Appendix: Convolutional Imputation of Matrix Networks

Qingyun Sun*1 Mengyuan Yan*2 David Donoho3 Stephen Boyd2

Exact recovery guarantee

**Theorem 1.** We assume that $A$ is a matrix network on a graph $G$, and its graph Fourier transform $\hat{A}(k)$ are a sequence of matrices, each of them is at most rank $r$, and $\hat{A}$ satisfy the incoherence condition with coherence $\mu$. And we observe a matrix network $A^\Omega$ on the graph $G$, for a subset of node in $\Omega$ random sampled from the network, node $i$ on the network is sampled with probability $p_i$, we define the average sampling rate $p = \frac{1}{N} \sum_{i=1}^{N} p_i = |\Omega|/(Nn^2)$, and define $\mathcal{R} = \frac{1}{p} P_t d_t^\star$.

Then we prove that for any sampling probability distribution $\{p_i\}$, as long as the average sampling rate $p > C n^2 \log^2(N)$ for some constants $C$, the solution to the optimization problem

$$\begin{align*}
\text{minimize} & \quad \|\hat{M}\|_{*,1}, \\
\text{subject to} & \quad A^\Omega = \mathcal{R} \hat{M}
\end{align*}$$

is unique and is exactly $\hat{A}$ with probability $1 - (Nn)^{-\gamma}$, where $\gamma = \frac{\log(Nn)^2}{10}$.

**Proof.** We define a inner product: $\langle \hat{M}_1, \hat{M}_2 \rangle = \sum_k \langle \hat{M}_1(k), \hat{M}_2(k) \rangle$. Then we have the following two inequalities

$$\|\hat{M}(k)\|_* = \text{Tr}(\text{sgn}(\hat{M}(k))\hat{M}(k)) = \langle \text{sgn}(\hat{M}(k)), \hat{M}(k) \rangle.$$ 

Therefore,

$$\|\hat{M}\|_{*,1} = \langle \text{sgn}(\hat{M}), \hat{M} \rangle.$$

Here $\text{sgn}(\hat{M}) = V_1 V_2^\star$ is the sign matrix of the singular values of $\hat{M}$ under the singular vector basis.

We consider $\Delta = \hat{M} - \hat{A}$, then either $\mathcal{R} \Delta \neq 0$, or $\|\hat{A} + \Delta\|_{*,1} > \|\hat{A}\|_{*,1}$.

---

*Equal contribution 1Department of Mathematics, Stanford University, California, USA 2Department of Electrical Engineering, Stanford University, California, USA 3Department of Statistics, Stanford University, California, USA. Correspondence to: Qingyun Sun <qysun@stanford.edu>.

First, we show that if
\[ \|\Delta_T\|_2 \geq (2nN)\|\Delta_T\|_2, \]
then \( \|R\Delta_T\|_2 > \|R\Delta_T\|_2 \),
\[
\|R\Delta_T\|_2 = \|R\Delta_T + R\Delta_T\|_2 \\
\geq \|R\Delta_T\|_2 - \|R\Delta_T\|_2 > 0.
\]

We have a lower bound on \( \|R\Delta_T\|_2 \) and upper bound on \( \|R\Delta_T\|_2 \).
\[
\|R\Delta_T\|_2 = \langle R\Delta_T, R\Delta_T \rangle \\
\geq \|R\|^2/(nN)^2(1 - \|P_T - P_T R P_T\|)\|\Delta_T\|_2^2.
\]

Since \( E(P_T R P_T) = P_T \), we only need to control the deviation, we could use a concentration inequality called operator-Bernstein inequality [1].
\[
P[|P_T - P_T R P_T| > t] \leq \exp(-\frac{nt^2}{4\mu r^2}).
\]

Using the condition that \( p = C\mu \frac{1}{n} \log^2(Nn) \), let \( t = 1/4 \), we have
\[
P[|P_T - P_T R P_T| > t] \leq \exp(-\frac{n\mu \log^2(Nn)}{16\mu r}) = \exp(-\frac{\log^2(Nn)}{16(nN)^{-\gamma}}).
\]

where \( \gamma = \frac{\log(Nn)}{16} \). Therefore, with probability \( 1 - (nN)^{-\gamma} \), the the inequality holds for \( t = 1/2 \). When the inequality holds, \( \|P_T - P_T R P_T\| < 1/2, R\Delta \neq 0 \).

Second, we construct the dual certificate \( K \) by the following construction: We decompose \( \Omega \) as the union of \( k \) subset \( \Omega_t \), where each entry is sampled independently so that \( E(|\Omega_t|) = p_t = 1 - (1 - p)^{1/k} \), and define \( R_t = \frac{1}{p_t} P_\Omega U^* \).

Define \( H_0 = (\hat{A}), K_t = \sum_{j=1}^{t} R_j H_{j-1}, H_t = (\hat{A}) - P_T K_t \).

Then the dual certificate is defined as \( K = K_k \).

This construction is called goffing scheme, which is invented in [1]. Since \( p_t = p/k = C\mu \frac{1}{n} \log^2(Nn) \), we can assume \( \|P_T - P_T R P_T\| < 1/2 \), which is true with probability \( 1 - \exp(\frac{C\exp^2}{nk}) \).

And
\[
\|P_T(K) - (\hat{A})\|_2 = \|H_k\| \leq \frac{1}{2} \|H_k\|_2 \leq \frac{1}{2} \|H_{k-1}\|_2.
\]

Then
\[
\|P_T(K) - (\hat{A})\|_2 \leq \frac{1}{2} \|H_{k-1}\|_2 \leq \frac{1}{2} \|H_{k-1}\|_2.
\]

Also, \( \|P_T - R_j H_{j-1}\| \leq t_j \|H_{j-1}\|_2 \), and since \( \|H_{j-1}\|_2 \leq \sqrt{T} \), then
\[
\|P_T - (\hat{A})\|_2 \leq \frac{1}{2} \|H_{k-1}\|_2 \leq \frac{1}{2} \|H_{k-1}\|_2 \leq \frac{1}{2} \|H_{k-1}\|_2 < 1/2.
\]

Therefore, \( K \) is the dual certificate, the whole proof is done.

\( \square \)

**Imputation algorithm convergence** Now we show that the solution of our imputation algorithm converges asymptotically to a minimizer of the previously defined objective \( L_\lambda(M) \).

Each step of our imputation algorithm is minimizing a surrogate \( Q_\lambda(\hat{M}|\hat{M}^{\text{old}}) \) of the above objective function as
\[
\|A\Omega + P_\Omega U^{-1} \hat{M}^{\text{old}} - U^{-1} \hat{M}\|^2 + \sum_{k=1}^{N} \lambda_k \|\hat{M}(k)\|_*.\n\]

The resulting minimizer forms a sequence \( \hat{M}_\lambda^t \) with starting point \( \hat{M}_\lambda^0 \)
\[
\hat{M}_\lambda^{t+1} = \text{argmin } Q_\lambda(\hat{M}|\hat{M}_\lambda^t).\n\]

**Theorem 2.** The imputation algorithm produces a sequence of iterates \( \hat{M}_\lambda^t \) that converges to the minimizer of \( L_\lambda(M) \).

The main idea of the proof is to show that \( Q_\lambda \) decreases after every iteration and \( \hat{M}_\lambda^t \) is a Cauchy sequence, and the limit point is a stationary point of \( L_\lambda \).

**Proof.** For each iteration in our algorithm, we are solving for a surrogate of the objective function as
\[
Q_\lambda(\hat{M}|\hat{M}^{\text{old}}) = \|A\Omega + P_\Omega U^{-1} \hat{M}^{\text{old}} - U^{-1} \hat{M}\|^2 + \sum_{k=1}^{N} \lambda_k \|\hat{M}(k)\|_*.
\]

And the sequence \( \hat{M}_\lambda^t \) with any starting point \( \hat{M}_\lambda^0 \) is given by
\[
\hat{M}_\lambda^{t+1} = \text{argmin } Q_\lambda(\hat{M}|\hat{M}_\lambda^t).\n\]

The sequence satisfies
\[
L_\lambda(\hat{M}_\lambda^{t+1}) \leq Q_\lambda(\hat{M}_\lambda^{t+1}|\hat{M}_\lambda^t) \leq L_\lambda(\hat{M}_\lambda^t).\n\]
Because
\[ Q(\hat{M}^{t+1}\|\hat{M}^{t+1}) = L(\hat{M}^{t+1}) \]
and
\[ Q(\hat{M}\|\hat{M}^{\text{old}}) \]
\[ = \| P_{\Omega} A + P_{\perp \Omega} U^{-1} M - U^{-1} \hat{M} \|^2 + \sum_{k=1}^{N} \lambda_k \| \hat{M}(k) \| \]
\[ \geq \| P_{\Omega} A + P_{\perp \Omega} U^{-1} M - U^{-1} \hat{M} \|^2 + \sum_{k=1}^{N} \lambda_k \| \hat{M}(k) \| \]
\[ = Q(\hat{M}\|\hat{M}) \]

Below we prove the following successive differences are monotonically decreasing
\[ \| \hat{M}^{t+1} - \hat{M}^{t} \|^2 \leq \| \hat{M}^{t} - \hat{M}^{t-1} \|^2. \]
and the difference sequence converges to zero,
\[ \hat{M}^{t+1} - \hat{M}^{t} \to 0. \]
The successive differences are monotonically decreasing because the soft threshold operator is a contraction in \( L_2 \) norm \[ (2) \]. And when there are positive singular values smaller than the threshold, the successive differences will strictly decrease until the algorithm converges.

Then \( \hat{M}_\lambda^t \) is a Cauchy sequence, therefore we have a set of limit points. Also by monotonic convergence theorem, since \( \hat{M}^{t+1} - \hat{M}^{t} \) converges to zero monotonically, the Cauchy sequence \( \hat{M}_\lambda^t \) has an unique limit \( \hat{M}_\lambda^\infty \). Moreover, we can verify that \( \hat{M}_\lambda^\infty \) is a solution to the fixed point equation \( \nabla L(\hat{M}_\lambda) = 0 \), and a stationary point of \( L(\hat{M}_\lambda) \). Since \( L(\hat{M}_\lambda) \) is convex, each stationary point is a minimizer. Therefore, the convergence is proved.

References
