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 $t \in \{1, 2, \dots, T\}$, 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]:

$$Reg_T^d(x_1^\star, \dots, x_T^\star) := \sum_{t=1}^{T} f_t(x_t) - f_t(x_t^\star)$$

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

In the OCO literature, the action typically is a gradient descent step [15]. Let $C_T := \sum_{t=2}^{T} \|x_t^\star - x_{t-1}^\star\|_2$. In [15], it has been proved that $Reg_T^d(x_1^\star, \dots, x_T^\star) = O(\sqrt{T}(1 + C_T))$. More recently, in [16] this result has been improved to $Reg_T^d(x_1^\star, \dots, x_T^\star) = O(1 + C_T)$, under the hypothesis that f_t is strongly convex, and assuming that ∇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 $\widetilde{x}_t \in \mathbb{R}^N$, $t \in \{1, 2, \dots, T\}$, be the sparse signal to be tracked. According to the CS paradigm, we acquire compressed measurements $y_t = A\widetilde{x}_t$, where $A \in \mathbb{R}^{M,N}$ 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:

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

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 $\widetilde{C}_T := \sum_{t=2}^{T} \|\widetilde{x}_{t-1} - \widetilde{x}_t\|_2$. Our main results are summarized in the following theorems (whose proofs are omitted for brevity):

Theorem 1. *If $\tau\|A\|_2^2 < 1$ (where τ is the gradient parameter, see Table I), the online IST for dynamic Elastic-net has $Reg_T^d(x_1^\star, \dots, x_T^\star) = O(1 + \widetilde{C}_T)$.*

Theorem 2. *If $\|\widetilde{x}_t\| \le \beta$ for some $\beta > 0$, the online ADMM for dynamic Elastic-net has $Reg_T^d(x_1^\star, \dots, x_T^\star) = O(1 + \widetilde{C}_T)$.*

These theorems state in particular that the regret stabilizes when the signal $\widetilde{x}_t$ 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 $\widetilde{x}(t) \in \mathbb{R}^N$:

(M1) constant support $\Omega \subset \{1, \dots, N\}$ (chosen uniformly at random), $|\Omega| = k$, non-zero values: $\left(\widetilde{x}|_\Omega\right)_i(t) = \eta_{i,0} + \frac{\eta_{i,t}}{t}$, where $\eta_{i,j} \sim \mathcal{N}(0,1)$, $i \in \{1, \dots, N\}$, $j \in \{0, 1, \dots, T\}$;

(M2) support Ω_t chosen uniformly at random at each t, with $|\Omega_t| = k$ constant; non-zero values: $\left(x|_{\Omega_t}\right)_i(t) = \left(x|_{\Omega_t}\right)_i(t-1) + \frac{\eta_{i,t}}{t}$ where $\eta_{i,t} \sim \mathcal{N}(0,1)$, $i \in \{1, \dots, N\}$, $t \in \{1, \dots, T\}$.

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

This work was supported by the European Research Council under FP7 / ERC, Grant agreement n.279848 - CRISP project, by the Spanish Government through the project INTENSYV (TEC2013-44591-P) and by the Catalan Government (2014 SGR 1567).

Table I
PROPOSED ALGORITHMS

Notation and initialization

$\mathbb{S}_\alpha$ denotes the soft thresholding operator with parameter α, defined as follows: given $x \in \mathbb{R}$, if $|x| \leq \alpha$, $\mathbb{S}_\alpha(x) := 0$; otherwise, $\mathbb{S}_\alpha(x) := x - \text{sgn}(x)\alpha$. The definition is extended component-wise to vectors.

$\tau > 0$ is the gradient parameter of IST; it must be chosen such that $\tau \|A\|_2^2 < 1$ (see Theorem 1). μ and τ are defined in (1).

For any $t \in \{1, \dots, T\}$, the measurement $y_t = A\widetilde{x}_t$ (hence, f_t) is revealed after that action x_t is played.

$B = \left[A^T A + (1+\mu)I\right]^{-1}; \quad x_1 = z_1 = u_1 = 0 \in \mathbb{R}^N.$

Online IST for dynamic Elastic-net

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

1. $x_t = \mathbb{S}_{\frac{\lambda\tau}{\mu\tau+1}}\left[\frac{x_{t-1}+\tau A^T(y_{t-1}-Ax_{t-1})}{1+\mu\tau}\right]$
2. read y_t

Online ADMM for dynamic Elastic-net

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

1. $x_t = B\left[A^T y_{t-1} + z_{t-1} - u_{t-1}\right]$
2. $z_t = \mathbb{S}_\lambda\left[x_t + u_{t-1}\right]$
3. $u_t = u_{t-1} + x_t - z_t$
4. read y_t

Table II
DYNAMIC REGRET ANALYSIS (C = CONVEX; SC = STRONGLY CONVEX)

	f_t	Reg_T^d	Assumptions
[15]	C	$O(\sqrt{T}(1+C_T))$	$\|\nabla f_t\| \leq \beta$, $\mathcal{X}$ compact
[16]	SC	$O(1+C_T)$	$\|\nabla f_t\| \leq \beta$
This work	Elastic-net (SC)	$O(1+\widetilde{C}_T)$	For ADMM: $\|\widetilde{x}_t\| \leq \beta$

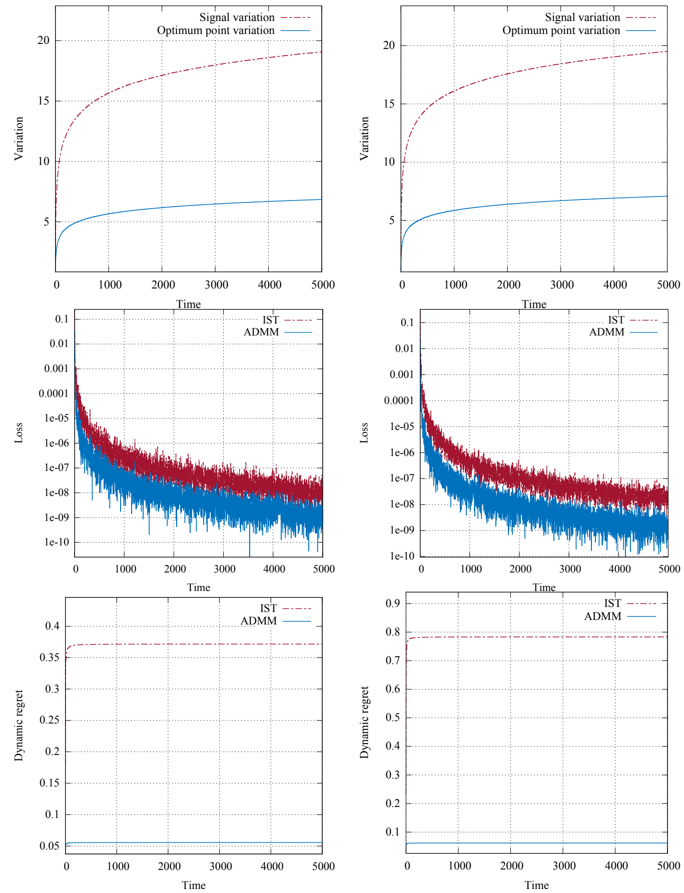

Figure 1. Simulations on models M1 (left column) and M2 (right column); $N = 100$, $M = 20$, $k = 5$. The graphs represent (from top to down) the signal variation $\sum_{t=1}^T \widetilde{x}_{t+1} - \widetilde{x}_t$ and the optimum point variation = $\sum_{t=1}^T x_{t+1}^\star - x_t^\star$, the loss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

REFERENCES

[1] J. Ziniel, L. C. Potter, and P. Schniter, "Tracking and smoothing of time-varying sparse signals via approximate belief propagation," in *Asilomar Conf. Signals Syst. Comp.*, 2010, pp. 808–812.

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

[3] E. Karseras, K. Leung, and W. Dai, "Tracking dynamic sparse signals using hierarchical bayesian kalman filters," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2013, pp. 6546–6550.

[4] A. Balavoine, C. J. Rozell, and J. Romberg, "Discrete and continuous-time soft-thresholding for dynamic signal recovery," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3165–3176, 2015.

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

[6] A. Majumdar, R. K. Ward, and T. Aboulnasr, "Compressed sensing based real-time dynamic mri reconstruction," *IEEE Trans. Med. Imaging*, vol. 31, no. 12, pp. 2253–2266, 2012.

[7] S. G. Lingala and M. Jacob, "Blind compressive sensing dynamic MRI," *IEEE Trans. Med. Imaging*, vol. 32, no. 6, pp. 1132–1145, 2013.

[8] W. Yin, Z. Wen, S. Li, J. Meng, and Z. Han, "Dynamic compressive spectrum sensing for cognitive radio networks," in *Conf. Inf. Sci. Syst. (CISS)*, 2011.

[9] N. Michelusi and U. Mitra, "Cross-layer estimation and control for cognitive radio: Exploiting sparse network dynamics," *IEEE Trans. Cogn. Commun. Netw.*, vol. 1, no. 1, pp. 128–145, 2015.

[10] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, 2006.

[11] A. Y. Carmi, L. Mihaylova, and S. J. Godsill, *Compressed sensing & sparse filtering*. Berlin Heidelberg: Springer, 2014.

[12] D. Zachariah, S. Chatterjee, and M. Jansson, "Dynamic iterative pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4967–4972, 2012.

[13] J. Mota, N. Deligiannis, A. C. Sankaranarayanan, V. Cevher, and M. Rodrigues, "Dynamic sparse state estimation using $\ell_1 - \ell_1$ minimization: Adaptive-rate measurement bounds, algorithms and applications," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2015, pp. 3332–3336.

[14] S. Shalev-Shwartz, "Online learning and online convex optimization," *Found. Trends Machine Learn.*, vol. 4, no. 2, pp. 107–194, 2011.

[15] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *Proc. Int. Conf. Machine Learning (ICML)*. Carnegie Mellon University, 2003.

[16] A. Mokhtari, S. Shahrampour, A. Jadbabaie, and A. Ribeiro, "Online optimization in dynamic environments: Improved regret rates for strongly convex problems," in *Proc. IEEE Conf. Decision and Control (CDC)*, 2016, pp. 7195–7201.

[17] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[18] V. Umanità and S. Villa, "Elastic-net regularization: iterative algorithms and asymptotic behavior of solutions," *Numerical functional analysis and optimization*, vol. 31, no. 12, pp. 1406–1432, 2010.

[19] M. Fornasier, "Numerical methods for sparse recovery," in *Theoretical Foundations and Numerical Methods for Sparse Recovery*. Radon Series Comp. Appl. Math., de Gruyter, 2010, pp. 93–200.

[20] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Machine Learn.*, vol. 3, no. 1, pp. 1–122, 2010.