Online convex optimization meets sparsity

Original

Availability:
This version is available at: 11583/2673368 since: 2018-03-16T17:45:03Z

Publisher:
SPARS

Published
DOI:

Terms of use:
openAccess
This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)
Online convex optimization meets sparsity

Sophie M. Fosson
ISMB, Torino (Italy) sophie.fosson@ismb.it

Javier Matamoros, Maria Gregori
CTTC/CERCA, Barcelona (Spain) javier.matamoros@cttc.es, maria.gregori@cttc.es

Enrico Magli
DET, Politecnico di Torino (Italy) enrico.magli@polito.it

Abstract—Tracking time-varying sparse signals is a novel problem, with broad applications. Techniques merging compressed sensing and Kalman filtering have been proposed in the related literature, which typically rely on specific dynamic models. In this work, we propose a new perspective on the problem, based on elements of online convex optimization. In particular, we design a suitable optimization problem and develop algorithms which do not assume any specific dynamic model. For these algorithms, we analytically evaluate the behavior of their dynamic regrets that serve as their performance measure.

The problem of tracking dynamic (i.e., time-varying) sparse signals has arisen in the last few years in the literature of sparse signal estimation [1], [2], [3], [4], [5]. The dynamic environment is more natural in a number of applications, e.g., magnetic resonance imaging [6], [7] and spectrum sensing in cognitive radios [8], [9], where the signal of interest is subject to variations. The acquired measurements are then time series, for which online processing is generally preferred for promptness of response and computational complexity: at each new measurement (or bunch of measurements) the current signal is estimated. Sparsity may be inherent in the signal itself (i.e., each signal frame has a sparse representation) or it may occur in the difference between consecutive frames assuming a sufficiently slow dynamics.

In the static environment, compressed sensing (CS) [10] has introduced a rigorous theory and efficient algorithms to recover sparse signals from linear, compressed measurements. It has then been natural to try to extend the CS paradigm to the dynamic environment. On the one hand, iterative CS algorithms have been revisited for the dynamic framework (e.g., approximate message passing in [1], [2], iterative soft thresholding in [4]); on the other hand, Kalman filtering approach has been merged with CS and sparsity models [11], [12], [13], [5]. In both cases, numerical results are encouraging, while theoretical results are lacking or strongly related to the knowledge of a specific signal evolution model.

This work aims at filling this gap, by providing a theoretical analysis untied from specific evolution models. For this purpose, we resort to the online convex optimization (OCO) theory [14], recently developed within the machine learning community. OCO can be described as a game in which, at each time step $t \in \{1, \ldots, T\}$, a learner incurs in a convex cost functional $f_t$ revealed by an adversary. Then, the learner aims at minimizing $f_t$, which may be not computationally feasible. To circumvent that, a low-complex tracking strategy is adopted instead, that keeps as close as possible to the desired optimum. A suitable performance metric to evaluate such a strategy is the so-called dynamic regret, defined as follows [15], [16]:

$$Reg^f_t(x^*_1, \ldots, x^*_T) := \sum_{t=1}^{T} f_t(x_t) - f_t(x^*_t)$$

where $x^*_t = \arg \min_{x \in \mathcal{X}} f_t(x)$ ($\mathcal{X}$ being the feasibility set), and $x_t$ is the action played by the learner at time $t$, before the revelation of $f_t$. $f_t(x_t) - f_t(x^*_t)$ is usually referred to as loss.

In the OCO literature, the action typically is a gradient descent step [15]. Let $C_T := \sum_{t=2}^{T} \|x_t - x_{t-1}\|_2$. In [15], it has been proved that $Reg^f_t(x^*_1, \ldots, x^*_T) = O(\sqrt{T}(1 + C_T))$. More recently, in [16] this result has been improved to $Reg^f_t(x^*_1, \ldots, x^*_T) = O(1 + C_T)$, under the hypothesis that $f_t$ is strongly convex, and assuming that $\nabla f_t$ is Lipschitz continuous and bounded [16, Assumptions 2-3].

Our aim is to design a suitable optimization problem and an online algorithm for our sparse signal tracking problem, and obtain a dynamic regret result analogous to [16]. Let $\tilde{x}_t \in \mathbb{R}^N$, $t \in \{1, \ldots, T\}$, be the sparse signal to be tracked. According to the CS paradigm, we acquire compressed measurements $y_t = A\tilde{x}_t$, where $A \in \mathbb{R}^{M \times N}$ is a suitable sensing matrix with $M < N$. As cost functional, we consider the Elastic-net, which supports sparsity with a grouping effect [17], [18], and reads as follows:

$$f_t(x) = \frac{1}{2} \|y_t - Ax_t\|_2^2 + \lambda \|x\|_1 + \mu \|x\|_2^2, \quad t \in \{1, \ldots, T\}$$

where $\lambda > 0$ and $\mu > 0$ are parameters to be fixed. The Elastic-net is strongly convex, but does not fulfill the assumptions of [16] (actually $f_t$ is even not differentiable), which prevents us to use the methods and the analysis proposed in that paper. Our contribution consists then of (a) the development of algorithms to tackle the dynamic Elastic-net, and (b) their corresponding dynamic regret analysis. The algorithms that we propose to tackle such dynamic Elastic-net are online versions of the well-known iterative soft thresholding (IST) algorithm [19] and alternating direction method of multipliers (ADMM) [20], see Table I. Let $C_T := \sum_{t=2}^{T} \|\tilde{x}_{t-1} - \tilde{x}_t\|_2$. Our main results are summarized in the following theorems (whose proofs are omitted for brevity):

**Theorem 1.** If $\tau \|A\|_2^2 < 1$ (where $\tau$ is the gradient parameter, see Table I), the online IST for dynamic Elastic-net has $Reg^f_t(x^*_1, \ldots, x^*_T) = O(1 + C_T)$.

**Theorem 2.** If $\|\tilde{x}_t\| \leq \beta$ for some $\beta > 0$, the online ADMM for dynamic Elastic-net has $Reg^f_t(x^*_1, \ldots, x^*_T) = O(1 + C_T)$.

These theorems state in particular that the regret stabilizes when the signal $\tilde{x}_t$ stabilizes. Table II puts these results into perspective with previous OCO works. To conclude, in Figure 1 we show the results of some numerical tests. We consider two models for $\tilde{x}(t) \in \mathbb{R}^N$:

- (M1) constant support $\Omega \subset \{1, \ldots, N\}$ (chosen uniformly at random), $|\Omega| = k$, non-zero values: $(\tilde{x}_{\Omega})_i(t) = \eta_{i,0} + \eta_{i,t}$, where $\eta_{i,j} \sim N(0,1), i \in \{1, \ldots, N\}, j \in \{0,1, \ldots, T\}$;
- (M2) support $\Omega_t$ chosen uniformly at random at each time $t$, with $|\Omega_t| = k$ constant; non-zero values: $(\tilde{x}_t|_{\Omega_t})_i(t) = \left(x_t\right|_{\Omega_t})_i(t-1) + \eta_{i,t}$ where $\eta_{i,t} \sim N(0,1), i \in \{1, \ldots, N\}, t \in \{1, \ldots, T\}$.

In Figure 1, we see that, for both algorithms, $Reg^f_t$ stabilizes as expected when the signal stabilizes, and accordingly the loss tends to zero. Moreover, ADMM turns out to be quicker than IST.
**Table I**

**PROPOSED ALGORITHMS**

| $S_\alpha$ denotes the soft thresholding operator with parameter $\alpha$, defined as follows: given $x \in \mathbb{R}$, $|x| \leq \alpha$, $S_\alpha(x) := 0$; otherwise, $S_\alpha(x) := x - \text{sign}(x) \alpha$. The definition is extended component-wise to vectors. $
\tau > 0$ is the gradient parameter of IST; it must be chosen such that $\tau \|A\|^2 < 1$ (see Theorem 1). $\mu$ and $\tau$ are defined in (1). For any $t \in \{1, \ldots, T\}$, the measurement $y_t = A\tilde{x}_t$ (hence, $f_t$) is revealed after that action $x_t$ is played.

**Online IST for dynamic Elastic-net**

For any $t \in \{2, \ldots, T\}$,

1. $x_t = S_{\alpha \tau^{-1}} \left[ x_{t-1} + \tau A^T (y_{t-1} - u_{t-1}) \right]$
2. read $y_t$.

**Online ADMM for dynamic Elastic-net**

For any $t \in \{2, \ldots, T\}$,

1. $x_t = B \left[ A^T y_{t-1} + x_{t-1} - u_{t-1} \right]$
2. $z_t = S_{\alpha \tau^{-1}} \left[ x_t + u_{t-1} \right]$
3. $u_t = u_{t-1} - x_t - z_t$
4. read $y_t$.

<table>
<thead>
<tr>
<th>$J_t$</th>
<th>$\text{Reg}_f^2$</th>
<th>Assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>[15] C</td>
<td>$O(\sqrt{T}(1+C_T))$</td>
<td>$|f_t| \leq \beta$, $\mathcal{X}$ compact</td>
</tr>
<tr>
<td>[16] SC</td>
<td>$O(1+C_T)$</td>
<td>$|f_t| \leq \beta$</td>
</tr>
<tr>
<td><strong>This work</strong> Elastic-net (SC)</td>
<td>$O(1+C_T)$</td>
<td>For ADMM: $|x_1| \leq \beta$</td>
</tr>
</tbody>
</table>

**References**


---

**Figure 1.** Simulations on models M1 (left column) and M2 (right column); $N = 100$, $M = 20$, $k = 5$. The graphs represent (from top to down) the signal variation $\sum_{t=1}^{T} \sum_{i=1}^{d} \tilde{x}_{i+1} - \tilde{x}_t$ and the optimum point variation $\sum_{t=1}^{T} \sum_{i=1}^{d} x_{i+1} - x_t^\star$, the loss $f_t(x_t) - f_t(x_t^\star)$, and the dynamic regret $\text{Reg}_f$. 

<table>
<thead>
<tr>
<th>$J_t$</th>
<th>$\text{Reg}_f^2$</th>
<th>Assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>[15] C</td>
<td>$O(\sqrt{T}(1+C_T))$</td>
<td>$|f_t| \leq \beta$, $\mathcal{X}$ compact</td>
</tr>
<tr>
<td>[16] SC</td>
<td>$O(1+C_T)$</td>
<td>$|f_t| \leq \beta$</td>
</tr>
<tr>
<td><strong>This work</strong> Elastic-net (SC)</td>
<td>$O(1+C_T)$</td>
<td>For ADMM: $|x_1| \leq \beta$</td>
</tr>
</tbody>
</table>