Guaranteed Rank Minimization via Singular Value Projection: Supplementary Material

Prateek Jain
Microsoft Research Bangalore
Bangalore, India
prajain@microsoft.com

Raghu Meka
UT Austin Dept. of Computer Sciences
Austin, TX, USA
raghu@cs.utexas.edu

Inderjit Dhillon
UT Austin Dept. of Computer Sciences
Austin, TX, USA
inderjit@cs.utexas.edu

Appendix A: Analysis for Affine Constraints Satisfying RIP

We first give a proof of Lemma 2.1 which bounds the error of the $(t+1)$-st iterate ($\psi(X^{t+1})$) in terms of the error incurred by the t-th iterate and the optimal solution.

Proof of Lemma 2.1 Recall that $\psi(X) = \frac{1}{2}\|\mathcal{A}(X) - b\|_2^2$. Since $\psi(\cdot)$ is a quadratic function, we have

$$\begin{aligned}\psi(X^{t+1}) - \psi(X^t) &= \langle \nabla\psi(X^t), X^{t+1} - X^t\rangle + \frac{1}{2}\|\mathcal{A}(X^{t+1} - X^t)\|_2^2 \\ &\leq \langle \mathcal{A}^T(\mathcal{A}(X^t) - b), X^{t+1} - X^t\rangle + \frac{1}{2}\cdot(1+\delta_{2k})\cdot\|X^{t+1} - X^t\|_F^2, \qquad (0.1)\end{aligned}$$

where the inequality follows from RIP applied to the matrix $X^{t+1} - X^t$ of rank at most $2k$. Let $Y^{t+1} = X^t - \frac{1}{1+\delta_{2k}}\mathcal{A}^T(\mathcal{A}(X^t) - b)$ and

$$f_t(X) = \langle \mathcal{A}^T(\mathcal{A}(X^t) - b), X - X^t\rangle + \frac{1}{2}\cdot(1+\delta_{2k})\cdot\|X - X^t\|_F^2.$$

Now,

$$\begin{aligned}f_t(X) &= \frac{1}{2}(1+\delta_{2k})\left[\|X - X^t\|_F^2 + 2\left\langle \frac{\mathcal{A}^T(\mathcal{A}(X^t) - b)}{1+\delta_{2k}}, X - X^t\right\rangle\right] \\ &= \frac{1}{2}(1+\delta_{2k})\|X - Y^{t+1}\|_F^2 - \frac{1}{2(1+\delta_{2k})}\cdot\|\mathcal{A}^T(\mathcal{A}(X^t) - b)\|_F^2.\end{aligned}$$

Thus, by definition, $\mathcal{P}_k(Y^{t+1}) = X^{t+1}$ is the minimizer of $f_t(X)$ over all matrices $X \in \mathcal{C}(k)$ (of rank at most k). In particular, $f_t(X^{t+1}) \leq f_t(X^*)$ and,

$$\begin{aligned}\psi(X^{t+1}) - \psi(X^t) \leq f_t(X^{t+1}) \leq f_t(X^*) &= \langle \mathcal{A}^T(\mathcal{A}(X^t) - b), X^* - X^t\rangle + \frac{1}{2}(1+\delta_{2k})\|X^* - X^t\|_F^2 \\ &\leq \langle \mathcal{A}^T(\mathcal{A}(X^t) - b), X^* - X^t\rangle + \frac{1}{2}\cdot\frac{1+\delta_{2k}}{1-\delta_{2k}}\|\mathcal{A}(X^* - X^t)\|_2^2 \qquad (0.2) \\ &= \psi(X^*) - \psi(X^t) + \frac{\delta_{2k}}{(1-\delta_{2k})}\|\mathcal{A}(X^* - X^t)\|_2^2,\end{aligned}$$

where inequality (0.2) follows from RIP applied to $X^* - X^t$.

We now prove Theorem 1.2.

Proof of Theorem 1.2 Let the current solution X^t satisfy $\psi(X^t) \geq G\|e\|^2/2$, where $G \geq 0$ is a universal constant. Using Lemma 2.1 and the fact that $b - \mathcal{A}(X^*) = e$,

$$\psi(X^{t+1}) \leq \frac{\|e\|_2^2}{2} + \frac{\delta_{2k}}{(1-\delta_{2k})}\|b - \mathcal{A}(X^t) - e\|_2^2 \leq \frac{\|e\|_2^2}{2} + \frac{2\delta_{2k}}{(1-\delta_{2k})}(\psi(X^t) - e^T(b - \mathcal{A}(X^t)) + \|e\|^2/2)$$
$$\leq \frac{\psi(X^t)}{G^2} + \frac{2\delta_{2k}}{(1-\delta_{2k})}\left(\psi(X^t) + \frac{2}{G}\psi(X^t) + \frac{1}{G^2}\psi(X^t)\right) \leq D\psi(X^t),$$

where $D = \left(\frac{1}{G^2} + \frac{2\delta_{2k}}{(1-\delta_{2k})}\left(1 + \frac{1}{G}\right)^2\right)$. Recall that $\delta_{2k} < 1/3$. Hence, selecting $G > (1 + \delta_{2k})/(1 - 3\delta_{2k})$, we get $D < 1$. Also, $\psi(X^0) = \psi(0) = \|b\|^2/2$. Hence, $\psi(X^\tau) \leq C\|e\|^2 + \epsilon$ where $\tau = \left\lceil \frac{1}{\log(1/D)} \log \frac{\|b\|^2}{2(C\|e\|^2+\epsilon)} \right\rceil$ and $C = G^2/2$.

Appendix B: Proof of Weak RIP for Matrix Completion

We now give a detailed proof of the weak RIP property for the matrix completion problem, Theorem 2.2.

Proof of Lemma 2.4 Let $X = U\Sigma V^T$ be the singular value decomposition of X. Then,

$$|X_{ij}| = |e_i^T U\Sigma V^T e_j| = |\sum_{l=1}^k U_{il}\Sigma_{ll}V_{jl}| \leq \sum_{l=1}^k \Sigma_{ll}|U_{il}||V_{jl}|.$$

Since X is μ-incoherent,

$$|X_{ij}| \leq \sum_{l=1}^k \Sigma_{ll}|U_{il}||V_{jl}| \leq \frac{\mu}{\sqrt{mn}} \cdot (\sum_{l=1}^k \Sigma_{ll}) \leq \frac{\mu}{\sqrt{mn}} \cdot \sqrt{k} \cdot (\sum_{l=1}^k \Sigma_{ll}^2)^{1/2} = \frac{\mu\sqrt{k}}{\sqrt{mn}} \cdot \|X\|_F.$$

We also need the following technical lemma for our proof.

Lemma 0.1 *Let $a, b, c, x, y, z \in [-1, 1]$. Then,*

$$|abc - xyz| \leq |a - x| + |b - y| + |c - z|.$$

Proof We have,

$$\begin{aligned} |abc - xyz| &= |(abc - xbc) + (xbc - xyc) + (xyc - xyz)| \\ &\leq |a - x||bc| + |b - y||xc| + |c - z||xy| \\ &\leq |a - x| + |b - y| + |c - z|. \end{aligned}$$

We now prove Lemma 2.5. We use the following form of the classical Bernstein's large deviation inequality in our proof.

Lemma 0.2 (Bernstein's Inequality, see [3]) *Let $X_1, X_2, \ldots, X_n$ be independent random variables with $E[X_i] = 0$ and $|X_i| \leq M$ for all i. Then,*

$$P[\sum_i X_i > t] \leq \exp\left(-\frac{t^2/2}{\sum_i Var(X_i) + Mt/3}\right).$$

Proof of Lemma 2.5 For $(i, j) \in [m] \times [n]$, let ω_{ij} be the indicator variables with $\omega_{ij} = 1$ if $(i, j) \in \Omega$ and 0 otherwise. Then, ω_{ij} are independent random variables with $\mathsf{Pr}[\omega_{ij} = 1] = p$. Let random variable $Z_{ij} = \omega_{ij}X_{ij}^2$. Note that,

$$E[Z_{ij}] = pX_{ij}^2,\ Var(Z_{ij}) = p(1-p)X_{ij}^4.$$

Since X is α-regular, $|Z_{ij} - E[Z_{ij}]| \leq |X_{ij}|^2 \leq (\alpha^2/mn) \cdot \|X\|_F^2$. Thus,

$$M = \max_{i,j} |Z_{ij} - \mathsf{E}[Z_{ij}]| \leq \frac{\alpha^2}{mn}\|X\|_F^2. \tag{0.3}$$

Now, define random variable $S = \sum_{i,j} Z_{ij} = \sum_{i,j} \omega_{ij} X_{ij}^2 = \|\mathcal{P}_\Omega(X)\|_F^2$. Note that, $E[S] = p\|X\|_F^2$. Since, Z_{ij} are independent random variables,

$$Var(S) = \sum_{i,j} p(1-p)\, X_{ij}^4 \leq p\,(\max_{i,j} X_{ij}^2) \cdot \sum_{i,j} X_{ij}^2 \leq \frac{p\,\alpha^2}{mn}\|X\|_F^4. \tag{0.4}$$

Using Bernstein's inequality (Lemma 0.2) for S with $t = \delta p \|X\|_F^2$ and Equations (0.3) and (0.4) we get,

$$\begin{aligned}
\Pr[|S - \mathsf{E}[S]| > t] &\leq 2\exp\left(\frac{-t^2/2}{Var(Z) + Mt/3}\right) \\
&\leq 2\exp\left(-\frac{\delta^2 pmn}{\alpha^2(1+\delta/3)}\right) \\
&\leq 2\exp\left(-\frac{\delta^2 pmn}{3\alpha^2}\right).
\end{aligned}$$

Proof of Lemma 2.6 We construct $S(\mu,\epsilon)$ by discretizing the space of low-rank incoherent matrices. Let $\rho = \epsilon/\sqrt{9k^2mn}$ and $D(\rho) = \{\rho\, i : i \in \mathbb{Z}, |i| < \lfloor 1/\rho \rfloor\}$. Let

$$\begin{aligned}
U(\rho) &= \{U \in \mathbb{R}^{m\times k} : U_{ij} \in (\sqrt{\mu/m})\cdot D(\rho)\ \}, \\
V(\rho) &= \{V \in \mathbb{R}^{n\times k} : V_{ij} \in (\sqrt{\mu/n})\cdot D(\rho)\ \}, \\
\Sigma(\rho) &= \{\Sigma \in \mathbb{R}^{k\times k} : \Sigma_{ij} = 0, i \neq j,\ \Sigma_{ii} \in D(\rho)\}, \\
S(\mu,\epsilon) &= \{\, U\Sigma V^T : U \in U(\rho), \Sigma \in \Sigma(\rho), V \in V(\rho)\,\}.
\end{aligned}$$

We will show that $S(\mu,\epsilon)$ satisfies the conditions of the Lemma. Observe that $|D(\rho)| < 2/\rho$. Thus,

$$|U(\rho)| < (2/\rho)^{mk}, \quad |V(\rho)| < (2/\rho)^{nk}, \quad |\Sigma(\rho)| < (2/\rho)^k.$$

Hence, $|S(\mu,\epsilon)| < (2/\rho)^{mk+nk+k} < (mnk/\epsilon)^{3(m+n)k}$.

Fix a μ-incoherent $X \in \mathbb{R}^{m\times n}$ of rank at most k with $\|X\|_2 = 1$. Let $X = U\Sigma V^T$ be the singular value decomposition of X. Let U_1 be the matrix obtained by rounding entries of U to integer multiples of $\sqrt{\mu}\,\rho/\sqrt{m}$ as follows: for $(i,l) \in [m]\times[k]$, let

$$(U_1)_{il} = \frac{\sqrt{\mu}\rho}{\sqrt{m}} \cdot \left\lfloor U_{il}\frac{\sqrt{m}}{\sqrt{\mu}\,\rho}\right\rfloor.$$

Now, since $|U_{il}| \leq \sqrt{\mu}/\sqrt{m}$, it follows that $U_1 \in U(\rho)$. Further, for all $i \in [m], l \in [k]$,

$$|(U_1)_{il} - U_{il}| < \frac{\sqrt{\mu}}{\sqrt{m}}\,\rho \leq \rho.$$

Similarly, define V_1, Σ_1 by rounding entries of V, Σ to integer multiples of $\sqrt{\mu}\,\rho/\sqrt{n}$ and ρ respectively. Then, $V_1 \in V(\rho)$, $\Sigma_1 \in \Sigma(\rho)$ and for $(j,l) \in [n]\times[k]$,

$$|(V_1)_{jl} - V_{jl}| < \frac{\sqrt{\mu}\rho}{\sqrt{n}} \leq \rho, \quad |(\Sigma_1)_{ll} - \Sigma_{ll}| < \rho.$$

Let $X(\rho) = U_1\Sigma_1V_1^T$. Then, by the above equations and Lemma 0.1, for $i \in [m], l \in [k], j \in [n]$,

$$|(U_1)_{il}(\Sigma_1)_{ll}(V_1)_{jl} - U_{il}\Sigma_{ll}V_{jl}| < 3\rho.$$

Thus, for $i,j \in [m]\times[n]$,

$$\begin{aligned}
|X(\rho)_{ij} - X_{ij}| &= |\sum_{l=1}^{k}(U_1)_{il}(\Sigma_1)_{ll}(V_1)_{jl} - U_{il}\Sigma_{ll}V_{jl}| \\
&\leq \sum_{l=1}^{k} |(U_1)_{il}(\Sigma_1)_{ll}(V_1)_{jl} - U_{il}\Sigma_{ll}V_{jl}| \\
&< 3k\rho.
\end{aligned} \tag{0.5}$$

Using Lemma 2.4 and Equation (0.5)

$$\max_{i,j} |X(\rho)_{ij}| < \max_{i,j} |X_{ij}| + 3k\rho \leq \frac{\mu\sqrt{k}}{\sqrt{mn}} \cdot \|X\|_F + \frac{\epsilon}{\sqrt{mn}}.$$

Also, using (0.5),

$$\|X(\rho) - X\|_F^2 = \sum_{i,j} |X(\rho)_{ij} - X_{ij}|^2 < 9k^2 mn\rho^2 = \epsilon^2.$$

Further, by triangle inequality, $\|X(\rho)\|_F > \|X\|_F - \epsilon > \|X\|_F/2$. Since, $\epsilon < 1$ and $\mu\sqrt{k}\|X\|_F \geq 1$,

$$\max_{i,j} |X(\rho)_{ij}| < \frac{2\mu\sqrt{k}}{\sqrt{mn}} \cdot \|X\|_F < \frac{4\mu\sqrt{k}}{\sqrt{mn}} \cdot \|X(\rho)\|_F.$$

Thus, $X(\rho)$ is $4\mu\sqrt{k}$-regular. The lemma now follows by taking $Y = X(\rho)$.

We now prove Theorem 2.2 by combining Lemmas 2.5 and 2.6.

Proof of Theorem 2.2 Let $m \leq n$, $\epsilon = \delta/9mnk$ and

$$S'(\mu, \epsilon) = \{Y : Y \in S(\mu, \epsilon), Y \text{ is } 4\mu\sqrt{k}\text{-regular}\},$$

where $S(\mu, \epsilon)$ is as in Lemma 2.6. Then, by Lemma 2.5 and a union bound,

$$\begin{aligned}\Pr\left[\,\left|\|\mathcal{P}_\Omega(Y)\|_F^2 - p\|Y\|_F^2\right| \geq \delta p\|Y\|_F^2 \text{ for some } Y \in S'(\mu, \epsilon)\,\right] &\leq 2\left(\frac{mnk}{\epsilon}\right)^{3(m+n)k} \exp\left(\frac{-\delta^2 pmn}{16\mu^2 k}\right) \\ &\leq \exp(C_1 nk \log n) \cdot \exp\left(\frac{-\delta^2 pmn}{16\mu^2 k}\right),\end{aligned}$$

where $C_1 \geq 0$ is a constant independent of m, n, k.

Thus, if $p > C\mu^2 k^2 \log n/\delta^2 m$, where $C = 16(C_1 + 1)$, with probability at least $1 - \exp(-n \log n)$, the following holds:

$$\forall Y \in S'(\mu, \epsilon), \quad |\|\mathcal{P}_\Omega(Y)\|_F^2 - p\|Y\|_F^2| \leq \delta p\|Y\|_F^2. \tag{0.6}$$

As the statement of the theorem is invariant under scaling, it is enough to show the statement for all μ-incoherent matrices X of rank at most k and $\|X\|_2 = 1$. Fix such a X and suppose that (0.6) holds. Now, by Lemma 2.6 there exists $Y \in S'(\mu, \epsilon)$ such that $\|Y - X\|_F \leq \epsilon$. Moreover,

$$\|Y\|_F^2 \leq (\|X\|_F + \epsilon)^2 \leq \|X\|_F^2 + 2\epsilon\|X\|_F + \epsilon^2 \leq \|X\|_F^2 + 3\epsilon k.$$

Proceeding similarly, we can show that

$$|\|X\|_F^2 - \|Y\|_F^2| \leq 3\epsilon k. \tag{0.7}$$

Further, starting with $\|\mathcal{P}_\Omega(Y - X)\|_F \leq \|Y - X\|_F \leq \epsilon$ and arguing as above we get that

$$|\|\mathcal{P}_\Omega(Y)\|_F^2 - \|\mathcal{P}_\Omega(X)\|_F^2| \leq 3\epsilon k. \tag{0.8}$$

Combining inequalities (0.7), (0.8) above, we have

$$\begin{aligned}|\|\mathcal{P}_\Omega(X)\|_F^2 - p\|X\|_F^2| &\leq |\|\mathcal{P}_\Omega(X)\|_F^2 - \|\mathcal{P}_\Omega(Y)\|_F^2| + p\,|\|X\|_F^2 - \|Y\|_F^2| + |\|\mathcal{P}_\Omega(Y)\|_F^2 - p\|Y\|_F^2| \\ &\leq 6\epsilon k + \delta p\|Y\|_F^2 && \text{from (0.6), (0.7), (0.8)} \\ &\leq 6\epsilon k + \delta p(\|X\|_F^2 + 3\epsilon k) && \text{from (0.7)} \\ &\leq 9\epsilon k + \delta p\|X\|_F^2 \\ &\leq 2\delta p\|X\|_F^2. && \text{since } \|X\|_F^2 \geq 1\end{aligned}$$

The theorem now follows.

Appendix C: SVP-Newton

The affine rank minimization problem is a natural generalization to matrices of the following compressed sensing problem (CSP) for vectors:

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0, \quad \text{s.t. } A\boldsymbol{x} = \boldsymbol{b}, \tag{0.9}$$

where $\|\boldsymbol{x}\|_0$ is the l_0 norm (size of the support) of $\boldsymbol{x} \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$ is the sensing matrix and $\boldsymbol{b} \in \mathbb{R}^m$ are the measurements. Similar to ARMP, the compressed sensing problem is also NP-hard in general. However, similar to ARMP, compressed sensing can also be solved for measurement matrices that satisfies restricted isometry property.

Restricted Isometry Property (RIP) for the Compressed Sensing problem is similar to the corresponding definition for matrices:

$$(1-\delta_k)\|\boldsymbol{x}\|_2^2 \le \|A\boldsymbol{x}\|_2^2 \le (1+\delta_k)\|\boldsymbol{x}\|_2^2, \tag{0.10}$$

where support$(x) \le k$. Note that support of a vector is same as the rank of the corresponding diagonal matrix whose diagonal is given by the same vector.

Since, Compressed Sensing is a special case of Affine Rank Minimization Problem (ARMP) with diagonal matrices, i.e., with fixed basis vectors $U_k = I_k$, $V_k = I_k$ (see step 4, SVP algorithm). Hence, SVP-Newton can be directly applied to Problem 0.9.

The corresponding updates for SVP-Newton applied to compressed sensing are given by:

$$\boldsymbol{y}^{t+1} \leftarrow \boldsymbol{x}^t - \eta_t A^T(A\boldsymbol{x} - \boldsymbol{b}), \tag{0.11}$$

$$S \leftarrow \text{ Set of top } k \text{ elements of} \boldsymbol{y}^{t+1}, \tag{0.12}$$

$$\boldsymbol{x}_S^{t+1} \leftarrow \operatorname*{argmin}_{\boldsymbol{x}_S} \|A_S\boldsymbol{x}_S - \boldsymbol{b}\|_2^2, \tag{0.13}$$

We now present a Theorem that shows that the SVP-Newton method applied to the Compressed Sensing problem converges to the optimal solution in $O(\log k)$ iterations where k is the optimal support.

Theorem 0.3 *Suppose the isometry constant of A in CSP satisfies $\delta_{2k} < 1/3$. Let $b = A\boldsymbol{x}^*$ for a support-k vector $\boldsymbol{x}^*$, i.e., $\|\boldsymbol{x}^*\|_0 = k$ and $b_i \le 1, \forall i$. Then, SVP-Newton algorithm for CSP (Updates 0.13) with step-size $\eta_t = 4/3$ converges to the* exact $\boldsymbol{x}^*$ *in at most $\frac{2}{\log 1/\alpha}\log_\alpha \frac{k}{\sqrt{1-\delta_{2k}}c}$ iterations, where $c = \min_i \|x_i^*\|$ and $\alpha = \frac{1/3+\delta_{2k}}{1-\delta_{2k}}$.*

Proof Assume that $\boldsymbol{x}^0 = 0$. Now, using Lemma 2.1 specialized to CSP (or see Theorem 2.1 in Garg and Khandekar [1]),

$$(1-\delta_{2k})\|\boldsymbol{x}^t - \boldsymbol{x}^*\| \le \|A\boldsymbol{x}^t - \boldsymbol{b}\|^2 \le \alpha^t\|\boldsymbol{b}\|^2,$$

where $\alpha = \frac{1/3+\delta_{2k}}{1-\delta_{2k}}$.

Let us assume that the support set S discovered by SVP-Newton at the t-th step differs from the optimal set S^* (support set for $\boldsymbol{x}^*$) on at least one element. Now, note that if the support set S discovered by SVP-Newton is the optimal set S^* (support set for $\boldsymbol{x}^*$) then $\boldsymbol{x}^t = \boldsymbol{x}^*$ as $\boldsymbol{x}^t$ minimizes $\|A_S\boldsymbol{x}_S - \boldsymbol{b}\|_2^2$.

Since, $S \cap \bar{S}^* \neq \phi$,

$$\|\boldsymbol{x}^t - \boldsymbol{x}^*\|^2 \ge c^2.$$

Hence, $(1-\delta_{2k})c^2 \le \alpha^t\|\boldsymbol{b}\|^2$, implying,

$$t \le \frac{2}{\log 1/\alpha}\log_\alpha \frac{k}{\sqrt{1-\delta_{2k}}c}.$$

Note that SVP-Newton converges to the optimal solution in logarithmic number of queries rather than ϵ-approximate solution obtained by GradeS [1] which is just a specialization of SVP to the compressed sensing problem. Note that the number of iterations required is an improvement over

$O(k)$ bound provided by [2]. [2] requires *incoherence* property for the measurement matrix[1], which is a stronger condition to RIP. In fact, one can construct measurement matrices A with bounded RIP constant but unbounded incoherence constant.

References

[1] Rahul Garg and Rohit Khandekar. Gradient descent with sparsification: an iterative algorithm for sparse recovery with restricted isometry property. In *ICML*, 2009. doi: 10.1145/1553374.1553417.

[2] Arian Maleki. Coherence analysis of iterative thresholding algorithms. *CoRR*, abs/0904.1193, 2009.

[3] Wikipedia. Bernstein inequalities (probability theory) — wikipedia, the free encyclopedia, 2009. URL `http://en.wikipedia.org/w/index.php?title=Bernstein_inequalities_(probability_theory)&oldid=282206343`. [Online; accessed 6-April-2009].

[1] Note that incoherence property mentioned here is different than the one used in the context of Matrix Completion