Estimating Multiple Frequency-Hopping Signal Parameters via Sparse Linear Regression

Daniele Angelosante, *Member, IEEE*, Georgios B. Giannakis, *Fellow, IEEE*, and Nicholas D. Sidiropoulos, *Fellow, IEEE*

***Abstract*—Frequency hopping (FH) signals have well-documented merits for commercial and military applications due to their near-far resistance and robustness to jamming. Estimating FH signal parameters (e.g., hopping instants, carriers, and amplitudes) is an important and challenging task, but optimum estimation incurs an unrealistic computational burden. The spectrogram has long been the starting non-parametric estimator in this context, followed by line spectra refinements. The problem is that hop timing estimates derived from the spectrogram are coarse and unreliable, thus severely limiting performance. A novel approach is developed in this paper, based on sparse linear regression (SLR). Using a dense frequency grid, the problem is formulated as one of under-determined linear regression with a dual sparsity penalty, and its exact solution is obtained using the alternating direction method of multipliers (ADMoM). The SLR-based approach is further broadened to encompass polynomial-phase hopping (PPH) signals, encountered in chirp spread spectrum modulation. Simulations demonstrate that the developed estimator outperforms spectrogram-based alternatives, especially with regard to hop timing estimation, which is the crux of the problem.**

***Index Terms*—Compressive sampling, frequency hopping signals, sparse linear regression, spectrogram, spread spectrum signals.**

I. Introduction

FREQUENCY-HOPPING spread-spectrum signaling is widely adopted in tactical communications due to its low probability of detection and interception, agility, and robustness to jamming [34]. Estimating and tracking the parameters of multiple superimposed FH signals are important tasks with applications in both military and civilian domains: from interception of noncooperative communications, to collision avoidance and cognitive radio. The problem is particularly challenging when the hopping patterns of the constituent signals are unknown, and, in addition to dwell frequency, hop timing is randomized as well for added protection. Maximum likelihood (ML) estimation is practically intractable in this context, which motivates the pursuit of alternative low- to moderate-complexity solutions.

Starting from *coarse channelization* techniques based on the spectrogram and related non-parametric time-frequency estimation tools, there is considerable literature on the subject of FH signal parameter estimation and tracking. Non-parametric methods based on the spectrogram are simple but suffer from limited resolution and require further refinements [1], [29]. Time-frequency distribution techniques have been investigated in [2] for acquisition of FH signals.

Parametric methods for FH signal estimation model the active frequency as piecewise-constant and achieve improved estimation accuracy at the cost of higher complexity. The crux of the overall problem is *hop timing estimation*: given the hop instants, what remains is essentially a sequence of harmonic retrieval problems. When the hops are periodic, the timing problem reduces to estimating the hopping period(s) and offset(s) [1], [2], [29]. Hop timing estimators for the more difficult case of aperiodic hop timing have been developed based on dynamic programming (DP) [22], [23], and the expectation-maximization (EM) algorithm [24]. The algorithms in [22], [23] and [24] require multiple receive antennas, and rely on the spectrogram for coarse acquisition.

When only one FH signal is present, an effective particle filtering solution based on a stochastic dynamical system formulation has recently appeared in [36]. Different from [22], [23] and [24], the approach in [36] allows for sequential processing, and is robust to various sources of mismatch in the probabilistic model adopted. The limitation of [36] is that it does not generalize to multiple FH signals, due to the *curse of dimensionality*: the required number of particles grows fast with the dimensionality of the state-space. The complexity of DP-based approaches [22], [23], on the other hand, increases rapidly also with the number of temporal samples—thus only short data records can be processed.

While sparse linear regression (SLR) has been advocated in [10] and [16] for harmonic retrieval without carrier hopping, in this paper a novel SLR-based technique is developed for multiple FH signals. Relative to [10] and [16], here we also take advantage of sparsity in terms of carrier hopping, which is effected through a dual sparsity penalty. The developed estimator is also generalized to handle polynomial-phase hopping (PPH) signals that emerge in chirp spread spectrum communications [6], [21],

Manuscript received October 26, 2009; accepted June 01, 2010. Date of publication June 10, 2010; date of current version September 15, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Sofia C. Olhede. The work in this paper was supported by NSF Grants CCF 0830480 and CON 0824007, and also through collaborative participation in the Communications and Networks Consortium sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon. Part of the results in this paper was presented at the International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, April 2010.

D. Angelosante and G. B. Giannakis are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: angel129@umn.edu; georgios@umn.edu).

N. D. Sidiropoulos is with the Department of Electronic and Computer Engineering, Technical University of Crete, 73100, Chania, Crete, Greece (e-mail: nikos@telecom.tuc.gr).

Digital Object Identifier 10.1109/TSP.2010.2052614

[28]. A pertinent sparsity-aware optimization problem is formulated and solved using the alternating direction method of multipliers (ADMoM). Simulations illustrate that the developed technique is robust to model mismatch, and far outperforms spectrogram-based methods, especially with regard to hop timing estimation.

Due to the non-convexity of the FH estimation problem, parametric techniques based on the likelihood function (such as the EM algorithm) can be trapped in local minima if the initialization is far from a global minimum. Due to its low complexity and high accuracy, the novel SLR-based estimator can be used both as a stand-alone FH signal estimation algorithm, and as an excellent initialization for iterative refinement algorithms, such as the one in [24].

Interestingly, the closely related problem of identifying the parameters of a piecewise-sinusoidal mixture model from (generalized) samples has been studied in [5] using a finite-rate of innovation (FRI) approach. SLR and FRI are different tools dealing with similar problems; see also [9] for a tutorial on FRI and its relation with SLR and compressive sampling. While introducing interesting identifiability conditions and algorithms for perfect reconstruction of the underlying continuous-time signal, the approach in [5] is not directly applicable to the present context. The switching instants ("hops") in [5] are continuous variables, and the measurements are obtained through *analog* pre-filtering with a properly chosen kernel waveform, which may or may not be affordable. Since the hopping steps are not instantaneous in practice [34], the present algorithm (as well as all existing alternatives for acquiring FH signals [22]–[24], [36]) do not attempt to localize the *exact* hopping instants, but rather aim to detect hops. Also, PPH signals are not considered in [5]. On the other hand, numerical simulations suggest that a suitable modification of the SLR estimator for the noiseless case can perfectly recover the sampled FH signals as well.

The rest of this paper is structured as follows. Section II contains preliminaries, and the problem statement. The novel SLR formulation is introduced in Section III, where sparsity tuning and extensions to PPH signals are also presented. An efficient solution based on the ADMoM is developed in Section IV. Simulations are presented in Section V, and conclusions are drawn in Section VI.

Notation: Column vectors (matrices) are denoted with lower-case (upper-case) boldface letters and sets with calligraphic letters; $(\cdot)^T$ stands for transposition, $(\cdot)^H$ for conjugate transposition, and $(\cdot)^\dagger$ for pseudoinverse; $\mathcal{CN}(\mu,\sigma^2)$ denotes the complex Gaussian probability density function with mean μ and variance σ^2; $\otimes$ denotes the Kronecker product; $\Re\{x\}$ and $\Im\{x\}$ denote the real and imaginary part of $x \in \mathbb{C}$, respectively; $\mathbf{0}_P$ is the P-dimensional column vector with all zeros and $\mathbf{I}_P$ is the P-dimensional identity matrix, while $\mathbf{0}_{P\times P}$ denotes the $P \times P$ matrix with all zeros. The (pseudo) ℓ_0-norm of $\mathbf{x}$ is defined as the number of nonzero elements of $\mathbf{x}$. The ℓ_1-, ℓ_2-, and ℓ_∞-norms of $\mathbf{x} \in \mathbb{C}^P$ are defined, respectively, as $\|\mathbf{x}\|_1 := \sum_{p=1}^P \sqrt{\Re\{x_p\}^2 + \Im\{x_p\}^2}$, $\|\mathbf{x}\|_2 := \sqrt{\sum_{p=1}^P (\Re\{x_p\}^2 + \Im\{x_p\}^2)}$, and $\|\mathbf{x}\|_\infty := \max_{p=1,\ldots,P} \sqrt{\Re\{x_p\}^2 + \Im\{x_p\}^2}$.

II. Preliminaries and Problem Statement

Consider the noiseless signal $s(t)$, which at time $t \in [t_{k-1}, t_k)$ consists of M_k pure tones; that is

$$s(t) := \sum_{m=1}^{M_k} a_{m,k} e^{j2\pi f_{m,k} t}, \quad t_{k-1} \le t < t_k \tag{1}$$

where $t_k \in [0,T]$ is the kth system-wise hopping instant[1] [22], $[t_{k-1}, t_k)$ is the kth system-wise dwell, and $a_{m,k} \in \mathbb{C}$, $f_{m,k} \in [-f_{\max}, f_{\max}]$ are the complex amplitude and frequency of the mth tone in the kth system-wise dwell, respectively. The number of tones, M_k, can also vary with k, due to emitter (de)activation or bandwidth mismatch [24]. The entire observation interval is $[0,T] (t_0 = 0)$. A noncooperative asynchronous scenario is considered; hop timing is aperiodic, and independent across transmitters. Our approach is geared toward slow FH signals, and offsets due to frequency modulation can be accommodated as well. The measured continuous-time waveform $y(t)$ is corrupted by additive circularly-symmetric complex white Gaussian noise $v(t)$, i.e.,

$$y(t) := s(t) + v(t), \quad 0 \le t \le T. \tag{2}$$

Let K denote the total number of system-wise hops in $[0,T]$, and $T_s := (1)/(2f_{\max})$ the period with which $y(t)$ is sampled at the receiving end. The discrete-time FH signal can be written as [cf. (1)]

$$s_n := s(nT_s) = \sum_{m=1}^{M_k} a_{m,k} e^{j\omega_{m,k} n}, \quad n_{k-1} \le n < n_k \tag{3}$$

where $n \in \{0,1,\ldots,N-1\}$, $N-1 := \lfloor (T)/(T_s) \rfloor$, $\omega_{m,k} \in [-\pi,\pi]$, $\omega_{m,k} := 2\pi f_{m,k} T_s$, and $n_k := \lceil (t_k)/(T_s) \rceil$. Correspondingly, the discrete-time noisy observations are [cf. (2)]

$$y_n := s_n + v_n, \quad n \in \{0,1,\ldots,N-1\} \tag{4}$$

where $\{v_n\}$ is white, and $v_n \sim \mathcal{CN}(0,\sigma^2)$.

Given $\mathbf{y} := [y_0, \ldots, y_{N-1}]^T$, the objective is to estimate K, $\{n_k\}$, $\{M_k\}$, $\{a_{m,k}\}$, and $\{\omega_{m,k}\}$. Since ML estimation of FH signal parameters is practically intractable, non-parametric estimators based on the spectrogram have been traditionally employed. These are outlined briefly in the next subsection in order to establish notation and context for the novel approach we will develop in Section III.

A. Spectrogram-Based Estimators

The spectrogram of $\{y_n\}_{n=0}^{N-1}$ is the squared modulus of the short-term Fourier transform defined as

$$Y[m,\ell] := \left| \sum_{n_1=0}^{N_1-1} y_{\ell L + n_1} w_{n_1} e^{-j2\pi \frac{n_1 m}{N_2}} \right|^2 \tag{5}$$

[1]The set of system-wise hopping instants is the union of all individual emitter hopping instants, splitting the time axis in system-wise dwells.

for $m=0,1,\ldots,N_2-1$, $\ell=0,1,\ldots,(N-N_1)/(L)$, and $N_2 \geq N_1 \geq L$. Specifically, one splits the observed data into $(N-N_1)/L+1$ overlapping segments, windows with w_n, and computes the discrete Fourier transform (DFT) evaluated at N_2 frequencies. Parameters N_1, N_2, L, and the window w_n highly affect the performance of spectrogram-based FH parameter estimators. A large N_1 yields improved frequency resolution, but poor temporal resolution which blurs hop timing. Small N_1 blurs the frequency axis, and close-by hops become indistinguishable. This unyielding tradeoff is the major limitation of spectrogram-based estimation, and it also affects parametric techniques which employ the spectrogram for coarse acquisition. Two types of spectrogram-based techniques have been proposed in the literature for (aperiodic) hop timing estimation.

1) *Entropy-based techniques* [22]. The columns of the spectrogram matrix formed with entries as in (5) are normalized to sum to unity, and the entropy of each column is computed. With $\{H_\ell\}_{\ell=0}^{(N-N_1)/L}$ denoting the sequence of entropies, FH causes spectral spreading that translates to higher entropy. This suggests obtaining the hopping instants by picking the peaks of H_ℓ; and
2) *Gradient techniques* [24]. After setting to zero the entries $Y[m,\ell]$ which are smaller than a predefined threshold (typically equal to the sample mean of the spectrogram), the sum of the difference of consecutive columns is evaluated, i.e., $\Delta_\ell := \sum_{m=0}^{N_2-1}|Y[m,\ell]-Y[m,\ell-1]|$, for $\ell=1,2,\ldots,(N-N_1)/L$. The system-wise hopping instants can then be estimated by picking the peaks of Δ_ℓ.

These estimates are subsequently processed for further refinement. Once hopping instants are acquired, the parameters within each dwell are estimated via harmonic retrieval techniques [22], [24].

The method developed in the sequel can be used as an effective stand-alone solution that jointly recovers hop timing and the remaining parameters of interest, namely K, M_k, dwell frequencies, and amplitudes. Alternatively, the novel method can be used to extract timing estimates, to be passed on to successive stages (e.g., those described in [22] and [24]) for further refinement.

III. Estimation via SLR

Suppose that the true frequencies $\{\omega_{m,k}\}$ in (3) belong to a known finite set $\mathcal{W} := \{\omega_1,\ldots,\omega_P\}$ with cardinality $P \gg \max_k M_k$. Note that this is not a limiting assumption for civilian applications, provided that Doppler is negligible. In cases where this information is not available, the set $\mathcal{W}$ can be a dense grid such that the separation between two consecutive frequencies in $\mathcal{W}$ is less than the desired resolution (in the same spirit of [10], [16] for harmonic retrieval). Clearly, as the preselected P increases the density of the grid increases, and so does the frequency resolution—what in the sparse linear regression parlance is referred to as super-resolution [16].

With $\{\omega_{m,k}\} \subset \mathcal{W}$, the received noisy samples can be rewritten as

$$y_n = \boldsymbol{\omega}_n^T \boldsymbol{x}_n + v_n, \quad n \in \{0,1,\ldots,N-1\} \tag{6}$$

where $\boldsymbol{\omega}_n := [e^{j\omega_1 n},\ldots,e^{j\omega_P n}]^T$, and $\boldsymbol{x}_n := [x_{n,1},\ldots,x_{n,P}]^T \in \mathbb{C}^P$. Observe that $x_{n,p}$ represents the amplitude and phase of the pth frequency bin at time n. Since $P \gg \max_k M_k$, a few of the coefficients $\{x_{n,p}\}$, representing the active frequencies at each time, are nonzero. Letting $\boldsymbol{x}^* := [\boldsymbol{x}_0^T,\ldots,\boldsymbol{x}_{N-1}^T]^T \in \mathbb{C}^{PN}$, and $\mathbf{w}_n := [\underbrace{\mathbf{0}_P^T,\ldots,\mathbf{0}_P^T}_{n},\boldsymbol{\omega}_n^T,\underbrace{\mathbf{0}_P^T,\ldots,\mathbf{0}_P^T}_{N-n-1}]^T \in \mathbb{C}^{PN}$, the model in (3) and (4) can be expressed in vector-matrix form as

$$\mathbf{y} = \mathbf{W}\boldsymbol{x}^* + \mathbf{v} \tag{7}$$

where $\mathbf{W} := [\mathbf{w}_0,\ldots,\mathbf{w}_{N-1}]^T$, and $\mathbf{v} := [v_0,\ldots,v_{N-1}]^T$. The FH signal parameters to estimate can be obtained from $\boldsymbol{x}^*$, which obeys the linear regression model in (7). Matrix $\boldsymbol{X}^* := [\boldsymbol{x}_0,\ldots,\boldsymbol{x}_{N-1}] \in \mathbb{C}^{P\times N}$ represents the time-localized frequency content of the signal, and is *related* to the spectrogram.

The key advantage of introducing the grid of candidate frequencies $\mathcal{W}$ is that the nonlinear parameter estimation task at hand is converted to a linear one [cf. (7)]. This is possible by increasing the problem dimensionality through the selection of $P \gg \max_k M_k$. Note also that as the $N \times PN$ matrix $\mathbf{W}$ is fat, the least-squares (LS) solution with minimum ℓ_2 norm, namely $\hat{\mathbf{x}}_{\text{min-norm}}^{LS} := \mathbf{W}^\dagger \mathbf{y}$, does not yield an accurate estimate of $\boldsymbol{x}^*$ even when the signal-to-noise ratio (SNR) is high. Improved alternatives are possible however, if one capitalizes on the fact that the unknown vector $\boldsymbol{x}^*$ exhibits the following two *sparsity* properties.

1) *Active carrier-domain sparsity.* Only a few of the coefficients $\{x_{n,p}\}$ are nonzero, which implies that $\boldsymbol{x}^*$ in (7) is sparse.
2) *Differential time-domain sparsity (smoothness).* Since FH is assumed slow, $x_{n+1,p} = x_{n,p}$ most of the time; hence, each row of $\boldsymbol{X}^*$ is piecewise constant. This means that adjacent row-wise differences are sparse.

Consider now the $(N-1)P \times NP$ matrix

$$\mathbf{D} := \begin{bmatrix} \mathbf{d}_1^T \\ \mathbf{d}_1^{T(1)} \\ \cdots \\ \mathbf{d}_1^{T((N-1)P-1)} \end{bmatrix} \tag{8}$$

where $\mathbf{d}_1 := [-1,\underbrace{0,\ldots,0}_{P-1},1,\underbrace{0,\ldots,0}_{(N-1)P-1}]^T$, and the notation $(\cdot)^{(m)}$ represents the right cyclic shift of m positions. From the definition in (8), the $(nP+p)$th entry of $\mathbf{D}\boldsymbol{x}^*$ contains the difference $x_{n+1,p}-x_{n,p}$; hence, as mentioned earlier, $\mathbf{D}\boldsymbol{x}^*$ is a sparse vector.

Ideally, one would form a sparse and piecewise constant estimate of $\boldsymbol{x}^*$ by solving the following optimization problem:

$$\check{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathbb{C}^{NP}} \left[\frac{1}{2}\|\mathbf{y}-\mathbf{W}\mathbf{x}\|_2^2 + \mu_1\|\mathbf{x}\|_0 + \mu_2\|\mathbf{D}\mathbf{x}\|_0\right]. \tag{9}$$

The first term of the cost function in (9) takes into account the observed signal while the positive scalars μ_1 and μ_2 control the intrinsic sparsity and smoothness of the estimate, respectively. However, the problem in (9) is non-convex and NP-hard.

Motivated by recent advances in variable selection [33] and compressive sampling [13], the ℓ_0-norm is relaxed with the convex ℓ_1-norm. Hence, the advocated formulation becomes

$$\hat{\mathbf{x}} := \arg\min_{\mathbf{x}\in\mathbb{C}^{NP}} \left[\frac{1}{2}\|\mathbf{y}-\mathbf{W}\mathbf{x}\|_2^2 + \lambda_1\|\mathbf{x}\|_1 + \lambda_2\|\mathbf{D}\mathbf{x}\|_1\right]. \tag{10}$$

Large λ_1 effects sparsity, and large λ_2 effects smoothness. Since $\|\mathbf{D}\mathbf{x}\|_1 = \sum_{p=1}^{P}\sum_{n=1}^{N-1}|x_{n,p}-x_{n-1,p}|$, the second ℓ_1-norm penalty in (10) captures the sum of total variation penalties.

A couple of remarks are now in order.

Remark 1. One motivation behind SLR-based harmonic retrieval in [10] and [16] is that non-uniform sampling can be accommodated—a case of interest, e.g., in astronomy or when observations are missing. Parametric and subspace-based high-resolution algorithms (such as ESPRIT and MUSIC) can afford non-uniform sampling only under suitable identifiability conditions (such as shift invariance); see e.g., [30]. Similar to [10], [16], the novel FH parameter estimator in (10) remains operational even with non-uniformly sampled data, thanks to the grid-based formulation. In this case, $\boldsymbol{\omega}_n := [e^{j\omega_1\tau_n},\ldots,e^{j\omega_P\tau_n}]^T$, where τ_n denotes the acquisition time of the nth sample.

Remark 2. If $\lambda_2 = 0$, (10) is known as the least-absolute shrinkage and selection operator (Lasso) [33]. With $\lambda_2 \neq 0$, the cost in (10) is similar to the one utilized by the fused Lasso in [19].

The optimization problem in (10) is convex because the cost comprises the sum of an ℓ_1-norm term and an ℓ_2-norm term, both of which are convex by definition; hence, the cost in (10) can be minimized via interior point solvers, which are computationally affordable for small-to-medium size problems [32]. Since the non-differentiable part in (10) is not separable coordinate-wise, convergence to a global optimum of coordinate-descent solvers [35] cannot be invoked for large-size problems. An iterative algorithm to *approximate* the solution of the fused Lasso is developed in [19]. On the other hand, a low-complexity algorithm to solve (10) *exactly* will be derived in Section IV. Before presenting this solution, it is of interest to explore useful properties of the estimator in (10) as a function of the scalars λ_1 and λ_2.

A. Guidelines for Choosing λ_1 and λ_2

Selection of the regularization parameters (λ_1,λ_2) affects critically the performance of the estimator in (10). While under-regularizing may not be sufficient to retrieve the signal of interest, over-regularization can result in poor and biased estimates. Of course, if the number of tones present can be provided *a priori* by other means, e.g., by inspecting the spectrogram, (λ_1,λ_2) can be tuned accordingly by trial and error. But in general, analytical methods to automatically choose the "best" values of λ_1 and λ_2 are not available. In essence, selecting the regularization parameters is more a matter of engineering art, rather than systematic science.

In this subsection, heuristic but useful guidelines will be provided to choose (λ_1,λ_2) based on rigorously established lower bounds of these parameters. To bound λ_1, we will rely on the following result, which was derived in [25].

Proposition 1. *If $\lambda_2 = 0$, then $\hat{\mathbf{x}} = \mathbf{0}_{NP}$ if and only if $\lambda_1 \geq \lambda_1^* := \|\mathbf{W}^H\mathbf{y}\|_\infty$.*

Proposition 1 asserts that if λ_1 is greater than a threshold specified by the regression matrix and the observations, and $\lambda_2 = 0$, then (10) yields estimates that are identically zero. This property of the Lasso has been exploited in [10] to select the penalty parameter λ_1. In the present context of FH signal estimation, the implication is that λ_1 must be chosen strictly less than λ_1^* in order to prevent the all-zero solution. Our extensive simulations suggest that setting λ_1 equal to a small percentage of λ_1^*, say 5%–10%, results in satisfactory estimates; see also Section V.

Turning our attention to bound the selection of λ_2, let $\mathbf{T}_N^\ell$ denote the $N\times N$ lower triangular matrix with all nonzero entries equal to one. Define $\boldsymbol{\Sigma} := \mathbf{T}_N^\ell \otimes \mathbf{I}_P$, and partition the matrix product $\mathbf{W}\boldsymbol{\Sigma}$ into $\mathbf{M}_0 \in \mathbb{C}^{N\times P}$ and $\mathbf{M} \in \mathbb{C}^{N\times(N-1)P}$, so that $[\mathbf{M}_0,\mathbf{M}] := \mathbf{W}\boldsymbol{\Sigma}$. Using these definitions, we have established the following property of the SLR estimator in (10); see Appendix A for the proof.

Proposition 2. *If $\lambda_1 = 0$, and $\mathbf{M}_0$ has full column rank, then $\hat{\mathbf{x}} = [\mathbf{x}_c^T,\ldots,\mathbf{x}_c^T]^T$ with $\mathbf{x}_c := (\mathbf{M}_0^H\mathbf{M}_0)^{-1}\mathbf{M}_0^H\mathbf{y}$, if and only if $\lambda_2 \geq \lambda_2^* := \|\mathbf{M}^H(\mathbf{M}_0\mathbf{x}_c - \mathbf{y})\|_\infty$.*

If λ_2 exceeds a threshold which is specified by the regression matrix and the observations, and $\lambda_1 = 0$, Proposition 2 implies that the estimates in (10) are constant in time; that is, all frequency bins are hop-free. To avoid this trivial (non-FH) solution, the guideline provided by Proposition 2 is that λ_2 must be chosen strictly less than λ_2^*. As with λ_1, the simulations of Section V will demonstrate that setting λ_2 to a small percentage of λ_2^* yields satisfactory estimation performance.

Remark 3. The scalars weighting the regularization terms also affects the bias present in the estimators obtained as in (10). Specifically, note that $\lambda_1\|\mathbf{x}\|_1$ biases $\hat{\mathbf{x}}$ towards zero, which may render the complex exponential amplitude estimates unreliable. While the proposed back-off in selecting the regularization parameters relative to the bounds in Propositions 1–2 can limit this bias, several strategies can be adopted to correct it. A simple way for correcting the bias is to first acquire the hops (or hops plus frequencies) via (10), and then solve a line spectrum (correspondingly, amplitude) estimation problem for each dwell in-between the detected hops. A drawback of this per-dwell approach is that it does not exploit the possible correlation present across adjacent dwells [cf. Section V-C].

Another approach to correct the bias in sparse regression is to retain only the support of (10) and re-estimate the amplitudes via, e.g., LS. Notice that this approach is not directly applicable here because the number of non-zero entries of $\hat{\mathbf{x}}$ in (10) is generally in the order of MN, while the number of equations in (7) is N; that is, the resultant linear regression model is still under-determined. However, one can take advantage of the fact that the vector estimate $\hat{\mathbf{x}}$ is not only sparse but also piecewise constant. To this end, summing the columns of $\mathbf{W}$ corresponding to the entries of $\hat{\mathbf{x}}$ that are equal, it is possible to reduce the number of unknowns.

An alternative approach to reducing the bias is through nonconvex regularization using e.g., the smoothly clipped absolute deviation (SCAD) scheme [18]. SCAD reduces bias without suffering from the inherent limitations of per-dwell processing. Its limitation is that the cost is nonconvex, thus rendering exact minimization problematic due to the presence of local minima. A viable way for retaining the efficiency of convex optimization while simultaneously limiting the bias due to the regularizing term, is to resort to *weighted* ℓ_1 norms [14], [37], [38]. Larger weights are given to terms that are most likely to be zero, while smaller weights are assigned to those that are most likely to be nonzero. Given an initial solution $\hat{\mathbf{x}}^{(0)}$, the weighted ℓ_1 norm is defined as $||\mathbf{x}||_{1,w} := \sum_{k=1}^{NP} w(|\hat{x}_k^{(0)}|)|x_k|$, where $w(\cdot)$ is a decreasing function of its argument (see [14], [37], [38] for three different weight functions). LS, ridge regression, or the (unweighted) estimator in (10) can be used for initialization.

B. Noiseless Case and Perfect Reconstruction

In this subsection, the SLR-estimator is tailored to the case of noiseless data. Consider the model

$$\mathbf{y} = \mathbf{W}\boldsymbol{x}^*. \tag{11}$$

For this noise-free model, the idea is to replace the LS part of the cost in (10) with an exact constraint involving the linear system of equations in (11). Specifically, the proposed modification of (10) is

$$\begin{aligned} \widehat{\mathbf{x}}^* = \arg \min_{\mathbf{x}\in\mathbb{C}^{NP}} & \left[(1-\gamma)||\mathbf{x}||_1 + \gamma||\mathbf{D}\mathbf{x}||_1\right] \\ \text{s.t.} \quad & \mathbf{y} = \mathbf{W}\mathbf{x}. \end{aligned} \tag{12}$$

Since $\mathbf{W}$ is fat, the linear system $\mathbf{y} = \mathbf{W}\mathbf{x}$ admits an infinite number of solutions. The rationale behind (12) is to select the solution that minimizes the cost $(1-\gamma)||\mathbf{x}||_1 + \gamma||\mathbf{D}\mathbf{x}||_1$. The parameter γ is tuned to strike a desirable tradeoff between sparsity and smoothness. Indeed, the larger the γ the smoother the solution, and the smaller the γ the sparser the solution.

The question that arises at this point is whether $\widehat{\mathbf{x}}^*$ coincides with $\boldsymbol{x}^*$. Introducing an auxiliary variable, $\mathbf{g} \in \mathbb{C}^{(N-1)P}$, the problem in (12) can be rewritten as

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{C}^{NP},\mathbf{g}\in\mathbb{C}^{(N-1)P}} \quad & \left[(1-\gamma)||\mathbf{x}||_1 + \gamma||\mathbf{g}||_1\right] \\ \text{s.t.} \quad & \mathbf{y} = \mathbf{W}\mathbf{x} \\ & \mathbf{g} = \mathbf{D}\mathbf{x}. \end{aligned} \tag{13}$$

Defining $\mathbf{s} := [(1-\gamma)\mathbf{x}^T, \gamma\mathbf{g}^T]^T \in \mathbb{C}^{(2N-1)P}$, $\mathbf{r} := [\mathbf{y}^T, \mathbf{0}_{(N-1)P}^T]^T \in \mathbb{C}^{(N-1)P+N}$ and

$$\mathbf{A} := \begin{bmatrix} \frac{1}{1-\gamma}\mathbf{W} & \mathbf{0}_{NP\times NP} \\ -\frac{1}{1-\gamma}\mathbf{D} & \frac{1}{\gamma}\mathbf{I}_{(N-1)P} \end{bmatrix} \in \mathbb{C}^{(N-1)P+N\times(2N-1)P}$$

the optimization in (13) can be recast as a standard sparse signal reconstruction problem, namely

$$\begin{aligned} \min_{\mathbf{s}} \quad & ||\mathbf{s}||_1 \\ \text{s.t.} \quad & \mathbf{r} = \mathbf{A}\mathbf{s}. \end{aligned} \tag{14}$$

Sufficient conditions ensuring equivalence of (14) with the ℓ_0-norm based optimization for exact recovery are based either on the restricted isometry property (RIP) or the incoherence conditions on the columns of $\mathbf{A}$; see [4], [11], [12] and references therein. Having shown that (12) reduces to (14) establishes that *any* scheme available for checking the RIP or incoherence conditions applies here too. In addition, the simulations of Section V indicate that if the γ parameter is chosen properly, the formulation in (12) can perfectly reconstruct the true $\boldsymbol{x}^*$.

It is also worth stressing that matrix $\mathbf{W}$ (and thus $\mathbf{A}$) in certain applications is prescribed, and it is not up to the designer's choice. For these applications, one focuses on the ℓ_1-norm based sparse recovery and the aforementioned equivalence as well as the RIP and incoherence conditions are not an issue.

Nonetheless, when the designer has the freedom to select $\mathbf{W}$, it is certainly interesting to know how the choice of $\mathbf{W}$ affects these sufficient conditions. With regards to checking their validity, it is also pertinent to underscore that RIP analysis entails the nonzero support (S) of the vector $\mathbf{s}$, as well as a "sufficiently small" constant (δ_S). Hence, whether $\mathbf{A}$ satisfies the RIP depends on the underlying S and the chosen δ_S; and this is NP-hard to check [12], [13]. Checking the incoherence conditions is feasible in polynomial time, but even when the columns of $\mathbf{A}$ are "sufficiently incoherent," the implied RIP (bounds) may yield values for S and δ_S, which may not be always practical [4].

Yet another major consideration constraining the choice of $\mathbf{A}$ in practice is the density of the grid points forming the entries of $\mathbf{W}$. This density affects the attainable frequency resolution, which has to be balanced with the size of $\mathbf{W}$ and the associated complexity in solving the optimization problem (12).

C. Generalization to PPH Signals

Signals described by (3) and (6) are encountered in many engineering applications. However, certain modulation types induce both continuous and abrupt frequency changes that do not obey this model. The goal of this section is to broaden the scope of the novel SLR-based FH estimation approach to polynomial-phase hopping (PPH) signals.

Polynomial-phase models are very important in radar signal processing, where relative velocity and acceleration are key parameters of interest; e.g., see [3], [20] and references therein. Due to inertia, however, model parameters change slowly in radar applications. Instead of radar, the motivation for PPH comes from *chirp modulation*, a digital communication technique originally proposed during the 1960s. Various generalizations and applications of chirp modulation have appeared since, including *chirp spread spectrum multiplexing*—see [6], [21], [28] and references therein. Instead of using a windowed carrier as the basic pulse, chirp modulation uses a windowed chirp, whose frequency is linearly swept up or down to represent a logical 0 or 1. Multiple slopes (and offsets) can be used for M-ary modulation, and/or to multiplex different users. Chirp modulation has a number of desirable properties relative to traditional FH, including robustness to Doppler and fading. In chirp modulation (multiplexing), abrupt changes of the PPH parameters occur at the boundary between symbol periods ("dwells").

The discrete-time model of a PPH signal can be written as

$$\begin{aligned} \widetilde{s}_n := \sum_{m=1}^{M_k} a_{m,k} e^{j\sum_{g=1}^{G} q_{m,k}^{(g)} n^g/g!}, \\ n_{k-1} \le n < n_k, \quad n = 0,\ldots,N-1. \end{aligned} \tag{15}$$

Observe that the model in (15) coincides with (3) when $G = 1$, while for $G = 2$ it includes also a linear-chirp hopping signal. For simplicity in exposition, the case $G = 2$ is detailed next.

Suppose that the parameters $\{q_{m,k}^{(1)}\}$ and $\{q_{m,k}^{(2)}\}$ in (15) belong to finite sets $\mathcal{W}_1 := \{\omega_1^{(1)}, \ldots, \omega_{P_1}^{(1)}\}$ and $\mathcal{W}_2 := \{\omega_1^{(2)}, \ldots, \omega_{P_2}^{(2)}\}$, respectively. Again, if this is not the case, $\mathcal{W}_1$ and $\mathcal{W}_2$ represent dense grids that approximate the true parameters $\{q_{m,k}^{(1)}\}$ and $\{q_{m,k}^{(2)}\}$. If this is the case, define $\boldsymbol{\omega}_n^{(1)} := [e^{j\omega_1^{(1)}n}, \ldots, e^{j\omega_{P_1}^{(1)}n}]^T$, $\boldsymbol{\omega}_n^{(2)} := [e^{j\omega_1^{(2)}n^2/2}, \ldots, e^{j\omega_{P_2}^{(2)}n^2/2}]^T$, and $\widetilde{\boldsymbol{\omega}}_n := \boldsymbol{\omega}_n^{(1)} \otimes \boldsymbol{\omega}_n^{(2)}$, with $\widetilde{\boldsymbol{\omega}}_n \in \mathbb{C}^{P_1P_2}$. Upon properly defining $\widetilde{\boldsymbol{x}}_n \in \mathbb{C}^{P_1P_2}$, the discrete-time signal in (15) can be written as

$$\widetilde{s}_n := \widetilde{\boldsymbol{\omega}}_n^T \widetilde{\boldsymbol{x}}_n. \tag{16}$$

At the receiver, $\widetilde{s}_n$ is corrupted by additive noise $\widetilde{v}_n$, and observed as $\widetilde{y}_n = \widetilde{s}_n + \widetilde{v}_n$. Defining $\widetilde{\mathbf{w}}_n := [\underbrace{\mathbf{0}_{P_1P_2}^T, \ldots, \mathbf{0}_{P_1P_2}^T}_{n}, \widetilde{\boldsymbol{\omega}}_n^T, \underbrace{\mathbf{0}_{P_1P_2}^T, \ldots, \mathbf{0}_{P_1P_2}^T}_{N-n-1}]^T \in \mathbb{C}^{P_1P_2N}$, $\widetilde{\mathbf{W}} := [\widetilde{\mathbf{w}}_0, \ldots, \widetilde{\mathbf{w}}_{N-1}]^T$, and letting $\widetilde{\mathbf{y}} := [\widetilde{y}_0, \ldots, \widetilde{y}_{N-1}]^T$ and $\widetilde{\mathbf{v}} := [\widetilde{v}_0, \ldots, \widetilde{v}_{N-1}]^T$ denote the observation and noise vectors, the received vector becomes

$$\widetilde{\mathbf{y}} = \widetilde{\mathbf{W}} \widetilde{\boldsymbol{x}}^* + \widetilde{\mathbf{v}} \tag{17}$$

where $\widetilde{\boldsymbol{x}}^* := [\widetilde{\boldsymbol{x}}_0^T, \ldots, \widetilde{\boldsymbol{x}}_{N-1}^T]^T \in \mathbb{C}^{P_1P_2N}$. Again, $\widetilde{\boldsymbol{x}}^*$ is sparse and piecewise constant. Letting $\widetilde{\mathbf{d}}_1 := [-1, \underbrace{0, \ldots, 0}_{P_1P_2-1}, 1, \underbrace{0, \ldots, 0}_{(N-1)P_1P_2-1}]^T$, and

$$\widetilde{\mathbf{D}} := \begin{bmatrix} \widetilde{\mathbf{d}}_1^T \\ \widetilde{\mathbf{d}}_1^{T(1)} \\ \cdots \\ \widetilde{\mathbf{d}}_1^{T((N-1)P_1P_2-1)} \end{bmatrix} \tag{18}$$

the proposed SLR-based estimator for PPH signals is

$$\widehat{\widetilde{\mathbf{x}}} := \arg \min_{\mathbf{x} \in \mathbb{C}^{NP_1P_2}} \left[\frac{1}{2} \|\widetilde{\mathbf{y}} - \widetilde{\mathbf{W}}\mathbf{x}\|_2^2 + \lambda_1 \|\mathbf{x}\|_1 + \lambda_2 \|\widetilde{\mathbf{D}}\mathbf{x}\|_1 \right]. \tag{19}$$

Clearly, by simply replacing the regression matrix $\mathbf{W}$ with $\widetilde{\mathbf{W}}$, the SLR estimator in (10) developed for FH signals carries over to the wider class of PPH signals.

IV. Efficient Implementation via ADMoM

A low-complexity algorithm is developed in this section to obtain $\hat{\mathbf{x}}$ in (10). The crux of the advocated solver of the optimization problem in (10) is to show how the alternating direction method of multipliers (ADMoM) [8, pp. 243–253] can be applied to the problem at hand.

Consider re-writing the minimization in (10) with the use of auxiliary variables $\mathbf{z}$ and $\mathbf{u}$, as

$$\begin{aligned} [\hat{\mathbf{x}}, \hat{\mathbf{z}}, \hat{\mathbf{u}}] &:= \arg \min_{\mathbf{x},\mathbf{z},\mathbf{u}} \left[\frac{1}{2} \|\mathbf{y} - \mathbf{W}\mathbf{x}\|_2^2 + \lambda_1 \|\mathbf{z}\|_1 + \lambda_2 \|\mathbf{u}\|_1 \right] \\ \text{s.t.} \quad & \mathbf{z} = \mathbf{x}, \mathbf{u} = \mathbf{D}\mathbf{x}. \end{aligned} \tag{20}$$

Associating Lagrange multipliers $(\boldsymbol{\zeta}, \boldsymbol{\mu})$ with the equality constraints, the quadratically augmented Lagrangian of the problem in (20) is

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \mathbf{z}, \mathbf{u}, \boldsymbol{\zeta}, \boldsymbol{\mu}) &= \frac{1}{2} \|\mathbf{y} - \mathbf{W}\mathbf{x}\|_2^2 + \lambda_1 \|\mathbf{z}\|_1 + \lambda_2 \|\mathbf{u}\|_1 \\ &\quad + \Re\{\boldsymbol{\zeta}^H(\mathbf{x} - \mathbf{z}) + \boldsymbol{\mu}^H(\mathbf{D}\mathbf{x} - \mathbf{u})\} \\ &\quad + \frac{c}{2} \left(\|\mathbf{x} - \mathbf{z}\|_2^2 + \|\mathbf{D}\mathbf{x} - \mathbf{u}\|_2^2 \right). \end{aligned} \tag{21}$$

Selecting any positive number c as well as arbitrary initial vectors $\mathbf{z}^{(0)}, \mathbf{u}^{(0)}, \boldsymbol{\zeta}^{(0)}, \boldsymbol{\mu}^{(0)}$, the ADMoM algorithm iterates over the following steps:

$$\mathbf{x}^{(i)} = \arg \min_{\mathbf{x}} \mathcal{L}\left(\mathbf{x}, \mathbf{z}^{(i-1)}, \mathbf{u}^{(i-1)}, \boldsymbol{\zeta}^{(i-1)}, \boldsymbol{\mu}^{(i-1)}\right) \tag{22}$$

$$\left[\mathbf{z}^{(i)}, \mathbf{u}^{(i)}\right] = \arg \min_{\mathbf{z},\mathbf{u}} \mathcal{L}\left(\mathbf{x}^{(i)}, \mathbf{z}, \mathbf{u}, \boldsymbol{\zeta}^{(i-1)}, \boldsymbol{\mu}^{(i-1)}\right) \tag{23}$$

$$\boldsymbol{\zeta}^{(i)} = \boldsymbol{\zeta}^{(i-1)} + c\left(\mathbf{x}^{(i)} - \mathbf{z}^{(i)}\right) \tag{24}$$

$$\boldsymbol{\mu}^{(i)} = \boldsymbol{\mu}^{(i-1)} + c\left(\mathbf{D}\mathbf{x}^{(i)} - \mathbf{u}^{(i)}\right). \tag{25}$$

With the auxiliary variables and the multipliers available from the $(i-1)$st iteration, the wanted vector $\mathbf{x}^{(i)}$ at iteration (i) is obtained as in (22). Because $\mathcal{L}$ in (21) is linear-quadratic in $\mathbf{x}$, in Appendix B is shown that this convex minimization problem accepts a closed-form solution, namely

$$\begin{aligned} \mathbf{x}^{(i)} = (\mathbf{W}^H\mathbf{W} + c\mathbf{D}^H\mathbf{D} + c\mathbf{I}_{NP})^{-1} \Big(& \mathbf{W}^H\mathbf{y} - \boldsymbol{\zeta}^{(i-1)} \\ & - \mathbf{D}^H\boldsymbol{\mu}^{(i-1)} + c\mathbf{z}^{(i-1)} + c\mathbf{D}^H\mathbf{u}^{(i-1)} \Big). \end{aligned} \tag{26}$$

Having found $\mathbf{x}^{(i)}$ and with the multipliers fixed from the $(i-1)$st iteration, the auxiliary variables $[\mathbf{z}^{(i)}, \mathbf{u}^{(i)}]$ at iteration (i) are subsequently obtained as in (23). After neglecting irrelevant terms, the pertinent minimization problem reduces to

$$\begin{aligned} \left[\mathbf{z}^{(i)}, \mathbf{u}^{(i)}\right] &= \arg \min_{\mathbf{z},\mathbf{u}} \left[\left(\lambda_1 \|\mathbf{z}\|_1 - \Re\left\{ {\boldsymbol{\zeta}^{(i-1)}}^H \mathbf{z} \right\} + \frac{c}{2} \left(\|\mathbf{x}^{(i)} - \mathbf{z}\|_2^2 \right) \right) \right. \\ &\quad \left. + \left(\lambda_2 \|\mathbf{u}\|_1 - \Re\{{\boldsymbol{\mu}^{(i-1)}}^H \mathbf{u}\} + \frac{c}{2} \left(\left\| \mathbf{D}\mathbf{x}^{(i)} - \mathbf{u} \right\|_2^2 \right) \right) \right]. \end{aligned} \tag{27}$$

Clearly, the cost in (27) can be minimized separately in $\mathbf{z}$ and $\mathbf{u}$. Since the resulting minimizers w.r.t. $\mathbf{z}$ and $\mathbf{u}$ are found analogously, only the minimization over $\mathbf{z}$ is detailed for brevity. Noting that $\lambda_1\|\mathbf{z}\|_1 - \Re\{{\boldsymbol{\zeta}^{(i-1)}}^H \mathbf{z}\} + (c)/(2)(\|\mathbf{x}^{(i)} - \mathbf{z}\|_2^2) = \sum_{k=1}^{NP} (\lambda_1 |z_k| - \Re\{{\zeta_k^{(i-1)}}^* z_k\} + (c)/(2)|x_k^{(i)} - z_k|^2)$, the minimization in (27) over $\mathbf{z}$ can be solved coordinate-wise; that is, for each coordinate $k = 1, \ldots, NP$, the problem to solve is

$$z_k^{(i)} = \arg \min_{z_k} \left[\lambda_1 |z_k| - \Re\left\{ {\zeta_k^{(i-1)}}^* z_k \right\} + \frac{c}{2} \left(|x_k^{(i)} - z_k|^2 \right) \right]. \tag{28}$$

Albeit non-differentiable, the scalar convex cost in (28) can be solved in closed form. Specifically, we show in Appendix B that the solution of (28) is given by

$$z_k^{(i)} = \begin{cases} 0, & \text{if } x_k^{(i)} + \frac{\zeta_k^{(i-1)}}{c} = 0 \\ \frac{x_k^{(i)} + c^{-1}\zeta_k^{(i-1)}}{|x_k^{(i)} + c^{-1}\zeta_k^{(i-1)}|} \max\left(\left| x_k^{(i)} + \frac{\zeta_k^{(i-1)}}{c} \right| - \frac{\lambda_1}{c}, 0 \right), & \text{otherwise} \end{cases}$$
$$:= \operatorname{shrink}\left(x_k^{(i)} + \frac{\zeta_k^{(i-1)}}{c}, \frac{\lambda_1}{c} \right) \tag{29}$$

which corresponds to the complex version of the soft shrinkage operator in, e.g., [17]. Collecting the coordinate minimizers in a vector, the closed-form solution of (23) can be compactly expressed using the vector shrinkage operator with entries as in (29), namely

$$\mathbf{z}^{(i)} = \operatorname{shrink}\left(\mathbf{x}^{(i)} + \frac{\boldsymbol{\zeta}^{(i-1)}}{c}, \frac{\lambda_1}{c} \right) \tag{30}$$

$$\mathbf{u}^{(i)} = \operatorname{shrink}\left(\mathbf{D}\mathbf{x}^{(i)} + \frac{\boldsymbol{\mu}^{(i-1)}}{c}, \frac{\lambda_2}{c} \right). \tag{31}$$

Furthermore, note that the Lagrange multipliers are subsequently updated as in (24) and (25), which are first-order, least mean-square (LMS)-like iterations.

In a nutshell, the primal problem in (10) can be decoupled in the minimization problems (22) and (23), which entail closed-form solutions per iteration plus simple Lagrange multiplier updates implemented as in (24) and (25). Apart from simplicity in implementation, this iterative algorithm for SLR-based FH parameter estimation is provably convergent to $\widehat{\mathbf{x}}$ in (10), since the ADMoM is guaranteed to converge to a global minimizer for convex functions [8, p. 253]. Summarizing, we have established the following.

Proposition 3. *For any* $c > 0$, $\mathbf{z}^{(0)}$, $\mathbf{u}^{(0)}$, $\boldsymbol{\zeta}^{(0)}$ *and* $\boldsymbol{\mu}^{(0)}$, *the iterates* $\mathbf{x}^{(i)}$ *in (26),* $\mathbf{z}^{(i)}$ *and* $\mathbf{u}^{(i)}$ *in (30) and (31), as well as* $\boldsymbol{\zeta}^{(i)}$ *and* $\boldsymbol{\mu}^{(i)}$ *in (24) and (25), are all convergent. Specifically,* $\mathbf{x}^{(i)}$ *converges to the solution of (10); that is,* $\lim_{i\to\infty} \mathbf{x}^{(i)} = \widehat{\mathbf{x}}$.

It is worth stressing at this point that the ADMoM solver of the SLR problem in (10) and the associated convergence result in Proposition 3 are not confined to the FH/PPH signal estimation problem dealt with here. In fact, they carry over to all problems that fused Lasso can be applied [19]. An extra attractive feature of the ADMoM algorithm in (26)–(31) is that the matrix to be inverted in (26) remains fixed during the iterations; hence, the matrix inversion in (26) can be performed off-line. With $(\mathbf{W}^H\mathbf{W} + c\mathbf{D}^H\mathbf{D} + c\mathbf{I}_{NP})^{-1}$ obtained off-line, the computational complexity per iteration is dominated by the multiplication in (26), that is $\mathcal{O}(N^2P^2)$. Furthermore, since the matrix $\mathbf{W}^H\mathbf{W} + c\mathbf{D}^H\mathbf{D} + c\mathbf{I}_{NP}$ is very sparse [cf. (7) and (8)], solving (26) for large N and P can be facilitated via computationally efficient solvers of sparse linear systems of equations, such as the conjugate gradient algorithm [7, p. 130]. In addition, the ADMoM can afford a convergent *distributed* implementation which is also robust to noisy links [27]—a useful attribute when estimation is to be performed using wireless sensor networks, where observations are spatially distributed.

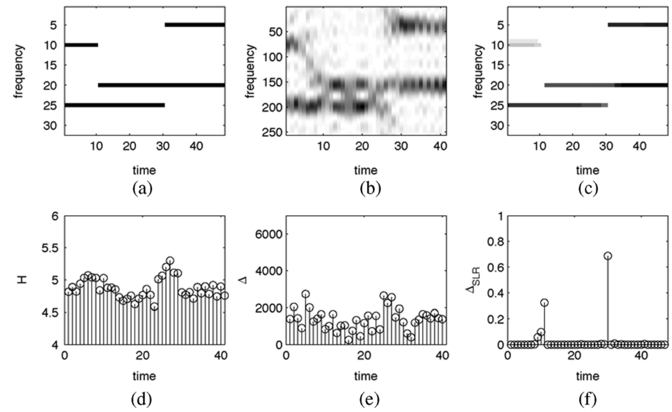

Fig. 1. Two hopping complex exponentials. (a) True time-frequency pattern; (b) spectrogram; (c) sparse linear regression estimates; (d) entropy of the (normalized) spectrogram estimates, H; (e) sum of the difference of consecutive columns of the spectrogram, Δ; (f) $\Delta_{\mathrm{SLR}\,n} = \sum_{p=1}^{P} |\widehat{x}_{n+1,p} - \widehat{x}_{n,p}|^2$ $(M = 2)$.

A. Noiseless Case

Similar to (10), the estimator in (12) admits an efficient implementation via the ADMoM. Indeed, mimicking the steps used to solve the minimization problem in (10), the following holds true.

Proposition 4. *For any* $c > 0$, $\mathbf{z}^{(0)}$, $\mathbf{u}^{(0)}$, $\boldsymbol{\zeta}^{(0)}$, $\boldsymbol{\mu}^{(0)}$, $\boldsymbol{\rho}^{(0)}$ *the iterates*

$$\mathbf{x}^{(i)} = \frac{1}{c}(\mathbf{W}^H\mathbf{W} + \mathbf{D}^H\mathbf{D} + \mathbf{I}_{NP})^{-1}(c\mathbf{W}^H\mathbf{y} + c\mathbf{z}^{(i-1)} + c\mathbf{D}^H\mathbf{u}^{(i-1)} - \boldsymbol{\zeta}^{(i-1)} - \mathbf{D}^H\boldsymbol{\mu}^{(i-1)} - \mathbf{W}^H\boldsymbol{\rho}^{(i-1)}) \tag{32}$$

$$\mathbf{z}^{(i)} = \operatorname{shrink}\left(\mathbf{x}^{(i)} + \frac{\boldsymbol{\zeta}^{(i-1)}}{c}, \frac{1-\gamma}{c} \right) \tag{33}$$

$$\mathbf{u}^{(i)} = \operatorname{shrink}\left(\mathbf{D}\mathbf{x}^{(i)} + \frac{\boldsymbol{\mu}^{(i-1)}}{c}, \frac{\gamma}{c} \right) \tag{34}$$

$$\boldsymbol{\zeta}^{(i)} = \boldsymbol{\zeta}^{(i-1)} + c\left(\mathbf{x}^{(i)} - \mathbf{z}^{(i)} \right) \tag{35}$$

$$\boldsymbol{\mu}^{(i)} = \boldsymbol{\mu}^{(i-1)} + c\left(\mathbf{D}\mathbf{x}^{(i)} - \mathbf{u}^{(i)} \right) \tag{36}$$

$$\boldsymbol{\rho}^{(i)} = \boldsymbol{\rho}^{(i-1)} + c(\mathbf{W}\mathbf{x}^{(i)} - \mathbf{y}) \tag{37}$$

converge to the solution of (12); that is, $\lim_{i\to\infty} \mathbf{x}^{(i)} = \widehat{\mathbf{x}}^*$.

V. Simulations

In this section, the developed algorithms are tested in several scenarios.

A. Frequency Hopping and Hop Timing Estimation

The signal of interest in (3) and (6) consists of two hopping tones, while the grid of carriers is chosen to be $\mathcal{W} = \{\frac{2p-P-1}{P}\pi\}_{p=1}^{P}$ with $P = 32$, and $N = 48$. The first FH tone is generated to be active on the 10th carrier in the interval [0, 9], and then hops to the 20th carrier during the interval [10, 47]. The second hopping tone occupies the 25th carrier in the interval [0, 29], and the 5th carrier in the interval [30, 47]. The two FH signals are in-phase and have equal amplitude.

The true time-frequency pattern of the signal of interest is depicted in Fig. 1(a). (Here and in what follows the squared modulus of the $\boldsymbol{X}^*$ entries is plotted.) The spectrogram obtained with

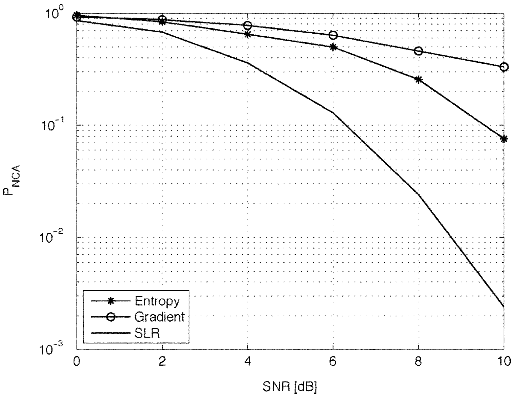

Fig. 2. Two hopping complex exponentials. Probability of incorrect detection versus SNR ($M = 2$).

$N_1 = 8, N_2 = 256$, $L = 1$, and using a rectangular window is shown in Fig. 1(b) at $\mathrm{SNR} := 10\log_{10}(\|\boldsymbol{x}^*\|_2^2)/(N\sigma^2) = 10$ dB. In Fig. 1(c), the modulus of the estimate in (10) rearranged in matrix form, i.e., $\widehat{\mathbf{X}} = [\widehat{\mathbf{x}}_0, \ldots, \widehat{\mathbf{x}}_{N-1}]$, is depicted for $\lambda_1 = \lambda_1^*/20$ and $\lambda_2 = \lambda_2^*/10$, with λ_1^* (λ_2^*) as in Proposition 1 (correspondingly 2). Here and in what follows these scaling parameters (λ_1, λ_2) are used unless specified otherwise. ADMoM updates in (22)–(25) are terminated either after a fixed number of (here 10^3) iterations, or, by using the following stopping criterion: $(\|\mathbf{x}^{(i)} - \mathbf{x}^{(i-1)}\|_2^2)/(\|\mathbf{x}^{(i)}\|_2^2) < 10^{-8}$. Observe that $\widehat{\mathbf{X}}$ is a far better estimate of the true time-frequency pattern than the spectrogram.

Fig. 1(d) and (e) depicts, respectively, the entropy sequence H_ℓ [22], and the gradient sequence Δ_ℓ [24] versus time. Notice that the peaks of these statistics provide estimates of the system-wise hopping instants. In Fig. 1(f), the statistic $\Delta_{\mathrm{SLR}n} = \sum_{p=1}^{P} |\widehat{x}_{n+1,p} - \widehat{x}_{n,p}|^2$ for $n = 0, \ldots, N-2$ is plotted. Clearly, $\Delta_{\mathrm{SLR}n}$ represents a better statistic than H_ℓ and Δ_ℓ to estimate the hopping instants.

Performance of the spectrogram- and SLR-based hop timing estimators is next assessed via Monte Carlo simulations. The signal of interest is the one in Fig. 1(a) and, for simplicity, the number of system-wise hops ($K = 2$) is assumed known. The hop timing estimates are obtained by picking the K peaks of H_ℓ, Δ_ℓ, and $\Delta_{\mathrm{SLR}n}$. To pick the K peaks of those statistics, the following steps are repeated K times: i) The maximum value of the statistic is found; ii) its index is stored; and iii) the value of this entry and the adjacent N_z entries are set to zero. Correct acquisition (CA) corresponds to having each of the K estimates of the hopping instants less than N_w samples away from the associated true hopping instants. Fig. 2 depicts the probability of incorrect acquisition ($P_{\mathrm{NCA}} = 1 - P_{\mathrm{CA}}$) versus SNR (averaged over 10^4 noise realizations) for the two spectrogram-based estimators, and the novel estimator in (10) with $N_z = 5$ and $N_w = 3$. Observe that the entropy-based technique outperforms the gradient-based one, and the SLR estimator achieves the best overall performance.

Next, the tested signal of interest comprises $M = 3$ FH tones: the two of Fig. 1(a) plus a third one that occupies the 15th carrier in the interval [0, 19], and then hops to the 30th carrier in the interval [20, 47]. With the parameters used in Fig. 1, the resulting signal together with the spectrogram, the SLR estimates and the decision statistics are depicted in Fig. 3. Fig. 4 shows the probability of incorrect acquisition versus SNR. Observe that the performance of the entropy-based estimator degrades while the SLR estimator achieves satisfactory performance.

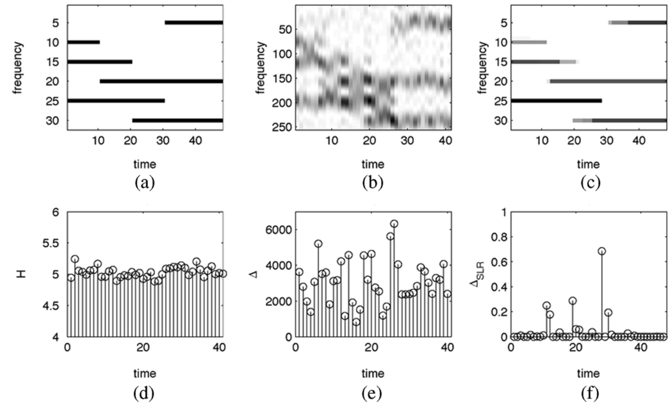

Fig. 3. Three hopping complex exponentials. (a) True time-frequency pattern; (b) spectrogram; (c) sparse linear regression estimates; (d) entropy of the (normalized) spectrogram estimates, H; (e) Sum of the difference of consecutive columns of the spectrogram, Δ; (f) $\Delta_{\mathrm{SLR}n} = \sum_{p=1}^{P} |\widehat{x}_{n+1,p} - \widehat{x}_{n,p}|^2$ ($M = 3$).

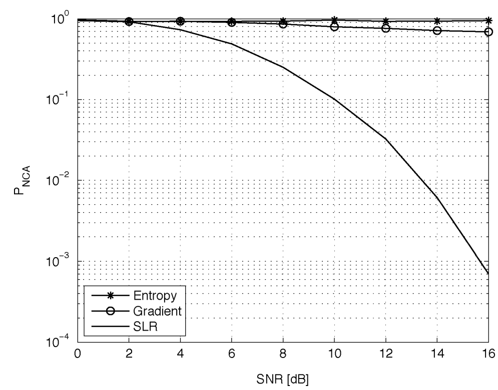

Fig. 4. Three hopping complex exponentials. Probability of incorrect detection versus SNR versus SNR ($M = 3$).

So far, the true signal comprised a fixed number of complex exponentials hopping only once. Next, a case is tested where the number of complex exponentials varies across dwells and more hops occur. The first complex exponential occupies the 10th carrier over the interval [0,9], then hops to the 20th carrier over [10, 34], and to the 30th carrier over [35, 47]. The second complex exponential occupies the 15th carrier over [0,19] and then it disappears, while the third complex exponential occupies the 25th carrier over the interval [0, 29], and the 5th carrier over [30, 47]. With the parameters identical to those used in Fig. 1, the resulting signal along with the spectrogram, the SLR estimates, and the decision statistics are depicted in Fig. 5. The selection strategies advocated in Section III-A are seen effective also in this case of multiple hops and a varying number of tones per dwell.

B. Robustness to Sources of Model Mismatch

In Section V-A the signal of interest was a superposition of ideal complex exponentials that hopped within a known frequency grid. In this subsection, the estimator in (10) is tested in the presence of various sources of mismatch between the model in (3), (4), and (6) and the signal of interest. First, a carrier mismatch is considered.

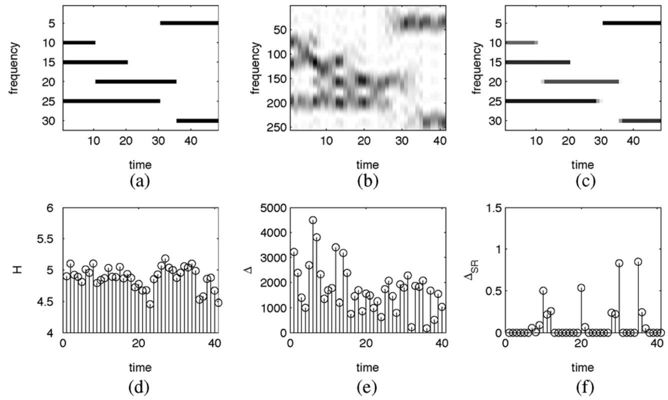

Fig. 5. Time-varying number of hopping complex exponentials. (a) True time-frequency pattern; (b) spectrogram; (c) sparse linear regression estimates; (d) entropy of the (normalized) spectrogram estimates, H; (e) sum of the difference of consecutive columns of the spectrogram, Δ; (f) $\Delta_{\mathrm{SLR}\,n} = \sum_{p=1}^{P} |\widehat{x}_{n+1,p} - \widehat{x}_{n,p}|^2$.

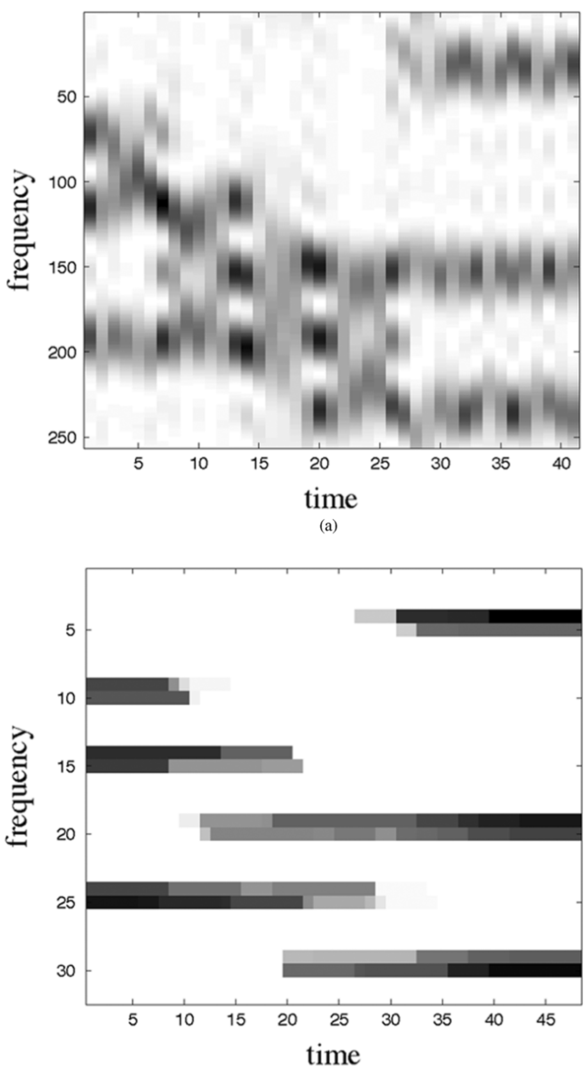

Fig. 6. Mismatch due to carrier shift: true frequencies are not within the grid. Spectrogram (top) and SLR estimates (bottom).

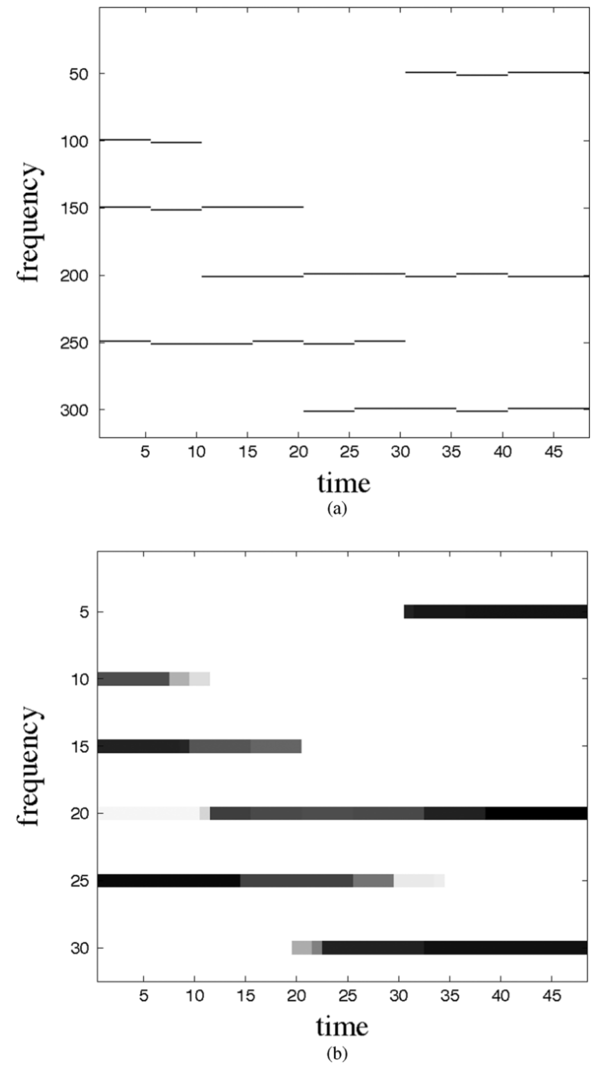

Fig. 7. Mismatch due to B-FSK modulation: signal frequencies not within the grid. True frequencies (top) and SLR estimates (bottom).

It is first assumed that each carrier in Fig. 3(a) is shifted by $-\frac{\pi}{P}$ so that none of the actual frequencies is within the grid; instead, they lie in the middle of two grid carriers. As a consequence, (6) is not exact but only an approximation. Fig. 6(a) and (b) shows the spectrogram and the SLR estimates for SNR = 10 dB, respectively. Notice that the SLR estimator picks the two closest carriers. While the time and frequency resolution of the SLR estimator *is* still better than the spectrogram, increasing the density of the frequency grid can further improve performance.

Next, a mismatch due to frequency modulation is considered. Each exponential signal in Fig. 3(a) is the carrier of a binary frequency shift keying (B-FSK) modulation, where each symbol lasts $N_s = 5$ sampling instants, and the two symbols undergo a frequency shift of $\pm\pi/5P$, which corresponds to $1/10$ of the carrier spacing. Despite the fact that the estimator in (10) may recover such a signal if a frequency grid 10 times denser than $\mathcal{W} = \{(2p-P-1)/(P)\pi\}_{p=1}^{P}$ was adopted, the question considered here is whether the SLR estimator with $\mathcal{W}$ can "filter out" the modulation and recover the actual tones. Fig. 7(a) and (b) shows the true time-frequency pattern along with the SLR estimate for SNR = 10 dB, respectively. It is clear that the SLR estimator recovers the carrier hops only, because frequency variations due to modulation are negligible relative to the grid spacing.

Next, a *near–far* scenario is considered. Wireless propagation may cause fluctuations of the received signal amplitude due to time- and frequency-selective fading. Frequency selectivity

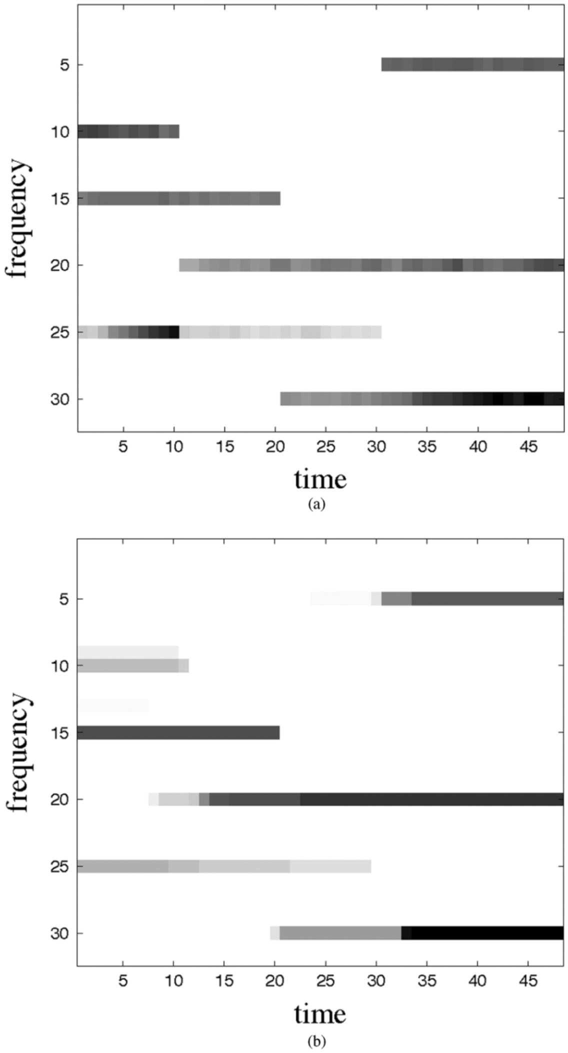

Fig. 8. Mismatch due to fading. The complex exponential amplitudes are time varying. True signal (top) and SLR estimates (bottom).

means that different tones are subject to different attenuation and phase shift; time selectivity means that the attenuation and phase shift of a given tone vary with time. If this is the case, exact reconstruction of the signals of interest is impossible, because the number of unknowns is much larger than the number of observations. In many cases, however, fading only induces relatively small fluctuations around a nominal amplitude. Indeed, Fig. 8(a) and (b) depicts the signal of interest affected by time-varying fading, and its SLR estimate for SNR = 10 dB. The nonzero time-varying amplitude coefficients in (6) are generated as a first-order Gauss–Markov process, i.e., $x_{n,p} = \beta x_{n-1,p} + e_n$ with $\beta = 0.99$, $x_{0,p} \sim \mathcal{CN}(0,1)$, and $e_n \sim \mathcal{CN}(0, 1-\beta^2)$. Interestingly, the developed estimator is able to recover the true time-localized frequency pattern, and "average out" small amplitude variations due to fading.

C. Noiseless Case

In this subsection, the noiseless reconstruction algorithm of (12) is tested. The signal of interest is the one in Fig. 3(a). Fig. 9 shows the squared error (SE), $||\mathbf{x}^{(i)} - \boldsymbol{x}^*||_2^2$, versus the iteration index (i) of the algorithm in Proposition 4 for $c = 0.25$, and various values of γ. Surprisingly, if properly tuned, the SLR estimator in (12) can perfectly recover the signal $\boldsymbol{x}^*$.

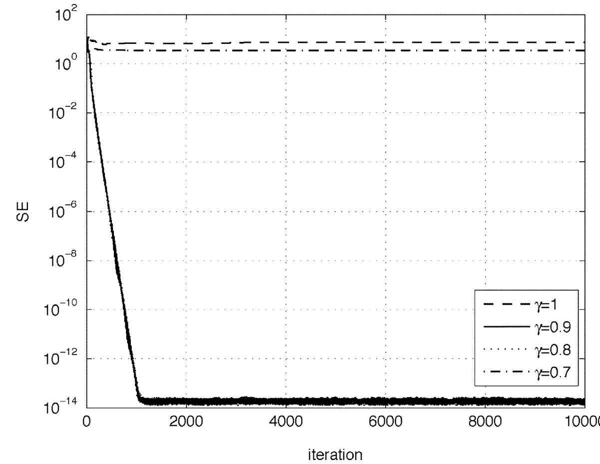

Fig. 9. Evolution of the squared-error in the noise-free case.

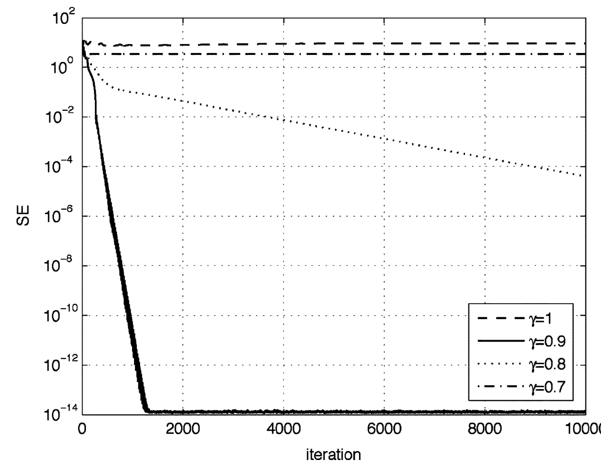

Fig. 10. Evolution of the squared-error in the noise-free case (dwell-wise identifiability is not met).

Once hop timing is acquired, one can solve a set of harmonic retrieval problems on a per-dwell basis to obtain refined frequency and complex amplitude estimates [22], [24]. This kind of processing is clearly suboptimum, since it does not take into account observations of adjacent dwells—which contain information about the frequency content in the dwell of interest. Parameter identifiability for the per-dwell harmonic retrieval problem goes back to Caratheodory [15]; see also [31]. Assuming distinct dwell frequencies, it turns out that identifiability of this nonlinear problem boils down to counting equations-versus-unknowns: for M complex exponentials within the dwell, one needs at least $M + \lceil M/2 \rceil$ observations (length of the dwell).[2] Next, a case similar to Fig. 3(a) is considered except that the first signal hops to the 20th carrier at time 18, so that only one sample is taken during the second dwell. In this case, per-dwell processing fails to recover the signal within the second dwell even with perfect knowledge of the hop timing. Fig. 10 shows the SE versus the iteration index (i) of the algorithm in Proposition 4 with $c = 0.25$. Observe that the SLR estimator is capable

[2]There are three real unknowns per complex exponential, and two real equations per complex measurement.

of recovering the signal of interest perfectly. This is possible because the estimator in (12) exploits the frequency smoothness along adjacent dwells. Clearly, this does not constitute an identifiability claim; what it does demonstrate, however, is that the SLR estimator is capable of perfect recovery in situations where per-dwell processing unequivocally fails.

D. PPH Signal Estimation

In this subsection, the generalization of the SLR estimator to PPH signals is tested. A mixture of FH and linear chirp hopping signals is considered. Specifically, the chosen parameters are: $G = 2$, $N = 48$, $P_1 = 20$, $P_2 = 3$, $\mathcal{W}_1 = \{(2p - P_1 - 1)/(P_1)\pi\}_{p=1}^{P_1}$, and $\mathcal{W}_2 = \{-(4\pi)/(P_1), 0, (4\pi)/(P_1)\}$. The signal of interest is a superposition of FH signals in $\mathcal{W}_1$, and hopping chirp signals in $\mathcal{W}_2$. The particular choice of $\mathcal{W}_2$ is not instrumental in any way other than allowing for easy visualization: it guarantees that the instantaneous frequency of the chirp signals at every sample point belongs to $\mathcal{W}_1$.

The signal of interest was generated as the superposition of two signals. The first occupies the 7th carrier in the interval [0,24], and then hops to the 12th carrier in the interval [25, 47]. The second signal occupies the 15th carrier in the interval [0, 14], and then turns into a linearly decreasing chirp starting from the 18th carrier in the interval [15,31], and finally to a linearly increasing chirp starting from the 5th carrier in the interval [32, 47]. Fig. 11(a) and (b) shows the true time-frequency pattern along with the SLR estimate for SNR = 10 dB, $\lambda_1 = 0.3$, and $\lambda_2 = 1.5$. It is worth noting that since $P_1 P_2 > N$ the assumption of Proposition 2 is not met. As expected, the SLR estimator correctly recovers the frequency content of the signal of interest.

VI. Conclusion

A novel technique was introduced to estimate FH signal parameters based on sparse linear regression. Earlier approaches rely upon the spectrogram of the received signal, at least for coarse acquisition. The estimation task was formulated here as an under-determined linear regression problem with a dual sparsity penalty. Its exact solution was obtained using the ADMoM. Guidelines were provided to select the regularization parameters, and the estimation approach was generalized to PPH signals. Simulations demonstrated that the novel technique outperforms spectrogram-based estimators by a significant margin, especially with regard to hop-timing estimation. A modification of the novel estimator in the noiseless case revealed that the SLR estimator can perfectly recover the signal of interest, even when per-dwell identifiability fails—thus holding greater promise than per-dwell processing approaches. The ADMoM-based algorithm developed here for FH/PPH signal estimation can be ported to other problems, such as applications of fused Lasso [19]. Interesting extensions of this work can be pursued in *slowly* time-varying line spectrum estimation.[3]

[3]The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U. S. Government.

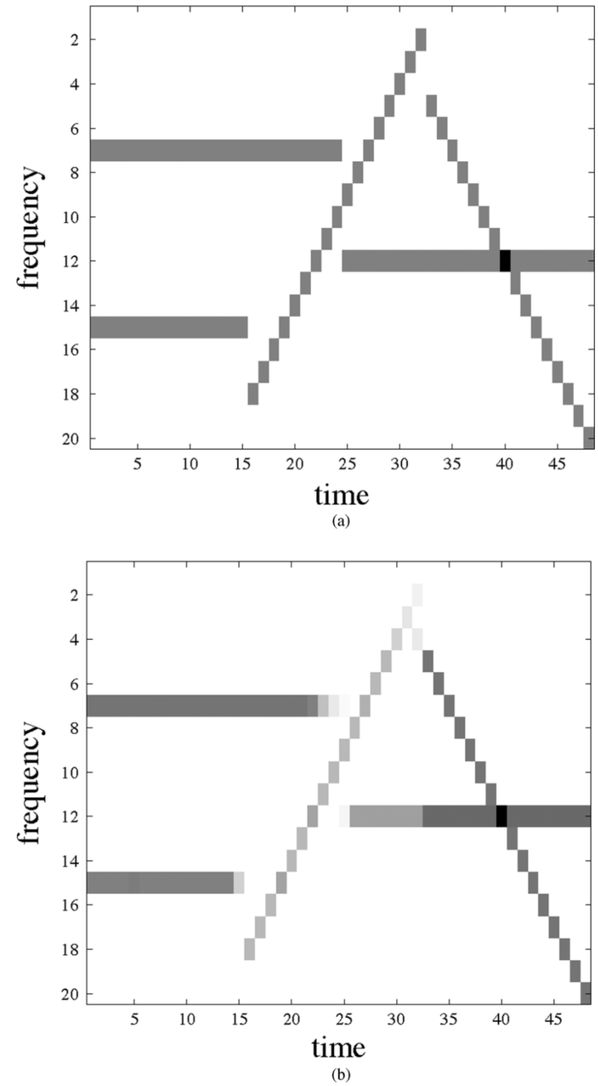

Fig. 11. Estimation of polynomial-phase hopping signals. True signal (top) and SLR estimates (bottom).

Appendix

A. Proof of Proposition 2

With $\lambda_1 = 0$, the problem in (10) simplifies to

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{C}^{NP}} \left[\frac{1}{2} \|\mathbf{y} - \mathbf{W}\mathbf{x}\|_2^2 + \lambda_2 \|\mathbf{D}\mathbf{x}\|_1 \right]. \tag{38}$$

Recall that $\mathbf{x} := [\mathbf{x}_0^T, \ldots, \mathbf{x}_{N-1}^T]^T$, and let

$$\mathbf{h}_n := \mathbf{x}_n - \mathbf{x}_{n-1}, \quad n = 1, \ldots, N-1. \tag{39}$$

Defining $\mathbf{h} := [\mathbf{h}_1^T, \ldots, \mathbf{h}_{N-1}^T]^T$, it holds that

$$\mathbf{x} = \boldsymbol{\Sigma} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{h} \end{bmatrix} \tag{40}$$

and

$$\|\mathbf{D}\mathbf{x}\|_1 = \sum_{n=1}^{N-1} \sum_{p=1}^{P} |x_{n,p} - x_{n-1,p}| = \|\mathbf{h}\|_1. \tag{41}$$

Hence, an equivalent form of (38) is

$$[\widehat{\mathbf{x}}_0, \widehat{\mathbf{h}}] = \arg \min_{\mathbf{x}_0 \in \mathbb{C}^P, \mathbf{h} \in \mathbb{C}^{(N-1)P}} J(\mathbf{x}_0, \mathbf{h}) \tag{42}$$

where

$$J(\mathbf{x}_0, \mathbf{h}) := \frac{1}{2} \left\| \mathbf{y} - \mathbf{W}\boldsymbol{\Sigma} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{h} \end{bmatrix} \right\|_2^2 + \lambda_2 \|\mathbf{h}\|_1. \tag{43}$$

The necessary and sufficient first-order optimality condition for $[\widehat{\mathbf{x}}_0, \widehat{\mathbf{h}}]$ to be the (unconstrained) minimum of $J(\mathbf{x}_0, \mathbf{h})$, is that the subgradient of $J(\mathbf{x}_0, \mathbf{h})$ evaluated at $[\widehat{\mathbf{x}}_0, \widehat{\mathbf{h}}]$ contains the zero vector [26, p. 92], i.e.,

$$\check{\nabla} J(\widehat{\mathbf{x}}_0, \widehat{\mathbf{h}}) \ni \mathbf{0}_{NP}. \tag{44}$$

Defining

$$\mathbf{r} := \boldsymbol{\Sigma}^H \mathbf{W}^H \left(\mathbf{W}\boldsymbol{\Sigma} \begin{bmatrix} \widehat{\mathbf{x}}_0 \\ \widehat{\mathbf{h}} \end{bmatrix} - \mathbf{y} \right) \tag{45}$$

the subgradient of $J(\mathbf{x}_0, \mathbf{h})$ evaluated at $[\widehat{\mathbf{x}}_0, \widehat{\mathbf{h}}]$ can be expressed as

$$\check{\nabla} J(\widehat{\mathbf{x}}_0, \widehat{\mathbf{h}}) = \mathbf{r} + \lambda_2 \mathbf{b} \tag{46}$$

where the kth entry of $\mathbf{b} \in \mathbb{C}^{NP}$ is

$$b_k := \begin{cases} 0, & k = 1, \ldots, P \\ \frac{\hat{h}_k}{|\hat{h}_k|}, & k = P+1, \ldots, NP, \hat{h}_k \neq 0 \\ s_k, & k = P+1, \ldots, NP, \hat{h}_k = 0 \end{cases} \tag{47}$$

with $s_k \in \mathbb{C}$ such that $|s_k| \leq 1$.

From (46) and (47), (44) translates to the following conditions:

c1) $r_k = 0$, for $k = 1, \ldots, P$; and,

c2) $\begin{cases} r_k + \lambda_2 \frac{\hat{h}_k}{|\hat{h}_k|} = 0, & \text{if } \hat{h}_k \neq 0 \\ |r_k| \leq \lambda_2, & \text{if } \hat{h}_k = 0 \end{cases}$ for $k = P + 1, \ldots, NP$.

The *constant* (i.e., hop-free) solution corresponds to having $\widehat{\mathbf{x}}_0$ and $\widehat{\mathbf{h}} = \mathbf{0}_{(N-1)P}$. Thus, c1) implies that

$$\mathbf{M}_0^H (\mathbf{M}_0 \widehat{\mathbf{x}}_0 - \mathbf{y}) = \mathbf{0}_P \tag{48}$$

which is uniquely satisfied by $\widehat{\mathbf{x}}_0 = \mathbf{x}_c$ since $\mathbf{M}_0$ has full column rank. Hence, $\widehat{\mathbf{x}}_0 = \mathbf{x}_c$ and $\widehat{\mathbf{h}} = \mathbf{0}_{(N-1)P}$ if and only if c2) is satisfied, which corresponds to $|r_k| \leq \lambda_2$ for $k = P+1, \ldots, NP$, or equivalently, $\lambda_2 \geq \lambda_2^*$.

B. Proof of Proposition 3

It suffices to show that the problems in (22) and (23) admit the closed-form solution in (26) and (30)–(31), respectively. After skipping constant terms, (22) can be written as

$$\mathbf{x}^{(i)} = \arg\min_{\mathbf{x}} \left[\frac{1}{2} \|\mathbf{y} - \mathbf{W}\mathbf{x}\|_2^2 + \Re \left\{ \boldsymbol{\zeta}^{(i-1)^H} \mathbf{x} + \boldsymbol{\mu}^{(i-1)^H} \mathbf{D}\mathbf{x} \right\} \right.$$
$$\left. + \frac{c}{2} \left(\left\| \mathbf{x} - \mathbf{z}^{(i-1)} \right\|_2^2 + \left\| \mathbf{D}\mathbf{x} - \mathbf{u}^{(i-1)} \right\|_2^2 \right) \right]. \tag{49}$$

Upon equating the gradient of the convex differentiable cost to zero, the expression in (26) is readily obtained.

In Section IV we have showed that (23) can be separated in scalar problems of the form in (28). Next, we show that (28) admits the closed-form solution in (29). Because the cost in (28) is convex but non-differentiable, the necessary and sufficient condition for $z_k^{(i)}$ to attain its minimum is [26, p. 92]

$$\begin{cases} \lambda_1 \frac{z_k^{(i)}}{\left|z_k^{(i)}\right|} - \zeta_k^{(i-1)} - c\left(x_k^{(i)} - z_k^{(i)}\right) = 0, & \text{if } z_k^{(i)} \neq 0 \\ \left|\zeta_k^{(i-1)} + c x_k^{(i)}\right| \leq \lambda_1, & \text{if } z_k^{(i)} = 0. \end{cases} \tag{50}$$

Substituting $z_k^{(i)}$ with the expression in (29), it is easy to verify that the conditions in (50) are satisfied. This completes the proof of the proposition.

REFERENCES

[1] L. Aydin and A. Polydoros, "Hop-timing estimation for FH signals using a coarsely channelized receiver," *IEEE Trans. Commun.*, vol. 44, no. 4, pp. 516–526, Apr. 1996.

[2] S. Barbarossa and A. Scaglione, "Parameter estimation of spread spectrum frequency-hopping signals using time-frequency distributions," *Proc. Signal Processing Advances in Wireless Commun.*, pp. 213–216, Apr. 1997.

[3] S. Barbarossa, A. Scaglione, and G. B. Giannakis, "Product high-order ambiguity function for multicomponent polynomial phase signal modeling," *IEEE Trans. Signal Process.*, vol. 46, no. 3, pp. 691–708, Mar. 1998.

[4] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Trans. Signal Process.*, 10.1109/TSP.2010.2052460, to be published.

[5] J. Berent, P. L. Dragotti, and T. Blu, "Sampling piecewise sinusoidal signals with finite rate of innovation methods," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 613–625, Feb. 2010.

[6] A. Berni and W. Gregg, "On the utility of chirp modulation for digital signaling," *IEEE Trans. Commun.*, vol. 21, no. 6, pp. 748–751, Jun. 1973.

[7] D. Bertsekas, *Non Linear Programming*, 2nd ed. Singapore: Athena Scientific, 2003.

[8] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, 2nd ed. Singapore: Athena Scientific, 1999.

[9] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations: Theory, algorithms, and performance bounds," *IEEE Signal Process. Mag.*, vol. 25, pp. 31–40, Mar. 2008.

[10] S. Bourguignon, H. Carfantan, and J. Idier, "A sparsity-based method for the estimation of spectral lines from irregularly sampled data," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 575–585, Dec. 2007.

[11] T. T. Cai, G. Xu, and J. Zhang, "On recovery of sparse signals via ℓ_1 minimization," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3388–3397, Jul. 2009.

[12] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[13] E. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, pp. 21–30, Mar. 2008.

[14] E. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, pp. 877–905, 2007.

[15] C. Carathéodory and L. Fejér, "Uber den Zusammenghang der Extemen von harmonischen Funktionen mit ihren Koeffizienten und uber den Picard-Landauschen Satz," *Rendiconti del Circolo Matematico di Palermo*, vol. 32, pp. 218–239, 1911.

[16] S. S. Chen and D. L. Donoho, "Application of basis pursuit in spectrum estimation," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Seattle, WA, May 1998, vol. 3, pp. 1865–1868.

[17] M. Elad, B. Matalon, J. Shtok, and M. Zibulevsky, "A wide-angle view at iterated shrinkage algorithms," presented at the SPIE (Wavelet XII), San Diego, CA, Aug. 2007.

[18] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Stat. Assoc.*, vol. 96, pp. 1348–1360, 2001.

[19] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, "Pathwise coordinate optimization," *Ann. Appl. Stat.*, vol. 1, pp. 302–332, Dec. 2007.

[20] F. Gini and G. B. Giannakis, “Hybrid FM-polynomial phase signal modeling: Parameter estimation and Cramer–Rao bounds,” *IEEE Trans. Signal Process.*, vol. 47, no. 2, pp. 363–377, Feb. 1999.

[21] S. Hengstler, D. P. Kasilingam, and A. H. Costa, “A novel chirp modulation spread spectrum technique for multiple access,” in *Proc. 7th IEEE Int. Symp. Spread Spectrum Techniques Applications*, 2002, vol. 1, pp. 73–77.

[22] X. Liu, N. D. Sidiropoulos, and A. Swami, “Blind high-resolution localization and tracking of multiple frequency hopped signals,” *IEEE Trans. Signal Process.*, vol. 50, no. 4, pp. 889–901, Apr. 2002.

[23] X. Liu, N. D. Sidiropoulos, and A. Swami, “Joint hop timing and frequency estimation for collision resolution in frequency-hopped networks,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 3063–3074, Nov. 2005.

[24] X. Liu, J. Li, and X. Ma, “An EM algorithm for blind hop timing estimation of multiple FH signals using an array system with bandwidth mismatch,” *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 2545–2554, Sep. 2007.

[25] M. R. Osborne, B. Presnell, and B. A. Turlach, “On the LASSO and its dual,” *J. Comput. Graph. Stat.*, vol. 9, pp. 319–337, Jun. 2000.

[26] A. Ruszczynski, *Nonlinear Optimization*. Princeton, NJ: Princeton Univ. Press, 2006.

[27] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, “Consensus in Ad Hoc WSNs with noisy links—Part I: Distributed estimation of deterministic signals,” *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 350–364, Jan. 2008.

[28] H. Shen, S. Machineni, C. Gupta, and A. Papandreou-Suppappola, “Time-varying multichirp rate modulation for multiple access systems,” *IEEE Signal Process. Lett.*, vol. 11, no. 5, pp. 497–500, May 2004.

[29] M. K. Simon, U. Cheng, L. Aydin, A. Polydoros, and B. K. Levitt, “Hop timing estimation for noncoherent frequency-hopped M-FSK intercept receivers,” *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1144–1154, Feb./Mar./Apr. 1995.

[30] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Englewood Cliffs, NJ: Prentice-Hall, 2005.

[31] P. Stoica and T. Soderstom, “Parameter identifiability problems in signal processing,” *Proc. Inst. Electr. Eng. F—Radar Sonar Navig.*, vol. 141, pp. 133–136, 1994.

[32] J. Sturm, “Using Sedumi 1.02, a matlab toolbox for optimization over symmetric cones,” *Optim. Methods Softw.*, vol. 11, no. 12, pp. 625–653, 1999.

[33] R. Tibshirani, “Regression shrinkage and selection via the Lasso,” *J. Roy. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[34] D. J. Torrieri, “Mobile frequency-hopping CDMA systems,” *IEEE Trans. Commun.*, vol. 48, no. 8, pp. 1318–1327, Aug. 2000.

[35] P. Tseng, “Convergence of block coordinate descent method for non-differentiable minimization,” *J. Optim. Theory Appl.*, vol. 109, pp. 475–494, Jun. 2001.

[36] A. Valyrakis, E. E. Tsakonas, N. D. Sidiropoulos, and A. Swami, “Stochastic modelling and particle filtering algorithms for tracking a frequency-hopped signal,” *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3108–3118, Aug. 2009.

[37] H. Zou, “The adaptive Lasso and its oracle properties,” *J. Amer. Stat. Assoc.*, vol. 101, no. 476, pp. 1418–1429, 2006.

[38] H. Zou and R. Li, “One-step sparse estimates in nonconcave penalized likelihood models,” *Ann. Stat.*, vol. 36, no. 4, pp. 1509–1533, 2008.

Daniele Angelosante (M’09) was born in Frosinone, Italy, on June 1, 1981. He received the B.Sc., Laurea Magistrale, and Ph.D. degrees in telecommunication engineering from the University of Cassino, Italy, in 2003, 2005, and 2009, respectively, and the M.Sc. degree in electrical engineering with specialization in signal and information processing for communication from the University of Aalborg, Denmark, in 2005.

From April to June 2007, he worked at the University Pompeu Fabra, Spain. In 2008, he was a visiting scholar at the University of Minnesota, Minneapolis. He is currently a Postdoctoral Associate with the Department of Electrical and Computer Engineering of the University of Minnesota. His research interests lie in the areas of statistical signal processing, with emphasis on wireless communications, tracking and compressed sensing.

Georgios B. Giannakis (F’97) received the Diploma degree in electrical engineering from the National Technical University of Athens, Greece, in 1981 and the M.Sc. degree in electrical engineering, the M.Sc. degree in mathematics, and the Ph.D. degree in electrical engineering from the University of Southern California (USC) in 1983, 1986, and 1986, respectively.

Since 1999, he has been a Professor with the University of Minnesota, where he now holds an ADC Chair in Wireless Telecommunications in the Electric and Computer Engineering Department and serves as Director of the Digital Technology Center. His general interests span the areas of communications, networking and statistical signal processing subjects on which he has published more than 300 journal papers, 500 conference papers, two edited books, and two research monographs. Current research focuses on compressive sensing, cognitive radios, network coding, cross-layer designs, mobile ad hoc networks, wireless sensor, and social networks.

Dr. Giannakis is the (co-)inventor of 20 patents issued and the (co)recipient of seven paper awards from the IEEE Signal Processing (SP) and Communications Societies, including the G. Marconi Prize Paper Award in Wireless Communications. He also received Technical Achievement Awards from the SP Society (2000), from EURASIP (2005), a Young Faculty Teaching Award, and the G. W. Taylor Award for Distinguished Research from the University of Minnesota. He is a Fellow of EURASIP, and has served the IEEE in a number of posts, including that of Distinguished Lecturer for the IEEE SP Society.

Nicholas D. Sidiropoulos (F’09) received the Diploma degree from the Aristotelian University of Thessaloniki, Greece, and M.Sc. and Ph.D. degrees from the University of Maryland at College Park (UMCP), in 1988, 1990, and 1992, respectively, all in electrical engineering.

He has been a Postdoctoral Fellow (1994–1995) and Research Scientist (1996–1997) at the Institute for Systems Research, UMCP, and has held positions as Assistant Professor, Department of Electrical Engineering, University of Virginia-Charlottesville (1997–1999), and Associate Professor, Department of Electrical and Computer Engineering, University of Minnesota—Minneapolis (2000–2002). Since 2002, he has been a Professor in the Department of Electronic and Computer Engineering at the Technical University of Crete, Chania-Crete, Greece, and Adjunct Professor at the University of Minnesota. His current research interests are primarily in signal processing for communications, convex approximation of NP-hard problems, cross-layer resource allocation for wireless networks, and multiway analysis.

Prof. Sidiropoulos received the U.S. NSF CAREER Award in 1998 and the IEEE Signal Processing Society (SPS) Best Paper Award in 2001 and in 2007. He has served as a Distinguished Lecturer of the IEEE SPS (2008–2009), and as Chair (2007–2008) of the Signal Processing for Communications and Networking Technical Committee of the IEEE SPS. He has also served as Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2000–2006) and the IEEE SIGNAL PROCESSING LETTERS (2000–2002), and currently serves on the Editorial Board of the *IEEE Signal Processing Magazine*.