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 estimation [1], [2], [3], [4], [5]. The dynamic environment is more natural in a number of applications, e.g., magnetic resonance imaging estimation [1], [2], [3], [4], [5]. The dynamic environment is more natural in a number of applications, e.g., magnetic resonance imaging estimation [1], [2], [3], [4], [5]. The dynamic environment is more natural in a number of applications, e.g., magnetic resonance imaging estimation [1], [2], [3], [4], [5].

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 \( t \in \{1, 2, \ldots, T\} \), a learner incurs in a convex cost functional \( f_t \) revealed by an adversary. Then, the learner aims at minimizing \( f_t \), which may not be 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]:

\[
Reg^d_t(x_1^t, \ldots, x_T^t) := \sum_{t=1}^{T} f_t(x_t) - f_t(x_t^\*)
\]

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 \( Reg^d_t(x_1^t, \ldots, x_T^t) = O(\sqrt{T}(1+C_T)) \). More recently, in [16] this result has been improved to \( Reg^d_t(x_1^t, \ldots, x_T^t) = O(1+C_T) \), 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 \( \tilde{x}_t \in \mathbb{R}^N \), \( t \in \{1, 2, \ldots, T\} \), be the sparse signal to be tracked. According to the CS paradigm, we acquire compressed measurements \( y_t = \Lambda \tilde{x}_t \), where \( \Lambda \in \mathbb{R}^{M \times N} \) 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:

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

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 \( Reg^d_T(x_1^t, \ldots, x_T^t) = O(1 + C_T) \).

**Theorem 2.** If \( \| \tilde{x}_t \| \leq \beta \) for some \( \beta > 0 \), the online ADMM for dynamic Elastic-net has \( Reg^d_T(x_1^t, \ldots, x_T^t) = O(1 + C_T) \).

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: \( \tilde{x}_{\Omega^t} \), \( t = \eta_{t,0} + \frac{\gamma}{\tau} t \), where \( \eta_{t,i} \sim N(0, 1), i \in \{1, \ldots, N\}, j \in \{0, 1, \ldots, T\} \);

- **(M2)** support \( \Omega_t \) chosen uniformly at random at each \( t \), with \( |\Omega_t| = k \) constant; non-zero values: \( \tilde{x}_{\Omega_{t_0}^t} \), \( t = \eta_{t,0} + \frac{\gamma}{\tau} t \), where \( \eta_{t,i} \sim N(0, 1), i \in \{1, \ldots, N\}, j \in \{0, 1, \ldots, T\} \).

In Figure 1, we see that, for both algorithms, \( Reg^d_T \) 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

<table>
<thead>
<tr>
<th>Assumptions</th>
<th>$J_t$</th>
<th>$\text{Reg}^2_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{SC}$</td>
<td>$O(\sqrt{T(1+C_T)})$</td>
<td>$|f_t| \leq \beta$, $\mathcal{X}$ compact</td>
</tr>
<tr>
<td>$\text{Elastic-net (SC)}$</td>
<td>$O(1+C_T)$</td>
<td>For ADMM: $|x_t| \leq \beta$</td>
</tr>
</tbody>
</table>

Table II

References


