Matrix completion

May 17, 2012

1 Problem setting

We study a generalization of the matrix completion problem

$$y = \mathcal{A}(\Theta^* + \Gamma^*) + w,$$

where $\Theta^*, \Gamma^* \in \mathbb{R}^{d_1 \times d_2}$ are unknown, and $y, w \in \mathbb{R}^n$ are respectively the observed vector and a random noise.

$$\mathcal{A} : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}^n, \quad \mathcal{A}(V)_k = V_{i_k j_k}$$

is an observation operator, with $\{i_1, \ldots, i_n\} \subseteq \{1, \ldots, d_1\}$ and $\{j_1, \ldots, j_n\} \subseteq \{1, \ldots, d_2\}$. We will assume that Θ^* is low rank and Γ^* is sparse and spiky. The goal is to recover accurate estimates of Θ^* and Γ^*.

Following [1], the proposed estimator is a minimizer of the regularized least squares problem

$$\begin{aligned} &\min_{\Theta,\Gamma} \left[\frac{1}{2} \left\| y - \mathcal{A}(\Theta + \Gamma) \right\|^2 + \tau \left\| \Theta \right\|_N + \mu \left\| \Gamma \right\|_1 \right] \\ &\text{s. t. } \left\| \Theta \right\|_\infty \leq \alpha \end{aligned} \tag{1}$$

2 A proximal alghorithm

To solve problem (1) we use an accelerated proximal method, FISTA algorithm [2]. We use the notations

$$I_B(\Theta) = \begin{cases} 0 & \text{if } \left\| \Theta \right\|_\infty \leq \alpha \\ +\infty & \text{otherwise} \end{cases}$$

$$\begin{aligned} \mathcal{E}(\Theta, \Gamma) &= \frac{1}{2} \left\| y - \mathcal{A}(\Theta + \Gamma) \right\|^2 + \lambda \left\| \Theta \right\|_N + \mu \left\| \Gamma \right\|_1 + I_B(\Theta), \quad \mathcal{E} : \mathbb{R}^{d_1 \times d_2} \times \mathbb{R}^{d_1 \times d_2} \to \mathbb{R} \\ F(\Theta, \Gamma) &= \frac{1}{2} \left\| y - \mathcal{A}(\Theta + \Gamma) \right\|^2, \quad \left\| \Theta \right\|_N + I_B(\Theta) = G(\Theta), \quad \left\| \Gamma \right\|_1 = H(\Gamma), \end{aligned}$$

and we compute all the necessary ingredients. First note that

$$\mathcal{A} : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}^n \Rightarrow \mathcal{A}^* : \mathbb{R}^n \to \mathbb{R}^{d_1 \times d_2},$$

where $\mathbb{R}^{d_1 \times d_2}$ is endowed with the Frobenious inner product (denoted by $\langle \cdot, \cdot \rangle_F$). Given $z \in \mathbb{R}^n$, defining

$$Z_{ij} = \begin{cases} z_k & \text{if } (i,j) = (i_k, j_k) \\ 0 & \text{otherwise} \end{cases}$$

for $V \in \mathbb{R}^{d_1 \times d_2}$, we have

$$\langle V, Z \rangle_F = \sum_{k=1}^{n} V_{i_k j_k} z_k = \langle \mathcal{A}(V), z \rangle,$$

therefore $Z = \mathcal{A}^*(z)$.

$$\nabla F(\Theta, \Gamma) =: (\nabla_\Theta F(\Theta, \Gamma), \nabla_\Gamma F(\Theta, \Gamma)) \in \mathbb{R}^{d_1 \times d_2} \times \mathbb{R}^{d_1 \times d_2},$$

with

$$\nabla_\Theta F(\Theta, \Gamma) = \nabla_\Gamma F(\Theta, \Gamma) = \mathcal{A}^*(\mathcal{A}(\Theta + \Gamma) - y).$$

Computation of the Lipschitz constant. On $\mathbb{R}^{d_1 \times d_2} \times \mathbb{R}^{d_1 \times d_2}$ we consider the inner product:

$$\langle (\Theta, \Gamma), (V, Z) \rangle = \langle \Theta, V \rangle_F + \langle \Gamma, Z \rangle_F$$

$$\begin{aligned}
\|\nabla F(\Theta, \Gamma) - \nabla F(\Theta', \Gamma')\|^2 &= \|\nabla_\Theta F(\Theta, \Gamma) - \nabla_\Theta F(\Theta', \Gamma')\|^2 + \|\nabla F_\Gamma(\Theta, \Gamma) - \nabla_\Gamma F(\Theta', \Gamma')\|^2 \\
&= 2 \|\mathcal{A}^*(\mathcal{A}(\Theta + \Gamma) - y) - \mathcal{A}^*(\mathcal{A}(\Theta' + \Gamma') - y)\|^2 \\
&\leq 2 \|\mathcal{A}^*\mathcal{A}(\Theta + \Gamma - \Theta' + \Gamma')\|^2 \\
&\leq 2 \|\mathcal{A}^*\mathcal{A}\|_{\text{op}}^2 \|\Theta - \Theta' + \Gamma - \Gamma'\|_F^2 \\
&\leq 2 \|\mathcal{A}^*\mathcal{A}\|_{\text{op}}^2 (\|\Theta - \Theta'\|_F + \|\Gamma - \Gamma'\|_F)^2 \\
&\leq 4 \|\mathcal{A}^*\mathcal{A}\|_{\text{op}}^2 \|(\Theta, \Gamma) - (\Theta', \Gamma')\|_F^2
\end{aligned}$$

From the last inequality it follows that $L = 2 \|\mathcal{A}^*\mathcal{A}\|_{\text{op}}$. Next, we estimate $\|\mathcal{A}^*\mathcal{A}\|_{\text{op}}$.

$$\|\mathcal{A}^*\mathcal{A}\|_{\text{op}}^2 = \sup_{\|V\| \leq 1} \|\mathcal{A}^*\mathcal{A}(V)\|^2 = \sup_{\|V\| \leq 1} \sum_k V_{i_k j_k}^2 = 1.$$

Then $L = 2$. By definition

$$\begin{aligned}
\text{prox}_{\lambda(\tau G + \mu H)}(\Theta, \Gamma) &= \operatorname*{argmin}_{\Theta', \Gamma'} \left[\tau G(\Theta') + \mu H(\Gamma') + \frac{1}{2\lambda} \|(\Theta, \Gamma) - (\Theta', \Gamma')\|^2 \right] \\
&= (\operatorname*{argmin}_{\Theta'} \left[\tau G(\Theta') + \frac{1}{2\lambda} \|\Theta - \Theta'\|_F^2 \right], \operatorname*{argmin}_{\Gamma'} \left[\mu H(\Gamma') + \frac{1}{2\lambda} \|\Gamma - \Gamma'\|^2 \right]) \\
&= (\text{prox}_{\lambda\tau G}(\Theta), \text{prox}_{\lambda\mu H}(\Gamma))
\end{aligned}$$

It is well-known that

$$\text{prox}_{\lambda\mu H}(\Gamma) = \mathcal{S}_{\lambda\mu}(\Gamma)$$

where $\mathcal{S}_{\lambda\mu}$ is the soft-thresholding operator, acting component-wise as

$$\mathcal{S}_{\lambda\mu}(\Gamma)_{ij} = s_{\lambda\mu}(\Gamma_{ij}) = \begin{cases} \Gamma_{ij} - \lambda\mu & \text{if } \Gamma_{ij} > \lambda\mu \\ 0 & \text{if } |\Gamma_{ij}| \leq \lambda\mu \\ \Gamma_{ij} + \lambda\mu & \text{if } \Gamma_{ij} < -\lambda\mu \end{cases}$$

$\text{prox}_{\lambda\mu H}(\Gamma)$ cannot be computed in closed form. We use the iterative algorithm presented in [4].

$$\text{prox}_{\lambda\mu\|\cdot\|_N}(Z) = U \mathcal{S}_{\lambda\mu}(D) V^T,$$

and $Z = UDV^T$ is the singular value decomposition of Z. On the other hand note that the Fenchel conjugate of i_B is the function $f(\Theta) = \alpha \|\Theta\|_1$. Applying Algorithm 3.5 in [4] we get

$$\left|\begin{array}{l} W_0 \in \mathbb{R}^{d_1 \times d_2},\ \beta \in]0,1[\quad \text{(initialization)} \\ \gamma_n \in [\beta, 2-\beta] \\ W_{n+1} = \mathcal{S}_{\alpha\gamma_n}(W_n + \gamma_n U_n \mathcal{S}_{\lambda\mu}(D_n) V_n^T), \qquad U_n D_n V_n^T = Z - W_n \end{array}\right. \tag{2}$$

The sequence $Z - W_n$ converges to $\text{prox}_{\lambda\mu G}(Z)$.

Alternatively we can use Algorithm 1.5.2 in [3].

$$\left|\begin{array}{l} T_0 = Z, P_0 = Q_0 = 0, \quad \text{(initialization)} \\ T_n = \text{Pr}_B(W_n + P_n), \\ P_{n+1} = X_n + P_n - T_n, \\ W_{n+1} = \mathcal{S}_{\lambda\mu}(T_n + Q_n) \\ Q_{n+1} = T_n + Q_n - T - W_{n+1} \end{array}\right. \tag{3}$$

The sequence W_n converges to $\text{prox}_{\lambda\mu G}(Z)$.

FISTA (using the (2) [or (3)] as internal algorithm) reads as

$$\left|\begin{array}{l} (\Theta_0, \Gamma_0) \in \mathbb{R}^{d_1 \times d_2} \times \mathbb{R}^{d_1 \times d_2}, L = 2 \\ \Theta_k = W_k - \frac{1}{L}\mathcal{A}^*(\mathcal{A}Z_k - y), \quad W_k = W_{n_k} \text{ computed via alg. (2) with } Z = Z_k - \frac{1}{L}\mathcal{A}^*(\mathcal{A}Z_k - y) \\ [or \\ \Theta_k = W_k, \quad W_k = W_{n_k} \text{ computed via alg. (3) with } Z = Z_k - \frac{1}{L}\mathcal{A}^*(\mathcal{A}Z_k - y)] \\ \Gamma_k = \mathcal{S}_{\mu/L}(C_k - \frac{1}{L}\mathcal{A}^*(\mathcal{A}C_k - y)) \\ t_{k+1} = (1 + \sqrt{1 + 4t_k^2})/2 \\ Z_{k+1} = \Theta_k + \left(\frac{t_k - 1}{t_{k+1}}\right)(\Theta_k - \Theta_{k-1}) \\ C_{k+1} = \Gamma_k + \left(\frac{t_k - 1}{t_{k+1}}\right)(\Gamma_k - \Gamma_{k-1}) \end{array}\right. \tag{4}$$

References

[1] A. Agarwal, S. Negahban, and M. Wainwright. Noisy matrix decomposition via convex relaxation:optimal rates in high dimensions. Technical report, Department of Statistics, UC Berkeley, 2012.

[2] A. Beck and Teboulle. M. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(11):2419–2434, 2009.

[3] P. Combettes and J.-C. Pesquet. *Proximal Splitting Methods in Signal Processing*. Springer-Verlag, Dec 2010.

[4] P. L. Combettes, D. Dũng, and B. C. Vũ. Dualization of signal recovery problems. *Set-Valued Var. Anal.*, 18(3-4):373–404, 2010.