Dynamic Filtering with Earth Mover's Distance Regularization

Adam S. Charles*, John Lee†, Nicholas P. Bertrand†, and Christopher J. Rozell†

*Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544
†School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332
*adamsc@princeton.edu, †{john.lee,nbertrand,crozell}@gatech.edu

Tracking dynamically evolving signals is a critical part of many applications, such as real-time video processing, radar and communications applications. Often in such applications, the signal of interest $\boldsymbol{x}_t \in \mathbb{R}^N$ is modeled as abiding by Markovian dynamic

$$\boldsymbol{x}_t = f\left(\boldsymbol{x}_{t-1}\right) + \boldsymbol{\nu}_t$$

where $f(\cdot) : \mathbb{R}^N \to \mathbb{R}^N$ is the dynamics function and $\boldsymbol{\nu}_t$ is the so-called *innovations*, or the accuracy to which the dynamics function is known. Often $\boldsymbol{x}$ is only observed indirectly as

$$\boldsymbol{y}_t = \boldsymbol{\Phi}\boldsymbol{x}_t + \boldsymbol{\epsilon}_t$$

where $\boldsymbol{y}_t \in \mathbb{R}^M$ is the observed measurement vector, $\boldsymbol{\Phi} \in \mathbb{R}^{M\times N}$ is the linear measurement operator and $\boldsymbol{\epsilon}_t \in \mathbb{R}^M$ is the measurement error. The filtering task is to recover $\boldsymbol{x}_t$ at each time t. Standard methods, such as the Kalman filter and its variants assume linearity in $f(\cdot)$, and do not leverage any additional information on $\boldsymbol{x}_t$ [1], [2]. More recently, a number of methods have been developed to leverage additional known structure in $\boldsymbol{x}$. Specifically, the case where $\boldsymbol{x}$ is sparse (i.e. mostly contains zeros), or is sparse in a basis (i.e. $\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{a}$ where $\boldsymbol{a}$ is sparse). For brevity, we focus here on the case where $\boldsymbol{x}$ is canonically sparse. For example, when tracking objects through a scene, there are often relatively few objects to track. While a number of methods have been developed to track sparse signals [3], [4], [5], [6], [7], [8], most make the assumption that the prediction error $\boldsymbol{\nu}_t$ is small in some ℓ_p norm. For example, basis-pursuit de-noising with dynamic filtering (BPDN-DF) [8] optimizes at each time-step

$$\widehat{\boldsymbol{x}}_t = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y}_t - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1 + \gamma\|\boldsymbol{x} - f(\widehat{\boldsymbol{x}}_{t-1})\|_p^p,$$

where λ and γ are constants dependent on the signal sparsity and signal-to-noise ratios. While these methods have had success in some areas, none of these methods take into account spatial information about the indices of $\boldsymbol{x}$. For example, consider the application where targets are being tracked in a 2D space. Predicting a target one pixel away from the truth has the same ℓ_p error as predicting a target 20 pixels away, even though we desire to consider the former as a much smaller error. Thus in these cases such methods would over-emphasize certain prediction errors and obtain sub-par performance. In this work we extend the dynamic filtering framework to capture dynamics related to the spatial locations. Specifically, we introduce the earth-mover's distance (EMD) as a replacement for the ℓ_p norm used in BPDN-DF. The EMD is a metric that measures the amount of "mass" needed to move to transform one signal into another. Typical applications of the EMD to linear-inverse problems uses the one-dimensional EMD [9], [10]. For one-dimensional signals, the EMD is easily calculated, however for higher-dimensional data (e.g. 2D images [11]), the EMD is calculated via an optimization program. Essentially, the EMD finds the minimum cost $\sum F_{i,j}r_{i,j}$ over all flows $\boldsymbol{F}$ that transfer mass from $\widetilde{\boldsymbol{x}}$ to $\boldsymbol{x}$, where $\boldsymbol{r}$ is the array of distances between signal indices (i.e. $r_{i,j}$ is the ℓ_p distance between signal index i and index j for an appropriate p). $\boldsymbol{r}$ thus weighs the cost of flows that transmit mass proportional to the distance transversed, penalizing long-distance movement of coefficinets.

Incorporating the EMD into the dynamic filtering, and defining $\widetilde{\boldsymbol{x}} = f(\widehat{\boldsymbol{x}}_{t-1})$ as the predicted signal, yields an ℓ_1-EMD optimization formulated as

$$\begin{aligned}\widehat{\boldsymbol{x}}_t = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y}_t - \boldsymbol{\Phi}_t\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1 + \gamma \min_{\mathbf{F}} \sum_{i,j} F_{i,j}r_{i,j}\\ \text{s.t.} \quad \boldsymbol{x} \geq \mathbf{0}, \quad F_{i,j} \geq 0, \quad \sum_j F_{i,j} \leq x_i, \quad \sum_i F_{i,j} \leq \widetilde{x}_j,\\ \sum_{i,j} F_{i,j} = \min(\|\boldsymbol{x}\|_1, \|\widetilde{\boldsymbol{x}}_{t-1}\|_1).\end{aligned} \tag{1}$$

This cost function can be jointly optimized over both $\boldsymbol{x}$ and $\boldsymbol{F}$ to find the overall new, current signal estimate $\widehat{\boldsymbol{x}}_t$. While the optimization here contains all convex terms and mostly linear constraints, there are two major impediments to efficient implementation: the final nonlinear constraint and the size of the flow matrix $\boldsymbol{F}$. To address the non-linear constraint, we instead solve the optimization program twice, once with $\sum_{i,j} F_{i,j} = \|\boldsymbol{x}\|_1$, and once with $\sum_{i,j} F_{i,j} = \|\widetilde{\boldsymbol{x}}_{t-1}\|_1$, and take as the solution the result with the minimum overall cost. Computationally, $\boldsymbol{F}$ has many parameters (N^2), and can be cumbersome to optimize. We utilize the sparsity of $\widetilde{x}$ to eliminate all but K rows of $\boldsymbol{F}$, reducing the parameter space to KN and greatly increasing computational efficiency.

We test our method by implementing our optimization program in CVX, and explore the utility of EMD-regularized BPDN over other sparsity-based tracking algorithms, BPDN-DF [6], [8] and re-weighted ℓ_1 dynamic filtering (RWL1-DF) [8]. To maximize performance of all algorithms in the experiments, a geometric grid search was performed over parameters λ and γ to minimize the relative mean-squared error (rMSE) given by $\|\widehat{\boldsymbol{x}} - \boldsymbol{x}\|_2^2/\|\boldsymbol{x}\|_2^2$ where $\widehat{\boldsymbol{x}}$ is the estimation and $\boldsymbol{x}$ is the true signal.

We first applied a one-step recovery using compressive measurements on a relatively large simulation example (Fig. 1) where K synthetic targets moved randomly within a 1-pixel neighborhood between frames. As the dynamics model was not well specified, the first frame's ground truth was used as the signal prediction. In this task, BPDN-EMD significantly outperforms BPDN-DF and RWL1-DF at exploiting dynamical information and recovering the target locations. When a full sequence of frames were run in succession (Fig. 2), BPDN-EMD quickly converges to a lower rMSE than any of the other methods, and tracks the targets more accurately over the entire time-course. Finally, we present Donoho-Tanner phase transition diagrams (Fig. 3) to illustrate algorithm performance across different numbers of measurements and signal sparsities. These diagrams contextualize the previous experiments (which was conducted at specific M, K), demonstrating that BPDN-EMD achieves superior performance over competing methods at lower measurement numbers and higher sparsities.

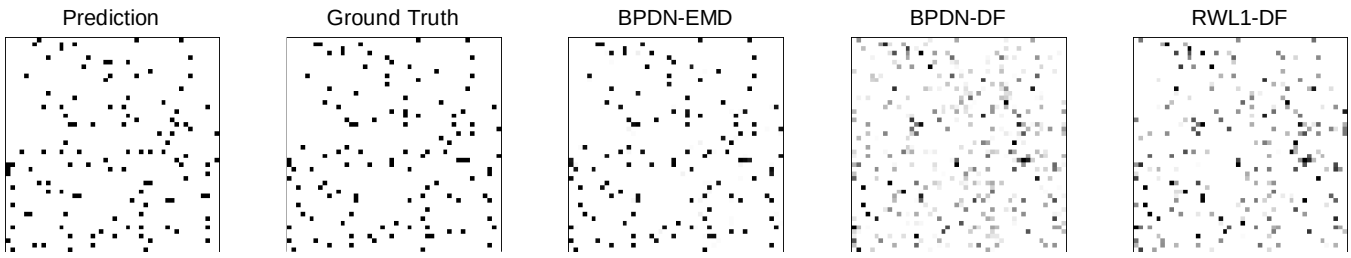

Fig. 1. Example recovery of sparse, moving targets from compressive measurements. Recovering 48×48 pixel-sparse images demonstrates high accuracy of recovery from the EMD-regularized BPDN algorithm. Here, dynamic-aware algorithms are given a prediction signal (i.e., the previous frame from a dynamic sequence) to assist at inferring the ground truth signal, both of which are shown above. In this simulation scenario ($N = 2304, M = 0.15N, K = 0.33M$), the EMD-regularized BPDN recovers the signal most accurately.

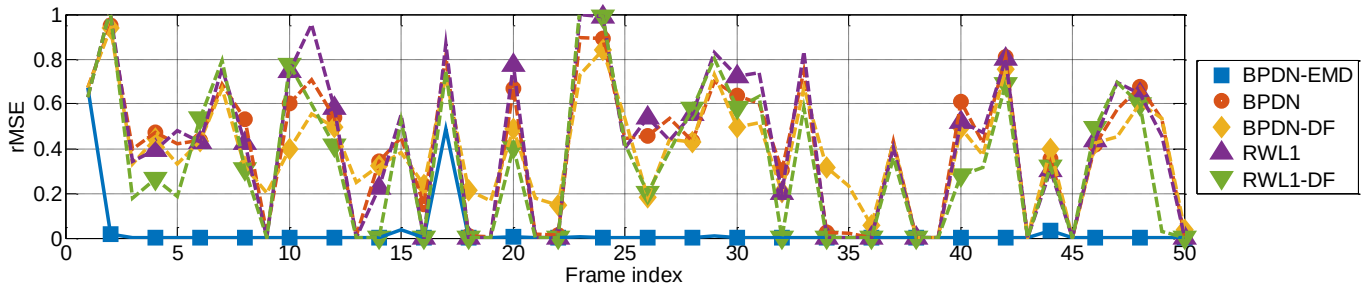

Fig. 2. Recovery of a sequence of moving-targets, simulated with pseudo-Brownian motion where targets move randomly within a one-pixel neighborhood (between frames). Compressive measurements were used to recover the sequence using basis-pursuit de-noising (BPDN), BPDN-DF, re-weighted ℓ_1 (RWL1) and RWL1-DF. Algorithmic accuracy was calculated with the mean-squared error. For dynamic-aware algorithms, the prediction signal is taken as the algorithms previous estimate. In this simulation scenario ($N = 256, M = 0.15N, K = 0.33M$), the EMD-regularized BPDN achieves the lowest overall rMSE with remarkable robustness. Although the other dynamic filtering algorithms (BPDN-DF and RWL1-DF) demonstrated slight improvements over their non-dynamic versions (BPDN and RWL1), they did not have the robustness of BPDN-EMD.

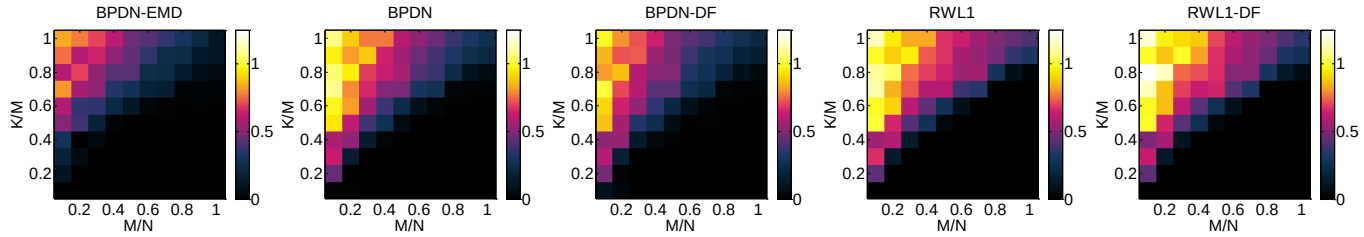

Fig. 3. One-step recovery results as Donoho-Tanner phase transition diagram. Here, $N = 100$ and each point in each image was generated from the mean rMSE of 10 independent one-step recovery simulations. These diagrams illustrate EMD-regularized BPDN's superior rMSE performance in the space of M, K, as compared to other algorithms.

References

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME-Journal of Basic Engineering*, vol. 82, no. D, pp. 35–45, 1960.

[2] S. Haykin, "Kalman filters," in *Kalman Filtering and Neural Networks*, S. Haykin, Ed. John Wiley & Sons, Inc., 2001, pp. 1–22.

[3] J. Ziniel and P. Schniter, "Dynamic compressive sensing of time-varying signals via approximate message passing," *IEEE Transaction on Signal Processing*, vol. 61, no. 21, pp. 5270–5284, 2013.

[4] D. Zachariah, S. Chatterjee, and M. Jansson, "Dynamic iterative pursuit," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4967–4972, 2012.

[5] N. Vaswani, "Kalman filtered compressed sensing," *Proc of ICIP 2008*, pp. 893–896, 2008.

[6] A. S. Charles, M. S. Asif, J. Romberg, and C. J. Rozell, "Sparsity penalties in dynamical system estimation," in *Proc of the CISS*, 2011, pp. 1–6.

[7] E. C. Hall and R. M. Willett, "Online convex optimization in dynamic environments," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 4, pp. 647–662, 2015.

[8] A. S. Charles, A. Balavoine, and C. J. Rozell, "Dynamic filtering of time-varying sparse signals via ℓ_1 minimization," *IEEE Transactions on Signal Processing*, vol. 64, no. 21, pp. 5644–5656, Nov 2016.

[9] L. Schmidt, C. Hegde, and P. Indyk, "The constrained earth mover distance model, with applications to compressive sensing," in *10th Intl. Conf. on Sampling Theory and Appl. (SAMPTA)*, 2013.

[10] R. Gupta, P. Indyk, and E. Price, "Sparse recovery for earth mover distance," in *48th Allerton Conference on Com., Cont., and Comp.*, 2010, pp. 1742–1744.

[11] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *International journal of computer vision*, vol. 40, no. 2, pp. 99–121, 2000.