp} \mathbf{y} = \frac{1}{N - \frac{1}{N} |\beta(f_{1}, f_{2})|^{2}} \cdot \left|\mathbf{s}_{2}^{H} \mathbf{y} - \frac{\mathbf{s}_{2}^{H} \mathbf{s}_{1} \mathbf{s}_{1}^{H} \mathbf{y}}{N}\right|^{2},$$

$$= \frac{1}{N - \frac{1}{N} |\beta(f_{1}, f_{2})|^{2}} \cdot \left\{|Y(f_{2})|^{2} - \frac{2}{N} \mathcal{R}e\left(Y^{*}(f_{1}) \beta(f_{1}, f_{2}) Y(f_{2})\right) + \frac{1}{N^{2}} |\beta(f_{1}, f_{2}) Y(f_{1})|^{2}\right\},$$

$$= \frac{1}{N} \cdot W(f_{1} - f_{2}) \cdot |R(f_{2}; f_{1})|^{2}.$$
(16)
This can be justified by the *equivalent* relation between the two under maximization while \hat{f}_{1} is fixed,

Based on the decomposition of CLF in formulae (15) and (16), we can *approximately* reduce the 2-D maximization in (3) into the following two 1-D maximizations,

$$\hat{f}_{1} = \arg \max_{f} \left\{ |Y(f)|^{2} \right\},$$

$$\hat{f}_{2} = \arg \max_{f} \left\{ W(f - \hat{f}_{1}) \cdot \left| R(f; \hat{f}_{1}) \right|^{2} \right\}.$$
(17)

As a matter of fact, our proposed algorithm uses the results of (17) in its initialization stage (see formula (5)). It can be seen that when $\mathbf{s}_1(f_1)$ and $\mathbf{s}_2(f_2)$ are not orthogonal, (17) yields a biased estimate of f_1 , hence a bisaed estimate of f_2 . However, under the condition that \hat{f}_1 is a unbiased estimate of f_1 , the \hat{f}_2 obtained from the weighted DFT of the residual signal in (17) should also be a unbiased estimate of f_2 . The effect of the weighting function $W(f - \hat{f}_1)$ is to debias [8] the effect of the $\mathbf{s}_1(\hat{f}_1)$ on the estimate \hat{f}_2 obtained from the $R(f; \hat{f}_1)$.

Using this observation in combination with the following forms for the second term in formula (14),

$$\mathbf{y}^{H} \mathbf{P}_{\mathbf{P}_{\mathbf{S}_{2}}^{\perp} \mathbf{S}_{1}} \mathbf{y} = \frac{1}{N} \cdot W(f_{1} - f_{2}) \cdot |R(f_{1}; f_{2})|^{2} , \quad (18)$$

we can refine our initial estimates, f_1 and f_2 , using formulae (17) and (18) iteratively, as proposed in (6). The actual effect of the iteration in (6) is to re-balance and de-bias the coupling between the non-orthogonal signal components, and finally to obtain the MLEs of signal frequencies.

It is noted that the proposed algorithm is different from previously proposed computationally efficient methods, such as the K-T [1], the KiSS/IQML [2, 3], the AP [4], and the FML [6]. For application of sinusoidal parameter estimation, a *different* initialization idea is used prior to the Newton search in [6],

$$\hat{f}_{1} = \arg \max_{f} \left\{ |Y(f)|^{2} \right\},$$

$$\hat{f}_{2} = \arg \max_{f} \left\{ L(\hat{f}_{1}, f) \right\},$$
(19)

$$W(f - \hat{f}_1) \cdot \left| R(f; \hat{f}_1) - \frac{1}{N^2} |\beta(\hat{f}_1, f) Y(\hat{f}_1)|^2 \right|^2 \stackrel{eqv}{=} L(\hat{f}_1, f)$$

5. SIMULATION EXAMPLES

Using the proposed algorithm, simulations have been conducted for different choice of signal strengthes and frequencies. The algorithm provides statistically and computationally *efficient* estimates for both signal frequencies. Simulation results are plotted in figure 1 for the cases of different signal strengthes and frequency spacings. Throughout our simulations, we assume that N = 25 data samples are available, and the noise samples are i.i.d. complex Gaussian $N(0, 2\sigma^2)$. We terminate the iteration in (6) whenever the difference between old estimate and updated estimate of either frequency is within half the width of frequency bin. The width of frequency quantization bin is chosen based on the value of the CR bound at each SNR. 1000 independent trials are run for each point of SNR.

In the figure 1, we plotted the MSE of estimates of both frequencies versus $SNR \stackrel{\Delta}{=} 10 \log_{10}(\frac{1}{2\sigma^2})$ in dB. Note that in calculating threshold SNR for each sinusoidal component, one needs to translate the SNR used in the figure into each signal's SNR using formula $SNR_i \stackrel{\Delta}{=} 10 \log_{10}(\frac{|a_i|^2}{2\sigma^2}) = SNR + 10 \log_{10}(|a_i|^2) dB$. For the case shown in figure 1(a), we have two or-

For the case shown in figure 1(a), we have two orthogonal signals $(f_2 - f_1 = 5/N)$ with very different strengthes embedded in WGN. The iterative algorithm stops right after the first iteration (i = 2). Notice that the threshold SNR for the first strong signal $\mathbf{s}_1(f_1)$ occurs approximately at -3dB. While the threshold SNR for the weak signal $\mathbf{s}_2(f_2)$ occurs approximately around $SNR_1 = -3dB$ too. For the case shown in figure 1(b), we have two closely spaced signals $(f_2 - f_1 << 1/N)$ with same strength embedded in WGN. The iterative algorithm stops right after a few iterations (from i = 2to i = 4 for different SNRs). Notice that the threshold SNR for both signal components $\mathbf{s}_1(f_1)$ and $\mathbf{s}_2(f_2)$ occurs approximately at 1dB. For the case shown in figure 1(c), two equal strength signal are closely spaced as $f_2 - f_1 = 1/N$. The iterative algorithm stops right after a few iterations (from i = 3 to i = 10 for different SNRs). In this case, the threshold SNR equals 0dB for both frequencies.

6. CONCLUSIONS

An iterative Gram-Schmidt based estimation method is proposed for estimating non-linear parameters of linear signal embedded in additive noise. For application of estimating sinusoid frequencies, computer simulations show that the method provides *efficient* estimates for both frequencies with low SNR thresholds under various signal strength and frequency spacing combinations. When the strengthes of two sinusoidal components are very different, we obtain the efficient estimates of both frequencies within very few iterations, regardless of the closed spacing between the two frequencies. When the two signal components are orthogonal $(|f_1 - f_2| = \frac{k}{N}$ with k being any integer), very few iterations are needed in getting efficient estimates of frequencies.

7. REFERENCES

- D. W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, pp. 975–989, 1982.
- [2] R. Kumaresan, L. L. Scharf, and A. Shaw, "An algorithm for pole-zero modeling and spectral analysis," *IEEE Trans. on Acoust., Speech, and Signal Processing*, vol. ASSP-34, pp. 637-640, Jun. 1986.
- [3] Y. Bresler, and A. Macovski "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. on Acoust., Speech,* and Signal Processing, vol. ASSP-34, pp. 1081–1089, Oct. 1986.
- [4] I. Ziskind and M. Wax, "Maximum likelihood locatization of multiple sources by alternating projection," *IEEE Trans. on Acoust., Speech, and Signal Processing*, vol. 37, pp. 1553-1560, Oct. 1988.
- [5] J. Cadzow, "Signal processing via least-squares error modeling," *IEEE Sig. Proc. Mag.*, vol. 7, pp. 12-31, Oct. 1990.
- [6] D. W. Tufts, H. Ge, and S. Umesh, "Fast maximum likelihood estimation of signal parameters using the shape of the compressed likelihood function," *IEEE J.* of Oceanic Eng., vol. 18, pp. 388-400, Oct. 1993.
- [7] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. on Signal Processing*, vo. 42, pp. 2146-2157, Aug. 1994.
- [8] D. W. Tufts, "Estimating frequencies of two sinusoids using Gram-Schmidt orthogonalization," *Re*search Notes of the DSP Group, Univ. of Rhode Island, 1993.
- [9] P. J. Davis, Interpolation & Approximation. New York:Dover, 1975.



(a). signal $s[t] = 1 e^{j 2 \pi 0.5 t} - 0.1 e^{j (2 \pi 0.7 t + \pi/4)}$.



Figure 1: Estimation accuracy of the proposed G-S based iterative estimators for both frequencies of sinusoidal components embedded in WGN. Horizontal axis is defined as $SNR = 10 \log_{10}(\frac{1}{2\sigma^2})$ in dB. While SNR for each sinusoidal component is defined as $SNR_i = 10 \log_{10}(\frac{|a_i|^2}{2\sigma^2})$. Asterisks denote the MSE of \hat{f}_1 in dB; circles denote the MSE of \hat{f}_2 in dB, and solid lines denote the corresponding CR bounds.