A JOINT ESTIMATION ALGORITHM FOR MULTIPLE SINUSOIDAL FREQUENCIES

Ta-Hsin Li

Department of Mathematical Sciences
IBM T. J. Watson Research Center
Yorktown Heights, NY 10598-0218, USA
E-mail: thl@us.ibm.com

Kai-Sheng Song

Department of Statistics
Florida State University
Tallahassee, FL 32306-4330, USA
E-mail: kssong@stat.fsu.edu

ABSTRACT

Accurate estimation of sinusoidal frequencies from noisy observations is an important problem in many applications including radar, sonar, and data communications. Among many algorithms is the iterative filtering algorithm (IFA), proposed by Kay, which provides a computationally simple procedure yet capable of accurate frequency estimation especially at low signal-to-noise ratio (SNR). However, the convergence and other numerical/statistical properties of IFA have not been established beyond simulation. This paper makes several important contributions: (a) It shows that the poles of the AR filter must be reduced via a shrinkage parameter to accommodate possibly poor initial values. (b) It shows that the AR estimates in each iteration must be bias-corrected to produce more accurate frequency estimates; a closed-form expression is provided for bias correction. (c) It shows that for a sufficiently large sample size, the resulting algorithm, called new IFA, or NIFA, converges to the desired fixed-point which constitutes a consistent frequency estimator. Numerical examples, including a radar data example, are provided to demonstrate the findings.

1. INTRODUCTION

Consider the problem of estimating the frequencies, $\{\omega_k\}$, of multiple complex sinusoids from

$$y_t := x_t + \varepsilon_t,\ x_t := \sum_{k=1}^{p} \beta_k e^{i(\omega_k t+\phi_k)} \quad (t = 1,\ldots,n), \tag{1}$$

where $p \geq 1$ is a known integer, $\beta_k > 0$, $\omega_k := 2\pi f_k \in (-\pi,\pi)\setminus\{0\}$, and $\phi_k \in (-\pi,\pi]$ are unknown constants, and $\{\varepsilon_t\}$ is a zero-mean white noise process with unknown variance σ^2. The iterative filtering algorithm (IFA) proposed by Kay [1] is based on the fact that $\{x_t\}$ is a special autoregressive (AR) process of order p satisfying $x_t + \sum_{k=1}^{p} a_k x_{t-k} = 0$, where the a_k have a one-to-one relationship with the ω_k such that

$$1 + \sum_{k=1}^{p} a_k z^{-k} = \prod_{k=1}^{p}(1 - e^{i\omega_k} z^{-1}). \tag{2}$$

So the frequency estimation problem can be reformulated as the problem of estimating $\mathbf{a} := [a_1,\ldots,a_p]^T$. To estimate the AR parameter $\mathbf{a}$, IFA iterates a cycle of estimation and filtering: It starts with an initial estimate $\hat{\mathbf{a}} := [\hat{a}_1,\ldots,\hat{a}_p]^T$ obtained from $\{y_t\}$ and uses it to construct an AR filter which is applied to $\{y_t\}$ to produce

$$\tilde{y}_t = -\sum_{k=1}^{p} \hat{a}_k \tilde{y}_{t-k} + y_t \quad (t = 1,\ldots,n), \tag{3}$$

where $\tilde{y}_{-p+1} = \cdots = \tilde{y}_0 = 0$. Then it re-estimates the AR parameter from the filtered time series $\{\tilde{y}_t\}$ and uses the new estimate to filter the original data $\{y_t\}$ in the same way as (3) to produce a new filtered time series. This cycle is repeated until a stopping criterion is satisfied. IFA is simple computationally yet capable of providing accurate estimates. But, the convergence of IFA has not been established beyond simulation and a special case.

It is well known that the Gaussian maximum likelihood method (a.k.a. nonlinear least squares) produces a statistically efficient estimator but suffers from the problem of numerous local extrema so that an initial value of precision $\mathcal{O}(n^{-1})$ is usually required in order for standard optimization algorithms to converge to the desired solution [2]–[4]. IFA suffers from the same initial value problem, as shown in Fig. 1.

In this paper, we provide a statistical analysis of IFA in the case of $p = 2$ and make several contributions:

(a) It is shown that the poles of the AR filter must be reduced via an extra shrinkage parameter in order to accommodate poor initial values and avoid being trapped into spurious solutions.

(b) It is shown that the AR estimates in each iteration must be bias-corrected in order to produce more accurate frequency estimates; a closed-form expression is derived for bias correction.

1-4244-0469-X/06/$20.00 ©2006 IEEE

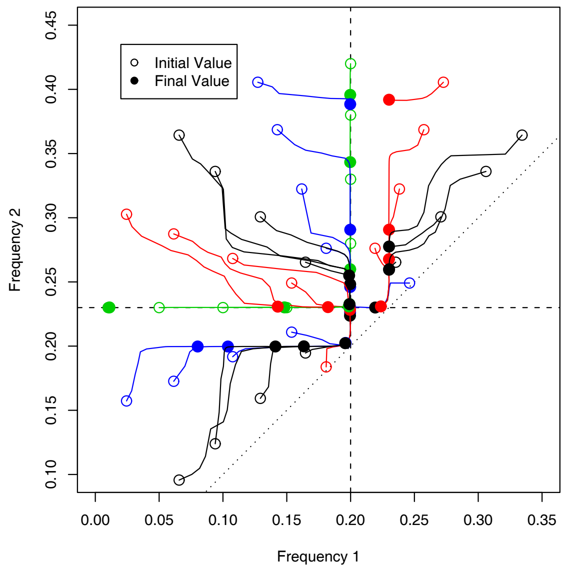

Fig. 1. Trajectory of the IFA estimates for two complex sinusoids in Gaussian white noise ($n = 100$ and SNR $= 0$ dB). Dashed lines indicate the location of the true frequencies ($f_1 = 0.2$ and $f_2 = 0.23$); dotted diagonal line shows the boundary $f_1 = f_2$. Open circles represent initial values; solid points represent final estimates after 15 iterations; intermediate estimates appear as lines that link open circles with solids points.

(c) It is shown that with probability tending to unity as the sample size grows, the resulting algorithm, which we call the *new* IFA, or NIFA, converges to the desired fixed-point which constitutes a consistent frequency estimator.

These results can be regarded as an extension of the earlier work such as [5]. It is also worth pointing out that by cascading the AR fitting with the AR filtering a notch filter can be obtained. It can be implemented as an adaptive filter for tracking time-varying frequencies. The results in this paper remain valid for the steady-state performance of the notch filtering algorithm.

2. THE NEW ITERATIVE FILTERING ALGORITHM (NIFA)

Let us assume in the remainder of the paper that $\{y_t\}$ is given by (1) with $p = 2$ and $\omega_1 < \omega_2$. Under this assumption, it follows from (2) that $\mathbf{a} = [a_1, a_2]^T$ and

$$a_1 := -(e^{i\omega_1} + e^{i\omega_2}), \quad a_2 := e^{i(\omega_1+\omega_2)}. \tag{4}$$

For any admissible variable $\boldsymbol{\alpha} := [\alpha_1, \alpha_2]^T$, which will be defined later, let $\{y_t(\boldsymbol{\alpha})\}$ denote the filtered time series

$$y_t(\boldsymbol{\alpha}) = -\sum_{k=1}^{2} \alpha_k \eta^k y_{t-k}(\boldsymbol{\alpha}) + y_t \quad (t = 1, \ldots, n) \tag{5}$$

with $y_{-1}(\boldsymbol{\alpha}) := y_0(\boldsymbol{\alpha}) := 0$, where $\eta \in (0,1)$ is the shrinkage parameter that contracts the poles of the filter towards the origin and thus stabilizes the filter.

Given $\{y_t(\boldsymbol{\alpha})\}$, we estimate $\mathbf{a}$ by the method of least squares (LS), i.e., by seeking $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ that minimizes $\|\mathbf{y}(\boldsymbol{\alpha}) + \mathbf{Y}(\boldsymbol{\alpha})\mathbf{D}\hat{\mathbf{a}}(\boldsymbol{\alpha})\|^2$, where

$$\mathbf{Y}(\boldsymbol{\alpha}) := \begin{bmatrix} y_2(\boldsymbol{\alpha}) & y_1(\boldsymbol{\alpha}) \\ \vdots & \vdots \\ y_{n-1}(\boldsymbol{\alpha}) & y_{n-2}(\boldsymbol{\alpha}) \end{bmatrix},$$

$$\mathbf{y}(\boldsymbol{\alpha}) := \begin{bmatrix} y_3(\boldsymbol{\alpha}) \\ \vdots \\ y_n(\boldsymbol{\alpha}) \end{bmatrix}, \quad \mathbf{D} := \mathrm{diag}(\eta, \eta^2).$$

This gives rise to an AR estimator

$$\hat{\mathbf{a}}(\boldsymbol{\alpha}) := -\mathbf{D}^{-1}\{\mathbf{Y}^H(\boldsymbol{\alpha})\mathbf{Y}(\boldsymbol{\alpha})\}^{-1}\mathbf{Y}^H(\boldsymbol{\alpha})\mathbf{y}(\boldsymbol{\alpha}), \tag{6}$$

where superscript H stands for Hermitian transpose. Note that the role of $\mathbf{D}$ is to compensate for η.

Unlike the true AR parameter $\mathbf{a}$, the AR estimator $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ does not necessarily correspond to a polynomial of the form (2) which has unit roots. To impose this constraint on each root of the polynomial of $\hat{\mathbf{a}}(\boldsymbol{\alpha})$, one can simply reset its modulus to unity while retaining its angle, thus projecting it on the unit circle. The resulting AR estimator is denoted by $\boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}))$, where $\boldsymbol{\psi}(\cdot)$ represents the unit-root (UR) projector. The UR projection not only stabilizes the AR filter but also eliminates the redundancy in the AR reparameterization.

With the AR estimator $\boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}))$ so defined, one seeks a fixed point of the mapping $\boldsymbol{\alpha} \mapsto \boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}))$ by the so-called fixed-point iteration

$$\boldsymbol{\alpha}_m = \boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}_{m-1})) \quad (m = 1, 2, \ldots). \tag{7}$$

The IFA in [1] can be regarded as a special case of (7) with $\eta = 1$, although, strictly speaking, it employs Burg's estimator rather than the LS estimator and it does not impose the UR constraint. Fig. 1 shows that with $\eta = 1$ the iteration in (7) may converge to spurious fixed points if the initial values are not near the desired solution. This problem can be overcome by choosing $\eta < 1$.

A careful analysis shows that in the case of $\eta < 1$ the mapping $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ contains a bias term at $\boldsymbol{\alpha} = \mathbf{a}$ that can be expressed as $\mathbf{b} := [b_1, b_2]^T$ where

$$b_k := \frac{(1-\eta)i^k(e^{i\omega_2} - e^{i\omega_1})^k \sin^{2-k}(\omega_2 - \omega_1)}{\eta^k\{1 - \cos(\omega_2 - \omega_1)\}} \quad (k = 1, 2). \tag{8}$$

By subtracting the bias from $\hat{\mathbf{a}}(\boldsymbol{\alpha})$, a new mapping $\boldsymbol{\alpha} \mapsto \boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}) - \mathbf{b})$ is formed. But the corresponding fixed-point iteration is not yet practical because the bias $\mathbf{b}$ given

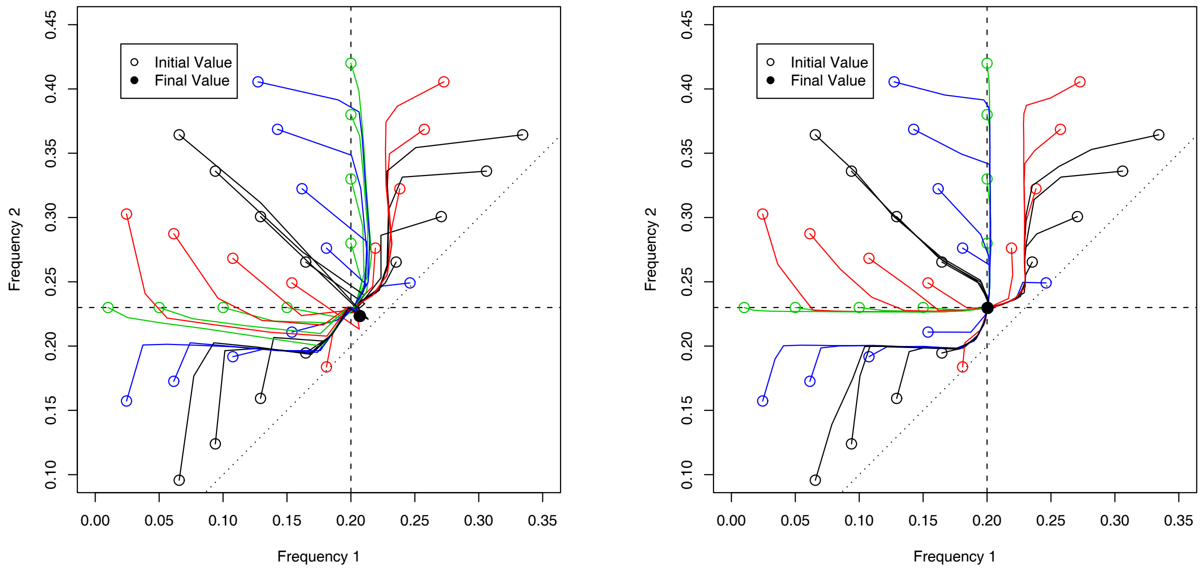

Fig. 2. Trajectory of the NIFA estimates for the same data as used in Fig. 1. Left, $\eta = 0.92$; right, $\eta = 0.96$.

by (8) depends on the true frequencies. One way to make it practical is to estimate the bias in each iteration using the estimate of $\mathbf{a}$ from the previous iteration. This gives rise to our NIFA algorithm

$$\boldsymbol{\alpha}_m := \boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}_{m-1}) - \mathbf{b}(\boldsymbol{\alpha}_{m-1})) \quad (m = 1, 2, \ldots), \tag{9}$$

where $\mathbf{b}(\boldsymbol{\alpha}) := [b_1(\boldsymbol{\alpha}), b_2(\boldsymbol{\alpha})]^T$ is defined in the same way as $\mathbf{b}$ by (8) except that ω_1 and ω_2 are replaced by λ_1 and λ_2 which are the angles of the roots of $1 + \alpha_1 z^{-1} + \alpha_2 z^{-2}$ satisfying $\lambda_1 \leq \lambda_2$. Let $\hat{\mathbf{a}} := [\hat{a}_1, \hat{a}_2]^T$ be the limiting value of $\{\boldsymbol{\alpha}_m\}$ in (9) as $m \to \infty$. Then, the final NIFA frequency estimator is given by $\hat{\boldsymbol{\omega}} := [\hat{\omega}_1, \hat{\omega}_2]^T$, where $\hat{\omega}_1 = 2\pi \hat{f}_1$ and $\hat{\omega}_2 = 2\pi \hat{f}_2$ are defined as the angles of the roots of $1 + \hat{a}_1 z^{-1} + \hat{a}_2 z^{-2}$ satisfying $\hat{\omega}_1 \leq \hat{\omega}_2$. Similarly, one can define the intermediate frequency estimates $\hat{\boldsymbol{\omega}}_m$ in terms of $\boldsymbol{\alpha}_m$. By definition, $\hat{\boldsymbol{\omega}}_m \to \hat{\boldsymbol{\omega}}$ as $m \to \infty$.

3. CONVERGENCE AND ACCURACY

Eq. (4) defines a one-to-one mapping from $\boldsymbol{\omega} := [\omega_1, \omega_2]^T$ to $\mathbf{a} = [a_1, a_2]^T$ which will be denoted by $\boldsymbol{\omega} \mapsto \boldsymbol{\phi}(\boldsymbol{\omega}) := \mathbf{a}$. Let Λ denote the set of $\boldsymbol{\lambda} := [\lambda_1, \lambda_2]^T$ with $-\pi < \lambda_1 < \lambda_2 \leq \pi$ and let $\mathcal{A}$ denote the set of $\boldsymbol{\alpha} := [\alpha_1, \alpha_2]^T$ such that $\boldsymbol{\alpha} = \boldsymbol{\phi}(\boldsymbol{\lambda})$ for some $\boldsymbol{\lambda} \in \Lambda$, i.e., $\mathcal{A} := \boldsymbol{\phi}(\Lambda)$. Let $\mathcal{A}_\delta$ denote a closed subset of $\mathcal{A}$ such that $\mathcal{A}_\delta := \{\boldsymbol{\alpha} \in \mathcal{A}\colon \|\boldsymbol{\alpha} - \mathbf{a}\| \leq \kappa \delta^\varepsilon\}$, where $\kappa > 0$ and $\varepsilon \in (1, \frac{3}{2})$ are constants and $\delta := 1 - \eta \in (0, 1)$ depends on n such that $\delta \to 0$ as $n \to \infty$.

Theorem. [6] Let $\mathcal{A}_\delta$ be the neighborhood of $\mathbf{a}$ defined above and assume that $n(1-\delta)^n = \mathcal{O}(1)$ and $n\delta^\varepsilon \to \infty$ as $n \to \infty$. Then, the following assertions are true.

(a) With probability tending to unity, the mapping $\boldsymbol{\alpha} \mapsto \boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}) - \mathbf{b}(\boldsymbol{\alpha}))$ is contractive in $\mathcal{A}_\delta$, with a contraction factor of the form $\mathcal{O}_P(\delta^{\varepsilon-1})$, and therefore has a unique fixed point $\hat{\mathbf{a}} \in \mathcal{A}_\delta$. Furthermore, with probability tending to unity, the sequence $\{\boldsymbol{\alpha}_m\}$ defined by (9) converges to $\hat{\mathbf{a}}$ as $m \to \infty$ for any initial value $\boldsymbol{\alpha}_0 \in \mathcal{A}_\delta$.

(b) Let $\hat{\boldsymbol{\omega}}$ be the frequency estimator corresponding to $\hat{\mathbf{a}}$, i.e., $\hat{\boldsymbol{\omega}} := \boldsymbol{\phi}^{-1}(\hat{\mathbf{a}})$. Then, for any constant $\beta \leq 3/2$ such that $n^{-1}\delta^{-\beta} = \mathcal{O}(1)$ as $n \to \infty$, $\hat{\boldsymbol{\omega}}$ is at least $\delta^{-\beta}$-consistent for estimating $\boldsymbol{\omega}$, i.e., $\delta^{-\beta}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}) = \mathcal{O}_P(1)$.

As an example, assume that $\delta = 1 - \eta = \mathcal{O}(n^{-\nu})$ for some $0 < \nu < \varepsilon^{-1} < 1$, then $n\delta^\varepsilon = \mathcal{O}(n^{1-\nu\varepsilon}) \to \infty$, so the conditions in the Theorem are satisfied. The resulting $\hat{\boldsymbol{\omega}}$ is at least $n^{\nu\beta}$-consistent for any $\beta \leq \min(\nu^{-1}, 3/2)$, as it satisfies $n^{-1}\delta^{-\beta} = \mathcal{O}(n^{-1+\nu\beta}) = \mathcal{O}(1)$. The required accuracy of initial values is $\mathcal{O}(n^{-\nu\varepsilon})$.

In particular, if $\nu = 2/3$ and $\beta = \nu^{-1} = 3/2$, the Theorem ensures that $\hat{\boldsymbol{\omega}}$ is at least n-consistent. By setting $\varepsilon = 1^+$, the initial requirement becomes nearly $\mathcal{O}(n^{-2/3})$. This means that it suffices to use a slightly better than $n^{2/3}$-consistent estimator as the initial value to obtain the n-consistent final estimator $\hat{\boldsymbol{\omega}}$. Such initial values can be produced by NIFA itself, with the choice of $\nu = (4/9)^+$. Indeed, with this choice, the Theorem guarantees a better than $n^{2/3}$ rate of consistency by taking $\beta = 3/2$ so that $\nu\beta = (2/3)^+$. To obtain this estimator, the initial values are required to be slightly more accurate than $\mathcal{O}(n^{-4/9})$, which can be satisfied by all $\sqrt{n}$-consistent estimators.

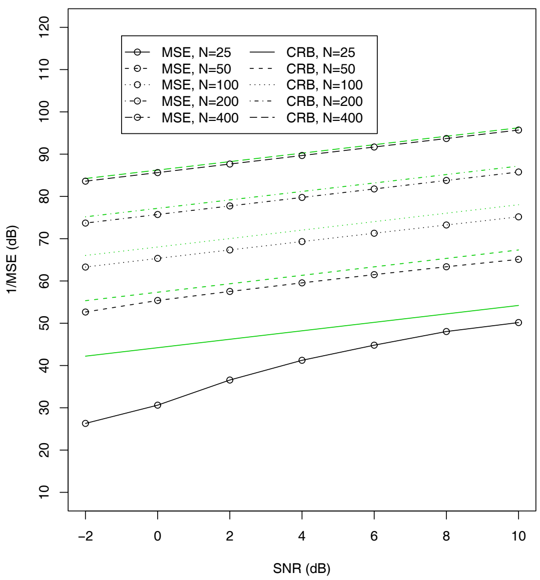

Fig. 3. Reciprocal MSE of the NIFA estimator for f_1. Lines without circle represent the CRLB under the assumption of complex Gaussian white noise. Results are based on 1,000 Monte Carlo runs. The true frequencies are $f_1 = 0.2$ and $f_2 = 0.23$ (the phases are $\phi_1 = 0$ and $\phi_2 = \pi/2$). The simulated noise is complex Gaussian.

In summary, by applying NIFA twice, once with a smaller η and once with a large η, one is able to improve the accuracy of the frequency estimator from $\mathcal{O}(n^{-4/9})$ to $\mathcal{O}(n^{-1})$ or better. The convergence of NIFA in both cases are guaranteed by the Theorem.

4. EXAMPLE AND APPLICATION

Fig. 2 shows the trajectory of $\{\hat{\boldsymbol{\omega}}_m\}$ with $\eta = 0.92$ and $\eta = 0.96$ for the same data as used in Fig. 1. As can be seen, the spurious fixed points in Fig. 1 no longer exist in Fig. 2 where all initial values lead to a single fixed point even if they are far away from the true frequencies. This implies that the initial requirement of NIFA, with $\eta < 1$, is much less stringent than that of IFA where $\eta = 1$. Generally speaking, η should not be made too close to unity to avoid spurious fixed points, which is consistent with the theoretical findings. A data-driven method of selecting appropriate η is proposed in [6].

The shrinkage parameter η plays a vital role not only in determining the requirement of initial values, but also in determining the accuracy of the final estimator. Fig. 2 shows that the fixed point of NIFA with the larger $\eta = 0.96$ is much closer to the true frequencies than that with the smaller $\eta = 0.92$ where the estimates are pushed towards a single value in the vicinity of $\frac{1}{2}(f_1 + f_2)$. These observations further justify the approach with multiple

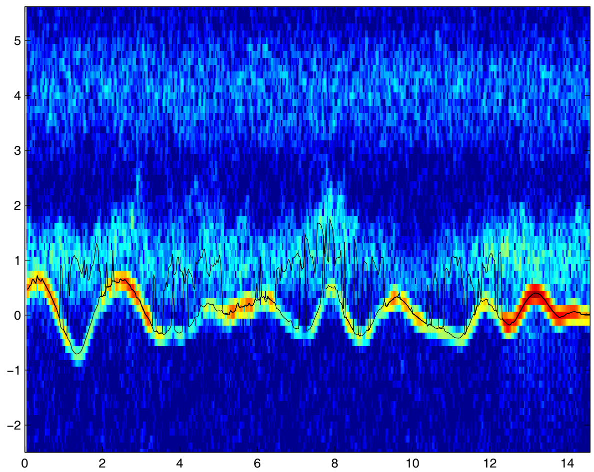

Fig. 4. Time-Doppler image and NIFA frequency estimates (dark lines). Horizontal: time; vertical: Doppler.

values of η discussed in the previous section.

Fig. 3 shows the mean-square error (MSE) of $\hat{f}_1$ and the corresponding CRLB for various sample sizes and SNRs. Two values of η are used for each sample size and Prony's estimator is used as the initial value.

Fig. 4 shows the result of estimating the Doppler frequency of a target. The data is taken from the 10th range bin of the Sea Clutter + Target Data File #283 of radar measurements collected with the McMaster IPIX Radar overlooking the Atlantic Ocean from a cliff-top in Dartmouth, Nova Scotia, Canada (http://soma.ece.mcmaster.ca/ipix). The target is a spherical block of styrofoam wrapped with wire mesh.

5. REFERENCES

[1] S. M. Kay, "Accurate frequency estimation at low signal-to-noise ratio," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 3, pp. 540–547, 1984.

[2] A. K. Shaw, "Maximum likelihood estimation of multiple frequencies with constraints to guarantee unit circle roots," *IEEE Trans. Signal Processing*, vol. 43, no. 3, pp. 796–799, 1995.

[3] D. Starer and A. Nehorai, "Newton algorithm for conditional and unconditional maximum likelihood estimation of the parameters of exponential signals in noise," *IEEE Transactions on Signal Processing*, vol. 40, no. 6, pp. 1528–1534, 1992.

[4] P. Stoica, R. L. Moses, B. Friedlander, and T. Söderström, "Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 3, pp. 378–392, 1989.

[5] T. H. Li and K. S. Song, "Asymptotic analysis of a fast algorithm for efficient multiple frequency estimation," *IEEE Transactions on Information Theory*, vol. 48, no. 10, pp. 2709–2720, 2002. Errata: vol. 49, no. 2, p. 529, 2003.

[6] K. S. Song and T. H. Li, "On convergence and bias correction of a joint estimation algorithm for multiple sinusoidal frequencies," IBM Technical Report RC23479, 2005.