SPARSE DYNAMIC FILTERING VIA EARTH MOVER'S DISTANCE REGULARIZATION

Nicholas P. Bertrand[†] *John Lee*[†] *Adam S. Charles*[⋆] *Pavel Dunn*[†] *Christopher J. Rozell*[†]

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

ABSTRACT

Tracking time-varying signals is an important task for practical systems working with large discretized domains. Under such settings, sparsity-based approaches improve tracking accuracy since typically few targets appear in the scene (i.e. few locations in the discretized space are occupied). Discretization introduces a unique challenge: the traditional ℓ_p-norm dynamic constraints produce significant errors when there is even a small spatial mismatch between the predicted and true state. To overcome this, we present a tracking algorithm leveraging concepts from optimal transport, namely utilizing the earth-movers distance (EMD) as a dynamic regularizer to the ℓ_1-regularized inference problem (i.e., LASSO [1], or BPDN [2]). We extend the problem formulation to complex valued signals and modify the optimization program to reduce the computational burden. We demonstrate the efficacy of our approach in imaging and frequency tracking applications.

Index Terms— Dynamic Filtering, Earth-mover's Distance, Compressive Sensing, Kalman Filtering

1. INTRODUCTION

Tracking a temporally changing signal is a classical problem in signal processing, often called *dynamic filtering*. Dynamic filtering combines noisy measurements of a time-varying signal with a predicted estimate, to accurately infer the new underlying signal. In classical tracking literature, the celebrated Kalman filter [3] provides optimal and efficient tracking under Gaussian model assumptions on the signal, measurement and dynamic model mismatch.

Another popular and more recent approach to tracking has been centered around the notion of exploiting sparse signal structure to regularize inverse problems [4, 5, 6, 7, 8, 9, 10, 11]. Numerous sparsity-aware tracking algorithms incorporate dynamic structure using the idea of prediction consistency via an ℓ_p-norm metric to encode the notion of distance between the prediction and the estimate. In some applications however, ℓ_p-norm metrics can disproportionately penalize "good" predictions that contain just small spatial mismatches. Consider, for example, the discretized scenario where a location mismatch of 1 pixel from its true location yields the same ℓ_p-norm error as a case that is mismatched by 10 pixel locations.

To alleviate the problem of spatial insensitivity in ℓ_p-norm based approaches, we leverage concepts from optimal transport (OT) which provides a natural geometric framework to proportionally penalize on spatial mismatches to robustly incorporate dynamical (predictive) information. In particular, we describe a tracking algorithm that uses the unbalanced earth-mover's distance (EMD) [12, 13] as a regularizer to infuse dynamical information into the tracking algorithm. The result is a new tracking framework we call EMD-regularized BPDN, or EMD-BPDN. We expand on the work discussed in [14] for nonnegative signals to allow for complex-valued signals. We further modify the formulation to remove a nonlinear constraint and in turn reduce computational complexity. Finally, we empirically demonstrate improvements due to the EMD-regularizer over current methods via target tracking and frequency tracking simulations.

2. BACKGROUND

2.1. Dynamic Filtering

Formally stated, an unknown signal $\boldsymbol{x}_n \in \mathbb{R}^N$ at time n may be observed through a measurement system

$$\boldsymbol{y}_n = \boldsymbol{A}_n \boldsymbol{x}_n + \boldsymbol{\epsilon}_n, \tag{1}$$

where $\boldsymbol{A}_n \in \mathbb{R}^{M\times N}$ is a measurement matrix, $\boldsymbol{\epsilon}_n \in \mathbb{R}^M$ represents measurement noise, and $\boldsymbol{y}_n \in \mathbb{R}^M$ are the resulting linear measurements. With some knowledge of the dynamical system governing the evolution of $\boldsymbol{x}_n$, we can linearly model dynamics as

$$\boldsymbol{x}_n = \boldsymbol{G}_n \boldsymbol{x}_{n-1} + \boldsymbol{\nu}_n, \tag{2}$$

where $\boldsymbol{G}_n \in \mathbb{R}^{N\times N}$ describes the dynamics and $\boldsymbol{\nu}_n$ the model error, sometimes called the *innovations*.

This work was supported in part by NSF grant CCF-1409422 and the James S. McDonnell Foundation.

The classical Kalman filter [3] may be concisely formulated as a regularized least-squares problem

$$\widehat{\boldsymbol{x}}_n = \arg\min_{\boldsymbol{x}} \Big[\|\boldsymbol{y}_n - \boldsymbol{A}_n\boldsymbol{x}\|^2_{2,\boldsymbol{R}_n} + \|\boldsymbol{x} - \boldsymbol{G}_n\widehat{\boldsymbol{x}}_{n-1}\|^2_{2,(\boldsymbol{Q}_n + \boldsymbol{G}_n\boldsymbol{P}_{n-1}\boldsymbol{G}_n^T)}\Big] \quad (3)$$

where $\widehat{\boldsymbol{x}}_{n-1}$ and $\boldsymbol{P}_{n-1}$ are estimates of the previous time step and its covariance, and $\boldsymbol{Q}_n$ and $\boldsymbol{R}_n$ are the covariances for the innovations and measurement noise, with the following norm-notation $\|\boldsymbol{v}\|_{2,\boldsymbol{C}} = \sqrt{\boldsymbol{v}^\top \boldsymbol{C}\boldsymbol{v}}$ for $\boldsymbol{C} \in \{\boldsymbol{x} \in \mathbb{R}^{N\times N} : \boldsymbol{x} \succeq 0\}$. Under a linear model (on the measurements and dynamics) with Gaussian assumptions (on the signal, innovations and measurement noise), the Kalman filter is guaranteed to converge to the same solution as if all past data was used at once.

Sparsity models, however, differ substantially from Gaussian assumptions and instead, are better modeled with high-kurtosis distributions such as the Laplace distribution [15]. Under such sparsity models, we express $\boldsymbol{x}_n$ as a linear combination of just a few elements from a large dictionary. Formally, $\boldsymbol{x}_n = \boldsymbol{\Phi}\boldsymbol{a}_n$, where the columns of $\boldsymbol{\Phi}$ contain the dictionary elements, and $\boldsymbol{a}_n$ is a sparse vector, i.e. only a small subset of its coefficients are non-zero.

A popular sparsity algorithm for inferring coefficients under noisy measurements is basis pursuit denoising (BPDN) [2], also known as the LASSO [1]:

$$\widehat{\boldsymbol{a}}_n = \arg\min_{\boldsymbol{a}} \left(\|\boldsymbol{y}_n - \boldsymbol{A}_n\boldsymbol{\Phi}\boldsymbol{a}\|_2^2 + \lambda \|\boldsymbol{a}\|_1 \right). \quad (4)$$

The minimization seeks to find the Pareto frontier on measurement fidelity versus coefficient sparsity using a trade-off parameter λ, but does not consider dynamic information.

One dynamical extension to BPDN is cast with an additional ℓ_p-norm tracking regularizer (BPDN-DF) [11, 16]:

$$\widehat{\boldsymbol{a}}_n = \arg\min_{\boldsymbol{a}} \|\boldsymbol{y}_n - \boldsymbol{A}_n\boldsymbol{\Phi}\boldsymbol{a}\|_2^2 + \lambda \|\boldsymbol{a}\|_1 + \gamma \|\boldsymbol{\Phi}\boldsymbol{a} - \boldsymbol{G}_n\boldsymbol{\Phi}\widehat{\boldsymbol{a}}_{n-1}\|_p^p. \quad (5)$$

This program incorporates dynamics by penalizing disagreements between the dynamics model and the current signal estimate, thereby encouraging the estimations to agree with the predictions. The free parameter γ trades-off between the dynamics prediction and the BPDN solution.

More recent methods, such as re-weighted ℓ_1 dynamic filtering (RWL1-DF), [16] have incorporated second order sparsity statistics into the model, by iteratively solving a weighted BPDN estimate

$$\widehat{\boldsymbol{a}}_n = \arg\min_{\boldsymbol{a}} \|\boldsymbol{y}_n - \boldsymbol{A}_n\boldsymbol{\Phi}\boldsymbol{a}\|_2^2 + \lambda_0 \sum_i \lambda_i |a_i|, \quad (6)$$

and a weight update that fuses dynamical and measurement information via

$$\lambda_i = \frac{\alpha}{\beta + |\widehat{a}_n[i]| + \kappa |(\boldsymbol{\Phi}^{-1}\boldsymbol{G}_n\boldsymbol{\Phi}\widehat{\boldsymbol{a}}_{n-1})[i]|}, \quad (7)$$

with free parameters α, β, κ. This model is more robust because it preserves sparsity deviations on dynamic model errors [16]. Although BPDN-DF and RWL1-DF exploit dynamical information, they have a strict spatial dependence on the previous time-step's estimate (e.g., ℓ_p-norm error metric), causing them to be sensitive to spatial mismatches.

2.2. Earth Mover's Distance

The earth mover's distance (EMD) was introduced in [12] and first applied to the machine learning task of histogram matching, but has since been applied to solving inverse problems [17, 18, 19]. The classical balanced EMD deals with transportation between two valid probabilities of equal masses, while the unbalanced setting deals with non-equal masses. The latter setting is sometimes called the optimal partial transport problem, with associated analyses found in [20, 21]. More generally, the masses need not be probabilities (e.g., they may represent intensities of pixels in images), therefore the unbalanced EMD can be useful in a variety of applications.

The EMD between two signals $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$, denoted by $d_{\text{EMD}}(\boldsymbol{x}, \tilde{\boldsymbol{x}})$ may be stated as solving the following constrained linear optimization program

$$\begin{aligned} \min_{\boldsymbol{F}} \sum_{ij} F_{ij} r_{ij} \quad \text{s.t.} \quad & F_{ij} \geq 0 \\ & \sum_j F_{ij} \leq x_i \\ & \sum_i F_{ij} \leq \widetilde{x}_j \\ & \sum_{ij} F_{ij} = \min\left(\sum_i x_i, \sum_j \widetilde{x}_j\right), \end{aligned} \quad (8)$$

The EMD may be interpreted as flows F_{ij} of "mass" travelling between pixel x_i and $\widetilde{x}_j$, with r_{ij} denoting its associated displacement cost. The second and third constraints describe the conservation of mass between $\boldsymbol{x}$ and $\widetilde{\boldsymbol{x}}$ while the final constraint motivates flows (thereby preventing the trivial solution of zero flows).

The EMD has been used in BPDN in place of the ℓ_2 norm [17], or to regularize differences between columns of a sparse matrix [18], but has not thus far been explored for tracking applications.

3. EMD AS A TRACKING REGULARIZER

We can incorporate the EMD into the tracking problem by replacing the ℓ_p dynamics term in (5):

$$\widehat{\boldsymbol{x}}_n = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y}_n - \boldsymbol{A}_n\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1 + \gamma d_{\text{EMD}}(\boldsymbol{x}, \tilde{\boldsymbol{x}}),$$

where $\tilde{\boldsymbol{x}} = \boldsymbol{G}_n \widehat{\boldsymbol{x}}_{n-1}$ represents the signal predicted from the dynamics model. Intuitively, the EMD dynamics penalty is more tolerant toward inaccuracies in the locations of the active elements in the signal compared to ℓ_p-based regularizers.

Since evaluation of $d_{\text{EMD}}(\cdot,\cdot)$ itself involves solving an optimization program, we can optimize jointly over the EMD flows and the signal solution:

$$
\begin{aligned}
\widehat{\boldsymbol{x}}_n = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y}_n - \boldsymbol{A}_n \boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1 + \gamma \min_{\boldsymbol{F}} \sum_{ij} F_{ij} r_{ij} \\
\text{s.t.} \quad & F_{ij} \geq 0 \\
& \sum_j F_{ij} \leq x_i \\
& \sum_i F_{ij} \leq \widetilde{x}_j \\
& \sum_{ij} F_{ij} = \min(\|\boldsymbol{x}\|_1, \|\widetilde{\boldsymbol{x}}\|_1). \qquad (9)
\end{aligned}
$$

There are two issues with this formulation that limit its utility in practice. First, the non-linearity in the last constraint complicates the computation of a solution. One way to deal with this non-linearity (which we explore in [14]) involves solving the optimization twice, using both possibilities of the $\min$ term. In this work, we introduce a slack variable u to replace the $\min$. In particular, we replace the last constraint with three new ones: $u \leq \sum \boldsymbol{x}$, $u \leq \sum \tilde{\boldsymbol{x}}$, and $\sum \boldsymbol{F} = u$. We then introduce an extra term into the objective function to encourage large values for u.

The second issue with the formulation is that we have thus far assumed that the elements of $\boldsymbol{x}$ are nonnegative. We would like to consider the more general signal class $\boldsymbol{x} \in \mathbb{C}^N$, such as signals represented with Fourier matrices. The natural extension would be to replace the second and third constraints with $\sum_j F \leq |x_i|$ and $\sum_i F \leq |\tilde{x}_j|$ respectively. Unfortunately the resulting optimization problem is non-convex. Instead, we introduce an approximation by decomposing the real and imaginary parts of the signal into positive and negative components $\boldsymbol{x}_{\text{re}}^+, \boldsymbol{x}_{\text{re}}^-, \boldsymbol{x}_{\text{im}}^+, \boldsymbol{x}_{\text{im}}^- \in \mathbb{R}^+$ such that $\boldsymbol{x} = (\boldsymbol{x}_{\text{re}}^+ - \boldsymbol{x}_{\text{re}}^-) + i(\boldsymbol{x}_{\text{im}}^+ - \boldsymbol{x}_{\text{im}}^-)$. Defining

$$
\boldsymbol{x}' := \begin{bmatrix} \boldsymbol{x}_{\text{re}}^+ \\ \boldsymbol{x}_{\text{re}}^- \\ \boldsymbol{x}_{\text{im}}^+ \\ \boldsymbol{x}_{\text{im}}^- \end{bmatrix}
$$

and

$$
\boldsymbol{A}' := \begin{bmatrix} \boldsymbol{A} & -\boldsymbol{A} & i\boldsymbol{A} & -i\boldsymbol{A} \end{bmatrix},
$$

one may verify that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}'\boldsymbol{x}'$ for all $\boldsymbol{x} \in \mathbb{C}^N$. We can thus approximate the solution to (9) by solving

$$
\begin{aligned}
\widehat{\boldsymbol{x}}_n = \arg\min_{\boldsymbol{x}', \boldsymbol{F}} & \|\boldsymbol{y}_n - \boldsymbol{A}'_n \boldsymbol{x}'\|_2^2 + \lambda \|\boldsymbol{x}'\|_1 \qquad (10) \\
& + \gamma \sum_{ij} F_{ij} r_{ij} - \mu u \\
\text{s.t.} \quad & F_{ij} \geq 0, \boldsymbol{x}' \geq 0 \\
& \sum_j F_{ij} \leq \left(x_{\text{re}}^+\right)_i + \left(x_{\text{re}}^-\right)_i + \left(x_{\text{im}}^+\right)_i + \left(x_{\text{im}}^-\right)_i \\
& \sum_i F_{ij} \leq |\widehat{x}_j| \\
& \sum_{ij} F_{ij} = u \\
& u \leq \sum_i \boldsymbol{x}'_i, u \leq \sum_i \tilde{\boldsymbol{x}}'_i \qquad (11)
\end{aligned}
$$

It is important to note that (11) does not solve (9) exactly. First, we observe that the representation $\boldsymbol{x}'$ is not unique; for example, the elements of $\boldsymbol{x}_{\text{re}}^+$ and $\boldsymbol{x}_{\text{re}}^-$ may both contain nonzero entries in the same position causing $\boldsymbol{x}_{\text{re}}^+ + \boldsymbol{x}_{\text{re}}^-$ to be a poor approximation for $|\Re(\boldsymbol{x})|$. However, the sparsity regularizer serves to discourage such solutions. Second, in the ideal case where the positive and negative components have disjoint support (i.e., $(x_{\text{re}}^+)_i (x_{\text{re}}^-)_i = (x_{\text{im}}^+)_i (x_{\text{im}}^-)_i = 0$ for $i = 1, \cdots, N$), we have

$$
\left(x_{\text{re}}^+\right)_i + \left(x_{\text{re}}^-\right)_i + \left(x_{\text{im}}^+\right)_i + \left(x_{\text{im}}^-\right)_i = |\Re(x_i)| + |\Im(x_i)| \neq |x_i|.
$$

Our experiments indicate that this approximate solution to the tracking problem is still of use in practice.

We end this section with a note on computational complexity. Initial inspection of (11) suggests that we must solve for N^2 flow variables in addition to the original N signal variables. However, if the predicted solution $\tilde{\boldsymbol{x}}$ is K sparse, then the inequality $\sum F \leq |\tilde{x}_j|$ implies that all but K columns of F are zero, reducing the number of unknown flow variables to NK. Furthermore, the replacement of the equality constraint on the flows allows us to solve the problem only once per iteration instead of twice. Future directions for further reducing the computational cost are addressed in the discussion.

4. RESULTS

To demonstrate the efficacy of EMD regularized dynamical filtering, we explore two applications: object tracking in a video stream and frequency tracking in a time series.

4.1. Target Tracking

First, we consider the scenario where a sparse collection of objects move throughout a scene. Our dataset consists of synthetically generated frames consisting of K active pixels which move randomly to adjacent locations at each time step.

The dynamics model predicts the next frame to be the same as the current estimate, i.e. $\boldsymbol{G}_n = \boldsymbol{I}$. We quantify the performance of BPDN, BPDN-DF, RWL1-DF, and BPDN-EMD at various sparsity levels using the Donoho-Tanner phase transition diagrams shown in Figure 1. BPDN-EMD shows superior performance in the regime corresponding to fewer measurements or more active pixels.

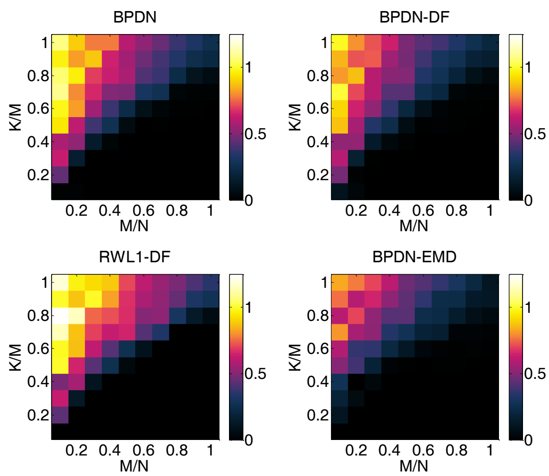

Fig. 1. One-step recovery results as Donoho-Tanner phase transition diagram. Here, $N = 100$ and each point value in each image was generated from the mean rMSE of 10 independent one-step recovery simulations (i.e., the lower the better). These diagrams illustrate EMD-regularized BPDN's superior rMSE performance in the space of M, K, as compared to other algorithms.

4.2. Frequency Tracking

The next experiment showcases the utility of the formulation of BPDN-EMD for complex valued signals by considering a frequency tracking application. We generate signals that consist of three frequencies which drift according to Brownian motion. We observe a noisy time-series of data and wish to recover the spectrum. The traditional spectrogram based on the Short-Time Fourier Transform is unable to simultaneously capture small changes in frequency and fine scale temporal dynamics. Instead, we employ sparse recovery methods with an overcomplete Discrete Fourier Transform (DFT) matrix, i.e., $A_{lk} = e^{i2\pi lk/N}$ for $l = 0, \cdots, M-1, k = 0, \cdots, N-1$ where $M < N$. The oversampling factor N/M controls how much additional frequency resolution the representation captures compared to the standard DFT matrix.

Our error metric is computed as follows: for each time sample, we project the ground truth frequencies and the spectrum estimate onto a high resolution grid. Each frequency present in the ground truth signal occupies a single element on this grid, whereas the frequency range covered by a spectral bin in the signal estimate occupies multiple elements (more for larger bins and fewer for smaller bins.) We then compute the normalized EMD between the ground truth and the signal estimate on this high frequency grid. The aggregate error is calculated by summing these distances over all time samples. Higher resolution estimates are more concentrated on the frequency grid, so less "work" must be done to transform them into the single peaks representing the ground truth. Thus, this metric assigns higher error to spectra with lower frequency resolution. Furthermore, since we use the EMD, estimates with frequencies close to the ground truth are assigned lower error than those that are far away. Figure 2 illustrates the benefit of using the EMD as a tracking regularizer.

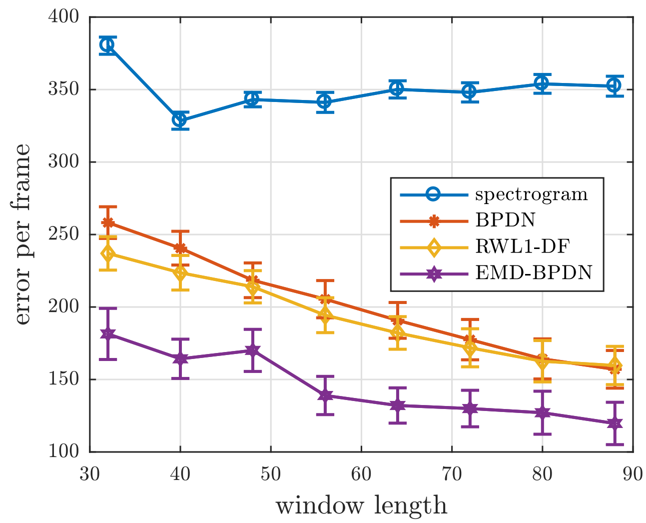

Fig. 2. Integrated EMD error. Error bars indicate confidence intervals computed using 150 trials. All sparse methods produce lower error than the spectrogram. The addition of dynamics information allows RWL1-DF and EMD-BPDN to outperform standard BPDN. EMD-BPDN is best able to take advantage of the dynamical model and thus produces the lowest error.

5. CONCLUSIONS

We investigate here how optimal transport can improve the performance of sparsity-aware dynamic filtering. Specifically, we describe an algorithm that exploits the EMD (or optimal partial transport) as a dynamical regularizer and empirically characterize its performance. We conclude that an EMD regularizer has the potential to improve performance in image processing and frequency tracking applications. Thus it a worthwhile goal to further explore EMD-regularized trackers in related fields such as computer vision. We formulate our algorithm as a convex optimization program that can handle general cost matrices, with strategies to reduce computational burden when the signal is sparse. In future work, we will study how to further improve computational efficiency by exploiting recent advances in the optimal transport literature.

6. REFERENCES

[1] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[2] Richard G Baraniuk, "Compressive sensing [lecture notes]," *IEEE signal processing magazine*, vol. 24, no. 4, pp. 118–121, 2007.

[3] 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.

[4] Adam S Charles and Christopher J Rozell, "Spectral superresolution of hyperspectral imagery using reweighted ℓ_1 spatial filtering," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 3, pp. 602–606, 2014.

[5] Namrata Vaswani, "Kalman filtered compressed sensing," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008, pp. 893–896.

[6] Adam S Charles and Christopher J Rozell, "Convergence of basis pursuit de-noising with dynamic filtering," in *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on*. IEEE, 2014, pp. 374–378.

[7] Zhilin Zhang and Bhaskar D Rao, "Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 912–926, 2011.

[8] M Salman Asif, Lei Hamilton, Marijn Brummer, and Justin Romberg, "Motion-adaptive spatio-temporal regularization for accelerated dynamic MRI," *Magnetic Resonance in Medicine*, vol. 70, no. 3, pp. 800–812, 2013.

[9] Justin Ziniel and Philip Schniter, "Dynamic compressive sensing of time-varying signals via approximate message passing," *IEEE transactions on signal processing*, vol. 61, no. 21, pp. 5270–5284, 2013.

[10] Eric C Hall and Rebecca M Willett, "Dynamical models and tracking regret in online convex programming," *arXiv preprint arXiv:1301.1254*, 2013.

[11] Adam Charles, M Salman Asif, Justin Romberg, and Christopher Rozell, "Sparsity penalties in dynamical system estimation," in *Information Sciences and Systems (CISS), 2011 45th Annual Conference on*. IEEE, 2011, pp. 1–6.

[12] 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.

[13] Ofir Pele and Michael Werman, "Fast and robust earth mover's distances," in *Computer vision, 2009 IEEE 12th international conference on*. IEEE, 2009, pp. 460–467.

[14] Adam S. Charles, Nicholas P. Bertrand, John Lee, and Christopher J. Rozell, "Earth-mover's distance as a tracking regularizer," in *2017 IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*. IEEE, Dec 2017.

[15] Pierre Garrigues and Bruno A Olshausen, "Group sparse coding with a laplacian scale mixture prior," in *Advances in neural information processing systems*, 2010, pp. 676–684.

[16] 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.

[17] Rishi Gupta, Piotr Indyk, and Eric Price, "Sparse recovery for earth mover distance," in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*. IEEE, 2010, pp. 1742–1744.

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

[19] Dian Mo and Marco F Duarte, "Compressive parameter estimation with earth movers distance via k-median clustering," *Proc. SPIE Wavelets and Sparsity XV*, vol. 8858, 2013.

[20] Luis A Caffarelli and Robert J McCann, "Free boundaries in optimal transport and monge-ampere obstacle problems," *Annals of mathematics*, pp. 673–730, 2010.

[21] Benedetto Piccoli and Francesco Rossi, "Generalized wasserstein distance and its application to transport equations with source," *Archive for Rational Mechanics and Analysis*, vol. 211, no. 1, pp. 335–358, 2014.