Recovery of Sparse Translation-Invariant Signals With Continuous Basis Pursuit

Chaitanya Ekanadham, Daniel Tranchina, and Eero P. Simoncelli, *Fellow, IEEE*

Abstract—We consider the problem of decomposing a signal into a linear combination of features, each a continuously translated version of one of a small set of elementary features. Although these constituents are drawn from a continuous family, most current signal decomposition methods rely on a finite dictionary of discrete examples selected from this family (e.g., shifted copies of a set of basic waveforms), and apply sparse optimization methods to select and solve for the relevant coefficients. Here, we generate a dictionary that includes auxiliary interpolation functions that approximate translates of features via adjustment of their coefficients. We formulate a constrained convex optimization problem, in which the full set of dictionary coefficients represents a linear approximation of the signal, the auxiliary coefficients are constrained so as to only represent translated features, and sparsity is imposed on the primary coefficients using an L1 penalty. The basis pursuit denoising (BP) method may be seen as a special case, in which the auxiliary interpolation functions are omitted, and we thus refer to our methodology as continuous basis pursuit (CBP). We develop two implementations of CBP for a one-dimensional translation-invariant source, one using a first-order Taylor approximation, and another using a form of trigonometric spline. We examine the tradeoff between sparsity and signal reconstruction accuracy in these methods, demonstrating empirically that trigonometric CBP substantially outperforms Taylor CBP, which, in turn, offers substantial gains over ordinary BP. In addition, the CBP bases can generally achieve equally good or better approximations with much coarser sampling than BP, leading to a reduction in dictionary dimensionality.

Index Terms—Basis pursuit, feature decomposition, interpolation, L1 optimization, sparsity, transformation-invariance.

I. INTRODUCTION

T HE decomposition of signals into sparse linear combinations of features is an important and well-studied problem, and plays a central role in many applications. A surge of re-

C. Ekanadham is with the Courant Institute of Mathematical Sciences (CIMS), New York University, New York, NY 10003 USA (e-mail: chaitu@math.nyu.edu).

D. Tranchina is with CIMS, the Center for Neural Science (CNS), and the Department of Biology, New York University, New York, NY 10003 USA (e-mail: tranchin@courant.nyu.edu).

E. P. Simoncelli is with the Howard Hughes Medical Institute (HHMI), CNS, and CIMS, New York University, New York, NY 10012 USA (e-mail: eero.simoncelli@nyu.edu).

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

cent effort focuses on representing a signal as a noisy superposition of the smallest possible subset of features drawn from a large finite dictionary. However, many real signals are generated by processes that obey natural invariances (e.g., translation-invariance, dilation-invariance, rotation-invariance). For example, the acoustic features of many sounds can occur at arbitrary times, dilations, or frequencies. In photographic images, features can occur at various spatial locations, scales, and orientations. Radar [1] and seismic [2] signals are often composed of elementary features that are translated and frequency-modulated. For these signals, a desirable goal is to identify features in the signal along with their associated amplitudes and transformation parameters.

In the majority of published examples, this problem is solved by constructing a finite dictionary that reflects the invariant structure: one discretely samples the transformation parameters and applies these to a finite set of elementary features. For example, dictionaries for sound processing, whether hand-constructed or learned, are commonly "convolutional", containing time-delayed copies of template waveforms (e.g., [3]–[5]). Dictionaries for image representation typically contain features that are translated, and in some cases, dilated and rotated (e.g., [6]–[9]).[1] Using this finite dictionary $\{\phi_i(t)\}$, one then attempts to optimize the corresponding coefficients:

$$\min_{\vec{x}\in\mathbb{R}^d} \|\vec{x}\|_0, \quad \text{s.t.} \quad \left\|y(t) - \sum_{i=1}^{d} x_i\phi_i(t)\right\|_2 \leq \epsilon \qquad (1)$$

where $y(t)$ is a noisy observation, $\vec{x}$ is a coefficient vector, ϵ bounds the L_2 norm of the noise, and $\|\cdot\|_0$ indicates the L_0 pseudonorm (number of nonzero elements). In general, this objective can only be minimized via exhaustive search of all 2^d subsets of the dictionary, making it infeasible in practice. However, two broad classes of approximate solutions have been widely studied in the literature. The first consists of greedy methods, dating back to variable selection methods in the 1970s [13]. These methods are exemplified by the well-known "matching pursuit" algorithm of Mallat and Zhang [14], and include a variety of more recent "iterative thresholding" methods [15]–[19]. The general idea is to solve sequentially for the nonzero coefficients, at each step choosing the dictionary element(s) that best explain the current residual. The second category of solutions arises from convex relaxations of (1), which typically employ the L_1 norm [20], [21].

---

[1]Many examples of sparse decomposition on images have been applied to nonoverlapping square blocks of pixels (e.g., [10]–[12]), but the effective dictionary for representing the entire image is the union of dictionary elements for each block, and thus consists of translated copies of the block dictionary elements.

The use of a finite dictionary allows one to approximate the full nonlinear problem with a more tractable linear inverse problem. However, the ability of this discrete dictionary to compactly represent signals depends critically on the spacing at which the dictionary is sampled. In general, a very fine sampling is required, resulting in a very large and ill-conditioned dictionary. This ill-conditioning, in turn, is unfavorable for the approximation methods mentioned above. Furthermore, it is unclear how to estimate the underlying amplitudes and transformation parameters from the coefficients.

Here, we propose an alternative linear approximation to the full nonlinear problem. We focus on the problem of translation-invariant, one-dimensional signals (although the methods generalize to other transformations, and higher dimensions). We construct a group of functions that can span local translations of the feature templates via continuous variation of their coefficients (as a concrete example, consider the original templates and their derivatives, which can approximate local translations through a first-order Taylor approximation). A signal of interest is then represented in this dictionary by "block-sparse" coefficients, where each nonzero coefficient block represents an amplitude-scaled and translated template. We formulate an objective function in which the coefficients are constrained so as to only represent scaled/transformed templates, and use an L_1 penalty to impose sparsity on the blocks. The advantage of this approach over ordinary BP is threefold: 1) better approximation of translation-invariant signals, 2) a smaller basis, which leads to sparser solutions via convex optimization, and 3) an explicit mapping from this representation to amplitudes and transformation parameters. An early version of this work was presented in [22].

II. PROBLEM FORMULATION

We begin by formulating a simple generative model for translation-invariant signals, as well as the maximum *a posteriori* (MAP) estimation framework for inferring the most likely parameters given the observed signal. Assume we observe a signal that is a noisy superposition of scaled time-shifted copies of a single known elementary waveform $f(t)$ on a finite interval $[0, T]$:

$$y(t) = \sum_{j=1}^{N} a_j f(t - \tau_j) + \eta(t) \qquad (2)$$

where $\eta(t)$ is a Gaussian white noise process with power σ^2, the *event times* $\{\tau_j\}$ are a Poisson process with rate μ, and the *event amplitudes* $\{a_j\}$ are drawn independently from a density $P_A(a)$. The inverse (inference) problem is to recover the most likely parameters $\{\tau_j, a_j\}$ given $y(t)$. This amounts to maximizing the posterior distribution $P(\{\tau_j, a_j\}|y(t))$, which reduces, on taking the negative log, to solving:

$$\min_{\substack{N, \{a_j\} \\ \{\tau_1 < \ldots < \tau_N\}}} \frac{1}{2\sigma^2} \left\| y(t) - \sum_{j=1}^{N} a_j f(t - \tau_j) \right\|_2^2$$
$$- N \log(\mu) - \sum_{j=1}^{N} \log P_A(a_j) \qquad (3)$$

Unfortunately, solving (3) directly is intractable, due to the discrete nature of N and the nonlinear embedding of the τ_j's within the argument of the waveform $f(\cdot)$. It is thus desirable to find alternative formulations that i) approximate the signal posterior distribution well, ii) have parameters that can be tractably estimated, and iii) have an intuitive mapping back to the original representation.

III. CONVENTIONAL SOLUTION: DISCRETIZATION AND BP

A standard simplification of the problem is to discretize the event times at a spacing that is fine enough that the Poisson process is well-approximated by a Bernoulli process. The interval $[0, T]$ is divided into $N_\Delta = \lceil \frac{T}{\Delta} \rceil$ time bins of size Δ, and $\mu\Delta$ approximates the probability of an event in each bin.[2] This discrete process is represented by a vector $\vec{x} \in \mathbb{R}^{N_\Delta}$, whose elements x_i are interpreted as the amplitude of any event in the interval $\left(\frac{(2i-1)\Delta}{2}, \frac{(2i+1)\Delta}{2} \right)$. The corresponding prior probability distribution on each x_i is a mixture of a point mass at zero, and $P_A(\cdot)$:

$$P(\vec{x}) = \prod_{i=1}^{N_\Delta} \left[(1 - \mu\Delta)\delta(x_i) + (\mu\Delta)P_A(x_i) \right]. \qquad (4)$$

By taking the log and applying standard approximations, the MAP estimation for this model can be approximated by

$$\min_{\vec{x}} \frac{1}{2\sigma^2} \|y(t) - (F_\Delta \vec{x})(t)\|_2^2 - \log(\mu\Delta)\|\vec{x}\|_0 - \sum_{i:x_i \neq 0} \log P_A(x_i) \qquad (5)$$

where F_Δ is a linear operator representing a fixed dictionary of time-shifted copies of $f(\cdot)$:

$$(F_\Delta \vec{x})(t) := \sum_{i=1}^{N_\Delta} x_i f(t - i\Delta). \qquad (6)$$

The advantage of the time discretization is that the data fidelity term is now a quadratic function of the parameters. Despite this, solving (5) exactly is NP-hard due to the L_0 term [23], and so we must either resort to approximate algorithms, or introduce further approximations in the objective. A well-known convex relaxation (known as the LASSO [20] or basis pursuit denoising [21]) replaces the nonconvex prior terms in (5) with an L_1 penalty term, resulting in

$$\min_{\vec{x}} \frac{1}{2\sigma^2} \|y(t) - (F_\Delta \vec{x})(t)\|_2^2 + \lambda\|\vec{x}\|_1 \qquad (7)$$

where $\lambda > 0$ is a parameter to be determined. This was previously applied in the convolutional setting, for example in [24] and [25]. Note that probabilistically, the second term indicates that we have effectively approximated the nonconvex mixture prior of (4) with a Laplacian distribution, $P(x) \propto e^{-\lambda|x|}$.

The global optimum of (7) can be found using standard quadratic programming methods. A well-known article by Candès *et al.* [26] provides sufficient conditions under which L_1-minimizing solutions are good approximations for L_0-minimizing solutions (up to a term that is linear in the L_2 norm of

---

[2]The probability of two or more events is $O(\Delta^2)$, which is negligible for Δ small.
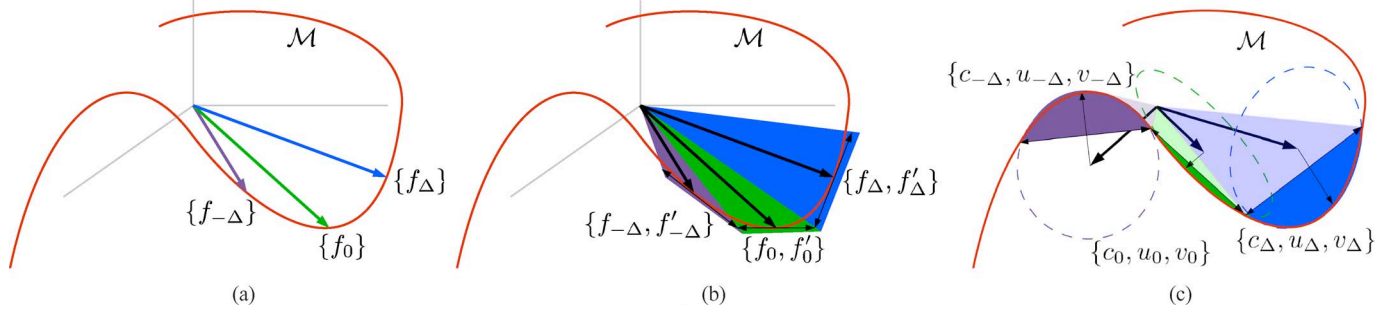
Fig. 1. Illustration of the three approximations of the manifold of translates of the waveform, $\mathcal{M}_{f,T}$. (a) The standard basis pursuit (BP) dictionary, F_Δ, as used in (7), consists of discrete time-shifts of the waveform $f(t)$. (b) Continuous basis pursuit with first-order Taylor interpolator (CBP-T), as specified by (13). Each pair of functions, $(f_{i\Delta}, f'_{i\Delta})$, with properly constrained coefficients, represents a triangular region of the space (shaded regions). (c) Continuous basis pursuit with polar interpolation (CBP-P), as specified by (20). Each triplet of functions, $(c_{i\Delta}, u_{i\Delta}, v_{i\Delta})$, represents a section of a cone [see Fig. 3(b) for parameterization].

the noise). Roughly speaking, the condition limits the correlations between small subsets of dictionary elements. However, this condition is rarely satisfied in the convolutional setting (even when f is smooth) because the validity of the discrete approximation typically requires Δ to be quite small, resulting in a highly correlated dictionary F_Δ. In addition to these concerns about the tradeoff between the quality of discrete approximation and the convex relaxation, the discretization of time also leaves us with no well-defined mapping between the solution of (5) and the continuous parameters $\{\tau_j, a_j\}$ that optimize the original objective of (3).

One can understand the extent of these problems more concretely by focusing on a single time-shifted waveform $f(t - \tau)$ with $\tau \in (0, \Delta)$, and a dictionary containing the feature at two gridpoints: $F_\Delta = [f(t), f(t - \Delta)]$. Equation (7) is then a 2-D problem, as illustrated in Fig. 2(a). The optimal solution corresponds to a point at which an elliptical level curve of the first (L_2) term is tangent to a straight-line level curve of the L_1 term. Note that the ellipses are stretched in the direction parallel to the L_1 level lines, because trading off amplitude between the two coefficients does not substantially change the reconstruction error when Δ is small (specifically, one can show mathematically that the two right singular vectors of the basis matrix are parallel and orthogonal to the L_1 level curves). Also shown in Fig. 2(a) is the family of solutions that are obtained as λ varies from 0 to $+\infty$ (red path). Notice that these solutions are not sparse in the L_0 sense (i.e., they do not intersect either of the two axes) until λ is so large that the signal reconstruction is quite poor. This is the case regardless of how small Δ is, and is due both to the failure of the L_1 norm to approximate the L_0 pseudonorm and to the inability of the discrete model to account for continuous event times. In the following sections, we develop and demonstrate a proposed solution for these problems.

IV. CONTINUOUS BASIS PURSUIT

The motivation for the discretization of the inference problem was tractability. Equation (5) approximates the original model well in the limit as Δ goes to 0, but it is only tractably solvable in regimes where Δ is large enough so that correlations are limited and L_1-based relaxations can be employed. In this section, we augment the discrete model by incorporating variables to account for the continuous nature of the event times,

and adapt the LASSO to solve this augmented representation. By accounting for continuous time-shifts, the augmented model not only approximates the original model better (and for a larger range of Δ), but also admits sparser solutions via an L_1-based recovery method. We refer to this method as continuous basis pursuit (CBP).

We return to the original formulation of (3), but we assume (without loss of generality) that the waveform is normalized, $\|f(t)\|_2 = 1$, and that the amplitudes, $\{a_j\}$, are all nonnegative. Our noisy observation arises from a linear superposition of time-shifted waveforms, $f(t - \tau)$ abbreviated as $f_\tau(t)$. The set of all time-shifted and amplitude-scaled waveforms constitute a nonlinear 2-D manifold:

$$\mathcal{M}_{f,T} := \{a f_\tau(t) : a \geq 0, \tau \in [0, T]\} \subset L_2([0, T]). \quad (8)$$

The discretized dictionary, F_Δ, provides a linear subspace approximation of this manifold, as illustrated in Fig. 1(a). But the representation of a single element of the manifold (corresponding to a translated scaled copy of the waveform) will typically be approximated by the superposition of several elements from the dictionary F_Δ. We can remedy this by augmenting the dictionary to include *interpolation* functions, that allow better approximation of the continuously shifted waveforms. We describe two specific examples of this method, and then provide a general form.

A. Taylor Interpolation

If $f(t)$ is sufficiently smooth, one can approximate local shifts of $f(t)$ by linearly combining $f(t)$ and its derivative via a first-order Taylor expansion:

$$f_\tau(t) = f(t) - \tau f'(t) + O(\tau^2) \quad (9)$$

This motivates a dictionary consisting of the original shifted waveforms, $\{f_{i\Delta}(t)\}$, and their derivatives, $\{f'_{i\Delta}(t)\}$. We choose a basis function spacing, Δ, as twice the maximal time-shift such that the first-order Taylor approximation holds within a desired accuracy δ:

$$\Delta := \max \left\{ \Delta' : \max_{|\tau| < \frac{\Delta'}{2}} \|f_\tau(t) - (f(t) - \tau f'(t))\|_2 \leq \delta \right\}. \quad (10)$$
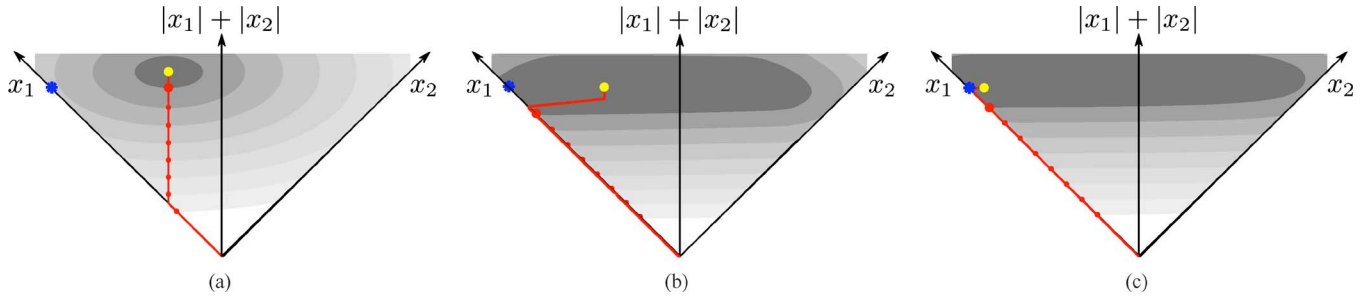
Fig. 2. Geometry of convex objective functions for BP, CBP-T, and CBP-P methods, illustrated in two dimensions. Signal consists of a single waveform, $f(t - 0.65\Delta)$, and the dictionary contains two shifted copies of the waveform, $\{f(t), f(t - \Delta)\}$, along with interpolating functions appropriate for each method. Each plot shows the space of coefficients, $\{x_1, x_2\}$, associated with the two shifted waveforms. Contours of gray-scale regions indicate level surfaces of reconstruction error (L_2 norm term of the objective function). The blue star indicates the true L_0-minimizing solution. The L_1 sparsity measure is the sum of the two coefficient values, corresponding to the vertical position on the plot. Red curve indicates family of solutions traced out as λ is increased from 0 (yellow dot) to infinity. Red points along this path indicate equal increments in reconstruction accuracy, with enlarged point indicating a common value for comparison across all three methods. (a) Standard BP solution (equation (7)). (b) Continuous basis pursuit with Taylor interpolator (equation (13)), using a basis of $\{f_0, f'_0, f_\Delta, f'_\Delta\}$. In this case, the iso-accuracy contours are computed by minimizing the L_2 error over all feasible values of the derivative coefficients, and are thus no longer elliptical. (c) CBP with polar interpolation (equation (20)).

We can then approximate the manifold of scaled and time-shifted waveforms using *constrained* linear combinations of dictionary elements:

$$\mathcal{M}_{f,T} \approx \left\{ x f_{i\Delta}(t) + d f'_{i\Delta}(t) : \begin{array}{l} x \geq 0, \\ |d| \leq \frac{\Delta}{2} x, \\ i = 1, \ldots, N_\Delta \end{array} \right\}. \quad (11)$$

There is a one-to-one correspondence between sums of waveforms on the manifold $\mathcal{M}_{f,T}$ and their respective approximations with this dictionary:

$$\sum_i x_i f_{i\Delta}(t) + d_i f'_{i\Delta}(t) \approx \sum_i x_i f_{(i\Delta - d_i/x_i)}(t). \quad (12)$$

This correspondence holds as long as $\left| \frac{d_i}{x_i} \right| \neq \frac{\Delta}{2}$ (which corresponds to the situation where the waveform is displaced exactly halfway in between two lattice points, and can thus be equally well represented by the basis function and associated derivative on either side). This is illustrated in Fig. 1(b). The inequality constraint on d is essential, and serves two purposes. First, it ensures that the Taylor approximation is only used when it is accurate, so that only time-shifted and scaled waveforms are used in reconstructing the signal. Second, it discourages neighboring pairs $(f_{i\Delta}, f'_{i\Delta})$ and $(f_{(i+1)\Delta}, f'_{(i+1)\Delta})$ from explaining the same event. The inference problem is now solved by optimizing a *constrained* convex objective function:

$$\min_{\vec{x}, \vec{d}} \frac{1}{2\sigma^2} \left\| y(t) - (F_\Delta \vec{x})(t) - \left(F'_\Delta \vec{d}\right)(t) \right\|_2^2 + \lambda \|\vec{x}\|_1$$

$$\text{s.t.} \quad \left\{ \begin{array}{l} x_i \geq 0, \\ |d_i| \leq \frac{\Delta}{2} x_i \end{array} \right\} \text{ for } i = 1, \ldots, N_\Delta \quad (13)$$

where the dictionary F_Δ is defined as in (6), and F'_Δ is a dictionary of time-shifted waveform derivatives $\{f'_{i\Delta}(t)\}$. Equation (12) provides an explicit mapping from appropriately constrained coefficient configurations to event amplitudes and timeshifts. Fig. 2(b) illustrates this objective function for the same single-waveform example described previously. The

shaded regions are the level sets of the L_2 term of (13) visualized in the (x_1, x_2)-plane by minimizing over the derivative coefficients (d_1, d_2). Note that unlike the corresponding BP level sets shown in Fig. 2(a), these are no longer elliptical, and that they allow sparse solutions (i.e., points on the x_1 axis) with low reconstruction error. As a result, for λ sufficiently large, the solution of (13) is not only sparse in the L_0 sense, but also provides a good reconstruction of the signal.

B. Polar Interpolation

Although the Taylor series provides the most intuitive and well-known method of approximating time-shifts, we have developed an alternative interpolator that is much more accurate. The solution is motivated by the observation that the manifold of time-shifted waveforms, $f_\tau(t)$, must lie on the surface of a hypersphere (because the waveform L_2-norm is preserved under translation), and furthermore, must have a constant curvature (by symmetry). This leads to the notion that it might be well-approximated by an arc of a circle. As such, we approximate a segment of the manifold, $\{f_\tau : |\tau| \leq \Delta/2\}$, by the unique circular arc that contains the three points $\{f_{-\Delta/2}, f_0, f_{\Delta/2}\}$, as illustrated in Fig. 3(a). The resulting interpolator is an example of a trigonometric spline [27], in which the three time-shifted functions are linearly combined using trigonometric coefficients to approximate intermediate translates of $f(t)$:

$$f_\tau(t) \approx c(t) + r \cos\left(\frac{2\tau}{\Delta}\theta\right) u(t) + r \sin\left(\frac{2\tau}{\Delta}\theta\right) v(t) \quad (14)$$

The functions $\{c(t), u(t), v(t)\}$ are computed from linear combinations of $\{f_{-\Delta/2}, f_0, f_{\Delta/2}\}$:

$$\begin{pmatrix} f_{-\frac{\Delta}{2}}(t) \\ f_0(t) \\ f_{\frac{\Delta}{2}}(t) \end{pmatrix} = \begin{pmatrix} 1 & r\cos(\theta) & -r\sin(\theta) \\ 1 & r & 0 \\ 1 & r\cos(\theta) & r\sin(\theta) \end{pmatrix} \begin{pmatrix} c(t) \\ u(t) \\ v(t) \end{pmatrix}. \quad (15)$$

The constants r and θ are the radius and half the subtended angle of the arc, respectively, both of which depend on $f(t)$ and can
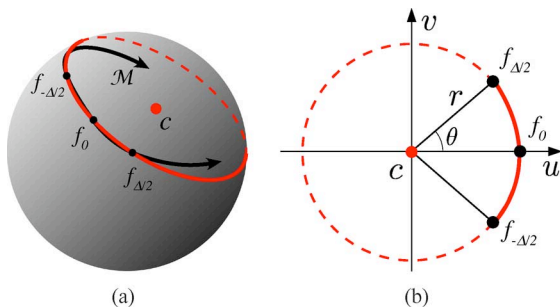
Fig. 3. Illustration of the polar interpolator. (a) The manifold of time shifts of $f(t)$ (black line) lies on the surface of a hypersphere. We approximate a segment of this manifold, for time shifts $\tau \in [-\Delta/2, \Delta/2]$, with a portion of a circle (red), with center defined by $c(t)$. Parameterization of the circular arc approximation.
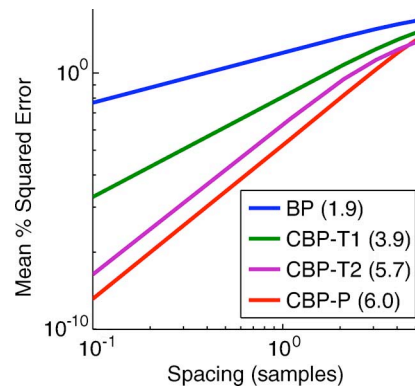

Fig. 4. Comparison of the nearest neighbor, first-order Taylor, and polar interpolators (as used in BP, CBP-T, and CBP-P, respectively) for a waveform $f(t) \propto t e^{-\alpha t^2}$. Second-order Taylor interpolation are also shown. The estimated log-domain slopes (asymptotic rates of convergence) are indicated in parenthesis in the legend.

be computed in closed form. These relationships are illustrated in Fig. 3(b). Combining (14), (15) gives

$$f_\tau(t) = \begin{pmatrix} 1 \\ r \cos\left(\frac{2\tau}{\Delta}\theta\right) \\ r \sin\left(\frac{2\tau}{\Delta}\theta\right) \end{pmatrix}^T A^{-1} \begin{pmatrix} f_{-\frac{\Delta}{2}}(t) \\ f_0(t) \\ f_{\frac{\Delta}{2}}(t) \end{pmatrix} \quad (16)$$

where A is the matrix in (15). The approximation can be more clearly understood in the frequency domain by taking the Fourier transform and dividing both sides by $\hat{f}(\omega)$:

$$e^{-i\omega\tau} \approx (1 - 2a(\tau)) + e^{i\omega\frac{\Delta}{2}}(a(\tau) - b(\tau)) \\ + e^{-i\omega\frac{\Delta}{2}}(a(\tau) + b(\tau)) \quad (17)$$

where

$$a(\tau) = \frac{\cos\left(\frac{2\tau\theta}{\Delta}\right) - 1}{2(\cos(\theta) - 1)} \quad \text{and} \quad b(\tau) = \frac{\sin\left(\frac{2\tau\theta}{\Delta}\right)}{2\sin(\theta)}.$$

Fig. 4 compares nearest neighbor (implicitly used in BP), first-order Taylor, and polar interpolation in terms of their accuracy in approximating time-shifts of a Gaussian derivative waveform, $f(t) \propto t e^{-\alpha t^2}$. For reference, the second-order Taylor interpolator is also included. The polar interpolator is seen to be much more accurate than nearest-neighbor and first-order Taylor, and even surpasses second-order Taylor by an order of magnitude (although they have the same asymptotic rate of convergence). This allows one to choose a much larger Δ for a given desired accuracy.

We now construct a dictionary of time-shifted copies of the functions used to represent the polar interpolation, $\{c_{i\Delta}, u_{i\Delta}, v_{i\Delta}\}$, and form a convex set from these to approximate the manifold:

$$\mathcal{M}_{f,T} \approx \left\{ xc_{i\Delta}(t) + yu_{i\Delta}(t) + zv_{i\Delta}(t) : \begin{array}{l} x \geq 0, \\ y^2 + z^2 \leq x^2 r^2, \\ xr\cos(\theta) \leq y \leq xr, \\ i = 1, \dots N_\Delta \end{array} \right\}. \quad (18)$$

The constraints on the coefficients (x, y, z) ensure that they represent a scaled translate of $f(t)$, except for the second, which is a convex relaxation of the true constraint, $y^2 + z^2 = x^2 r^2$ (see below). As with the Taylor approximation, we have a one-to-one

correspondence between event amplitudes/time-shifts and the constrained coefficients:

$$\sum_i x_i c_{i\Delta}(t) + y_i u_{i\Delta}(t) + z_i v_{i\Delta}(t) \\ \approx \sum_i x_i f_{\left(i\Delta - \frac{\Delta}{2\theta}\tan^{-1}(z_i/y_i)\right)}(t) \quad (19)$$

as long as $z_i/y_i \neq \tan(\theta)$ for all i. The inference problem again boils down to minimizing a constrained convex objective function:

$$\min_{\vec{x}, \vec{y}, \vec{z}} \frac{1}{2\sigma^2} \|y(t) - (C_\Delta\vec{x})(t) - (U_\Delta\vec{y})(t) - (V_\Delta\vec{z})(t)\|_2^2 \\ + \lambda\|\vec{x}\|_1 \\ \text{s.t.} \begin{cases} x_i \geq 0, \\ \sqrt{y_i^2 + z_i^2} \leq x_i r, \\ x_i r \cos(\theta) \leq y_i \leq x_i r, \end{cases} \text{ for } i = 1, \dots N_\Delta \quad (20)$$

where $C_\Delta, U_\Delta,$ and V_Δ are dictionaries containing Δ-shifted copies of $c(t), u(t), v(t)$, respectively. Equation (20) is an example of a "second-order cone program" for which efficient solvers exist [28]. After the optimum values for $\{\vec{x}, \vec{y}, \vec{z}\}$ are obtained, time-shifts and amplitudes can be inferred by first projecting the solution back to the original constraint set:

$$(x_i, y_i, z_i) \leftarrow \left(x_i, \frac{y_i x_i r}{\sqrt{y_i^2 + z_i^2}}, \frac{z_i x_i r}{\sqrt{y_i^2 + z_i^2}} \right) \quad (21)$$

and then using (19) to solve for the event times.

Fig. 2(c) illustrates the optimization of (20) for the simple example described in the previous section. Notice that the solution corresponding to $\lambda = 0$ (yellow dot) is substantially sparser relative to both the CBP-T and BP solutions, and that the solution becomes L_0 sparse if λ is increased by just a small amount, giving up very little reconstruction accuracy.

C. General Interpolation

We can generalize the CBP approach to use any linear interpolation scheme. Suppose we have a set of basis functions

TABLE I
EXAMPLES OF INTERPOLATORS

Property	BP (nearest-neighbor)	CBP - Taylor interp	CBP - polar interp
basis: $\{\phi_n(t)\}_{n=1}^m$	$[f(t)]$	$[f(t) \,,\; f'(t)]$	$[c(t) \,,\; u(t) \,,\; v(t)]$
interpolator: $\vec{D}(\tau)$	1	$[1 \,,\; \tau]^T$	$[1 \,,\; r\cos(\theta\frac{2\tau}{\Delta}) \,,\; r\sin(\theta\frac{2\tau}{\Delta})]^T$
constrained coefficient set: S	$\{x_1 \geq 0\}$	$\{x_1 \geq 0 \,,\; |x_2| \leq x_1\frac{\Delta}{2}\}$	$\{x_1 \geq 0 \,,\; x_2^2 + x_3^2 = r^2 x_1^2 \,,\; rx_1\cos(\theta) \leq x_2 \leq rx_1\}$
convex relaxation: $\mathcal{H}$	S	S	$\{x_1 \geq 0 \,,\; x_2^2 + x_3^2 \leq r^2 x_1^2 \,,\; rx_1\cos(\theta) \leq x_2 \leq rx_1\}$
projection operator: $P_S(\vec{x})$	$\vec{x}$	$\vec{x}$	$[x_1 \,,\; rx_1\frac{x_2}{\sqrt{x_2^2+x_3^2}} \,,\; rx_1\frac{x_3}{\sqrt{x_2^2+x_3^2}}]^T$

$\{\phi_n(t)\}_1^m$ in $L_2([0,T])$ and a corresponding interpolation map $D : \left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right] \to \mathbb{R}^m$ such that local shifts can be approximated as

$$f_\tau(t) \approx \sum_{n=1}^m D_n(\tau)\phi_n(t), \quad |\tau| \leq \frac{\Delta}{2} \tag{22}$$

[e.g., (9) and (14)]. Let S be the set of all nonnegative scalings of the image of $[-\Delta/2, \Delta/2]$ under the interpolator

$$S = \{a\vec{x} : a \geq 0, \quad \vec{x} \in \text{Range}(D)\}.$$

In general, S may not be convex (as in the polar case), and we denote its convex hull by $\mathcal{H}$. If we have a set of coefficients $\vec{x} \in \mathbb{R}^{N_\Delta \times m}$ where each block $\vec{x}_i := [x_{i1}, \dots, x_{im}]$ is in $\mathcal{H}$, then the signal given by

$$\sum_{i=1}^{N_\Delta} \sum_{n=1}^m x_{in}\phi_n(t - i\Delta) \tag{23}$$

approximates a sum of scaled and translated waveforms. If D is invertible, the amplitudes and shifts are obtained as follows:

$$a_i \leftarrow a \quad \text{s.t.} \quad \frac{P_S(\vec{x}_i)}{a} \in \text{Range}(D) \tag{24}$$

$$\tau_i \leftarrow i\Delta + D^{-1}\left(P_S(\vec{x}_i)/a\right) \tag{25}$$

where $P_S(.)$ projects points in $\mathcal{H}$ onto S. Note that in this general form, the L_2 norm of (the projection of) each group $\vec{x}_i$ governs the amplitude of the corresponding time-shifted waveform.[3] Finally, we can obtain the coefficients by solving

$$\min_{\vec{x}} \frac{1}{2\sigma^2}\|y(t) - (\Phi_\Delta\vec{x})(t)\|_2^2 + \lambda \sum_{i=1}^{N_\Delta} \|\vec{x}_i\|_2$$
$$\text{s.t.} \quad \vec{x}_i \in \mathcal{H} \quad \text{for} \quad i = 1, \dots, N_\Delta \tag{26}$$

where the linear operator Φ_Δ is defined as

$$(\Phi_\Delta\vec{x})(t) := \sum_{i=1}^{N_\Delta} \sum_{n=1}^m x_{in}\phi_n(t - i\Delta) \tag{27}$$

Equation (26) can be solved efficiently using standard convex optimization methods (e.g., interior point methods [28]). It is similar to the objective functions used to recover so-called

---

[3] Our specific examples used the amplitude of a single coefficient as opposed to the group L_2 norm. However, the constraints in these examples make the two formulations equivalent up to $O(\Delta)$. For the Taylor interpolator, $x_i^2 + d_i^2 \approx x_i^2$. For the polar interpolator, $c_i^2 + u_i^2 + v_i^2 \approx 2c_i^2$.

"block-sparse" signals (e.g., [12] and [29]), but includes auxiliary constraints on the coefficients to ensure that only signals close to $span(\mathcal{M}_{f,T})$ are represented. Table I summarizes the Taylor and polar interpolation examples within this general framework, along with the case of nearest-neighbor interpolation (which corresponds to standard BP).

The quality of the solution relies on the accuracy of the interpolator, the convex approximation $\mathcal{H} \approx S$, and the ability of the block-L_1 based penalty term in (26) to achieve L_0-sparse solutions that reconstruct the signal accurately. The first two of these are relatively straightforward, since they depend solely on the properties of the interpolator (see Fig. 4). The last is difficult to predict, even for the simple examples illustrated in Fig. 2. The level sets of the L_2 term can have a complicated form when taking the constraints into account, and it is not clear *a priori* whether this will facilitate or hinder the L_1 term in achieving sparse solutions. Nevertheless, our empirical results clearly indicate that solving (26) with Taylor and polar interpolators yields substantially sparser solutions than those achieved with standard BP.

V. EMPIRICAL RESULTS

We evaluate our method on data simulated according to the generative model of (2). Event amplitudes were drawn uniformly from the interval [0.5,1.5]. We used a single template waveform $f(t) \propto te^{-\alpha t^2}$ (normalized, so that $\|f\|_2 = 1$), for which the interpolator performances are plotted in Fig. 4. We compared solutions of (7), (13), and (20). Amplitudes were constrained to be nonnegative (this is already assumed for the CBP methods, and amounts to an additional linear inequality constraint for BP). Each method has two free parameters: Δ controls the spacing of the basis, and λ controls the tradeoff between reconstruction error and sparsity. We varied these parameters systematically and measured performance in terms of two quantities: 1) signal reconstruction error (which decreases as λ or Δ decreases), and 2) sparsity of the estimated event amplitudes, which increases as λ increases. The former is simply the first term in the objective function (for all three methods). For the latter, to ensure numerical stability, we used the L_p norm with $p = 0.1$ (results were stable with respect to the choice of p, as long as $p < 1$ and p was not below the numerical precision of the optimizations). Computations were performed numerically, by sampling the functions $f(t)$ and $y(t)$ at a fine constant spacing. We used the convex solver package CVX [30] to obtain numerical solutions.
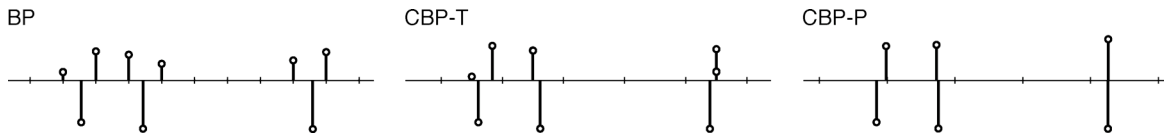
Fig. 5. Sparse signal recovery using BP (equation (7)), CBP-T (equation (13)), and CBP-P (equation (20)). The source was a sparse superposition of arbitrarily time-shifted copies of the function $f(t) \propto te^{-\alpha t^2}$. The SNR (defined as $\|f\|_\infty / \sigma$) was 12 with identical signal and noise for all three examples. For each method, the values of Δ and λ were chosen to minimize the sum of squares of sparsity and reconstruction error (large dots in Fig. 6). Upward stems indicate the estimated magnitudes, placed at locations determined by the interpolation coefficients via (24). Amplitudes less than 0.01 were eliminated for clarity. Ticks denote the location of the basis functions corresponding to each upward-pointing stem. Downward stems indicate true locations and magnitudes.
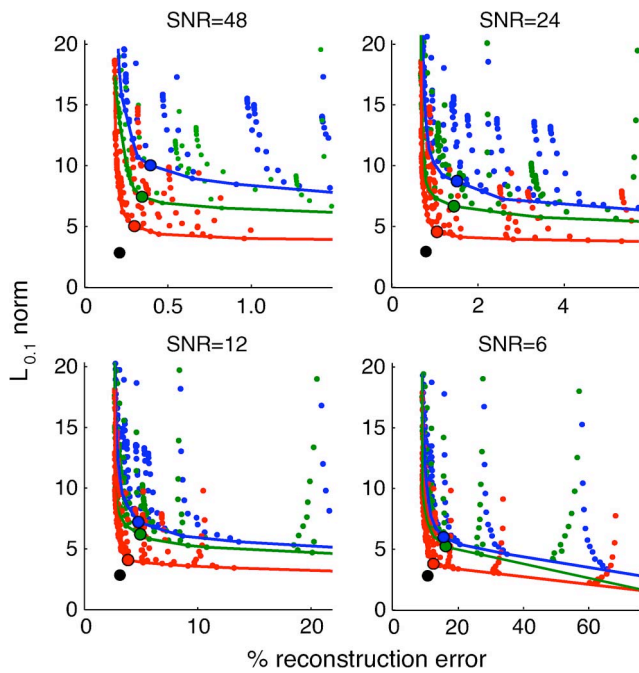

Fig. 6. Error plots for 4 noise levels (SNR is defined as $\|f\|_\infty / \sigma$). Each graph shows the tradeoff between the average reconstruction error and sparsity (measured as average $L_{0.1}$ norm of estimated amplitudes). Each point represents the error values for one of the methods, applied with a particular setting of (Δ, λ), averaged over 500 trials. Colors indicate the method used (blue: BP, green: CBP-T, red: CBP-P). Bold lines denote the convex hulls of all points for each method. The large dots indicate the "best" solution, as measured by Euclidean distance from the correct solution (indicated by black dots).
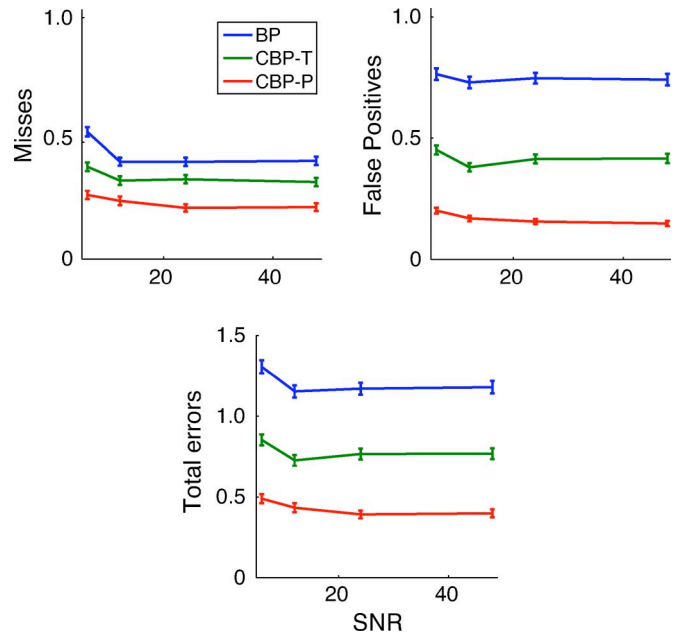
Fig. 7. Signal detection analysis of solutions at three SNR levels (see text). Misses, false positives, and total errors (sum of misses and false positives), as a fraction of the mean number of events, were computed over 500 trials for each method, and for each SNR (defined as $\|f\|_\infty / \sigma$).

A small temporal window of the events recovered by the three methods is provided in Fig. 5. The three plots show the estimated event times and amplitudes for BP, CBP-T, and CBP-P (upward stems) compared to the true event times/amplitudes (downward stems). The figure demonstrates that CBP, equipped with either Taylor or polar interpolators, is able to recover the event train more accurately, and with a larger spacing between basis functions (indicated by the tick marks on the x-axis). As predicted by the reasoning laid out in Fig. 2(a), basis pursuit tends to split events across two or more adjacent low-amplitude coefficients, thus producing less sparse solutions and making it hard to infer the number of events and their respective amplitudes and times. Sparsity can be improved by increasing λ, but at the expense of a substantial increase in approximation error.

Fig. 6 illustrates the tradeoff between sparsity and approximation error for each of the methods. Each panel corresponds to a different noise level. The individual points, color-coded for each method, are obtained by running the associated method 500 times for a given (Δ, λ) combination, and averaging the

errors over these trials. The solid curves are the (numerically computed) convex hulls of all points obtained for each method, and clearly indicate the tradeoff between the two types of error. We can see that the performance of BP is strictly dominated by that of CBP-T: For every BP solution, there is a CBP-T solution that has lower values for both error types. Similarly, CBP-T is strictly dominated by CBP-P, which can be seen to come close to the error values of the ground truth answer (which is indicated by a black dot). Note that the reconstruction error of the true solution and of all methods is bounded below by the variance of the noise that lies outside of the subspace spanned by the set of shifted copies of the waveform.

We performed a signal detection analysis of the performance of these methods, classifying identification errors as misses and false positives. We match an estimated event with a true event if the estimated location is within 3 samples of the true event time, the estimated amplitude is within a threshold $1/\sqrt{12}$ of the true amplitude (one standard deviation of the amplitude distribution Unif ([0.5,1.5])), and no other estimated event has been matched to the true event. We found that results were relatively stable with respect to the threshold choices. For each method and noise level we chose the (λ, Δ) combination yielding a solution closest to ground truth (corresponding to the large dots in
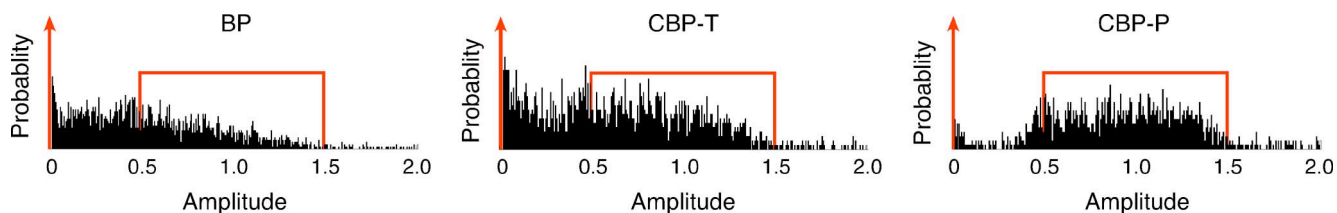
Fig. 8. Histograms of the estimated amplitudes for BP, CBP-T, and CBP-P, respectively. All methods were constrained to estimate only nonnegative amplitudes, but no upper bound was imposed. The true distribution of amplitudes is given by (4), and is indicated in red.
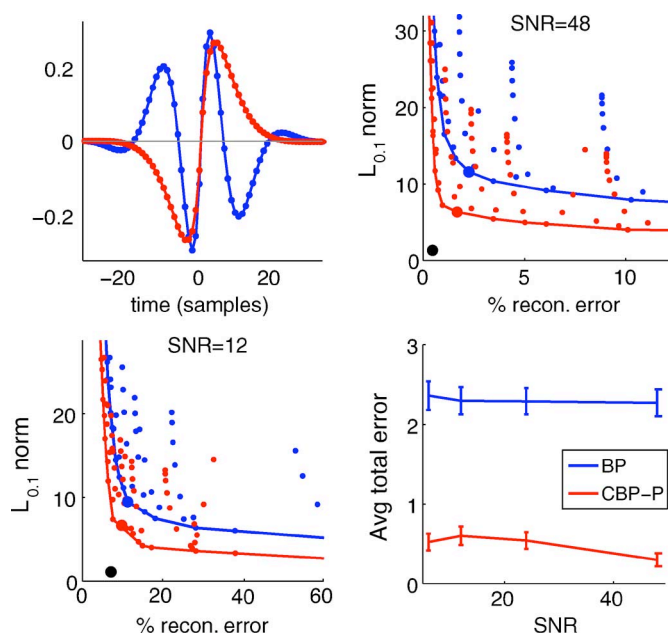

Fig. 9. Sparse signal decomposition with two waveforms. Upper-left: gammatone waveforms, $f_i(t) = at^{n-1}e^{-2\pi bt}\cos(2\pi\omega_i t)$ for $i = 1, 2$. Upper-right and lower-left: sparsity and reconstruction errors for BP (blue) and CBP-P (red), as in Fig. 6, with SNR's of 48 and 12, respectively (again, SNR is defined as $\frac{\|f\|_\infty}{\sigma}$). Lower right: total number of misses and false positives (with same thresholds as in bottom plot of Fig. 7) for each method.

Fig. 6). Fig. 7 shows the errors as a function of the noise level. We see that performance of all methods is surprisingly stable across SNR levels. We also see that BP performance is dominated at all noise levels by CBP-T, which has fewer misses as well as fewer false positives, and CBP-T is similarly dominated by CBP-P.

Finally, Fig. 8 shows the distribution of the amplitudes estimated by each algorithm, and compare with the distribution of the source, as given by (4). We see that CBP-P produces amplitude distributions that are far better-matched to the correct distribution of amplitudes.

A. Multiple Features

All the methods we described can be extended to the case of multiple templates by taking as a dictionary the union of dictionaries associated with each individual template. We performed a final set of experiments using two "gammatone" features (shown in upper left of Fig. 9), which are commonly used in audio processing. Event times were sampled from two correlated Poisson processes with the same marginal rate λ_0 and a correlation of $\rho = 0.5$. These were generated by independently

sampling 2 Poisson process with rate $\lambda_0(1 - \rho)$ and then superimposing a randomly jittered "common" Poisson process with rate $\lambda_0\rho$. As before, event amplitudes were drawn uniformly from the interval [0.5,1.5]. We compared the performance of BP and CBP-P, in both cases using dictionaries formed from the union of dictionaries for each template with a common spacing, Δ, for both templates. In general, the spacing could be chosen differently for each waveform, providing more flexibility at the expense of additional parameters that must be calibrated. The upper right and lower left plots in Fig. 9 show the error tradeoff for different settings of (Δ, λ) at SNR levels of 48 and 12, respectively (the results were qualitatively unchanged for SNR values of 24 and 6). The lower right plot in Fig. 9 plots the total number of event identification errors (misses plus false positives) for each method as a function of SNR, at each method's optimal (Δ, λ) setting.

VI. DISCUSSION

We have introduced a novel methodology for sparse signal decomposition in terms of continuously shifted features. The method can be seen as a continuous form of the well-known basis pursuit method, and we thus have dubbed it *continuous basis pursuit*. The method overcomes the limitation of basis pursuit in the convolutional setting caused by the tradeoff between discretization error and the effectiveness of the L_1 relaxation for obtaining sparse solutions. In particular, our method employs an alternative discrete basis (not necessarily the features themselves) which can explicitly account for the continuous time-shifts present in the signal. We derived a general convex objective function that can be used with any such basis. The coefficients are constrained so as to represent only transformed versions of the templates, and the objective function uses an L_1 norm to penalize the amplitudes of these transformed versions. We showed empirically that using simple first-order (Taylor) and second-order (polar) interpolation schemes yields superior solutions in the sense that 1) they are sparser, with improved reconstruction accuracy, 2) they produce substantially better identification of events (fewer misses and false positives), and 3) the amplitude statistics are a better match to those of the true generative model. We showed that these results are stable across a wide range of noise levels. We conclude that an interpolating basis, coupled with appropriate constraints on coefficients, provides a powerful and tractable tool for modeling and decomposing sparse translation-invariant signals.

We believe our method can be extended for use with other types of signal, as well as to transformations other than translation. For example, for one-dimensional signals such as audio, one might also include dilation or frequency-modulation of the

templates. Both the Taylor and polar basis constructions can be extended to account for multiple transformations. In the Taylor case, one only needs to add incorporate waveform derivatives with respect to each transformation (with a separate constraint for each of their coefficients). In the polar case, one can model the manifold of local (renormalized) translations and dilations as a 2-D patch on the surface of a sphere, ellipsoid, or torus (instead of a 1-D arc) which can be parameterized with two angles. For images, one could include rotation along with dilation (e.g., [31]). In general, the primary hurdles for such extensions are to specify 1) the form of the linear interpolation (for joint variables, this might be done separably, or using a multi-dimensional interpolator), 2) the constraints on coefficients (and a convex relaxation of these constraints), and 3) a means of inverting the interpolator so as to obtain transformation parameters from recovered coefficients. Another natural extension is to use CBP in the context of learning optimal features for decomposing an ensemble of signals, as has been previously done with BP (e.g., [4], [8], [10], [11], and [32]).

Although we have described the use of an interpolator basis in the context of L_1-based recovery methods, we believe the same representations can be used to improve other recovery methods (e.g., greedy or iterative thresholding methods with a convolutional basis). Furthermore, our polar approximation of a translational manifold can provide a substrate for new forms of sampling (e.g., [33]–[35]). By introducing a basis and constraints that explicitly model the local geometry of the manifold on which the signals lie, we expect our method to offer substantial improvements in many of the applications for which sparse signal decomposition has been found useful.

ACKNOWLEDGMENT

REFERENCES

[1] E. J. Rothwell and W. Sun, "Time domain deconvolution of transient radar data," *IEEE Trans. Antennas Propag.*, vol. 38, no. 4, pp. 470–475, Apr. 1990.

[2] J. Mendel, *Optimal Seismic Deconvolution: An Estimation Based Approach*. New York: Academic, 1983.

[3] R. D. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhan, and M. Allerhand, "Complex sounds and auditory images," in *Proc. 9th Int. Symp. Hearing Audit., Physiol. Perception,* , 1992, pp. 429–446.

[4] E. Smith and M. S. Lewicki, "Efficient coding of time-relative structure using spikes," *Neural Comput.*, vol. 17, no. 1, pp. 19–45, Jan. 2005.

[5] R. Grosse, R. Raina, H. Kwong, and A. Y. Ng, "Shift-invariant sparse coding for audio classification," in *Proc. Uncertainty in Artif. Intell. (UAI)*, 2007, pp. 149–158.

[6] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. 2nd IEEE Int. Conf. Image Process.*, Washington, DC, Oct. 23–26, 1995, vol. III, pp. 444–447.

[7] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1999.

[8] P. Sallee and B. A. Olshausen, "Learning sparse multiscale image representations," in *Proc. NIPS*, 2002, pp. 1327–1334.

[9] E. J. Candès, L. Demanet, D. L. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Modeling and Simulation*, vol. 5, 2005.

[10] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, Jun. 1996.

[11] A. J. Bell and T. J. Sejnowski, "The "independent components" of natural scenes are edge filters," *Vision Res.*, vol. 37, no. 23, pp. 3327–3338, Dec. 1997.

[12] A. Hyvärinen and P. Hoyer, "Emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspaces," *Neural Comput.*, vol. 12, no. 7, pp. 1705–1720, Jul. 2000.

[13] J. H. Friedman and J. W. Tukey, "A projection pursuit algorithm for exploratory data analysis," *IEEE Trans. Comput.*, vol. C-23, no. 9, pp. 881–890, 1974.

[14] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[15] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[16] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 629–654, Dec. 2008.

[17] M. Elad, "Why simple shrinkage is still relevant for redundant representations," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5559–5569, 2006.

[18] J. Bioucas-Dias and M. Figueiredo, "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2992–3004, 2007.

[19] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak, "Majorization-minimization algorithms for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2980–2991, 2007.

[20] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. Series B (Methodol.)*, vol. 58, no. 1, pp. 267–288, 1996.

[21] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[22] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "Sparse decomposition of transformation-invariant signals with continuous basis pursuit," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2011.

[23] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Construct. Approx.*, vol. 13, pp. 57–98, 1997, 10.1007/BF02678430.

[24] H. L. Taylor, S. C. Banks, and J. F. McCoy, "Deconvolution with the l1 norm," *Geophysics*, vol. 44, no. 1, p. 39, 1979.

[25] M. S. O'Brien, A. N. Sinclair, and S. M. Kramer, "Recovery of a sparse spike time series by l1 norm deconvolution," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3353–3365, Dec. 1994.

[26] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[27] I. J. Schoenberg, "On trigonometric spline interpolation," *J. Math Mech.*, vol. 13, pp. 795–825, 1964.

[28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[29] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.

[30] M. Grant and S. Boyd, CVX: Matlab Software for Disciplined Convex Programming, ver. 1.21 [Online]. Available: http://cvxr.com/cvx, Oct. 2010

[31] "Manifold models for signals and images," *Comput. Vis. Image Understand.*, vol. 113, no. 2, pp. 249–260, 2009.

[32] P. Berkes, R. E. Turner, and M. Sahani, "A structured model of video reproduces primary visual cortical organisation," *PLoS Comput. Biol.*, vol. 5, no. 9, Sep. 2009, e1000495.

[33] M. Unser, "Sampling-50 years after Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, 2000.

[34] 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, Jun. 2002.

[35] J. Uriguen, Y. C. Eldar, P. L. Dragotti, and Z. Ben-Haim, *Sampling at the Rate of Innovation: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

Chaitanya Ekanadham received the B.S. degree in mathematical/computational sciences and symbolic systems from Stanford University, Stanford, CA, in 2007. He is currently working towards the Ph.D. degree in applied math at the Courant Institute, New York University, working with E. P. Simoncelli and D. Tranchina.

His research interests lie at the intersection of statistical signal processing, machine learning, and applied mathematics.

Eero P. Simoncelli (S'90–M'91–SM'04–F'09) received the B.S. degree (*summa cum laude*) in physics from Harvard University, Cambridge, MA, in 1984 and the M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1988 and 1993, respectively. He studied applied mathematics at Cambridge University for a year and a half.

He was an Assistant Professor in the Computer and Information Science Department at the University of Pennsylvania, Philadelphia, from 1993 to 1996. In September 1996, he moved to New York University, New York, where he is currently an Associate Professor in neural science and mathematics. In August 2000, he became an Associate Investigator at the Howard Hughes Medical Institute, New York, under their new program in computational biology. His research interests span a wide range of topics in the representation and analysis of visual images, in both machine and biological systems.

Daniel Tranchina received the Ph.D. degree from The Rockefeller University, New York, in 1981 and the B.A. degree in chemistry from The State University of New York at Binghamton in 1975. His graduate training and postdoctoral studies were in mathematics and experimental neuroscience.

In 1984, he became a faculty member at New York University, where he is currently a Professor of biology, mathematics, and neural science. His research interests include computational neuroscience, phototransduction, neural network modeling, population density methods, stochastic gene expression, and statistical analysis and modeling of genomic data.