MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

A Short Note on Improved ROSETA

Mansour, H.

TR2017-154 October 2017

Abstract

This note presents a more efficient formulation of the robust online subspace estimation and tracking algorithm (ROSETA).

arXiv

A Short Note on Improved ROSETA

Hassan Mansour
Mitsubishi Electric Research Laboratories
Cambridge, MA 02139
mansour@merl.com

October 12, 2017

1 Problem Formulation

We consider the problem of identifying at every time t an r-dimensional subspace $\mathcal{U}_t$ in $\mathbb{R}^n$ with $r \ll n$ that is spanned by the columns of a rank-r matrix $U_t \in \mathbb{R}^{n\times r}$ from incomplete and noisy measurements

$$b_t = \Omega_t(U_t a_t + s_t), \tag{1}$$

where Ω_t is a selection operator that specifies the observable subset of entries at time t, $a_t \in \mathbb{R}^r$ are the coefficients specifying the linear combination of the columns of U_t, and $s_t \in \mathbb{R}^n$ is a sparse outlier vector.

When the subspace $\mathcal{U}_t$ is stationary, we drop the subscript t from U_t and the problem reduces to robust matrix completion or robust principal component analysis where the task is to separate a matrix $B \in \mathbb{R}^{n\times m}$ into a low rank component UA and a sparse component S using incomplete observations

$$B_\Omega = \Omega(UA + S).$$

Here the columns of the matrices A and S are respectively the vectors a_t and s_t stacked horizontally for all $t \in \{1 \dots m\}$, and the operator Ω specifies the observable entries for the entire matrix B.

2 Improved Robust Online Subspace Estimation and Tracking

We describe in this section an improved version of the robust online subspace estimation and tracking algorithm (ROSETA) first published in [1].

We consider the combined loss function $\mathcal{L}(U_t, s_t, a_t, e_t)$ that includes an ℓ_2 data misfit term and a one-norm regularizer applied to the outliers s_t

$$\mathcal{L}(U_t, s_t, a_t, e_t) = \frac{\mu}{2}\|b_t - (U_t a_t + s_t + e_t)\|_2^2 + \lambda\|s_t\|_1 \tag{2}$$

where e_t is supported on the complement of Ω_t, hereby denoted Ω_t^c, such that $\Omega_t(e_t) = 0$ and $\Omega_t^c(e_t) = -\Omega_t^c(U_t a_t)$.

2.1 Update the subspace coefficients

Following the ROSETA framework, we first fix the subspace matrix U to be the previous estimate U_{t-1} and update the triplet (s_t, a_t, e_t) by solving the LASSO problem

$$(s_t, a_t, e_t) = \arg \min_{s,a,e,y} \mathcal{L}(U_{t-1}, s, a, e). \tag{3}$$

The solution to (3) is obtained by applying the following sequence of updates:

$$\begin{array}{rcl} a_t^k & = & U_{t-1}^{\dagger}\left(b_t - s_t^{k-1} - e_t^{k-1}\right) \\ e_t^k & = & -\Omega_t^c\left(U_{t-1}a_t^k\right) \\ s_t^k & = & \mathcal{S}_{\lambda}\left(b_t - U_{t-1}a_t^k - e_t^k\right) \end{array} \tag{4}$$

where $\mathcal{S}_\tau(x) = \text{sign}(x)\cdot\max\{|x|-\tau, 0\}$ denotes the element-wise soft thresholding operator with threshold τ, k indicates the iteration number, and † is the Moore-Penrose pseudo-inverse of a matrix.

2.2 Update the subspace matrix

The subspace matrix U_t is then updated by minimizing with respect to U the quadratic approximation $Q(U)$ of $\mathcal{L}(U, s_t, a_t, e_t)$ around the estimate U_{t-1}, i.e. we define the quadratic function

$$Q(U) := \mathcal{L}(U_{t-1}) + \nabla\mathcal{L}(U_{t-1})^T(U - U_{t-1}) + \frac{\mu}{2}\text{Tr}\left[(U - U_{t-1})(I_r + a_t a_t^T)(U - U_{t-1})^T\right], \tag{5}$$

where μ is the step-size parameter, and $\text{Tr}[\cdot]$ is the trace function. Consequently, the update equation of the subspace matrix U is given by

$$\begin{aligned} U_t &= \arg\min_U Q(U) \\ &= U_{t-1} + \tfrac{1}{\mu}(b_t - (U_{t-1}a_t + s_t + e_t))a_t^T(I_r + a_t a_t^T)^{-1} \end{aligned} \tag{6}$$

2.3 Adaptive parameter selection

Inspired by the adaptive step size selection in GRASTA [2], we developed a corresponding adaptive step-size parameter for ROSETA.

The parameter μ_t controls the speed of convergence of the subspace estimate. In particular, a smaller value of μ allows for faster adaptability of U_t to a changing subspace (larger descent step), whereas a larger value of μ only permits a small variation in U_t. Consider the descent direction

$$D_t = (b_t - (Ua_t + s_t + e_t))a_t^T(1 + a_t^T a_t)^{-1}. \tag{7}$$

The parameter μ_t can then be updated according to

$$\mu_t = \frac{C}{1 + \eta_t}, \tag{8}$$

where $\eta_t = \min\{\eta_{\max}, \max\{C, \eta_{t-1} + \text{sigmoid}\left(\frac{\langle D_{t-1}, D_t\rangle}{\|D_{t-1}\|_F \|D_t\|_F}\right)\}\}$, where $\eta_{\max}$ is a control parameter. Here $\text{sigmoid}(x) = f + 2f/(1+e^{10x})$, for some predefined f.

Similar to GRASTA, the intuition behind choosing such an update rule comes from the idea that if two consecutive subspace updates D_{t-1} and D_t have the same direction, i.e. $\langle D_{t-1}, D_t\rangle > 0$, then the target subspace is still far from the current subspace estimate. Consequently, the new μ_t should be smaller to allow for fast adaptability which is achieved by increasing η_t. Similarly, when $\langle D_{t-1}, D_t\rangle < 0$, the subspace update seems to bounce around the target subspace and hence a larger μ_t is needed. The new ROSETA algorithm is summarized in Algorithm 1.

Algorithm 1 Robust Subspace Estimation and Tracking

Input Sequence of measurements $\{b_t\}$, η_{LOW}, η_{HIGH}
Output Sequences $\{U_t\}$, $\{a_t\}$, $\{s_t\}$
Initialize U_0, μ_0, η_0, $\eta_{\max}$
for $t = 1 \dots N$ **do**
  Solve for a_t, s_t, and e_t:
  while not converged **do**
    $a_t^k = U_{t-1}^{\dagger}\left(b_t - s_t^{k-1} - e_t^{k-1}\right)$
    $e_t^k = -\Omega_t^c\left(U_{t-1}a_t^k\right)$
    $s_t^k = \mathcal{S}_\lambda\left(b_t - U_{t-1}a_t^k - e_t^k\right)$
  end while
  Update subspace estimate:
    $D_t = (b_t - (Ua_t + s_t + e_t))a_t^T(1 + a_t^T a_t)^{-1}$
    $U_t = U_{t-1} + \frac{1}{\mu}D_t$
  Update parameter μ_t:
    $\eta_t = \min\{\eta_{\max}, \max\{C, \eta_{t-1} + \text{sigmoid}\left(\frac{\langle D_{t-1}, D_t\rangle}{\|D_{t-1}\|_F \|D_t\|_F}\right)\}\}$
    $\mu_t = \frac{C}{1+\eta_t}$
end for

References

[1] H. Mansour and X. Jiang, "A robust online subspace estimation and tracking algorithm," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 4065–4069.

[2] J. He, L. Balzano, and J. C. S. Lui, "Online robust subspace tracking from partial information," *preprint, http://arxiv.org/abs/1109.3827*, 2011.