Matrix Completion

\[
\min \frac{1}{2} \|P_X(X - O)\|_F^2 + \lambda \|X\|_1,
\]

where \([P_X(A)]_{ij} = A_{ij}\) if \(\Omega_{ij} = 1\), and 0 otherwise; and \(\|X\|_1\) is the nuclear norm of \(X\). An important application is recommendation systems.

Proximal Gradient Descent

\[
\min_{x,y} f(x) + \lambda g(x),\quad \text{where (i) both } f(\cdot) \text{ and } g(\cdot) \text{ are smooth; (ii) } g(\cdot) \text{ can be nonsmooth. Let } z_t = x_t - \nabla f(x_t), \text{ it generates } z_{t+1} \text{ as }
\]

\[
\arg \min_{z} \left[ \left( \frac{1}{2} \|x - z\|_2^2 + \frac{\lambda}{2} \|x - z\|_2^2 \right) + g(z) \right] = \arg \min_{z} \frac{1}{2} \|z - y_t\|_2^2 + \frac{\lambda}{2} \|z - y_t\|_2^2,
\]

- this is called the proximal step, often with closed-form solutions.
- convergence rate is \(O(1/T)\), but can be accelerated to \(O(1/T^2)\) by using

\[
y_t = (1 + \theta_t) x_t - \theta_t x_{t-1}, \quad z_t = y_t - \nabla f(y_t).
\]

Proposed Algorithm (AIS-Impute)

**Contributions**
- efficient acceleration is possible with sparse + low-rank structure;
- speeding up approximate SVT using power method.

**Fast convergence + Low iteration complexity**

\[
Z_t = \frac{P_t(O - Y_t)}{1 + \theta_t} + \frac{1}{2} X_t - \theta_t X_{t-1}.
\]

Similar to (1), for any \(u \in \mathbb{R}^n\), \(\mathbb{A} u\) can be performed as

\[
P_t(O - Y_t)u + \left( \frac{1}{2} U_t \Sigma_t V_t^\top u - \theta_t U_{t-1} \Sigma_{t-1} V_{t-1}^\top u \right) \Omega(t) \equiv \frac{1}{2} U_t \Sigma_t V_t^\top u + \theta_t U_{t-1} \Sigma_{t-1} V_{t-1}^\top u.
\]

- again, takes \(O(\|\Omega\|_{1+k} + (m+n)k^2)\) time, but convergence rate is improved to \(O(1/T^2)\).

**Approximate SVT**


**Require**: \(Z \in \mathbb{R}^{m \times n}\), initial \(R \in \mathbb{R}^{m \times n}\) for warm-start, tolerance \(\epsilon\);

1. Initialize \(Y_1 \leftarrow Z\);
2. for \(t = 1, 2, \ldots \) do
3. \(Q_{t+1} = QR(Y_t)\)
4. \(\hat{Z}_t = Y_t - Q_{t+1} R_{t+1}\)
5. \(Y_{t+1} = Y_t - Q_{t+1} R_{t+1}\)
6. end for
7. return \(Q_{t+1}\).

**Convergence Analysis**

Approximation quality of approximate SVT:

**Proposition 1** Let \(h_{\|\cdot\|_F}(X; \hat{Z}) \equiv \frac{1}{2} \|X - \hat{Z}\|_F^2 + \lambda \|X\|_1\). Assume \(k \geq \hat{k}, \hat{k} \geq \alpha n_k \sqrt{T + \tau}\). Then,

\[
h_{\|\cdot\|_F}(X; \hat{Z}) \leq h_{\|\cdot\|_F}(X_t; \hat{Z}_t) \leq h_{\|\cdot\|_F}(X_t; Z_t) + \frac{\alpha n_k}{\sqrt{T + \tau}}.
\]

**Theorem 3** (Convergence of AIS-Impute) Assume the conditions in Proposition 2, Algorithm 3 converges to the optimal with a rate of \(O(1/T^2)\).

**Soft-Impute [Mazumder et al., 2010]**

At iteration \(t\), it generates \(X_{t+1} = \text{SVT}(\hat{Z}_t)\) where

\[
\hat{Z}_t = P_{t}(O) + P_{t}^{\top}(X_t) = P_{t}(O - X_t) + X_t.
\]

'Sparse + Low-Rank' structure in \(Z_t\):
- let SVD of \(X_t\) be \(U_t \Sigma_t V_t^\top\),
- for any \(u \in \mathbb{R}^n\), \(\mathbb{A} u\) (same for \(Z_t^\top v\)) can be computed as

\[
Z_t u = \frac{P_t(O - X_t) u + U_t \Sigma_t V_t^\top u}{O(t) \Omega(t)} = \frac{1}{2} U_t \Sigma_t V_t^\top u + \theta_t U_{t-1} \Sigma_{t-1} V_{t-1}^\top u,
\]
- a rank-1 SVD on \(Z_t\) is \(\mathbb{R}^{m \times n}\) takes \(O(\|\Omega\| k + (m+n)k^2)\) instead of \(O(\text{rank}^2)\) time,
- \(Z_t X_t = X_t - \nabla f(X_t) = P_t(X_t) + P_{t}^{\top}(O)\), and so

Soft-Impute = Proximal Gradients
- acceleration is possible, but previous results suggested it is useless as the sparse + low-rank structure in \(Z_t\) is lost.

Proposed AIS-Impute can be downloaded at https://github.com/quanningyao/AIS-impute.

**MOVIELENS DATA**

Besides proximal algorithms, we also compare with (i) active subspace selection (‘active’) [Hsieh and Olsen, 2014]; (ii) Frank-Wolfe algorithm (‘boost’) [Zhang et al., 2012]; (iii) a recent variant of Soft-Impute [Hastie et al., 2014]; and (iv) a trust-region method (‘TR’) [Mishra et al., 2013].

All algorithms are equally good at recovering the missing entries, but AIS-Impute is the fastest.