

Estimating Signals With Finite Rate of Innovation From Noisy Samples: A Stochastic Algorithm

Vincent Y. F. Tan, *Student Member, IEEE*, and Vivek K Goyal, *Senior Member, IEEE*

Abstract—As an example of the recently introduced concept of rate of innovation, signals that are linear combinations of a finite number of Diracs per unit time can be acquired by linear filtering followed by uniform sampling. However, in reality, samples are rarely noiseless. In this paper, we introduce a novel *stochastic algorithm* to reconstruct a signal with finite rate of innovation from its *noisy samples*. Even though variants of this problem have been approached previously, satisfactory solutions are only available for certain classes of sampling kernels, for example, kernels that satisfy the Strang–Fix condition. In this paper, we consider the infinite-support Gaussian kernel, which does not satisfy the Strang–Fix condition. Other classes of kernels can be employed. Our algorithm is based on Gibbs sampling, a Markov chain Monte Carlo method. Extensive numerical simulations demonstrate the accuracy and robustness of our algorithm.

Index Terms—Analog-to-digital conversion, Gibbs sampling, Markov chain Monte Carlo, sampling.

I. INTRODUCTION

THE celebrated Nyquist–Shannon sampling theorem [1], [2]¹ states that a signal $x(t)$ known to be bandlimited to Ω_{\max} is uniquely determined by samples of $x(t)$ spaced $1/(2\Omega_{\max})$ apart. The textbook reconstruction procedure is to feed the samples as impulses to an ideal lowpass (sinc) filter. Furthermore, if $x(t)$ is not bandlimited or the samples are noisy, introducing prefiltering by the appropriate sinc *sampling kernel* gives a procedure that finds the orthogonal projection to the space of Ω_{\max} -bandlimited signals. Thus the noisy case is handled by simple, linear, time-invariant processing.

Sampling has come a long way since the sampling theorem, but until recently the results have mostly applied only to signals contained in shift-invariant subspaces [4]. Moving out of this restrictive setting, Vetterli *et al.* [5] showed that it is possible to develop sampling schemes for certain classes of nonbandlimited signals that are not subspaces. As described in [5], for reconstruction from samples, it is necessary for the class of signals to

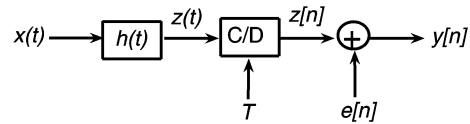


Fig. 1. Block diagram showing our problem setup. $x(t)$ is a signal with FRI given by (2) and $h(t)$ is the Gaussian filter with width σ_h given by (3). Discrete-time sequence $e[n]$ is i.i.d. Gaussian noise with standard deviation σ_e and $y[n]$ are the noisy samples. From $y[n]$, we will estimate the parameters that describe $x(t)$, namely, $\{(c_k, t_k)\}_{k=1}^K$, and the standard deviation of the noise σ_e .

have *finite rate of innovation* (FRI). The paradigmatic example is the class of signals expressed as

$$x(t) = \sum_k c_k \phi(t - t_k) \quad (1)$$

where $\phi(t)$ is some known function. For each term in the sum, the signal has two real parameters c_k and t_k . If the density of t_k s (the number that appear per unit of time) is finite, the signal has FRI. It is shown constructively in [5] that the signal can be recovered from (noiseless) uniform samples of $x(t) * h(t)$ (at a sufficient rate) when $\phi(t) * h(t)$ is a sinc or Gaussian function. Results in [6] are based on similar reconstruction algorithms and greatly reduce the restrictions on the sampling kernel $h(t)$.

In practice, though, acquisition of samples is not a noiseless process. For instance, an analog-to-digital converter (ADC) has several sources of noise, including thermal noise, aperture uncertainty, comparator ambiguity, and quantization [7]. Hence, samples are inherently noisy. This motivates our central question: Given a signal model (i.e., a class of signals with FRI) and a noise model, how well can we approximate the parameters that describe the signal and hence the signal itself? In this paper, we address this question and develop a novel algorithm to estimate the signal from the noisy samples, which we will denote $y[n]$ (see Fig. 1).

A. Related Work and Motivation

Signals with FRI were initially introduced by Vetterli *et al.* [5]. The reconstruction schemes hinged on identifying algebraically independent parameters of the signals, e.g., the weights $\{c_k\}$ and time locations $\{t_k\}$ in (1). In the seminal paper on FRI, the sampling kernel for finite signals was chosen to be either the sinc or the Gaussian. An annihilating filter approach led to an elegant algebraic solution via polynomial root finding and least squares. The authors alluded to the noisy case and suggested the use of the singular value decomposition (SVD) for dealing with noisy samples. We will show that, in fact, this method is ill-conditioned because root-finding is itself

Manuscript received December 21, 2007; revised May 29, 2008. Current version published September 17, 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Chong-Meng Samson See. The work of V. Y. F. Tan was supported by A*STAR, Singapore, and the U.S. Army Research Office under MURI Grant W911NF-06-0076. The work of V. K. Goyal was supported by the National Science Foundation under CAREER Award CCF-643836.

The authors are with the Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: vtan@mit.edu; vgoyal@mit.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2008.928510

¹A more expansive term could be the Whittaker–Nyquist–Kotelnikov–Shannon sampling theorem; see, e.g., [3] and [4].

not at all robust to noise. Thus it is not amenable to practical implementations, for instance, on an ADC.

Subsequently, Dragotti *et al.* [6] examined acquisition of the same signals with an eye toward implementability of the sampling kernel. Instead of using the sinc and Gaussian kernels (which do not have compact support), the authors limited the choice of kernels to functions satisfying the Strang–Fix conditions [8] (e.g., splines and scaling functions), exponential splines [9], and functions with rational Fourier transforms. They combined the moment-sampling and annihilating filter approaches to solve for the parameters. Somewhat analogous for multidimensional signals is the extension of work by Maravić and Vetterli [10] by Shukla and Dragotti [11].

In this paper, we use the Gaussian sampling kernel. Though the Gaussian has infinite support, it can be well approximated by its truncated version. Hence, we believe that one can draw general insights from the analysis of using Gaussian filters and the subsequent reconstruction of the signal from its noisy samples $y[n]$. More importantly, unlike with previous approaches, the sampling kernel plays no fundamental role in the reconstruction algorithm. We use the Gaussian kernel because of its prominence in earlier work and the intuitiveness of its information spreading properties.

For the noisy case, Maravić and Vetterli [12] and Ridolfi *et al.* [13] proposed and solved a related problem. Instead of modeling the noise at the *output*, they considered the scenario where $x(t)$, the signal in question, is corrupted by additive white noise $e(t)$. Clearly, $\tilde{x}(t) = x(t) + e(t)$ does not belong to the class of signals with FRI. However, in [12], novel algebraic/subspace-based approaches solve the sampling problem in the Laplace domain, and these methods achieve some form of optimality. Even though the models differ, we have also applied the technique described in [12] for comparison. In [13], various algorithms such as ESPRIT and MUSIC were used and comparisons made. The authors concluded that, in the noisy signal case, the parameters can be recovered when the signal is sampled at a rate below that prescribed by the Shannon–Nyquist theorem but at a factor above the critical rate.

Our algorithm is based on Gibbs sampling (GS) [14], [15], a Markov chain Monte Carlo (MCMC) method used to sample from multivariate densities. It is assumed that while it may be too complex to sample directly from a given multivariate density, it may be easier to sample from univariate (or conditional densities with lower dimensions). The Hammersley–Clifford theorem guarantees that after the burn-in period (i.e., upon convergence to the stationary distribution of the Markov chain), iterates of the Gibbs sampler are, in fact, samples drawn from the original density. In particular, this convergence implies that GS techniques are robust to initialization.

B. Our Contributions

In this paper, we model acquisition imperfection as additive noise $e[n]$ affecting the samples $y[n]$ (not a continuous-time noise affecting $x(t)$ directly) and use the *Gaussian sampling kernel*. We demonstrate that with a fully Bayesian approach based on Gibbs sampling, we are able to estimate signal parameters and hence the signal itself. In particular, our stochastic

approach effectively circumvents the ill-conditioning of the problem, whereas algebraic methods do not.

Importantly, our algorithm is not constrained to work with the Gaussian kernel. *Any* kernel can be employed because the formulation of the Gibbs sampler does not depend on the specific form of the kernel $h(t)$. Also, our approach is able to estimate the standard deviation of the noise process σ_e .

C. Paper Organization

The rest of this paper is organized as follows. In Section II, we will formally state the problem and define the notation to be used in the rest of this paper. We proceed to delineate our algorithm in Section III: a stochastic optimization procedure based on Gibbs sampling followed by least squares estimation. We report the results of extensive numerical experiments on synthetic as well as real data in Section IV. Also in Section IV we highlight some of the main deficiencies in [5], which motivate the need for new algorithms for recovering the parameters of a signal with FRI given noisy samples $y[n]$. We conclude our discussion in Section V and provide directions for further research.

II. PROBLEM DEFINITION AND NOTATION

The basic setup is shown in Fig. 1. As mentioned in the Introduction, we consider a class of signals characterized by a finite number of parameters. In this paper, similar to [5], [12], and [6], the class is the weighted sum of K Diracs²

$$x(t) = \sum_{k=1}^K c_k \delta(t - t_k). \quad (2)$$

The signal to be estimated $x(t)$ is filtered using a Gaussian low-pass filter

$$h(t) = \exp\left(-\frac{t^2}{2\sigma_h^2}\right) \quad (3)$$

with width σ_h to give the signal $z(t)$. Even though $h(t)$ does not have compact support, it can be approximated well by a version that is truncated to have compact support. The filtered signal $z(t)$ is uniformly sampled with period T to obtain $z[n] = z(nT)$ for $n = 0, 1, \dots, N-1$. Finally, additive white Gaussian noise $e[n]$ is added to $z[n]$ to give $y[n]$. Therefore, the whole acquisition process from $x(t)$ to $\{y[n]\}_{n=0}^{N-1}$ can be represented by the model \mathcal{M}

$$\mathcal{M}: \quad y[n] = \sum_{k=1}^K c_k \exp\left(-\frac{(nT - t_k)^2}{2\sigma_h^2}\right) + e[n] \quad (4)$$

for $n = 0, 1, \dots, N-1$. The amount of noise added is a function of σ_e . We define the signal-to-noise ratio (SNR) in decibels as

$$\text{SNR} \triangleq 10 \log_{10} \left(\frac{\sum_{n=0}^{N-1} |z[n]|^2}{\sum_{n=0}^{N-1} |z[n] - y[n]|^2} \right) \text{ dB}. \quad (5)$$

²The use of a Dirac delta simplifies the discussion. It can be replaced by a known pulse $g(t)$ and then absorbed into the sampling kernel $h(t)$, yielding an effective sampling kernel $g(t) * h(t)$.

In the sequel, we will use boldface to denote vectors. In particular

$$\mathbf{y} = [y[0], y[1], \dots, y[N-1]]^\top \quad (6)$$

$$\mathbf{c} = [c_1, c_2, \dots, c_K]^\top \quad (7)$$

$$\mathbf{t} = [t_1, t_2, \dots, t_K]^\top. \quad (8)$$

We will sometimes use $\boldsymbol{\theta} = \{\mathbf{c}, \mathbf{t}, \sigma_e\}$ to denote the complete set of decision variables. We will be measuring the performance of our reconstruction algorithms by using the normalized reconstruction error

$$\mathcal{E} \triangleq \frac{\int_{-\infty}^{\infty} |z_{\text{est}}(t) - z(t)|^2 dt}{\int_{-\infty}^{\infty} |z(t)|^2 dt} \quad (9)$$

where $z_{\text{est}}(t)$ is the reconstructed version of $z(t)$. By construction $\mathcal{E} \geq 0$; the closer \mathcal{E} is to zero, the better the reconstruction algorithm. The problem can be summarized as follows. Given $\mathbf{y} = \{y[n] \mid n = 0, \dots, N-1\}$ and the model \mathcal{M} , estimate the parameters $\{(c_k, t_k)\}_{k=1}^K$. Also estimate the noise variance σ_e^2 .

Ideally, we would like to minimize \mathcal{E} in (9) directly, but this does not seem to be tractable since the dependence of $y[n]$ on $\{t_k\}_{k=1}^K$ is highly nonlinear. Thus, we propose the use of a stochastic algorithm (known as the Gibbs sampler) for the maximum likelihood estimation of $\{t_k\}_{k=1}^K$. The Gibbs sampler is a proxy for minimizing \mathcal{E} . This is followed by linear least squared error (LLSE) estimation of $\{c_k\}_{k=1}^K$ as a tractable and effective means for approximate minimization of \mathcal{E} .

III. PRESENTATION OF THE GIBBS SAMPLER AND LEAST SQUARES ESTIMATION

In this section, we will describe the stochastic optimization procedure based on Gibbs sampling to estimate $\boldsymbol{\theta} = \{\mathbf{c}, \mathbf{t}, \sigma_e\}$. The overall procedure is given in Algorithm 1. We will first present a self-contained introduction of Gibbs sampling for readers who may not be familiar with this technique.

Algorithm 1: Parameter Estimation and Signal Reconstruction Algorithm

Require: Data \mathbf{y} , Model \mathcal{M}

- 1: Obtain estimates $\{\hat{t}_k\}_{k=1}^K$ and $\hat{\sigma}_e$ using the Gibbs sampler detailed in Algorithm 2 with the data \mathbf{y} and the model \mathcal{M} given in (4). This is elaborated in Section III-A and -B.
 - 2: Obtain estimates $\{\hat{c}_k\}_{k=1}^K$ using a linear least squares estimation procedure and $\{\hat{t}_k\}_{k=1}^K$ from the Gibbs sampler. This is elaborated in Section III-C.
 - 3: Compute $z_{\text{est}}(t) = \hat{x}(t) * h(t)$ given the parameters $\{(\hat{c}_k, \hat{t}_k)\}_{k=1}^K$ and the known pulse $h(t)$.
 - 4: Compute reconstruction error \mathcal{E} given in (9).
-

A. Gibbs Sampling (GS)

Gibbs sampling is an algorithm used to generate samples from any multivariate distribution. It falls under the class of algorithms known as MCMC methods, which were first studied by the statistical physics community [16] and then later in the statistics community [14], [17], [18]. In particular, Gibbs sampling has been used extensively and successfully in image [14] and audio restoration [15]. The basis for Gibbs sampling is the Hammersley–Clifford theorem [19], which states that given the data \mathbf{y} , the conditional densities $p_i(\theta_i \mid \boldsymbol{\theta}_{\{j \neq i\}}, \mathbf{y}, \mathcal{M})$ contain sufficient information to produce samples from the joint density $p(\boldsymbol{\theta} \mid \mathbf{y}, \mathcal{M})$. These samples can then be used to compute minimum mean squared error (MMSE) estimates for the parameters $\boldsymbol{\theta}$. Furthermore, the joint density $p(\boldsymbol{\theta} \mid \mathbf{y}, \mathcal{M})$ can be directly derived from the conditional densities $p_i(\theta_i \mid \boldsymbol{\theta}_{\{j \neq i\}}, \mathbf{y}, \mathcal{M})$.

1) *An Example of Gibbs Sampling at Work:* We now present an example to show how the Gibbs sampler works. For simplicity, we choose $p(\boldsymbol{\theta} \mid \mathbf{y}, \mathcal{M})$ to be a two-dimensional zero-mean Gaussian distribution ($\boldsymbol{\theta} = (\theta_1, \theta_2)$) with covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \quad (10)$$

where $\sigma_1 = 1, \sigma_2 = \sqrt{2}, \rho = 0.6\sqrt{2}$. We wish to sample from this distribution.³ It is well known that the posterior distributions are also Gaussian and are given by

$$p(\theta_2 \mid \theta_1 = a, \mathbf{y}, \mathcal{M}) = \mathcal{N}\left(\theta_2; \rho \frac{\sigma_1}{\sigma_2} a, (1 - \rho^2)\sigma_2^2\right) \quad (11)$$

and similarly for $p(\theta_1 \mid \theta_2 = b, \mathbf{y}, \mathcal{M})$. We use Gibbs sampling to obtain samples from the joint distribution by successively drawing from the distributions

$$\theta_1^{(i+1)} \sim p\left(\theta_1 \mid \theta_2^{(i)}, \mathbf{y}, \mathcal{M}\right) \quad (12)$$

$$\theta_2^{(i+1)} \sim p\left(\theta_2 \mid \theta_1^{(i+1)}, \mathbf{y}, \mathcal{M}\right) \quad (13)$$

where the superscript (i) denotes the iteration count. A contour plot of $\log p(\boldsymbol{\theta} \mid \mathbf{y}, \mathcal{M})$ and the samples are shown in Fig. 2. We observe that even though we initialized the Gibbs sampler far from the mode at $(\theta_1, \theta_2) = (-10, 10)$, samples eventually are drawn from the joint distribution $p(\boldsymbol{\theta} \mid \mathbf{y}, \mathcal{M})$ and are concentrated around the mode at $(\theta_1, \theta_2) = (0, 0)$.

2) *Overview of the Gibbs Sampling Algorithm:* In this section, we give a high-level overview of the Gibbs sampler that will be used to estimate $\boldsymbol{\theta} = \{\mathbf{c}, \mathbf{t}, \sigma_e\}$. The algorithm is detailed in Algorithm 2.

We will now derive the posterior distribution of the parameters given the data. We will first focus on the priors assigned to the parameters $\boldsymbol{\theta}$. We place a Jeffrey's (improper) noninformative prior [20], [21] on the standard deviation of the noise such that

$$p(\sigma_e) \propto \frac{1}{\sigma_e}. \quad (14)$$

³Of course, Gibbs sampling is not the usual method to perform such a simple task. One would typically take the Cholesky factorization of $\boldsymbol{\Sigma} = \mathbf{R}^\top \mathbf{R}$ and use the matrix \mathbf{R}^\top to linearly transform a white Gaussian random vector.

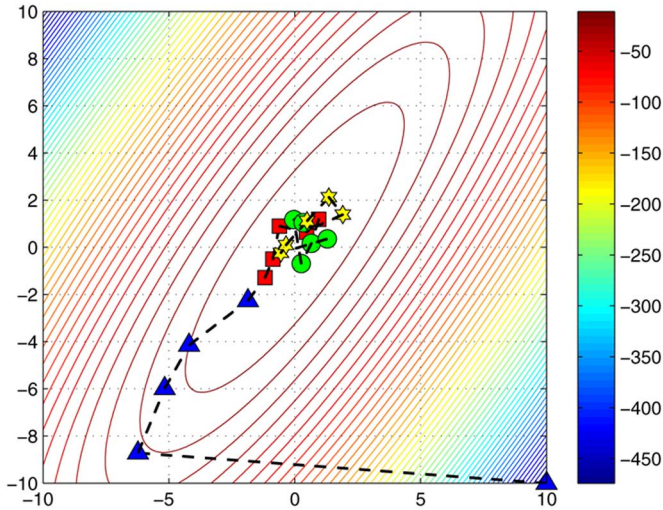


Fig. 2. Illustration of the Gibbs sampler for a zero-mean Gaussian distribution with covariance matrix given in (10). The Gibbs sampler is initialized at $(-10, 10)$ (bottom right) and iterates converges to samples from the joint distribution $p(\boldsymbol{\theta} | \mathbf{y}, \mathcal{M})$. The triangles (blue), circles (green), squares (red), and stars (yellow) are samples 1:5, 6:10, 11:15, and 16:20, respectively. The ellipses are contours of $\log p(\boldsymbol{\theta} | \mathbf{y}, \mathcal{M})$.

The remaining parameters $\mathbf{c} = \{c_k\}_{k=1}^K$ and $\mathbf{t} = \{t_k\}_{k=1}^K$ are treated as deterministic parameters so the prior of the parameters satisfies

$$p(\boldsymbol{\theta}) \propto p(\sigma_e). \quad (15)$$

One can also regard \mathbf{c} and \mathbf{t} as having uniform densities over a prespecified compact domain. Thus, Gibbs sampling serves to obtain maximum likelihood estimates of \mathbf{c} and \mathbf{t} . As an additional feature of our algorithm, unlike in other known algorithms [5], [12], [13], we estimate the standard deviation of the noise σ_e . This parameter—unlike \mathbf{c} and \mathbf{t} —has a prior distribution; thus we are in fact obtaining its maximum a posteriori (MAP) estimate. We adopt the standard principle (amongst Bayesians) to use a noninformative prior (14) for situations in which we want to make minimal assumptions.

Having established the prior over the parameters, we use Bayes's rule to write the posterior in terms of the likelihood and the prior

$$p(\boldsymbol{\theta} | \mathbf{y}, \mathcal{M}) \propto p(\mathbf{y} | \boldsymbol{\theta}, \mathcal{M}) \cdot p(\boldsymbol{\theta}). \quad (16)$$

The independent identically distributed (i.i.d.) Gaussian noise assumption and the model in (4) allow us to express the posterior distribution of the parameters (given the data) as

$$\log p(\mathbf{c}, \mathbf{t}, \sigma_e | \mathbf{y}, \mathcal{M}) \propto -(N+1) \log(\sigma_e) - \frac{1}{2\sigma_e^2} \sum_{n=0}^{N-1} \left[y[n] - \sum_{k=1}^K c_k \exp\left(-\frac{(nT - t_k)^2}{2\sigma_h^2}\right) \right]^2. \quad (17)$$

In the Gibbs sampling algorithm, as soon as a variate is drawn, it is inserted immediately into the conditional probability density function and remains there until being substituted in the next

iteration like in (12) and (13). This is further emphasized in Algorithm 2.⁴

In the algorithm, $\boldsymbol{\theta} \sim q(\boldsymbol{\theta})$ means that $\boldsymbol{\theta}$ is a random draw from the distribution $q(\boldsymbol{\theta})$. After I_b iterations,⁵ the Markov chain approximately reaches its stationary distribution $p(\boldsymbol{\theta} | \mathbf{y}, \mathcal{M})$. MMSE estimates can then be approximated by taking averages of the samples from the next I iterations $\{\boldsymbol{\theta}^{(I_b+1)}, \boldsymbol{\theta}^{(I_b+2)}, \dots, \boldsymbol{\theta}^{(I_b+I)}\}$, i.e.,

$$\hat{\boldsymbol{\theta}}_{\text{MMSE}} = \int \boldsymbol{\theta} p(\boldsymbol{\theta} | \mathbf{y}, \mathcal{M}) d\boldsymbol{\theta} \approx \frac{1}{I} \sum_{i=I_b+1}^{I_b+I} \boldsymbol{\theta}^{(i)}. \quad (18)$$

Algorithm 2 : The Gibbs Sampling Algorithm

Require: $\mathbf{y}, I, I_b, \boldsymbol{\theta}^{(0)} = \{\mathbf{c}^{(0)}, \mathbf{t}^{(0)}, \sigma_e^{(0)}\}$

- 1: **for** $i \leftarrow 1 : I + I_b$ **do**
 - 2: $c_1^{(i)} \sim p(c_1 | c_2^{(i-1)}, c_3^{(i-1)}, \dots, c_K^{(i-1)}, \mathbf{t}^{(i-1)}, \sigma_e^{(i-1)})$
 - 3: $c_2^{(i)} \sim p(c_2 | c_1^{(i)}, c_3^{(i-1)}, \dots, c_K^{(i-1)}, \mathbf{t}^{(i-1)}, \sigma_e^{(i-1)})$
 - 4: $\vdots \sim \vdots$
 - 5: $c_K^{(i)} \sim p(c_K | c_1^{(i)}, c_2^{(i)}, \dots, c_{K-1}^{(i)}, \mathbf{t}^{(i-1)}, \sigma_e^{(i-1)})$
 - 6: $t_1^{(i)} \sim p(t_1 | \mathbf{c}^{(i)}, t_2^{(i-1)}, t_3^{(i-1)}, \dots, t_K^{(i-1)}, \sigma_e^{(i-1)})$
 - 7: $t_2^{(i)} \sim p(t_2 | \mathbf{c}^{(i)}, t_1^{(i)}, t_3^{(i-1)}, \dots, t_K^{(i-1)}, \sigma_e^{(i-1)})$
 - 8: $\vdots \sim \vdots$
 - 9: $t_K^{(i)} \sim p(t_K | \mathbf{c}^{(i)}, t_1^{(i)}, t_2^{(i)}, \dots, t_{K-1}^{(i)}, \sigma_e^{(i-1)})$
 - 10: $\sigma_e^{(i)} \sim p(\sigma_e | \mathbf{c}^{(i)}, \mathbf{t}^{(i)})$
 - 11: **end for**
 - 12: Compute $\hat{\boldsymbol{\theta}}_{\text{MMSE}}$ using (18)
 - 13: **return** $\hat{\boldsymbol{\theta}}_{\text{MMSE}}$
-

B. Presentation of the Posterior Densities in the GS

We will now derive the conditional densities. In the sequel, we will use the notation $\boldsymbol{\theta}_{-\ell}$ to denote the set of parameters excluding the ℓ th parameter. It follows from Bayes's theorem that

$$p(\theta_\ell | \boldsymbol{\theta}_{-\ell}, \mathbf{y}, \mathcal{M}) \propto p(\mathbf{y} | \boldsymbol{\theta}, \mathcal{M}) \cdot p(\boldsymbol{\theta}). \quad (19)$$

Thus, the required conditional (posterior) distributions are proportional to the *likelihood* of the data times the *priors* on the parameters. We can easily calculate the posterior distributions of the parameters given the rest of the parameters. The parameters conditioned on $\boldsymbol{\theta}_\ell$ are taken as constant and can be left out of each posterior distribution $p(\theta_\ell | \boldsymbol{\theta}_{-\ell}, \mathbf{y}, \mathcal{M})$. We will sample from these posterior densities in the GS iterations as shown in the above algorithm.

We will now proceed to present the posterior densities. The derivations are provided in the Appendix.

⁴For brevity, the dependence on the model \mathcal{M} and data \mathbf{y} is omitted from the conditional density expressions in Algorithm 2.

⁵ I_b is also commonly known as the *burn-in* period in the Gibbs sampling and MCMC literature [15].

1) *Sampling c_k* : c_k is sampled from a Gaussian distribution given by

$$p(c_k | \boldsymbol{\theta}_{-c_k}, \mathbf{y}, \mathcal{M}) = \mathcal{N}\left(c_k; -\frac{\beta_k}{2\alpha_k}, \frac{1}{2\alpha_k}\right) \quad (20)$$

where

$$\alpha_k \triangleq \frac{1}{2\sigma_e^2} \sum_{n=0}^{N-1} \exp\left(-\frac{(nT - t_k)^2}{\sigma_h^2}\right) \quad (21)$$

$$\beta_k \triangleq \frac{1}{\sigma_e^2} \sum_{n=0}^{N-1} \exp\left(-\frac{(nT - t_k)^2}{2\sigma_h^2}\right) \times \left\{ \sum_{\substack{k'=1 \\ k' \neq k}}^K c_{k'} \exp\left(-\frac{(nT - t_{k'})^2}{2\sigma_h^2}\right) - y[n] \right\}. \quad (22)$$

It is easy to sample from Gaussian densities when the parameters (α_k, β_k) have been determined.

2) *Sampling t_k* : t_k is sampled from a distribution of the form

$$p(t_k | \boldsymbol{\theta}_{-t_k}, \mathbf{y}, \mathcal{M}) \propto \exp\left[-\frac{1}{2\sigma_e^2} \sum_{n=0}^{N-1} \gamma_k \times \exp\left(-\frac{(nT - t_k)^2}{\sigma_h^2}\right) + \nu_k \exp\left(-\frac{(nT - t_k)^2}{2\sigma_h^2}\right)\right] \quad (23)$$

where

$$\gamma_k \triangleq c_k^2 \quad (24)$$

$$\nu_k \triangleq 2c_k \left\{ \sum_{\substack{k'=1 \\ k' \neq k}}^K c_{k'} \exp\left(-\frac{(nT - t_{k'})^2}{2\sigma_h^2}\right) - y[n] \right\}. \quad (25)$$

It is not straightforward to sample from this distribution. We can sample t_k from a uniform grid of discrete values with probability masses proportional to (23). But in practice, and for greater accuracy, we used rejection sampling [22], [23] to generate samples $t_k^{(i)}$ from $p(t_k | \boldsymbol{\theta}_{-t_k}, \mathbf{y}, \mathcal{M})$. The proposal distribution $\tilde{q}(t_k)$ was chosen to be an appropriately scaled Gaussian, since it is easy to sample from Gaussians. This is shown in Algorithm 3.

In general, rejection sampling is not tractable for high-dimensional distributions since most of the samples will be discarded. However, since we sample from univariate distributions at each step, rejection sampling is tractable here.

Algorithm 3: Rejection Sampling

Require: $\tilde{p}(t_k) \triangleq p(t_k | \boldsymbol{\theta}_{-t_k}, \mathbf{y}, \mathcal{M})$

1: Select $\tilde{q}(t_k) \sim \mathcal{N}$ and c s.t. $\tilde{p}(t_k) < c\tilde{q}(t_k)$

2: $u \sim \mathcal{U}(0, 1)$

3: **repeat**

4: $t_k \sim \tilde{q}(t_k)$

5: **until** $u < \tilde{p}(t_k)/(c\tilde{q}(t_k))$

3) *Sampling σ_e* : σ_e is sampled from the *square-root inverted-gamma* [21] distribution $\mathcal{IG}^{1/2}(\sigma_e; \varphi, \lambda)$,⁶

$$p(\sigma_e | \boldsymbol{\theta}_{-\sigma_e}, \mathbf{y}, \mathcal{M}) = \frac{2\lambda^\varphi \sigma_e^{-(2\varphi+1)}}{\Gamma(\varphi)} \times \exp\left(-\frac{\lambda}{\sigma_e^2}\right) \mathbb{I}_{[0,+\infty)}(\sigma_e) \quad (26)$$

where

$$\varphi \triangleq \frac{N}{2} \quad (27)$$

$$\lambda \triangleq \frac{1}{2} \left[y[n] - \sum_{k=1}^K c_k \exp\left(-\frac{(nT - t_k)^2}{2\sigma_h^2}\right) \right]^2. \quad (28)$$

Thus the distribution of the variance of the noise σ_e^2 is inverted gamma, which corresponds to the conjugate prior of σ_e^2 in the expression of $\mathcal{N}(e; 0, \sigma_e^2)$ [21], and thus it is easy to sample from. In our simulations, we sampled from this density using the Matlab function `gamrnd` and applied the inverted square-root transformation.

C. Further Improvements via Linear Least Squares Estimation

We can perform an additional postprocessing step to improve on the estimates of c_k . We noted from our preliminary experiments (see Fig. 3) that the variance of the stationary distribution of the t_k s is smaller than that of the c_k s. This results in better estimates for the locations t_k s as compared to the magnitudes c_k s. Now, we observe that the observations $y[n]$ are linear in the c_k s once the t_k s are known. A natural extension to the Gibbs sampler algorithm is to estimate the weights c_k using an LLSE procedure given the data \mathbf{y} and the MMSE estimates of t_k . Equation (4) can be written as

$$y[n] = \sum_{k=1}^K c_k h(nT - t_k) + e[n], \quad 0 \leq n \leq N-1 \quad (29)$$

with $h(t)$, the Gaussian sampling kernel, given in (3). Given the set of estimates of the time locations $\{\hat{t}_k\}_{k=1}^K$, we can rewrite (29) as a matrix equation, giving

$$\mathbf{y} = \mathbf{H}\mathbf{c} + \mathbf{e} \quad (30)$$

where $[\mathbf{H}]_{nk} = h(nT - \hat{t}_k) = \exp(-(nT - \hat{t}_k)^2/2\sigma_h^2)$ and $1 \leq n \leq N, 1 \leq k \leq K$. We now minimize the square of the residual $\|\mathbf{e}\|_2^2 = \|\mathbf{H}\mathbf{c} - \mathbf{y}\|_2^2$, giving the normal equations $\mathbf{H}^\top \mathbf{H}\mathbf{c} = \mathbf{H}^\top \mathbf{y}$ and the least squares solution [24]

$$\hat{\mathbf{c}}_{\text{LS}} = (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{y}. \quad (31)$$

From our experiments, we found that, in general, using $\hat{\mathbf{c}}_{\text{LS}}$ as estimates for the magnitudes of the impulses provided a lower reconstruction error \mathcal{E} .

In summary, in order to approximately minimize \mathcal{E} in (9), we use the Gibbs sampler to estimate the time locations $\{t_k\}_{k=1}^K$

⁶ X follows a square-root inverted-gamma distribution if X^{-2} follows a gamma distribution.

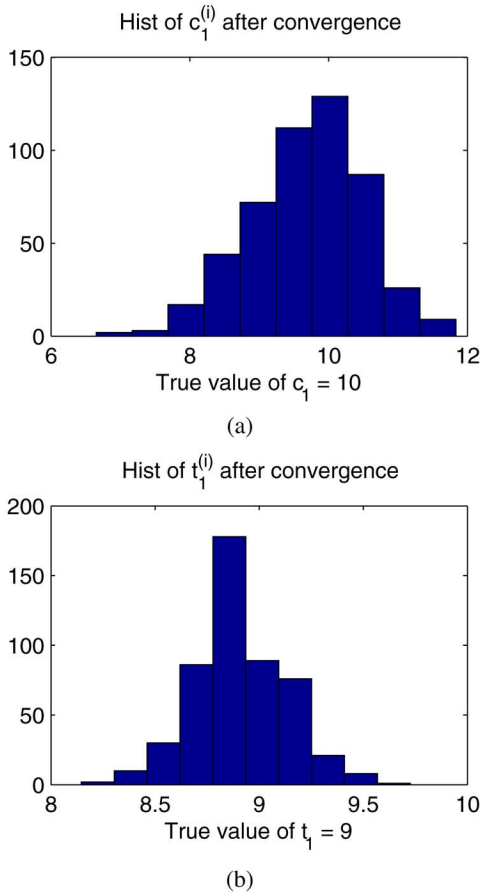


Fig. 3. Note that the variance of the stationary distribution of the t_k s is smaller than that of the c_k s after convergence of the Markov chain. Histogram of the samples of (a) $c_1^{(i)}$ and (b) $t_1^{(i)}$.

and the noise standard deviation σ_e . The Gibbs sampler maximizes the log-likelihood of the parameters $\{t_k\}_{k=1}^K$ given the observations \mathbf{y} . The Gibbs sampler also returns the MAP estimate for σ_e . Following the use of Gibbs sampling, we use a least squares procedure to estimate the magnitudes of the Gaussian kernels $\{c_k\}_{k=1}^K$. Note that given estimates of the locations $\{\hat{t}_k\}_{k=1}^K$, this step *exactly* minimizes the squared error \mathcal{E} .

IV. NUMERICAL RESULTS AND EXPERIMENTS

In this section, we will first review the annihilating filter and root-finding algorithm [5] for solving for the parameters of a signal with FRI. This algorithm provides a baseline for comparison. We also comment briefly on the applicability of [12] to our setup. Then we will provide extensive simulation results to validate the accuracy of the algorithm we proposed in Section III. In addition, to demonstrate that the reconstruction algorithm can also work on real signals, we apply the same GS algorithm to an audio signal. We show empirically that if the number of components K increases, the reconstruction error decreases.

A. Problems With Annihilating Filter and Root-Finding

We will now describe two existing algorithms for estimating a signal with FRI in the presence of noise.

1) *Comparison to Vetterli et al. [5]*: In [5], for signals of the form (2) and certain sampling kernels, the annihilating filter

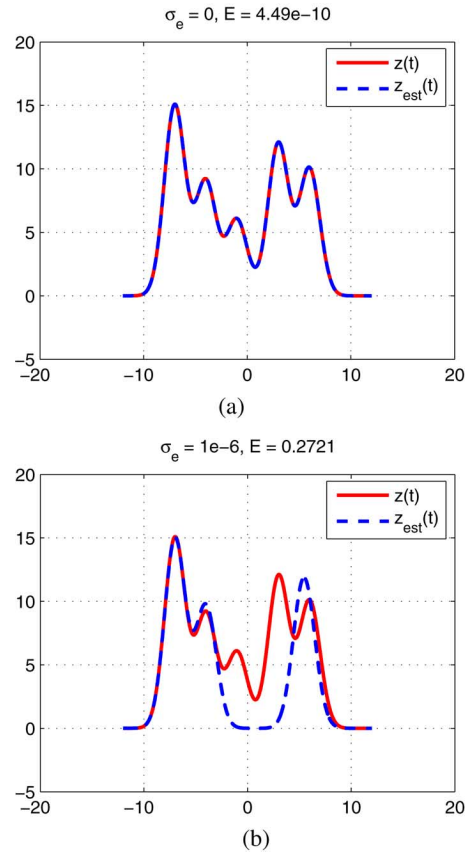


Fig. 4. Demonstration of the instability of annihilating filter/root-finding approach. (a) The annihilating filter approach reconstructs the signal exactly in the noiseless scenario. (b) The reconstruction completely breaks down when noise of a small standard deviation $\sigma_e = 10^{-6}$ (SNR = 137 dB) is added.

was used as a means to locate the t_k values. Subsequently, a least squares approach yielded the weights c_k . It was shown that in the noiseless scenario, this method recovers the parameters exactly [see Fig. 4(a)]. For completeness, we will briefly outline their method here. Denoting the noiseless samples by $z[n]$, (4) can be written as

$$p[n] = \sum_{k=1}^K a_k u_k^n, \quad n = 0, 1, \dots, N-1 \quad (32)$$

with the identifications

$$p[n] = w[n]z[n] \quad (33)$$

$$a_k = c_k \exp(-t_k^2 / (2\sigma_h^2)) \quad (34)$$

$$u_k = \exp(t_k T / \sigma_h^2) \quad (35)$$

where $w[n] = \exp(n^2 T^2 / (2\sigma_h^2))$. Now, since $p[n]$ is a linear combination of exponentials, we find the annihilating filter $a[n]$ such that

$$a[n] * p[n] = \sum_{\ell=0}^K a[\ell] p[n - \ell] = 0, \quad n \in \mathbb{Z}. \quad (36)$$

This can be written in matrix/vector form as $\mathbf{P}\mathbf{a} = \mathbf{0}$, where $[\mathbf{P}]_{nl} = p[n - l]$. This system will admit a solution \mathbf{a} when $\text{rank}(\mathbf{P}) = K$. Thus \mathbf{a} is a vector in the nullspace of \mathbf{P} . As suggested in [5], this is solved using an SVD where

TABLE I
PARAMETER VALUES FOR DEMONSTRATION OF THE ANNIHILATING
FILTER AND ROOT-FINDING ALGORITHM

Param.	K	σ_e	N	SNR
Value	5	0 and 10^{-6}	30	∞ and 137 dB

$\mathbf{P} = \mathbf{USV}^H$, $\mathbf{a} = \mathbf{V}\mathbf{e}_{K+1}$, and \mathbf{e}_{K+1} is a length- $(K+1)$ vector with “1” in position $(K+1)$ and “0” elsewhere. Now, once the coefficients $a[n]$ are found, the values u_k are simply the roots of the filter

$$A(z) = \sum_{n=0}^K a[n]z^{-n}. \quad (37)$$

The t_k s can then be determined from (35), and the solution for the c_k s essentially parallels the development in Section III-C.

In the same paper, it was suggested that to deal with the noisy samples $y[n] = z[n] + e[n]$, we can minimize $\|\tilde{\mathbf{P}}\mathbf{a}\|_2$, where $\tilde{\mathbf{P}}$ corresponds to the matrix with elements

$$\tilde{p}[n] = w[n]y[n]. \quad (38)$$

In this case, \mathbf{a} is the eigenvector that corresponds to the smallest eigenvalue of $\tilde{\mathbf{P}}^T\tilde{\mathbf{P}}$. Here, we provide three reasons to argue that this method is inherently ill-conditioned and thus not robust to noise.

- 1) Minimizing $\|\tilde{\mathbf{P}}\mathbf{a}\|_2$ involves finding the eigenvector \mathbf{v}_1 that corresponds to the largest eigenvalue λ_1 . Because computing eigenvalues and eigenvectors are essentially root-finding operations, this is ill-conditioned.
- 2) Even if the vector $\mathbf{a} = \mathbf{v}_1$ can be found, the zeros of the filter $A(z)$ have to be found. This again involves root finding, which is ill-conditioned.
- 3) From (38), any noise added to $z[n]$ will be exponentially weighted by $w[n]$ in the observations $p[n]$. We feel that this is the greatest source of ill-conditioning.

Because of the three reasons highlighted above, there is a need to explore new algorithms for finding the parameters. In Fig. 4, we show a simulation with the parameters as tabulated in Table I, but we varied the noise ($\sigma_e = 10^{-6}$ gives SNR = 137 dB, a very low noise level). We observe from Fig. 4(b) that (even with an oversampling factor of $N/(2K) = 3$), the annihilating filter and root-finding method is not robust even when a miniscule amount of noise is added.

Remark. The root-finding method is so unstable that, at times, even for low levels of noise, we obtain complex roots for the locations $\{t_k\}_{k=0}^{K-1}$. To solve this problem, we orthogonally projected the polynomial described by the filter coefficients $a[n]$ to the closest polynomial that belongs to the space of polynomials with only real roots.

2) *Comparison to Maravić and Vetterli [12]:* We emphasize that, in our setup (Fig. 1), we consider adding discrete-time white noise sequence $e[n]$ to the filtered and sampled signal $z[n]$. In [12] and [13], the authors considered the situation when the signal $x(t)$ was corrupted by continuous-time additive white Gaussian noise $e(t)$. Hence, the samples of $e(t) * h(t)$ are, in fact, *colored*. Despite the discrepancy in the setups, we applied

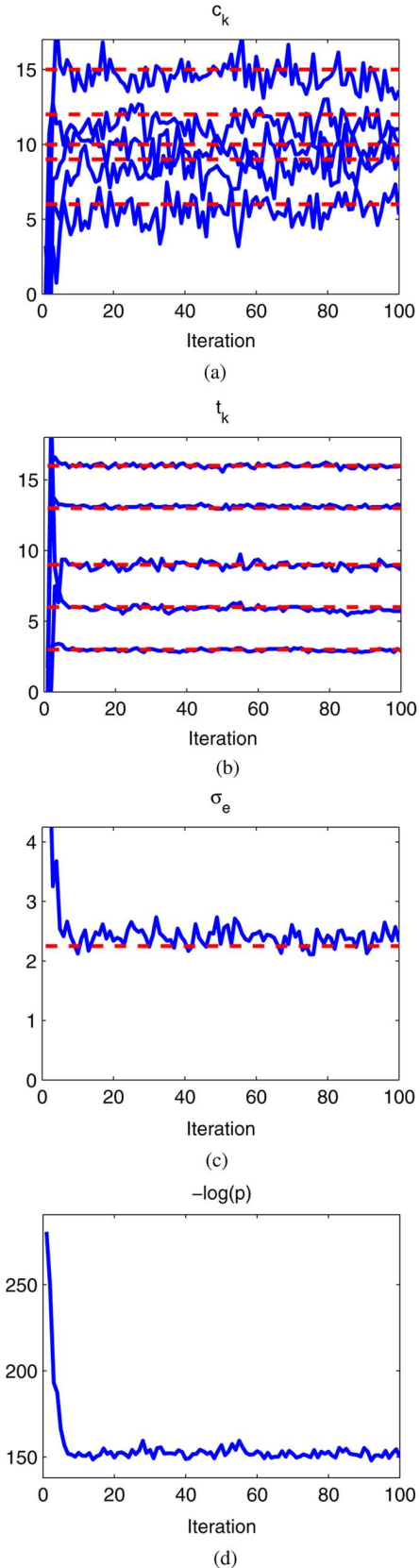


Fig. 5. Evolution of the GS algorithm. The iterates of the parameters $\{(c_k, t_k)\}_{k=1}^K$ and σ_e are shown. The true values are indicated by the broken red lines. In Fig. 5(d), we see that the negative log-posterior converges to the global minimum in fewer than 20 iterations for this problem size ($K = 5$). Evolution of the (a) c_k s, (b) t_k s, and (c) σ_e . (d) Reduction of the (negative) log-posterior $-\log p(\mathbf{c}, \mathbf{t}, \sigma_e | \mathbf{y}, \mathcal{M})$.

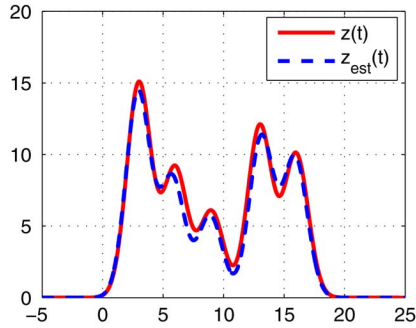


Fig. 6. Comparison between $z(t)$ and $z_{\text{est}}(t)$ using the GS algorithm. For this run, $\mathcal{E} = 0.0072$.

the algorithm suggested in [12] to the toy problem in this section. To partially ameliorate the ill-conditioning of $\tilde{\mathbf{P}}$, the authors suggested to replace the matrix $\tilde{\mathbf{P}}$ with another matrix

$$\tilde{\mathbf{P}}_s = \mathbf{A}\tilde{\mathbf{P}}\mathbf{B} \quad (39)$$

where the diagonal matrices \mathbf{A} and \mathbf{B} are defined as

$$[\mathbf{A}]_{ii} = \left(\frac{1}{N} \sum_{m=0}^{M-1} w[m+i] \right)^{-1}, \quad i = 0, \dots, L-1 \quad (40)$$

$$[\mathbf{B}]_{jj} = \left(\frac{1}{M} \sum_{l=0}^{L-1} [\mathbf{A}]_{ll} w[l+j] \right)^{-1}, \quad j = 0, \dots, M-1 \quad (41)$$

for integers $L, M > K$. Now, since the condition number of $\tilde{\mathbf{P}}_s$ is reduced by the preconditioners \mathbf{A} and \mathbf{B} , the SVD can more accurately be applied to $\tilde{\mathbf{P}}_s$ to find the K principal components $\mathbf{U}_s \mathbf{S}_s \mathbf{V}_s^H$, which correspond to the clean signal $z[n]$. The denoised data matrix is then given by $\tilde{\mathbf{P}}_d = \mathbf{A}^{-1} \mathbf{U}_s \mathbf{S}_s \mathbf{V}_s^H \mathbf{B}^{-1}$. The same root-finding procedure can then be applied to $\tilde{\mathbf{P}}_d$ in the hope that it will be close to \mathbf{P} , the matrix with $p[n]$ as its entries, where $p[n]$ is defined in (32) and (33).

We implemented this preconditioning step but realized that the performance as compared to [5] (described earlier in the section) did not improve. We surmise that this is due to two reasons.

- 1) Because of the exponential weights $w[n]$, not all the noise components were removed in the denoising step. Thus, $\tilde{\mathbf{P}}_d$ and \mathbf{P} are, in fact, not close. To quantify this, we used the toy example in the previous section and computed the Frobenius norm of the difference between the two matrices $\|\tilde{\mathbf{P}}_d - \mathbf{P}\|_F = 1.73 \times 10^{12}$. This large value implies that the subsequent root-finding step will be extremely unstable.
- 2) The two setups are intrinsically different—we modeled the noise as discrete-time white noise applied to the $z[n]$, whereas [12] applied continuous-time white noise to the signal $x(t)$.

We now turn our attention to our results.

B. Performance of Our Gibbs Sampling Algorithm

Clearly, the annihilating filter/root-finding algorithm is not robust to noise. We have suggested an alternative reconstruction algorithm in Section III and, in this section, will present

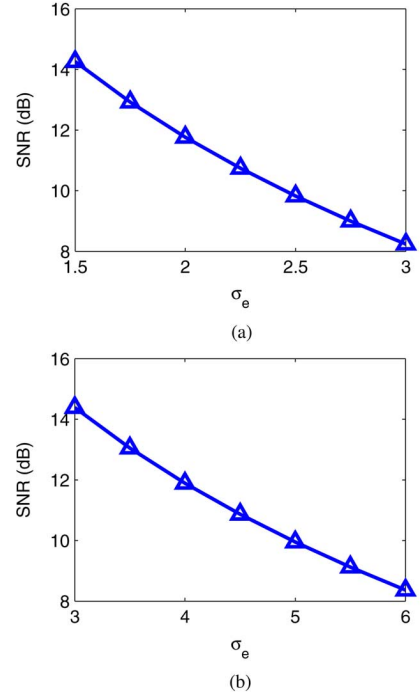


Fig. 7. SNR (dB) against σ_e for (a) experiment A ($K = 5$) and (b) experiment B ($K = 10$).

our results on several synthetic examples.⁷ Experiments on an audio signal will also be presented in this section to investigate our reconstruction algorithm on a real signal—one that may not exactly be modeled as a sum of Diracs.

Initial Demonstration: To demonstrate the evolution the Gibbs sampler, we performed an initial experiment and chose the parameters to be those in Table I, with the exception that the noise standard deviation was increased to $\sigma_e = 2.5$, giving an SNR of 10.2 dB. We plot the iterates in Fig. 5. The true filtered signal $z(t)$ and its estimate $z_{\text{est}}(t)$ are plotted in Fig. 6. Note the close similarity between $z(t)$ and $z_{\text{est}}(t)$.

We observe that the sampler converges in fewer than 20 iterations for this run, even though the parameter values were initialized far from their optimal values. We emphasize that as Gibbs sampling is essentially a stochastic optimization procedure (not unlike simulated annealing or genetic algorithms), it is insensitive to the choice of starting point $\theta^{(0)}$. The Markov chain is guaranteed to converge to the stationary distribution after the burn-in period [22].

1) Further Experiments on Simulated Data: To further validate our algorithm, we performed extensive simulations (experiments A and B) on two different problem sizes. For consistency, each experiment was repeated using 100 different random seeds and the means of \mathcal{E} [see (9)] taken. The parameters are chosen according to Table II. The unknown parameters were initialized as $\mathbf{c}^{(0)} = \mathbf{t}^{(0)} = [0, \dots, 0]$ and $\sigma_e^{(0)} = 0.01$. Plots of the SNR against σ_e is for experiments A and B shown in Fig. 7 and the results of the two experiments are shown in Fig. 8. We noted the following from these experiments.

⁷All the code, written in MATLAB, can be found at <http://web.mit.edu/~vtan/frimcmc>.

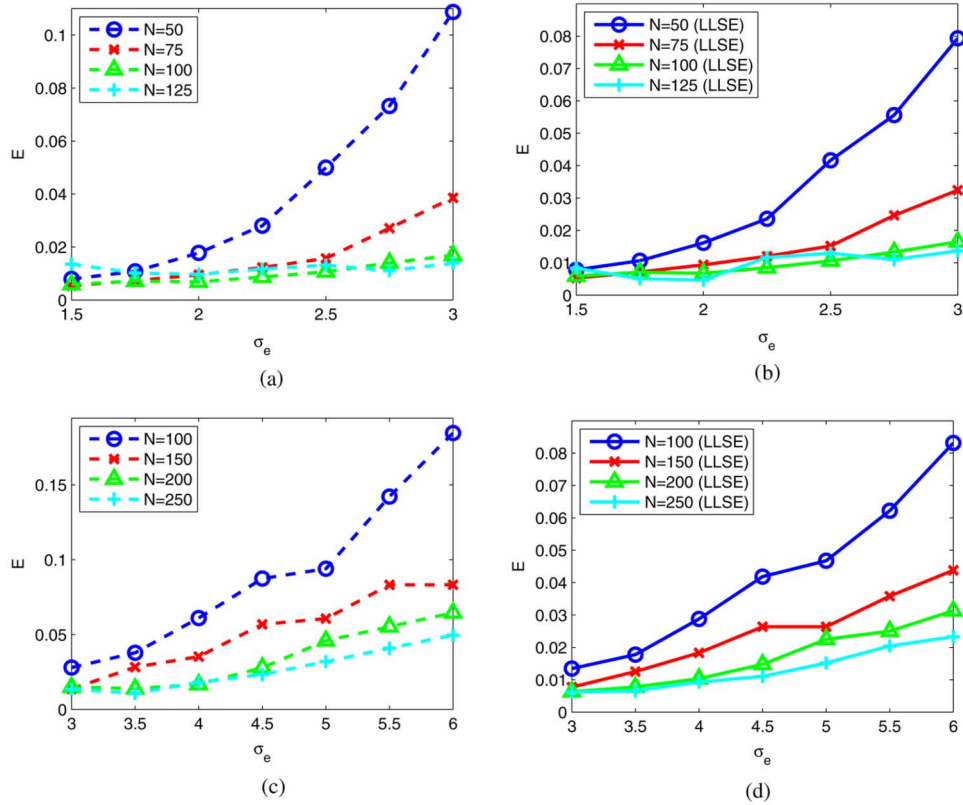


Fig. 8. Plots of \mathcal{E} against σ_e for various oversampling factors and problem sizes. In Fig. 8(a) and (c), the $\{c_k\}_{k=1}^K$ from the sampler was used directly. In Fig. 8(b) and (d), better estimates of $\{c_k\}_{k=1}^K$ were obtained by performing the LLSE postprocessing step as described in Section III-C. Errors \mathcal{E} against σ_e for (a) experiment A ($K = 5$) without LLSE postprocessing, (b) experiment A ($K = 5$) with LLSE postprocessing, (c) experiment B ($K = 10$) without LLSE postprocessing, and (d) experiment B ($K = 10$) with LLSE postprocessing.

TABLE II
PARAMETER VALUES FOR NUMERICAL SIMULATIONS

Param.	K	σ_e	σ_h	N	SNR
Expt A	5	1.5:0.25:3.0	1.0	50:25:150	Fig. 7(a)
Expt B	10	3.0:0.50:6.0	1.0	100:50:250	Fig. 7(b)

- The Gibbs sampler algorithm is insensitive to initialization. It *always* finds approximately optimal estimates from any starting point because the Markov chain provably converges to the stationary distribution [22].
- The LLSE postprocessing step in the Gibbs sampler algorithm reduces the reconstruction error \mathcal{E} . This is a consequence of using the (more accurate) t_k s from the sampler to estimate the c_k s via LLSE, instead of using the c_k s from the sampler directly.
- From Fig. 8(b) and (d), we observe that, if the problem size doubles (from $K = 5$ to $K = 10$), with corresponding doubling of σ_e and N , \mathcal{E} remains approximately constant, assuming $\{c_k\}_{k=1}^K$ are obtained from the least squares postprocessing instead of from the Gibbs sampler alone. This demonstrates scalability of the algorithm. For example, $\mathcal{E}(K = 5, \sigma_e = 2.5, N = 50) \approx \mathcal{E}(K = 10, \sigma_e = 5.0, N = 100) \approx 0.045$.
- The noise standard deviation σ_e can be estimated accurately as shown in Fig. 5(c). This may be important in some applications, e.g., in an ADC.

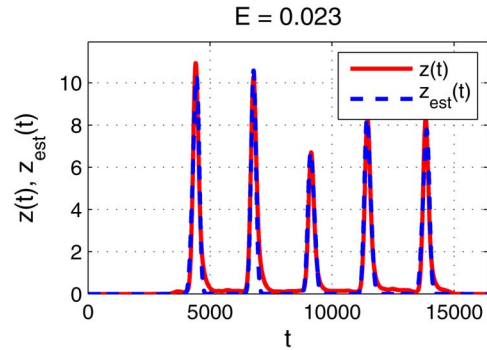


Fig. 9. Comparison between $z(t)$ and $z_{\text{est}}(t)$ using the GS algorithm for the audio signal. Because there can only be a maximum of $K = 9$ modes, the clean signal cannot be exactly modelled using a sum of $K = 9$ Gaussian kernels. Thus, the parts of the signal $z(t)$ with small but nonzero magnitudes (e.g., in sample locations 5000–6000) are effectively reduced to zero in $z_{\text{est}}(t)$. In this experiment, $\mathcal{E} = 0.023$.

To conclude, although our approach is more computationally demanding than the annihilating filter approach of [5], it is much more amenable to a scenario where noisy samples are acquired.

2) *Experiments on a Real Audio Signal:* In this section, we investigate efficacy of our algorithm on a real audio signal $x(t)$ that does not exactly meet the sum-of-Diracs model. This 2-s-long signal⁸ is a recording of a person clapping his/her hands five times in the 2-s interval. $x(t)$ is filtered with a

⁸The signal can be downloaded from <http://web.mit.edu/~vtan/frimcmc>.

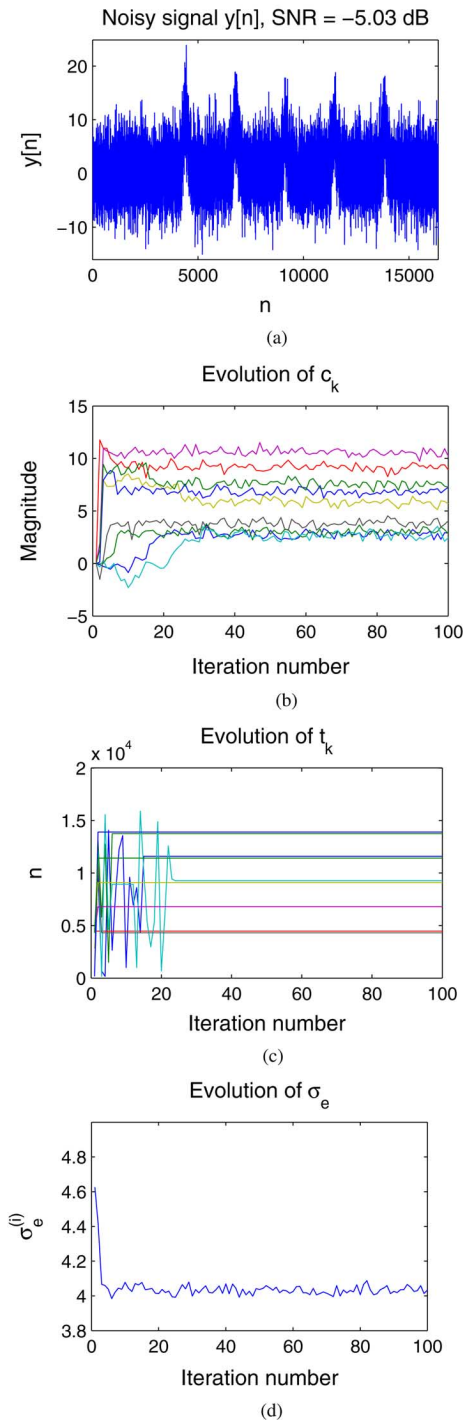


Fig. 10. Evolution of Gibbs sampler for the audio signal. We chose $K = 9$ for this experiment. The sampler converges in about 40 iterations. (a) The noisy signal. Evolution of (b) c_k , (c) t_k , and (d) σ_e . The true value of $\sigma_e = 4$.

Gaussian filter $h(t)$ of width $\sigma_h = 100$ samples. The resulting $z(t)$ is plotted in Fig. 9. Discrete-time additive white Gaussian noise $e[n]$ with standard deviation $\sigma_e = 4$ was added to the signal, giving an SNR of -5.03 dB. The noisy signal $y[n]$ is shown in Fig. 10(a).

In Fig. 10(b)–(d), we plot the evolution of the Gibbs sampler for the various parameters using $K = 9$. The sampler converges to the stationary distribution in about 40 iterations. In Fig. 9,

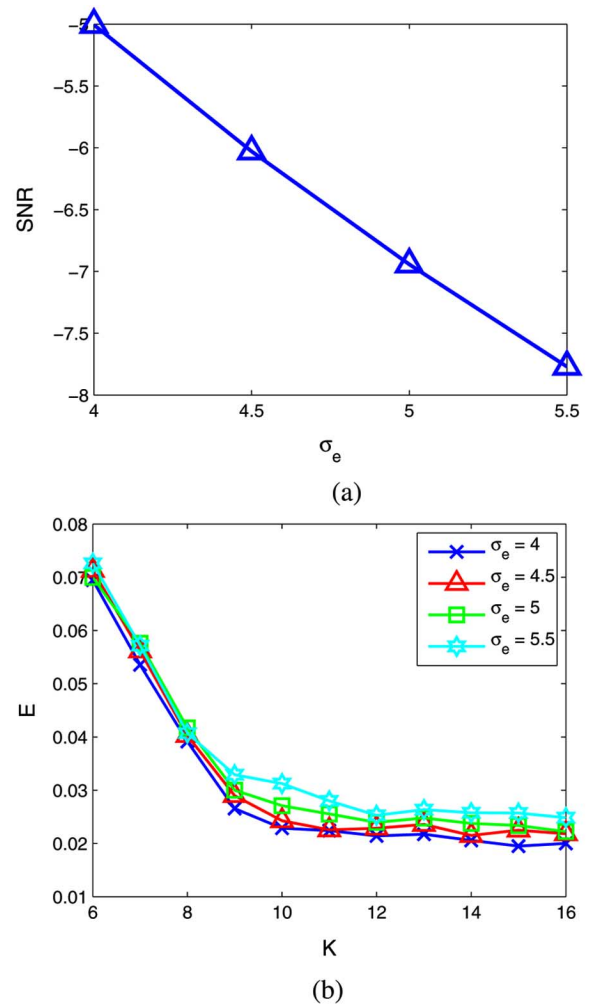


Fig. 11. Performance plots for the audio signal. (a) SNR (dB) against σ_e for the audio signal. (b) Errors \mathcal{E} against K for the audio signal.

we compare $z(t)$ and $z_{\text{est}}(t)$. Although the primary modes corresponding to the five claps have been correctly located, the samples with smaller magnitudes are not well modelled since there is a model mismatch: the audio signal does not meet the sum-of-Diracs model.

Finally, we varied both the noise level σ_e and the number of components K . Intuitively, when K increases, the modelling power of the FRI model improves and hence the error \mathcal{E} should also be reduced. This is indeed the case as shown in Fig. 11(b), where we plot the reconstruction error \mathcal{E} as a function of K for several values of σ_e with LLSE postprocessing. The results of the experiments are averaged over 100 independent runs. A plot of SNR against σ_e is also shown in Fig. 11(a). As usual, we observe that as the noise level increases, the reconstruction worsens. However, for this experiment, it seems that the fidelity of reconstruction is not very sensitive to the noise level.

This set of experiments on a real audio signal suggests that the finite-rate-of-innovation model can, in principle, approximate some real signals.

V. CONCLUSION

In this paper, we addressed the problem of reconstructing a signal with FRI given noisy samples. We showed that it is pos-

sible to circumvent some of the problems of the annihilating filter and root-finding approach [5], [12]. We introduced the Gibbs sampling algorithm to find the locations and augmented this with a least squares approach to find the weights. From the performance plots on both synthetic and real signals, we observe the Gibbs sampler performs very well as compared to the annihilating filter method, which is not robust to noise.

Perhaps the most important observation we made is the following: The success of the Gibbs sampling algorithm does not depend on the choice of kernel $h(t)$, but rather the i.i.d. Gaussian noise assumption. The formulation of the Gibbs sampler does not depend on the specific form of $h(t)$. In fact, we used a Gaussian sampling kernel to illustrate that our algorithm is not restricted to the classes of kernels considered in [6].

A natural extension to our work here is to assign structured priors to \mathbf{c} , \mathbf{t} , and σ_e . These priors can themselves be dependent on their own set of *hyperparameters*, giving a hierarchical Bayesian formulation. In this way, there would be greater flexibility in the parameter estimation process. We can also seek to improve on the computational load of the algorithms introduced here and, in particular, the sampling of t_k via rejection sampling.

A question that remains is: How well can other real-world signals (including natural images such as in [11]) be modeled as signals with FRI, possibly with different sampling kernels? We believe the answer will have profound ramifications for areas such as sparse approximation [25] and compressed sensing [26], [27].

APPENDIX

DERIVATION OF THE CONDITIONAL DENSITIES

For brevity, we define

$$g_{nk} \triangleq h(nT - t_k) = \exp\left(-\frac{(nT - t_k)^2}{2\sigma_h^2}\right).$$

We start from the log-likelihood of the parameters $\boldsymbol{\theta}$ given the data \mathbf{y} and model \mathcal{M} [see (17)]. To obtain $p(c_k | \boldsymbol{\theta}_{-c_k}, \mathbf{y}, \mathcal{M})$, we treat the other parameters $\boldsymbol{\theta}_{-c_k}$ as constant, giving

$$\begin{aligned} \log p(c_k | \boldsymbol{\theta}_{-c_k}, \mathbf{y}, \mathcal{M}) \\ \propto -\frac{1}{2\sigma_e^2} \sum_{n=0}^{N-1} \left[c_k^2 g_{nk}^2 \right. \\ \left. + 2c_k g_{nk} \left(\sum_{\substack{k'=1 \\ k' \neq k}}^K c_{k'} g_{nk'} - y[n] \right) \right]. \end{aligned}$$

Comparing this expression in c_k to the Gaussian distribution with mean μ and variance σ^2

$$\log p(c_k; \mu, \sigma^2) \propto -\frac{1}{2\sigma^2}(c_k - \mu)^2$$

and equating coefficients, we obtain (21) and (22). The distribution $p(t_k | \boldsymbol{\theta}_{-t_k}, \mathbf{y}, \mathcal{M})$ can be obtained similarly and is omitted. Finally, for the noise standard deviation σ_e

$$\log p(\sigma_e | \boldsymbol{\theta}_{-\sigma_e}, \mathbf{y}, \mathcal{M}) \propto -(N+1) \log(\sigma_e) - \frac{\lambda}{\sigma^2}$$

where λ is defined in (28). Taking the antilog on both sides yields

$$p(\sigma_e | \boldsymbol{\theta}_{-\sigma_e}, \mathbf{y}, \mathcal{M}) \propto \sigma_e^{-(N+1)} \exp\left(-\frac{\lambda}{\sigma^2}\right)$$

which is the square-root inverted-gamma distribution with parameters given by (27) and (28). All the densities have been derived.

REFERENCES

- [1] C. E. Shannon, "Communication in the presence of noise," *Proc. Inst. Radio Eng.*, vol. 37, no. 1, pp. 10–21, Jan. 1949.
- [2] H. Nyquist, "Certain topics in telegraph transmission theory," *Trans. Amer. Inst. Electr. Eng.*, vol. 47, pp. 617–644, Apr. 1928.
- [3] A. J. Jerri, "The Shannon sampling theorem—Its various extensions and applications: A tutorial review," *Proc. IEEE*, vol. 65, pp. 1565–1596, Nov. 1977.
- [4] M. Unser, "Sampling—50 years after Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, 2000.
- [5] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, 2002.
- [6] P. L. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang–Fix," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1741–1757, 2007.
- [7] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.
- [8] G. Strang and G. Fix, "A Fourier analysis of the finite element variational method," *Constr. Aspects Funct. Anal.*, pp. 796–830, 1971.
- [9] M. Unser and T. Blu, "Cardinal exponential splines: Part I—Theory and filtering algorithms," *IEEE Trans. Signal Process.*, vol. 53, pp. 1425–1438, Apr. 2005.
- [10] I. Maravić and M. Vetterli, "Exact sampling results for some classes of parametric nonbandlimited 2-D signals," *IEEE Trans. Signal Process.*, vol. 52, pp. 175–189, Jan. 2004.
- [11] P. Shukla and P. L. Dragotti, "Sampling schemes for multidimensional signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3670–3686, 2007.
- [12] I. Maravić and M. Vetterli, "Sampling and reconstruction of signals with finite rate of innovation in the presence of noise," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2788–2805, 2005.
- [13] A. Ridolfi, I. Maravić, J. Kusuma, and M. Vetterli, "Sampling signals with finite rate of innovation: The noisy case," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2002.
- [14] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, pp. 721–741, 1984.
- [15] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration: A Statistical Model Based Approach*. Berlin, Germany: Springer-Verlag, 1998.
- [16] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and W. Teller, "Equations of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, pp. 1087–1091, 1953.
- [17] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, 1970.
- [18] W. J. Fitzgerald, S. J. Godsill, A. C. Kokaram, and J. A. Stark, *Bayesian Methods in Signal and Image Processing*. Oxford, U.K.: Oxford Univ. Press, 1999, ch. Bayesian Statistics.
- [19] J. M. Hammersley and M. S. Clifford, "Markov fields on finite graphs and lattices," 1970, unpublished.
- [20] H. Jeffreys, *Theory of Probability*. Oxford, U.K.: Oxford Univ. Press, 1939.
- [21] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*, 1st ed. New York: Wiley, 2001.
- [22] L. Tierney, Markov chains for exploring posterior distributions, School of Statistics, Univ. of Minnesota, Tech. Rep. 560, Mar. 1994.
- [23] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer-Verlag, 2004.
- [24] G. Strang, *Introduction to Linear Algebra*, 3rd ed. Cambridge, MA: Wellesley Cambridge, 2001.
- [25] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, "Data compression and harmonic analysis," *IEEE Trans. Inf. Theory*, vol. 44, pp. 2435–2476, Oct. 1998.

- [26] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, Apr. 2006.
- [27] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5406–5425, Dec. 2006.



Vincent Y. F. Tan (S'07) received the B.A. (with first-class honors) and M.Eng. (with distinction) degrees from Electrical and Information Sciences Tripos (EIST), Sidney Sussex College, Cambridge University, U.K. He is currently pursuing the Ph.D. degree in electrical engineering in the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge.

He was a Research Engineer with the Defence Science Organization National Laboratories, Singapore, in 2005–2006; Research Officer with the Institute for Infocomm Research, Singapore, in 2006–2007; Teaching Assistant with the National University of Singapore, in 2006; and Research Intern with Microsoft Research Cambridge, U.K., in 2008. His current research interests include statistical signal processing, probabilistic graphical models, machine learning, and convex optimization.

Mr. Tan received the Public Service Commission Overseas Merit Scholarship and the National Science Scholarship from the Agency for Science Technology and Research. In 2005, he received the Charles Lamb Prize, a Cambridge University Engineering Department prize awarded annually to the candidate who demonstrates the greatest proficiency in the EIST.



Vivek K Goyal (S'92–M'98–SM'03) received the B.S. degree in mathematics and the B.S.E. degree in electrical engineering (both with highest distinction) from the University of Iowa, Iowa City, in 1993. He received the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1995 and 1998, respectively.

He was a Research Assistant with the Laboratoire de Communications Audiovisuelles, École Polytechnique Fédérale de Lausanne, Switzerland, in 1996; a Member of Technical Staff with the Mathematics of Communications Research Department, Bell Laboratories, Lucent Technologies, in 1998–2001; and a Senior Research Engineer with Digital Fountain, Inc., in 2001–2003. He is currently Esther and Harold E. Edgerton Associate Professor of Electrical Engineering and a Member of the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge. His research interests include source coding theory, sampling, quantization, and information gathering and dispersal in networks.

Dr. Goyal is a member of Phi Beta Kappa, Tau Beta Pi, Sigma Xi, Eta Kappa Nu, SIAM, and ACM. In 1998, he received the Eliahu Jury Award from the University of California, Berkeley, awarded to a graduate student or recent alumnus for outstanding achievement in systems, communications, control, or signal processing. He also received the 2002 IEEE Signal Processing Society Magazine Award and a National Science Foundation CAREER Award. He is a member of the IEEE Signal Processing Society's Image and Multiple Dimensional Signal Processing Technical Committee and permanent Cochair of the SPIE Wavelets conference series.