Supporting Information

Ba et al. 10.1073/pnas.1320637111

SI Text

Link Between Spectrotemporal Pursuit and Basis Pursuit Denoising.
Spectrotemporal pursuit builds upon the key idea, put forth by
Chen et al. (1), of using ℓ_1-regularized least squares to compute a
sparse representation $\mathbf{x}_1$ of a deterministic signal given noisy
observations $\mathbf{y}_1$ and a known dictionary $\mathbf{F}_1$ (e.g., obtained by
discretization in the frequency domain). BPDN refers to the
solution of the ℓ_1-regularized least-squares problem proposed
by Chen et al. (1).

Close inspection of Eq. **9** reveals that, when $n=1$ and
$f(\cdot)=f_1(\cdot)$, spectrotemporal pursuit approximates BPDN as
the solution of the regularized least-squares problem

$$\max_{\mathbf{x}_1} -\frac{1}{2\sigma^2}\|\mathbf{y}_1-\mathbf{F}_1\mathbf{x}_1\|_2^2 - \alpha_1 \sum_{k=1}^{K}\left(x_{1,k}^2+\epsilon^2\right)^{1/2}. \quad \textbf{[S1]}$$

The case of $\epsilon=0$ solves BPDN exactly. We termed our proce-
dure spectrotemporal pursuit to highlight this link.

When $n>1$, spectrotemporal pursuit is a time-varying gener-
alization of BPDN pursuit denoising that, given a sequence of
noisy observations $(\mathbf{y}_n)_{n=1}^N$ of deterministic signals, finds sparse
representations $(\mathbf{x}_n)_{n=1}^N$ of these signals with common spatio-
temporal structure in known dictionaries $(\mathbf{F}_n)_{n=1}^N$ (each possibly
overcomplete). In the case of $f(\cdot)=f_1(\cdot)$, spectrotemporal pursuit
(Eq. **9**) is a strictly concave optimization problem that, in prin-
ciple, can be solved using standard techniques. However, our
experience has shown that these techniques do not scale well
with N; this can be attributed to the form of the regularizer in
spectrotemporal pursuit when $n>1$, which is the key difference
with BPDN. More specifically, the goal of spectrotemporal
pursuit—to compute signal representations $(\mathbf{x}_n)_{n=1}^N$ that are
sparse in the spatial dimension (e.g., frequency) and smooth in
time—is predicated upon the use of regularizers that enforce
group sparsity (2), in particular, ones that are decomposable in
space but not in time. For each spatial dimension k, such regu-
larizers must penalize jointly the temporal variables $(x_{n,k})_{n=1}^N$
associated with this dimension, a property we refer to as the
nondecomposability of the regularizer over space and time.
Though nondecomposability allows spectrotemporal pursuit to
capture the desired (highly structured) dynamic behavior of the
deterministic signals, it suggests that, for each k, we must solve
for $(x_{n,k})_{n=1}^N$ simultaneously, i.e., in batch. For large N, such batch
computations become very challenging.

In our treatment, we demonstrated how to leverage the Bayesian
formulation of spectrotemporal pursuit, which is consistent with that
of BPDN, to overcome the computational challenges posed by the
nondecomposability of the regularizers and the batch nature of
spectrotemporal pursuit.

Advantages of IRLS over Gradient-Based Methods. IRLS presents
several advantages over gradient-based algorithms (3, 4) for
solving Eq. **9** with $\epsilon=0$. First, our experience has shown that the
IRLS solution is numerically more stable than that obtained
using gradient-based methods. Second, unlike gradient-based
algorithms, which only yield point estimates, IRLS is a second-
order method that can be used to perform statistical inference.
Indeed, one can combine the covariance matrices $(\Sigma_{n|N})_{n=1}^N$, at the
last iteration of the spectrotemporal pursuit algorithm, with a co-
variance smoothing algorithm (5) to compute a Gaussian approxi-
mation to the posterior distribution of $\mathbf{x}$ given the observations $\mathbf{y}$.
This joint distribution, in turn, can be used to approximate the joint

distribution of any function of $\mathbf{x}$ given $\mathbf{y}$, either in closed form or
by Monte Carlo (6). For instance, we can use this joint distribu-
tion to compute confidence intervals for point estimates that are
functions of $\mathbf{x}_{n_1}$ and $\mathbf{x}_{n_2}$, $n_1 \neq n_2 \in \{1, 2, \cdots, N\}$ (6). An important
goal of applications of spectral analysis to neural signal process-
ing is to compare the spectrotemporal representation of a time
series at different times and/or frequencies. The Bayesian for-
mulation, along with IRLS, allows us to accomplish this goal in
a seamless fashion.

Continuous-Time Variational Interpretation of Spectrotemporal Pursuit.
Spectrotemporal pursuit is a discrete-time algorithm for computing
structured time–frequency representations of signals whose time-
varying mean is the linear combination of a small number of
oscillatory components.

The objective function of spectrotemporal pursuit trades off an
error term, which represents the cost incurred by approximating
the data $(\mathbf{y}_n)_{n=1}^N$ as $(\mathbf{F}_n\mathbf{x}_n)_{n=1}^N$, and a prior that promotes estimates
$(\mathbf{x}_n)_{n=1}^N$ that are sparse in frequency and smooth in time. This
intuitive interpretation suggests that, for $f(\cdot)=f_1(\cdot)$, spec-
trotemporal pursuit approximates the solution in discrete time of
the following continuous-time variational problem

$$\min_{\tilde{x}(t,\omega)} \int_0^T \left| y(t) - \int_{\omega'} \tilde{x}(t,\omega')e^{j\omega' t}d\omega' \right|^2 dt$$
$$+ \alpha_1 \int \left(\int_0^T \left| \frac{\partial \tilde{x}(t,\omega)}{\partial t} \right|^2 dt \right)^{1/2} d\omega, \quad \textbf{[S2]}$$

where $\tilde{x}(t,\omega)$ is $(\tilde{\mathbf{x}}_n)_{n=1}^N$ in continuous time. Eq. **S2** resembles, but
is different from, the continuous-time variational problem that
the synchrosqueezed wavelet transform (7) approximates. The
connection between the continuous-time formulations of spec-
trotemporal pursuit and synchrosqueezed wavelet transform (7)
further strengthens the interpretation of spectrotemporal pursuit
as applying a data-dependent filter bank to a given time series.

For the case of a point process, as in the main text's *A Spec-
trotemporal Pursuit Analysis of Neural Spiking Activity*, the analog
of Eq. **S2** is

$$\min_{\tilde{x}(t,\omega)} -\left(\int_0^T \log \lambda(t|H_t)dN(t) - \lambda(t|H_t)dt \right)$$
$$+ \alpha_1 \int \left(\int_0^T \left| \frac{\partial \tilde{x}(t,\omega)}{\partial t} \right|^2 dt \right)^{1/2} d\omega, \quad \textbf{[S3]}$$

where $N(t)$ is the counting process of a point process with CIF
$\lambda(t|H_t)$ (8) and

$$\log \lambda(t|H_t) = \int_{\omega'} \tilde{x}(t,\omega')e^{j\omega' t}d\omega'.$$

**Alternate Approaches to Compute Spectral Representations of
Signals that Exhibit Dynamic Behavior.** A number of algorithms
for computing spectra of signals exhibiting dynamic behavior have
appeared in the literature.

For deterministic signal models, notable contributions are those in refs. 9 and 10 for a signal model consisting of a small number of amplitude-modulated oscillations. In both cases, the authors find a sparse representation of the signal in a dictionary of modulated and translated oscillations by solving an optimization problem that trades off the least-squares error and a sparsity-promoting prior. Spectrotemporal pursuit differs with these works in that, in our approach, the choice of prior fully dictates the structure in the time–frequency plane. In refs. 9 and 10, the dictionary (modulations and translations) enforces temporal smoothness, whereas the prior on the time–frequency plane yields sparsity. The effect of this two-stage approach is to decouple, in some sense, the goal of simultaneously achieving sparsity in frequency and smoothness in time. We believe our approach is more direct and easier to generalize to different specifications of the desired structure in the time–frequency plane.

In the context of stochastic signals, the authors in ref. 11 propose an algorithm to estimate the nonparametric, nonstationary spectrum of a Dahlhaus locally stationary process. The authors partition the signal of interest in small sections and assume that the log spectrum in each window follows a mixture of smoothing splines model. To model the nonstationarity of the spectrum, the mixture weights are allowed to evolve in time according to a logistic regression model. An obvious difference between this work and spectrotemporal pursuit is the choice of a stochastic, as opposed to a deterministic, signal model. Unlike spectrotemporal pursuit, this algorithm does not set out to model sparsity in frequency.

Lastly, dynamic models with time-varying sparsity have been proposed in the literature in various contexts. In ref. 12, the authors introduce one class of such models to study network interactions with sparsely time-varying changes in connectivity. More recently, a broad class of dynamic regression models with time-varying sparsity was introduced in ref. 13. These works suggest the use of sparsely time-varying autoregressive models to capture the nonstationarity of the spectrum of a stochastic signal. One key problem with this approach is that the sparsity of the coefficients in an autoregressive model does not necessarily translate to sparsity of the spectrum. It is therefore not obvious how to choose a prior on the coefficients in an autoregressive model so as to promote certain desired features of the spectrum. Moreover, a key novelty of our work, compared with refs. 12 and 13 is the idea of structured dynamic sparsity/sparsely time-varying models and the design of efficient algorithms to estimate such models. In particular, our work demonstrates that space–time priors that are not decomposable over space and time can provide a powerful framework to promote intricate forms of structured dynamic sparsity. We believe that these priors are better suited to capture intricate structure in dynamic regression models.

1. Chen SS, Donoho DL, Saunders MA (1996) Atomic decomposition by basis pursuit. *SIAM J Sci Comput* 20(1):33–61.
2. Yuan M, Lin Y (2006) Model selection and estimation in regression with grouped variables. *J R Stat Soc Series B Stat Methodol* 68(1):49–67.
3. Beck A, Teboulle M (2009) A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J Imaging Sci* 2(1):183–202.
4. Combettes PL, Pesquet J-C (2011) Proximal splitting methods in signal processing. *Fixed-Point Algorithms for Inverse Problems in Science and Engineering* (Springer, New York), pp 185–212.
5. De Jong P, Mackinnon MJ (1988) Covariances for smoothed estimates in state space models. *Biometrika* 75(3):601–602.
6. Smith AC, et al. (2004) Dynamic analysis of learning in behavioral experiments. *J Neurosci* 24(2):447–461.
7. Daubechies I, Lu J, Wu HT (2011) Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. *Appl Comput Harmon Anal* 30(2):243–261.
8. Daley DJ, Vere-Jones D (2003) *An Introduction to the Theory of Point Processes* (Springer, Berlin), 2nd Ed, Vol 1.
9. Flandrin P, Borgnat P (2010) Time-frequency energy distributions meet compressed sensing. *IEEE Trans Signal Proc* 58(6):2974–2982.
10. Wolfe PJ, Godsill SJ, Ng WJ (2004) Bayesian variable selection and regularization for time–frequency surface estimation. *J R Stat Soc Series B Stat Methodol* 66(3): 575–589.
11. Rosen O, Stoffer DS, Wood S (2009) Local spectral analysis via a Bayesian mixture of smoothing splines. *J Am Stat Assoc* 104(485):249–262.
12. Ahmed A, Xing EP (2009) Recovering time-varying networks of dependencies in social and biological studies. *Proc Natl Acad Sci USA* 106(29):11878–11883.
13. Kalli M, Griffin JE (2014) Time-varying sparsity in dynamic regression models. *J Econom* 178(2):779–793.