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 $\ell_p$-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 $\ell_1$-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 $\ell_p$-norm metric to encode the notion of distance between the prediction and the estimate. In some applications however, $\ell_p$-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 $\ell_p$-norm error as a case that is mismatched by 10 pixel locations.

To alleviate the problem of spatial insensitivity in $\ell_p$-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 non-linear 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 $x_n \in \mathbb{R}^N$ at time $n$ may be observed through a measurement system

$$y_n = A_n x_n + \epsilon_n,$$

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

$$x_n = G_n x_{n-1} + \nu_n,$$

where $G_n \in \mathbb{R}^{N \times N}$ describes the dynamics and $\nu_n$ the model error, sometimes called the innovations.
The classical Kalman filter [3] may be concisely formulated as a regularized least-squares problem
\[
\hat{x}_n = \arg\min_x \left[ \| y_n - A_n x \|_2^2 + \| x - G_n \hat{x}_{n-1} \|_2^2 (Q_n + G_n P_{n-1} G_n^\top) \right]
\]
where \(\hat{x}_{n-1}\) and \(P_{n-1}\) are estimates of the previous time step and its covariance, and \(Q_n\) and \(R_n\) are the covariances for the innovations and measurement noise, with the following notation \(\| v \|_2 C = \sqrt{v^\top C v}\) for \(C \in \{ x \in \mathbb{R}^{N \times N} : x \succeq 0 \}\). Under a linear model (on the measurements and dynamics), the Kalman filter is guaranteed to converge to the same solution as if all past data was used at once.

Dynamical extensions to BPDN-DF and RWL1-DF exploit dynamical information, they have a strict spatial dependence on 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., \(\ell_p\)-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 \(x\) and \(\tilde{x}\), denoted by \(d_{\text{EMD}}(x, \tilde{x})\), may be stated as solving the following constrained linear optimization program
\[
\min_F \sum_{ij} F_{ij} r_{ij} \quad \text{s.t.} \quad F_{ij} \geq 0, \sum_j F_{ij} \leq x_i, \sum_i F_{ij} \leq \tilde{x}_j
\]

The EMD may be interpreted as flows \(F_{ij}\) of “mass” traveling between pixel \(x_i\) and \(\tilde{x}_j\), with \(r_{ij}\) denoting its associated displacement cost. The second and third constraints describe the conservation of mass between \(x\) and \(\tilde{x}\) 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 \(\ell_2\) 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 \(\ell_p\) dynamics term in (5):
\[
\hat{x}_n = \arg\min_x \| y_n - A_n x \|_2^2 + \lambda \| x \|_1 + \gamma d_{\text{EMD}}(x, \tilde{x}),'\]
where \( \tilde{x} = G_n \hat{x}_{n-1} \) 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 \( \ell_p \)-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:

\[
\hat{x}_n = \arg \min_{\tilde{x}} \|y_n - A_n \tilde{x}\|_2^2 + \lambda \|\tilde{x}\|_1 + \gamma \min_{F} \sum_{ij} F_{ij} r_{ij}
\]

s.t. \( F_{ij} \geq 0 \)
\[\sum_j F_{ij} \leq x_i\]
\[\sum_i F_{ij} \leq \tilde{x}_j\]
\[\sum_i F_{ij} = \min(\|\tilde{x}\|_1, \|\tilde{x}\|_1). \tag{9}\]

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: \( u \leq \sum \tilde{x}, u \leq \sum \hat{x} \), and \( \sum F = u \). 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 \( \tilde{x} \) are nonnegative. We would like to consider the more general signal class \( \tilde{x} \in \mathbb{C}^N \), 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 \( \sum_i F \leq |	ilde{x}_j| \) 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 \( x_{\text{re}}^+, x_{\text{re}}^-, x_{\text{im}}^+, x_{\text{im}}^- \in \mathbb{R}^+ \) such that \( \tilde{x} = (x_{\text{re}}^+ - x_{\text{re}}^-) + i(x_{\text{im}}^+ - x_{\text{im}}^-) \). Defining

\[
\tilde{x}' := \begin{bmatrix} x_{\text{re}}^+ \\ x_{\text{re}}^- \\ x_{\text{im}}^+ \\ x_{\text{im}}^- \end{bmatrix}
\]

and

\[
A' := \begin{bmatrix} A & -A & iA & -iA \end{bmatrix},
\]

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

\[
\hat{x}_n = \arg \min_{\tilde{x}', F} \|y_n - A'_n \tilde{x}'\|_2^2 + \lambda \|\tilde{x}'\|_1 + \gamma \sum_{ij} F_{ij} r_{ij}
\]

s.t. \( F_{ij} \geq 0, \tilde{x}', x \geq 0 \)
\[\sum_j F_{ij} \leq (x_{\text{re}}^+)_i + (x_{\text{re}}^-)_i + (x_{\text{im}}^+)_i + (x_{\text{im}}^-)_i \]
\[\sum_i F_{ij} \leq |\tilde{x}_j|\]
\[\sum_{ij} F_{ij} = u \]
\[u \leq \sum \tilde{x}', u \leq \sum \hat{x}' \tag{11}\]

It is important to note that (11) does not solve (9) exactly. First, we observe that the representation \( \tilde{x}' \) is not unique; for example, the elements of \( x_{\text{re}}^+ \) and \( x_{\text{re}}^- \) may both contain non-zero entries in the same position causing \( x_{\text{re}}^+ + x_{\text{re}}^- \) to be a poor approximation for \( |\Re(\tilde{x})| \). 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 \( i = 1, \ldots, N \)), we have

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

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 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., \( G_n = I \). 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.

![Phase Transition Diagrams](image)

**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., \( A_{ik} = e^{i2\pi ik/M} \) 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.

![Error per Frame](image)

**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


