

OPTIMIZATION-BASED RECOVERY FROM RATE OF INNOVATION SAMPLES

Tomer Michaeli and Yonina C. Eldar

Department of Electrical Engineering,
Technion–Israel Institute of Technology, Haifa, Israel

ABSTRACT

We address the problem of recovering signals from samples taken at their rate of innovation. Our only assumption is that the sampling system is such that the parameters defining the signal can be stably determined from the samples. As such, our analysis subsumes previously studied nonlinear acquisition devices and nonlinear signal classes. Our strategy relies on minimizing a least-squares (LS) objective, which is generally non-convex and might possess many local minima. We show, though, that under the stability hypothesis, any optimization method designed to trap a stationary point necessarily converges to the true solution. We demonstrate the usefulness of our approach in recovering finite-duration and periodic pulse streams.

Index Terms— Finite rate of innovation, nonlinear distortion, generalized sampling.

1. INTRODUCTION

Sampling theory is concerned with recovery of continuous-time signals from their samples. Two important aspects of every sampling theorem are the *prior* on the signal and the *sampling mechanism*. For example, in the Shannon sampling theorem the prior is that the signal is π/T -bandlimited and the measurements are pointwise uniformly-spaced samples at a rate of $1/T$ [1].

Until recently, much of the sampling literature treated linear acquisition devices and linear signal priors, that is, families of signals that form subspaces of L_2 (see [1] and references therein). These include shift-invariant (SI) spaces, of which the bandlimited prior is a special case [2]. Subspace models and linear sampling result in linear recovery algorithms that are often easy to implement. However, many real-world signals do not conform to the subspace model and practical samplers often introduce nonlinear distortions [3].

One deviation from the linear setting concerns *nonlinear sampling of linear models*. This topic has been treated in several works (see *e.g.*, [3] and references therein) which primarily focused on systems with memoryless nonlinear distortions and SI signal priors.

Another departure from the classical setting, which has recently drawn much attention, corresponds to *linear sampling of nonlinear models*. Particularly, focus has been devoted to finite rate-of-innovation (FRI) signals [4], which are classes of functions defined by a finite number ρ of parameters per time unit. The quantity ρ , referred to as the rate of innovation, is often far lower than the Nyquist rate. Yet, various sampling settings allow for perfect recovery from samples taken at a rate of ρ . This has been demonstrated for several families of pulse streams [4, 5, 6, 7] as well as for multiband signals with unknown band locations [8].

Both lines of work treating nonlinear sampling of linear models and linear sampling of nonlinear models lack the full generality

required for deployment in a wide range of practical systems. In particular, common to all nonlinear sampling works is the assumption that the nonlinearity is memoryless, while this is not the case in many real-world applications. Similarly, all nonlinear models treated in the literature correspond to unions of subspaces [9], with the vast majority focusing on pulse streams. These do not include, for example, FRI signals such as continuous-phase modulation (CPM) transmissions. Furthermore, even within the restricted category of pulse streams, solutions are only available for a few special cases of signal structures and sampling devices. These solutions are very unstable in certain situations [10].

In this paper, we address the problem of reconstructing arbitrary FRI signals from possibly nonlinear measurements obtained at the rate of innovation. The only assumption we make on the sampling mechanism and signal prior is that the parameters defining the signal can be stably recovered from the samples. This assumption must be made by any practical sampling theorem that attempts to recover the signal parameters, whether explicitly or implicitly. Our approach is based on minimization of the error norm between the given set of samples and those of our signal estimate. Our main result is that, under the stability assumption, this least-squares criterion possesses a unique stationary point. Consequently, any optimization algorithm designed to trap a stationary point, will necessarily converge to the true parameters. In particular, we show that the steepest-descent and Gauss-Newton methods can be used to recover the signal parameters.

Our approach is suited to a family of problems, which supersedes those treated by existing techniques. In particular, we do not assume that the sampling mechanism is linear or that the class of feasible signals forms a union of subspaces. It also provides a unified framework for recovering signals from samples taken at their rate of innovation. Thus, rather than tailoring a different algorithm for every possible combination of sampling method and signal prior, we can always apply the same optimization technique to recover the signal parameters.

The proofs of the results presented here can be found in [11].

2. PROBLEM SETTING

The signal classes in focus are those that are determined by a finite number of parameters per time unit. The τ -local rate of innovation of a signal $x(t)$, denoted ρ_τ , is the minimal number of parameters defining any length- τ segment of $x(t)$, divided by τ . An FRI signal is one for which ρ_τ is finite, at least for large enough τ .

Perhaps the simplest class of FRI signals corresponds to functions that can be expressed as

$$x(t) = \sum_{m \in \mathbb{Z}} a_m g(t - mT), \quad (1)$$

where $\{a_m\}$ are unknown coefficients, $g(t)$ is a given pulse shape, and $T > 0$. This set of signals is a linear space, termed *shift-*

This work was supported in part by the Israel Science Foundation under Grant no. 170/10 and by a Google Research Award.

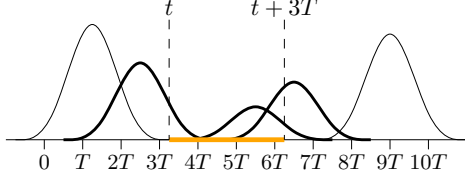


Fig. 1: The signal model (2) with a pulse $g(t)$ supported on $[-2T, 2T]$ and minimal separation of T . Bold pulses are those that affect the segment $[t, t + 3T]$. The maximal number of pulses that can affect the segment is 7 so that $\rho_{3T} = 2 \times 7 / (3T) = 14 / (3T)$.

invariant (SI) [2]. The model (1) can represent bandlimited signals, spline functions, pulse-amplitude modulation (PAM) transmissions and more [1]. If $\text{supp}\{g(t)\} \subseteq [t_a, t_b]$, then any segment $[t, t + \tau]$ is affected by no more than $\lceil (t_b - t_a + \tau) / T \rceil$ coefficients. Thus, the τ -local rate of innovation of (1) is¹ $\rho_\tau = \lceil (t_b - t_a + \tau) / T \rceil / \tau$ and the asymptotic rate is $\lim_{\tau \rightarrow \infty} \rho_\tau = 1/T$.

A more complicated model results when the location of the pulses are unknown a-priori, as often happens in channel sounding scenarios. In these cases,

$$x(t) = \sum_{m \in \mathbb{Z}} a_m g(t - t_m), \quad (2)$$

where both $\{a_m\}$ and $\{t_m\}$ are unknown parameters. If we fix the time-delays $\{t_m\}$ and vary only the amplitudes $\{a_m\}$ then we get a subspace. But different choices of time-delays result in different subspaces so that overall (2) corresponds to a *union of subspaces*. Assuming that the minimal separation between any two of the time delays $\{t_m\}$ is T , this model is determined by (at most) twice the number of parameters defining (1) per time unit, as demonstrated in Fig. 1. Therefore, the associated τ -local rate of innovation is twice that of (1) and the asymptotic rate is $2/T$.

The model (2) and several of its variants have received the largest amount of attention in the FRI literature. However, as shown in [11], other interesting FRI signal classes exist, all of which can be treated within our framework. These include multiband signals with unknown band locations and CPM transmissions, which do not conform to the union-of-subspace structure.

In this paper, we focus on the recovery of an arbitrary segment from an FRI signal. By definition, any length- τ interval is determined by no more than $K = \lceil \tau \rho_\tau \rceil$ parameters. Therefore, any such segment is a vector in some Hilbert space \mathcal{H} , which is known to lie within the set

$$\mathcal{X} = \{x = h(\theta) : \theta \in \mathcal{A}\}, \quad (3)$$

where \mathcal{A} is an open set in \mathbb{R}^K and $h : \mathcal{A} \rightarrow \mathcal{H}$ is some given function. For example, with a minimal separation of T seconds between any two time delays $\{t_m\}$, the segment $[T + t_b, MT + t_a]$ from (2) is affected by no more than M pulses. Indexing these pulses as $m = 1, \dots, M$, this setting corresponds to the $2M$ -dimensional parameter vector $\theta = (t_1, \dots, t_M, a_1, \dots, a_M)^T$ and to the function $h : \mathbb{R}^{2M} \rightarrow L_2([T + t_b, MT + t_a])$ given by

$$h : (t_1, \dots, t_M, a_1, \dots, a_M)^T \mapsto \sum_{m=1}^M a_m g(t - t_m). \quad (4)$$

¹According to our definition, if $g(t)$ is not compactly supported then the rate of innovation is infinite, unless there is only a finite number of pulses. Thus, for example, bandlimited signals (which correspond to $g(t) = \text{sinc}(t/T)$) are not considered FRI in this paper.

In addition to the recovery of x , it is often of interest to identify the parameters θ defining it. This goal cannot be achieved if the parametrization of the set \mathcal{X} is redundant in the sense that there exist parameters $\theta_1 \neq \theta_2$ such that $h(\theta_1) = h(\theta_2)$. To be able to recover θ in a stable manner, we require the slightly stronger condition that

$$\alpha_h \|\theta_1 - \theta_2\|_{\mathbb{R}^K} \leq \|h(\theta_1) - h(\theta_2)\|_{\mathcal{H}} \quad (5)$$

for some constant $\alpha_h > 0$ and for all $\theta_1, \theta_2 \in \mathcal{A}$.

As shown in [11], condition (5) has several implications to union-of-subspace models. Specifically, suppose that θ comprises a sub-vector² θ^N , which determines a subspace \mathcal{A}_{θ^N} in \mathcal{H} and a sub-vector θ^L that determines a vector within \mathcal{A}_{θ^N} . A special case is (2), in which θ^N comprises $\{t_\ell\}$ and θ^L comprises $\{a_\ell\}$. In this situation, condition (5) implies that the feasible set \mathcal{A} must be such that the elements of θ^L are bounded away from zero, the vector θ^N is restricted to a bounded set in \mathbb{R}^K and its elements are sufficiently separated. This can be achieved in (2) by requiring that

$$a_m > a_0, \quad T_{\min} < t_m - t_{m-1} < T_{\max}, \quad (6)$$

for every $m = 1 \dots, M$, and for some $a_0 > 0$, $0 < T_{\min} < T_{\max} < \infty$ and arbitrary t_0 .

Our goal is to recover x by observing N generalized samples $\mathbf{c} = (c_1, \dots, c_N)^T$ obtained as

$$\mathbf{c} = S(x), \quad (7)$$

where $S : \mathcal{H} \rightarrow \mathbb{R}^N$ is some (possibly nonlinear) Fréchet differentiable operator. This representation of the sampling mechanism is more general than the widely used linear setting, in which $c_n = \langle x, s_n \rangle$ for some set of vectors $\{s_n\}_{n=1}^N$ in \mathcal{H} . In particular, (7) may account for nonlinear distortions introduced by the sampling device. For example, S can represent the samples $c_n = f(\langle x, s_n \rangle)$, where $f(\cdot)$ is a nonlinear sensor response.

We require that

$$\alpha_s \|x_2 - x_1\|_{\mathcal{H}} \leq \|S(x_1) - S(x_2)\|_{\mathbb{R}^N} \leq \beta_s \|x_2 - x_1\|_{\mathcal{H}} \quad (8)$$

for all $x_1, x_2 \in \mathcal{X}$, where $0 < \alpha_s \leq \beta_s < \infty$. The left-hand inequality ensures that if two signals x_1 and x_2 are sufficiently different, then their samples $S(x_1)$ and $S(x_2)$ are different as well. In particular, it implies that two different signals $x_1, x_2 \in \mathcal{X}$ cannot produce the same set of samples, so that there is a unique recovery $x \in \mathcal{X}$ associated with every valid set of samples $\mathbf{c} = S(x) \in \mathbb{R}^N$.

Conditions (8) and (5) lie at the heart of any practical sampling theorem, whether implicitly or not.

3. LEAST SQUARES RECOVERY

As a first step towards devising a general reconstruction strategy, we determine the minimal number of samples N required for perfect recovery. Interestingly, conditions (8) and (5) implicitly impose a limitation on N .

Proposition 1 Assume that conditions (5) and (8) hold. Then

$$N \geq K + \max_{x_1 \in \mathcal{X}} \dim \left(\mathcal{N} \left(\left(\frac{\partial S}{\partial x} \Big|_{x_1} \right)^* \right) \right), \quad (9)$$

where $\partial S / \partial x$ is the Fréchet derivative of $S(x)$.

²The superscripts ‘N’ and ‘L’ stand for nonlinear and linear respectively, intending as a reminder that h is linear in θ^L and nonlinear in θ^N .

Proposition 1 shows that the minimal number of samples N required for perfect recovery is the number of parameters K defining x . In other words, stable recovery is impossible when sampling below the rate of innovation. Furthermore, we see that sampling at the rate of innovation is insufficient if the null space of $(\partial S/\partial x)^*$ is nonempty at some $x \in \mathcal{X}$. We therefore focus on the case in which $N = K$ samples of $x(t)$ are obtained with an operator S satisfying

$$\mathcal{N}\left(\left(\frac{\partial S}{\partial x}\bigg|_{x_1}\right)^*\right) = \{\mathbf{0}\}, \quad \forall x_1 \in \mathcal{X}. \quad (10)$$

This corresponds to sampling at the rate of innovation.

To recover a signal $x = h(\theta_0)$ from its samples $\mathbf{c} = S(x)$, where $\theta_0 \in \mathbb{R}^K$ is unknown, it is natural to seek the minimizer of

$$\varepsilon(\theta) = \frac{1}{2} \|S(h(\theta)) - \mathbf{c}\|_{\mathbb{R}^K}^2 = \frac{1}{2} \|\hat{\mathbf{c}}(\theta) - \mathbf{c}\|_{\mathbb{R}^K}^2, \quad (11)$$

where we defined $\hat{\mathbf{c}}(\theta) = S(h(\theta))$. The reasoning behind this choice follows from the following observation

Proposition 2 Assume that $N = K$ and conditions (5) and (8) hold. Then θ_0 is the unique global minimizer of $\varepsilon(\theta)$.

When the samples are perturbed by white Gaussian noise, (11) yields the maximum-likelihood (ML) estimate of θ from \mathbf{c} .

Unfortunately, the function $\varepsilon(\theta)$ is generally non-convex and might possess many local minima. It therefore seems that standard optimization techniques may fail in finding its global minimizer θ_0 . However, as we show next, when sampling at the rate of innovation, assumptions (5) and (8) guarantee that θ_0 is the unique stationary point of $\varepsilon(\theta)$. Thus, any algorithm designed to trap a stationary point, necessarily converges to the true parameter vector θ_0 .

Theorem 1 Assume that $N = K$ and conditions (5), (8) and (10) hold. Then $\nabla \varepsilon(\theta_1) = \mathbf{0}$ only if $\theta_1 = \theta_0$.

Theorem 1 shows that, rather than developing a different algorithm for every choice of signal family and sampling method, we can always employ the same general-purpose optimization technique to find the stationary point of (11).

4. ITERATIVE RECOVERY

There are numerous optimization algorithms that can be used to find the stationary point of the objective function $\varepsilon(\theta)$ over \mathcal{A} . For simplicity, we focus here on unconstrained optimization methods, namely those that can be applied when $\mathcal{A} = \mathbb{R}^K$. This does not limit the generality of the discussion since if $\mathcal{A} \neq \mathbb{R}^K$, then the constrained problem $\min_{\theta \in \mathcal{A}} \varepsilon(\theta)$ can be transformed into the unconstrained problem $\min_{\tilde{\theta} \in \mathbb{R}^K} \varepsilon(p(\tilde{\theta}))$, where $p: \mathbb{R}^K \rightarrow \mathcal{A}$ is one-to-one and onto. For example, the model (2) with the constraints (6) can be handled by defining

$$\tilde{\theta}_m^L = \ln(a_m - a_0), \quad \tilde{\theta}_m^N = \tan\left(\pi \frac{t_m - t_{m-1} - \bar{T}}{\Delta}\right), \quad (12)$$

where $\bar{T} = (T_{\max} + T_{\min})/2$ and $\Delta = T_{\max} - T_{\min}$, so that

$$a_m = e^{\tilde{\theta}_m^L} + a_0, \quad t_m = t_0 + m\bar{T} + \frac{\Delta}{\pi} \sum_{i=1}^m \arctan\left(\tilde{\theta}_i^N\right). \quad (13)$$

With this choice, $\tilde{\theta}^L$ and $\tilde{\theta}^N$ vary over the entire space \mathbb{R}^M .

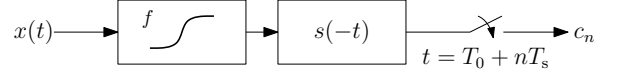


Fig. 2: Nonlinear and nonideal sampling.

Most unconstrained optimization methods start with an initial guess θ^0 and perform iterations of the form

$$\theta^{\ell+1} = \theta^\ell - \gamma^\ell \mathbf{B}^\ell \nabla \varepsilon(\theta^\ell), \quad (14)$$

where γ^ℓ is a scalar step size obtained by means of a one dimensional search and \mathbf{B}^ℓ is a positive definite matrix. As we show next, convergence guarantees for such methods can be obtained under assumptions on the line-search method and the behavior of h on the level-set $\mathcal{N} = \{\theta : \varepsilon(\theta) \leq \varepsilon(\theta^0)\}$.

Theorem 2 Suppose that $N = K$, conditions (5), (8) and (10) hold, and the Fréchet derivative $\partial h/\partial \theta$ is Lipschitz continuous over \mathcal{N} . Consider the iterations (14), with backtracking line search [12]. Then each of the following options guarantees that $\theta^\ell \rightarrow \theta_0$:

1. $\mathbf{B}^\ell = \mathbf{I}$.
2. $\mathbf{B}^\ell = ((\partial \hat{\mathbf{c}}/\partial \theta|_{\theta^\ell})^* (\partial \hat{\mathbf{c}}/\partial \theta|_{\theta^\ell}))^{-1}$ and h is Lipschitz continuous over \mathcal{N} .

The two options correspond, respectively, to the steepest-descent and Gauss-Newton methods.

5. APPLICATION TO CHANNEL SOUNDING

We now demonstrate our approach in the setting (2). We assume that $x(t)$ comprises only M pulses and focus on recovering them from observations of the segment $[0, 1]$. We adopt the assumptions (6) and transform the optimization problem into an unconstrained one by using the transformation described in (12) and (13).

Consider first the sampling system of Fig. 2, in which $x(t)$ is sampled after passing through an amplitude limiter $f(\cdot)$ and being convolved with a filter $s(-t)$. Fig. 3 demonstrates the behavior of the algorithm in recovering $M = 2$ pulses from $N = 4$ samples contaminated by white Gaussian noise. Here, $g(t)$ and $s(t)$ were taken to be Gaussian functions with standard deviations 0.05 and 0.1, respectively, and $f(c)$ was chosen to be $100 \arctan(0.01c)$. We chose $T_s = 1/4$ and $T_0 = 1/8$ so that $\{s_n(t)\}$ equally span the entire observation segment. The constraints (6) corresponded to $a_0 = 0.1$, $T_{\min} = 0.3$, $T_{\max} = 0.7$ and $t_0 = -0.3$. The true parameters were $t_1 = 0.2$, $t_2 = 0.8$, $a_1 = 1$ and $a_2 = 5$. The figure depicts the mean squared error (MSE) in $x(t)$, defined as $\mathbb{E}[\int_0^\tau |x(t) - \hat{x}(t)|^2 dt]$, as a function of the signal-to-noise (SNR) ratio. The solid line corresponds to the Cramér-Rao bound (CRB), developed in [10], which is a lower bound on the MSE attainable by any unbiased estimation technique. As can be seen, the MSE of our method coincides with the CRB in high SNR scenarios and outperforms it at low SNR levels. This is a result of the fact that our technique is biased.

Next, consider the situation in which $g(t)$ is a 1-periodic function and the sampling system is that depicted in Fig. 4 with

$$s_n(t) = \begin{cases} 1 & n = 0, \\ \cos(2\pi n t/\tau) & 1 \leq n \leq M, \\ \sin(2\pi n t/\tau) & M+1 \leq n \leq 2M. \end{cases} \quad (15)$$

Figure 5 compares between iterative recovery and the method of [7], which relies on the annihilating filter technique. Here, $M = 2$ pulses

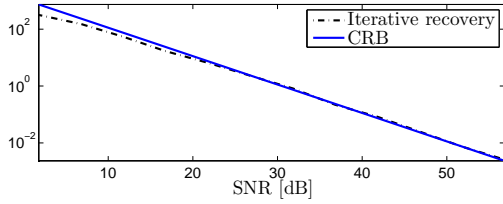


Fig. 3: MSE versus SNR for recovery with the system of Fig. 2.

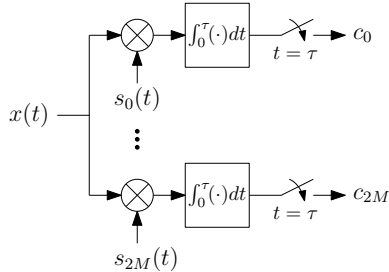


Fig. 4: Linear multichannel sampling.

were recovered from $N = 2M + 1 = 5$ samples. The pulse $g(t)$ had quadratically decaying Fourier coefficients. The true time delays were $t_1 = 1/\sqrt{15} \approx 0.2582$ and $t_2 = 1/\sqrt{2} \approx 0.7071$ and the true amplitudes were randomly generated to yield $a_1 \approx 0.5285$ and $a_2 \approx 0.14$. The figure depicts the performance of both approaches, as well as the CRB. As can be seen, the optimization-based approach outperforms the annihilating-filter method at all SNR levels.

Finally, we note that our algorithm can be used to detect situations in which not all possible time-delay constellations can be stably recovered. Indeed, we proved that if all $\theta \in \mathcal{A}$ can be stably recovered then the algorithm converges to the global minimum θ_0 at which $\varepsilon(\theta_0) = 0$. Therefore, termination of the algorithm at a point θ_1 at which $\varepsilon(\theta_1) \neq 0$ indicates that not all $\theta \in \mathcal{A}$ can be stably recovered. In fact, it can be shown that the problematic constellation is θ_1 itself. Namely, *no method can stably recover θ_1* in this setting.

To demonstrate this, we applied the algorithm in the setting of Fig. 4. We took four sinusoidal sampling functions (two sines and two cosines) with frequencies 1 and 3. While the true parameters were $(t_1, t_2, a_1, a_2) = (0.2, 0.8, 1, 5)$, the algorithm converged to $(0.34, 0.85, 0.41, 3.1)$. This means that the latter constellation cannot be recovered stably by any method. Figure 6 depicts the CRB for estimating θ as function of $t_2 \in [0.85, 1]$ for $t_1 = 0.34$. As can be seen, the CRB indeed tends to infinity as t_2 approaches 0.85.

6. CONCLUSION

We studied recovery of FRI signals from samples taken at their rate of innovation. We showed that when the parameters can be stably recovered, this can be achieved by unconstrained optimization methods. Our approach thus provides a simple means for treating a wide range of FRI signal classes and sampling methods. We demonstrated the usefulness of our strategy in reconstructing finite and periodic pulse streams from nonlinear and nonideal samples.

7. REFERENCES

[1] Y. C. Eldar and T. Michaeli, “Beyond bandlimited sampling,” *IEEE Signal Process. Mag.*, vol. 26, no. 3, pp. 48–68, 2009.

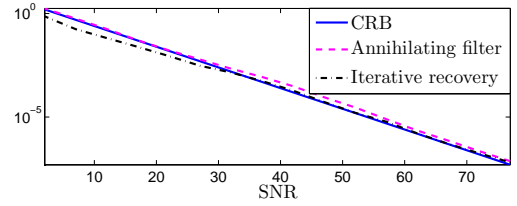


Fig. 5: MSE versus SNR for recovery with the system of Fig. 4.

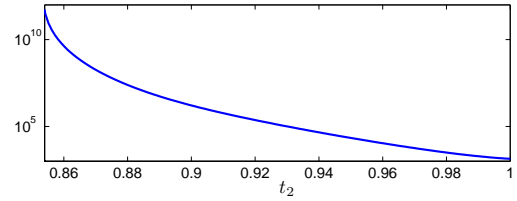


Fig. 6: CRB versus t_2 for fixed t_1 in a setting with sinusoidal sampling kernels with nonconsecutive frequencies.

- [2] M. Unser and A. Aldroubi, “A general sampling theory for nonideal acquisition devices,” *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 2915–2925, 1994.
- [3] T. G. Dvorkind, Y. C. Eldar, and E. Matusiak, “Nonlinear and non-ideal sampling: Theory and methods,” *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 5874–5890, 2008.
- [4] 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.
- [5] 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. 7, pp. 1741–1757, 2007.
- [6] R. Tur, Y. C. Eldar, and Z. Friedman, “Low rate sampling of pulse streams with application to ultrasound imaging,” *IEEE Trans. Signal Process.*, vol. 59, no. 4, 2011.
- [7] K. Gedalyahu, R. Tur, and Y. C. Eldar, “Multichannel sampling of pulse streams at the rate of innovation,” *IEEE Trans. Signal Process.*, vol. 59, no. 4, 2011.
- [8] M. Mishali and Y. C. Eldar, “Blind multi-band signal reconstruction: Compressed sensing for analog signals,” *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 993–1009, 2009.
- [9] Y. M. Lu and M. N. Do, “A theory for sampling signals from a union of subspaces,” *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2334–2345, 2008.
- [10] Z. Ben-Haim, T. Michaeli, and Y. C. Eldar, “Performance bounds and design criteria for estimating finite rate of innovation signals,” *IEEE Trans. Inf. Theory*, submitted. [Online]. Available: <http://arxiv.org/abs/1009.2221>
- [11] T. Michaeli and Y. C. Eldar, “Xampling at the rate of innovation,” *IEEE Trans. Signal Process.*, to appear.
- [12] J. Nocedal and S. J. Wright, *Numerical optimization*. Springer verlag, 1999.