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

Vincent Y. F. Tan and Vivek K Goyal

Massachusetts Institute of Technology, Cambridge, MA 02139 USA
vtan@mit.edu, vgoyal@mit.edu

Abstract:

As an example of the 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 not noiseless. In a recent paper, we introduced a novel *stochastic* algorithm to reconstruct a signal with finite rate of innovation from its *noisy* samples. Even though variants of this problem has been approached previously, satisfactory solutions are only available for certain classes of sampling kernels, for example kernels which satisfy the Strang–Fix condition. In our paper, we considered 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 (MCMC) method. This paper summarizes the algorithm and provides numerical simulations that demonstrate the accuracy and robustness of our algorithm.

1. Introduction

The celebrated Nyquist–Shannon sampling theorem [4, 6] states that a signal $x(t)$ known to be bandlimited to $\Omega_{\max}$ Hz is uniquely determined by samples of $x(t)$ spaced $1/(2\Omega_{\max})$ sec 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 pre-filtering by the appropriate sinc *sampling kernel* gives a procedure that finds the orthogonal projection to the space of $\Omega_{\max}$-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 [9]. Moving out of this restrictive setting, Vetterli *et al.* [10] showed that it is possible to develop sampling schemes for certain classes of non-bandlimited signals that are not subspaces. As described in [10], for reconstruction from samples it is necessary for the class of signals to have *finite rate of innovation* (FRI). The paradigmatic example is the class of signals expressed as

$$x(t) = \sum_k c_k \phi(t - t_k) \tag{1}$$

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_ks (the number that appear per unit of time) is finite, the signal has FRI. It is shown constructively in [10] 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 [2] 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 [11]. Hence, samples are inherently noisy. This motivates our central question: *Given the signal model (i.e. a signal with FRI) and the noise model, how well can we approximate the parameters that describe the signal and hence the signal itself?* In this work, we address this question by developing a novel algorithm to reconstruct the signal from the noisy samples. The main contribution is to show that a stochastic approach can effectively circumvent the ill-conditioning of algebraic techniques.

This paper is an abridged version of [7], where many additional details can be found.

2. 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 [2, 3, 10], the class is the weighted sum of K Diracs

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

(The use of a Dirac delta simplifies the discussion. It can be replaced by a known pulse $\phi(t)$ and then absorbed into the sampling kernel $h(t)$, yielding an effective sampling kernel $\phi(t) * h(t)$.) The signal to be estimated $x(t)$ is filtered using a Gaussian lowpass filter

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

with width σ_h to give the signal $z(t)$. Even though $h(t)$ does not have compact support, it can be well approximated by a truncated Gaussian, which does have compact

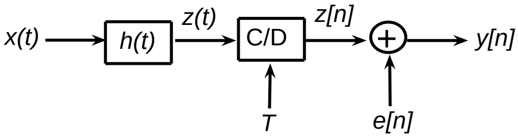


Figure 1: Block diagram showing our problem setup. $x(t)$ is a signal with FRI given by (2) and $h(t)$ is the Gaussian filter with width σ_h given by (3). $e[n]$ is i.i.d. Gaussian noise with standard deviation σ_e and $y[n]$ are the noisy samples. From $y[n]$ we will estimate the parameters that describe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

support. The filtered signal $z(t)$ is sampled at rate of $1/T$ Hz to obtain $z[n] = z(nT)$ for $n = 0, 1, \ldots, N-1$. Finally, additive white Gaussian noise (AWGN) $e[n]$ 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 $n = 0, 1, \ldots, N-1$. The amount of noise added is a function of σ_e. We define the signal-to-noise ratio (SNR) in dB as

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

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

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

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

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

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 $\mathcal{E} \geq 0$ and the closer $\mathcal{E}$ is to 0, the better the reconstruction algorithm. The problem can be summarized as: *Given* $\mathbf{y} = \{y[n] \,|\, n = 0, \ldots, N-1\}$ *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 (ML) 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 $\mathcal{E}$.

3. Presentation of the Gibbs Sampler

The algorithm introduced in [7] is a stochastic optimization procedure based on Gibbs sampling to estimate $\boldsymbol{\theta} = \{\mathbf{c}, \mathbf{t}, \sigma_e\}$. Detailed derivations and a self-contained introduction to Gibbs sampling are given in [7], and code written in MATLAB can be found at http://web.mit.edu/∼vtan/frimcmc. Here, we merely summarize the main steps of the algorithm and the intuition behind Gibbs sampling.

The overall procedure is given in Algorithm 1. The algorithm uses Gibbs sampling (Algorithm 2) to estimate the set of Dirac positions $\{t_k\}_{k=1}^K$. It then uses a least-squares procedure to estimate the weights $\{c_k\}_{k=1}^K$. The basic idea of Gibbs sampling is to exploit the fact that it is easier to compute samples drawn approximately according to the posterior distribution of the parameters given the data than it is to directly minimize $\mathcal{E}$. This is true when one can analytically determine the conditional distribution of one parameter given the remaining parameters and the data. (The required derivations are presented in [7].) After a number of iterations I_b called the *burn-in period*, samples drawn through Gibbs sampling can be treated as if they are drawn from the true posterior. Thus, samples drawn in I additional iterations can be averaged to obtain a good approximation of the mean of the posterior distribution.

Algorithm 1 Parameter Estimation and Signal Reconstruction Algorithm

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

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).

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.

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 $h(t)$.

Compute reconstruction error $\mathcal{E}$ given in (9).

Algorithm 2 The Gibbs Sampling Algorithm

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

for $i \leftarrow 1 : I + I_b$ **do**

$c_1^{(i)} \sim p(c_1 | c_2^{(i-1)}, c_3^{(i-1)}, \ldots, c_K^{(i-1)}, \mathbf{t}^{(i-1)} \sigma_e^{(i-1)})$

$c_2^{(i)} \sim p(c_2 | c_1^{(i)}, c_3^{(i-1)}, \ldots, c_K^{(i-1)}, \mathbf{t}^{(i-1)} \sigma_e^{(i-1)})$

$\vdots \sim \vdots$

$c_K^{(i)} \sim p(c_K | c_1^{(i)}, c_2^{(i)}, \ldots, c_{K-1}^{(i)}, \mathbf{t}^{(i-1)}, \sigma_e^{(i-1)})$

$t_1^{(i)} \sim p(t_1 | \mathbf{c}^{(i)}, t_2^{(i-1)}, t_3^{(i-1)}, \ldots, t_K^{(i-1)}, \sigma_e^{(i-1)})$

$t_2^{(i)} \sim p(t_2 | \mathbf{c}^{(i)}, t_1^{(i)}, t_3^{(i-1)}, \ldots, t_K^{(i-1)}, \sigma_e^{(i-1)})$

$\vdots \sim \vdots$

$t_K^{(i)} \sim p(t_K | \mathbf{c}^{(i)}, t_1^{(i)}, t_2^{(i)}, \ldots, t_{K-1}^{(i)}, \sigma_e^{(i-1)})$

$\sigma_e^{(i)} \sim p(\sigma_e | \mathbf{c}^{(i)}, \mathbf{t}^{(i)})$

end for

Compute $\hat{\boldsymbol{\theta}}_{\mathrm{MMSE}}$ using least squares

return $\hat{\boldsymbol{\theta}}_{\mathrm{MMSE}}$

Sampling c_k. c_k is sampled from a Gaussian distribution given by

$$p(c_k|\boldsymbol{\theta}_{-c_k}, \mathbf{y}, \mathcal{M}) = \mathcal{N}\left(c_k; -\frac{\beta_k}{2\alpha_k}, \frac{1}{2\alpha_k}\right), \qquad (10)$$

where

$$\alpha_k \triangleq \frac{1}{2\sigma_e^2} \sum_{n=0}^{N-1} \exp\left(-\frac{(nT-t_k)^2}{\sigma_h^2}\right), \qquad (11)$$

$$\beta_k \triangleq \frac{1}{\sigma_e^2} \sum_{n=0}^{N-1} \exp\left(-\frac{(nT-t_k)^2}{2\sigma_h^2}\right) \times \left\{ \sum_{\substack{k'=1 \\ k' \neq k}}^{K} c_{k'} \exp\left(-\frac{(nT-t_{k'})^2}{2\sigma_h^2}\right) - y[n] \right\}. \qquad (12)$$

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

Sampling t_k. t_k is sampled from a distribution of the form

$$p(t_k|\boldsymbol{\theta}_{-t_k}, \mathbf{y}, \mathcal{M}) \propto \exp\left[-\frac{1}{2\sigma_e^2} \sum_{n=0}^{N-1} \gamma_k \times \exp\left(-\frac{(nT-t_k)^2}{\sigma_h^2}\right) + \nu_k \exp\left(-\frac{(nT-t_k)^2}{2\sigma_h^2}\right) \right] \qquad (13)$$

where

$$\gamma_k \triangleq c_k^2, \qquad (14)$$

$$\nu_k \triangleq 2c_k \left\{ \sum_{\substack{k'=1 \\ k' \neq k}}^{K} c_{k'} \exp\left(-\frac{(nT-t_{k'})^2}{2\sigma_h^2}\right) - y[n] \right\}. \qquad (15)$$

It is not straightforward to sample from this distribution. We used rejection sampling [5,8] 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.

Sampling σ_e. σ_e is sampled from the 'Square-root Inverted-Gamma' [1] distribution $\mathcal{IG}^{-1/2}(\sigma_e; \varphi, \lambda)$,

$$p(\sigma_e|\boldsymbol{\theta}_{-\sigma_e}, \mathbf{y}, \mathcal{M}) = \frac{2\lambda^{\varphi}\sigma_e^{-(2\varphi+1)}}{\Gamma(\varphi)} \exp\left(-\frac{\lambda}{\sigma_e^2}\right) \mathbb{I}_{[0,+\infty)}(\sigma_e), \qquad (16)$$

where

$$\varphi \triangleq \frac{N}{2}, \qquad (17)$$

$$\lambda \triangleq \frac{1}{2}\left[y[n] - \sum_{k=1}^{K} c_k \exp\left(-\frac{(nT-t_k)^2}{2\sigma_h^2}\right) \right]^2 \qquad (18)$$

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 $\mathcal{N}(e; 0, \sigma_e^2)$ [1] and thus it is easy to sample from.

	K	N	σ_e	SNR
AF/RF (Fig. 2(a))	5	30	10^{-6}	137 dB
GS (Fig. 2(b))	5	30	2.5	10.2 dB

Table 1: Parameter values for comparing annihilating filter and root-finding (AF/RF) against Gibbs sampling (GS).

4. Numerical Results and Experiments

In this section, the annihilating filter and root-finding algorithm [10] provides a baseline for comparison. After exhibiting its instability, we provide simulation results to validate the accuracy of the algorithm we proposed in Section 3. More extensive experimentation, including comparisons to [3] and applications to an audio signal, is reported in [7].

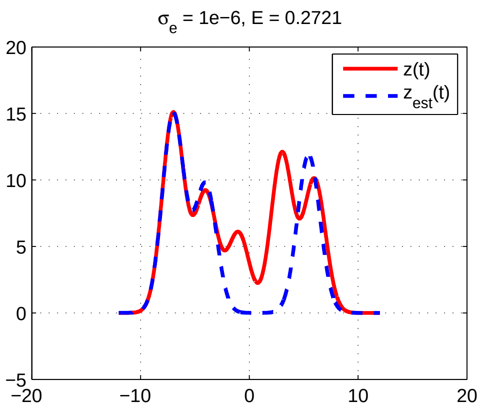
4.1 Annihilating Filter and Root-Finding

In [10], for signals of the form (2) and certain sampling kernels, the annihilating filter 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. In the same paper, a method for dealing with noisy samples is suggested. Unfortunately, this method seems to be inherently ill-conditioned. In Fig. 2, we show a pair of simulations with the parameters as tabulated in Table 1. We observe from Fig. 2(a) that (even with an oversampling factor of $N/(2K) = 3$) the annihilating filter and root-finding method is not robust to even a miniscule amount of added noise.

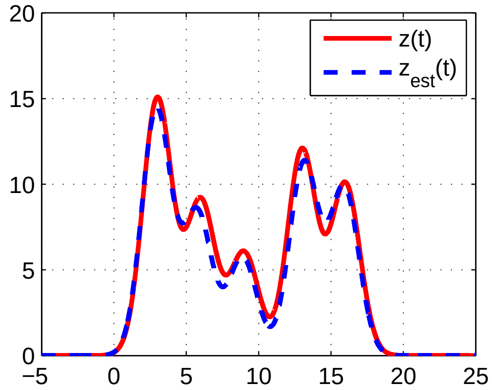
4.2 Gibbs Sampling Algorithm

Initial Demonstration. To demonstrate the evolution the Gibbs sampler, we performed an initial experiment with parameters as above, 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 of the most challenging parameters—the t_ks—in Fig. 3. 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. The true filtered signal $z(t)$ and its estimate $z_{\text{est}}(t)$ are plotted in Fig. 2(b). Note the close similarity between $z(t)$ and $z_{\text{est}}(t)$.

Further Experiments on Simulated Data. To further validate our algorithm, we performed extensive simulations on different problem sizes with different levels of noise [7]. These experiments support the conclusion that 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 [8]. We also find that the noise standard deviation σ_e can be estimated accurately; this may be important in some applications.



(a) The reconstruction using annihilating filter and root-finding completely breaks down when noise of a small standard deviation $\sigma_e = 10^{-6}$ (SNR = 137 dB) is added.



(b) The Gibbs sampling technique gives a much better reconstruction even at a higher noise level $\sigma_e = 2.5$ (SNR = 10.2 dB).

Figure 2: Demonstration of the instability of annihilating filter/root-finding approach and the improvement from Gibbs sampling.

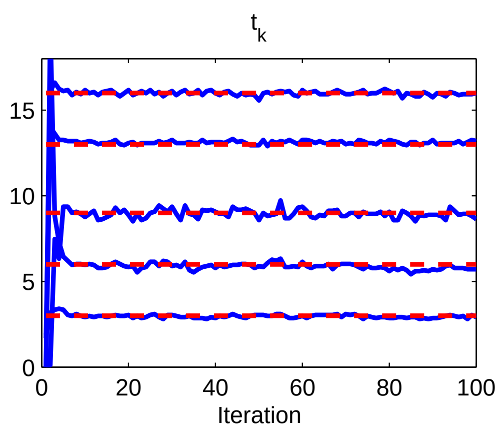


Figure 3: Evolution of the t_ks in the GS algorithm. The true values are indicated by the broken red lines.

5. Concluding Comments

We addressed the problem of reconstructing a signal with FRI given noisy samples. We showed that it is possible to circumvent some of the problems of the annihilating filter and root-finding approach [3, 10]. We introduced the Gibbs sampling algorithm to find the locations and augmented with a least squares approach to find the weights.
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 [2].
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.

References:

[1] J. M. Bernardo and A. F. M. Smith. *Bayesian Theory*. Wiley, 1st edition, 2001.

[2] 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 Processing*, 55(5):1741–1757, 2007.

[3] I. Maravic and M. Vetterli. Sampling and reconstruction of signals with finite rate of innovation in the presence of noise. *IEEE Trans. Signal Processing*, 53(8):2788–2805, 2005.

[4] H. Nyquist. Certain topics in telegraph transmission theory. *Trans. American Institute of Electrical Engineers*, 47:617–644, April 1928.

[5] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 2nd edition, 2004.

[6] C. E. Shannon. Communication in the presence of noise. *Proc. Institute of Radio Engineers*, 37(1):10–21, January 1949.

[7] V. F. Y. Tan and V. K. Goyal. Estimating signals with finite rate of innovation from noisy samples: A stochastic algorithm. *IEEE Trans. Signal Process.*, 56(10):5135–5146, October 2008.

[8] L. Tierney. Markov chains for exploring posterior distributions. Technical Report 560, School of Statistics, Univ. of Minnesota, March 1994.

[9] M. Unser. Sampling–50 years after Shannon. *Proc. IEEE*, 88(4):569–587, 2000.

[10] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE Trans. Signal Processing*, 50(6):1417–1428, 2002.

[11] R. H. Walden. Analog-to-digital converter survey and analysis. *IEEE J. Selected Areas of Communication*, 17(4):539–550, April 1999.