Strong Recovery Conditions for Low-rank Matrices

Maryam Fazel

University of Washington

Joint work with:
Samet Oymak (Caltech), Karthik Mohan (UW), Babak Hassibi (Caltech)

Banff Sparse and Low-rank Approximations Workshop, 3/8/11

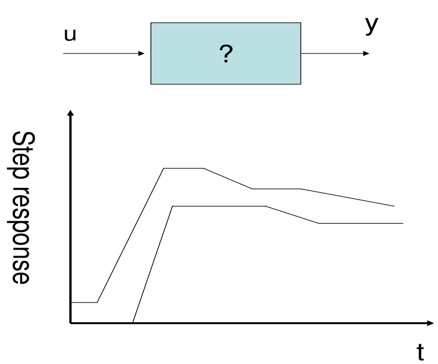
Low-rank matrix recovery problem

$$\begin{array}{ll} \text{minimize} & \operatorname{rank}(X) \\ \text{subject to} & \mathcal{A}(X) = b \end{array}$$

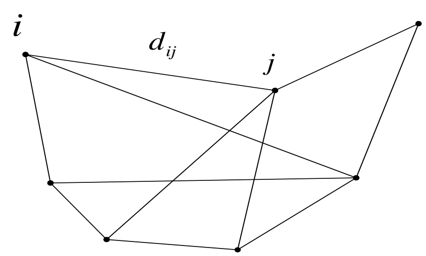
where $X \in \mathbb{R}^{n_1 \times n_2}$, $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ is a linear map, $m \ll n_1, n_2$

NP-hard in general.

system ID



distance geometry



min embedding dimension,
given pairwise distances

machine learning

users/movies database

$$\begin{bmatrix} 5 & ? & 8 & \dots & ? \\ ? & 10 & 3 & \dots & 5 \\ & & \ddots & & \\ 2 & 1 & ? & \dots & 6 \end{bmatrix}$$

movie recommendation
based on few features

quantum tomography, network tomography,. . .

ℓ_1 and nuclear norm minimization

goal: find k-sparse $\mathbf{x}_0 \in \mathbb{R}^n$ from m noisy measurements $\mathbf{y} = A\mathbf{x}_0 + \mathbf{z}$, $\|\mathbf{z}\|_2 \leq \epsilon$.
solve (convex relaxation)

$$\begin{array}{ll} \text{minimize} & \|\mathbf{x}\|_1 \\ \text{subject to} & \|A\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon \end{array}$$

goal: find rank-k $X_0 \in \mathbb{R}^{n_1 \times n_2}$ from $\mathbf{y} = \mathcal{A}(X_0) + \mathbf{z}$, $\|\mathbf{z}\|_2 \leq \epsilon$. solve

$$\begin{array}{ll} \text{minimize} & \|X\|_\star \\ \text{subject to} & \|\mathcal{A}(X) - \mathbf{y}\|_2 \leq \epsilon \end{array}$$

- both robust, even if not perfectly sparse/low-rank
- **def:** $\mathbf{x}^*$ is *as good as* $\mathbf{x}_0$ w.r.t $\mathbf{y}$, if $\|A\mathbf{x}^* - \mathbf{y}\|_2 \leq \epsilon$ and $\|\mathbf{x}^*\|_1 \leq \|\mathbf{x}_0\|_1$ ($\mathbf{x}^*$ in intersection of tube constraint and scaled ℓ_1 ball)
- similarly for matrices

This talk

- direct extension of strong recovery* conditions from vectors to matrices; eliminate the existing gap

- use a key singular value inequality and unitary invariance

- typical result: same restricted isometry conditions for k sparse vector recovery guarantees rank k matrix recovery—nothing lost in translation!

- nullspace based properties (NSP, SSP); robust recovery; recovery using ℓ_p, $p < 1$

* strong recovery: works for *all* matrices up to certain rank

Restricted isometry property

Restricted Isometry Constant (RIC) for $\mathcal{A}$: the smallest δ_k s.t.

$$(1-\delta_k)\|X\|_F^2 < \|\mathcal{A}(X)\|_2^2 < (1+\delta_k)\|X\|_F^2$$

holds for all X with $\operatorname{rank}(X) \leq k$.

Restricted Orthogonality Constant (ROC) for $\mathcal{A}$: the smallest $\theta_{k,k'}$ s.t.

$$|\langle \mathcal{A}(X), \mathcal{A}(X') \rangle| \leq \theta_{k,k'}\|X\|_F\|X'\|_F$$

holds for all X, X' where $\operatorname{rank}(X) \leq k$, $\operatorname{rank}(X') \leq k'$, and X, X' have orthogonal row and column spaces.

RIP: $\mathcal{A}$ satisfies inequalities of form $f(\delta_{i_1}, \ldots, \delta_{i_m}, \theta_{j_1,j_1'}, \ldots, \theta_{j_n,j_n'}) \leq c$, where $f(0) = 0$ and f is increasing.

Spherical section property

$\mathcal{A}$ satisfies the Δ-Spherical Section Property (SSP) if $\Delta(\mathcal{A}) \geq \Delta$, where

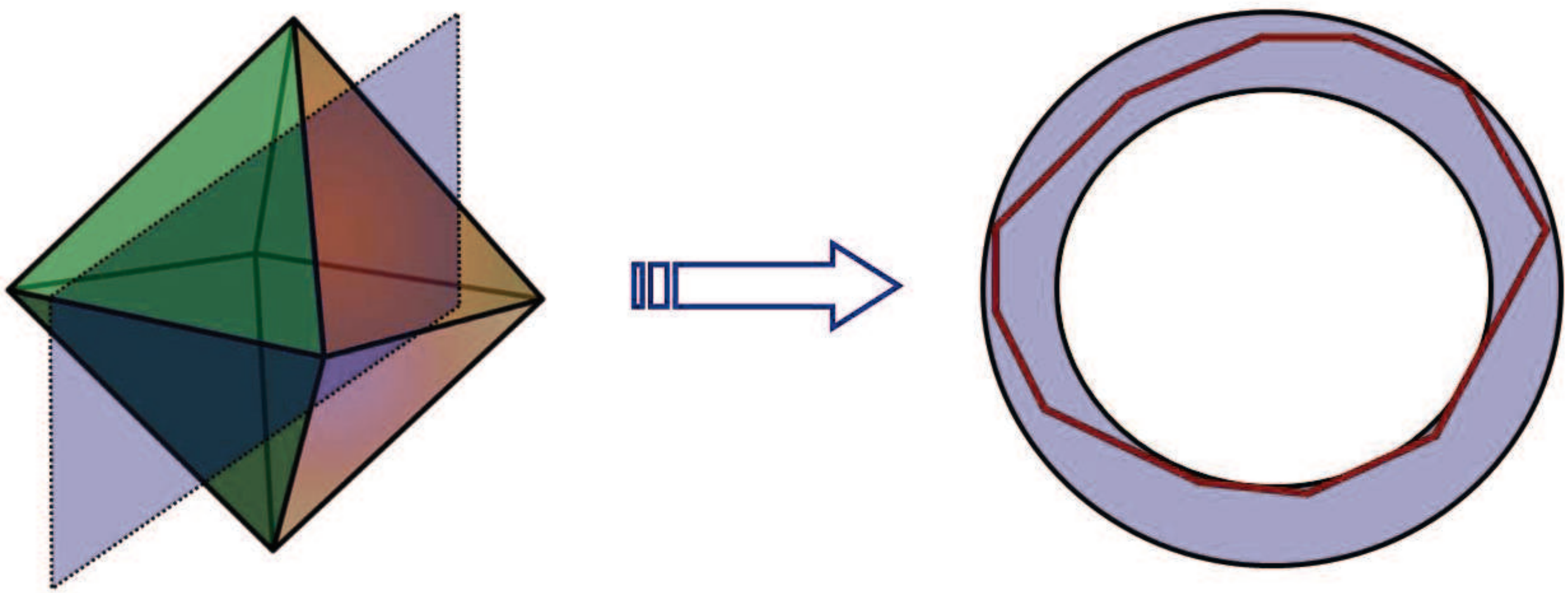
$$\Delta(\mathcal{A}) = \min_{W \in \mathcal{N}(\mathcal{A}) \setminus \{0\}} \frac{\|W\|_\star^2}{\|W\|_F^2}$$

Δ large $\Rightarrow$ nullspace doesn't include low rank matrices (also known as an *almost Euclidean* subspace)

vector case: ℓ_1/ℓ_2 [Kashin'77],[Gluskin,Garnaev'84],. . .

used in compressed sensing [Kashin,Temlyakov'07],[Zhang'08],[Vavasis'09], and matrix recovery [Dvijotham,F.'10] vector case: $\frac{\|\mathbf{w}\|_1}{\|\mathbf{w}\|_2} \geq \sqrt{\Delta} \quad (1 \leq \Delta \leq n)$

meaning: Δ large means if ℓ_1 ball is cut by a subspace, intersection looks spherical $(1/\sqrt{n} \leq \|\mathbf{w}\|_2 \leq 1/\sqrt{\Delta})$



in high dimensions, random subspaces should have large Δ

Key inequality

[HornJohnson'90] for any $X, Y \in \mathbb{R}^{n_1 \times n_2}$

$$\sum_{i=1}^{n_1} |\sigma_i(X) - \sigma_i(Y)| \leq \|X - Y\|_\star$$

LHS independent of singular vectors; equality when singular vectors are aligned.

Lemma. given W with SVD $W = U\Sigma_W V^T$, if there exists any X_0 for which $\|X_0 + W\|_\star \leq \|X_0\|_\star$, then $X_1 = -U\Sigma_{X_0}V^T$ also satisfies $\|X_1 + W\|_\star \leq \|X_1\|_\star$.

meaning: if there is a "bad" X_0 for a particular W, can construct other "bad" X_1, with the same singular values, that lies on W's subspace (given by U, V).

Nullspace property (NSP)

let $\epsilon = 0$ (no noise)

Sparse recovery: [Feuer,Nemirovsky'03] All $\mathbf{x}_0$ with $\|\mathbf{x}_0\|_0 \leq k$ can be recovered via ℓ_1 minimization **iff**

$$\sum_{i=1}^{k} \bar{w}_i < \sum_{i=k+1}^{n} \bar{w}_i, \quad \forall\, \mathbf{w} \in \mathcal{N}(A)$$

where $\bar{w}_i$ is ith largest entry of $|\mathbf{w}|$.

Low-rank recovery: [Oymak,Hassibi'10] All X_0 with $\mathrm{rank}(X_0) \leq k$ can be recovered via nuclear norm minimization **iff**

$$\sum_{i=1}^{k} \sigma_i(W) < \sum_{i=k+1}^{n_1} \sigma_i(W), \quad \forall\, W \in \mathcal{N}(\mathcal{A})$$

meaning: top k singular values contain no more than 1/2 of total ℓ_1 mass of σ

Main result

- V_1: A satisfies a property $\mathbf{P}$
- V_2: for any $\mathbf{x}_0$, $\|\mathbf{z}\|_2 \le \epsilon$, $\mathbf{y} = A\mathbf{x}_0 + \mathbf{z}$ and any $\mathbf{x}^*$ as good as $\mathbf{x}_0$, $\|\mathbf{x}^* - \mathbf{x}_0\| \ \le \ h(\bar{\mathbf{x}}_0, \epsilon)$
- V_3: any $\mathbf{w} \in \mathcal{N}(A)$ satisfies a property $\mathbf{Q}$

- M_1: $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ satisfies the (extension) property $\mathbf{P}$
- M_2: for any X_0, $\|\mathbf{z}\|_2 \le \epsilon$, $\mathbf{y} = \mathcal{A}(X_0) + \mathbf{z}$ and any X^* as good as X_0, $\|X^* - X_0\| \ \le \ h(\Sigma(X_0), \epsilon)$
- M_3: for any $W \in \mathcal{N}(\mathcal{A})$, $\sigma(W)$ satisfies property $\mathbf{Q}$

Main result: for a given $\mathbf{P}$, the following hold:

$$(V_1 \implies V_2) \implies (M_1 \implies M_2)$$

$$(V_1 \implies V_3) \implies (M_1 \implies M_3)$$

Application to RIP based recovery

Robust + noisy recovery: Let $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \longrightarrow \mathbb{R}^m$ satisfy RIP inequality described by $f(\delta_{i_1}, \ldots, \delta_{i_m}, \theta_{j_1, j_1'}, \ldots, \theta_{j_n, j_n'}) \leq c$. Then for all X_0, all $\|\mathbf{z}\|_2 \leq \epsilon$ and all X^* as good as X_0,

$$\begin{aligned} \|X_0 - X^*\|_F &\leq \frac{C_1}{\sqrt{k}} \|X_0 - X_0^k\|_\star + C_2 \epsilon \\ \|X_0 - X^*\|_\star &\leq C_3 \|X_0 - X_0^k\|_\star \end{aligned}$$

with exactly the same constants C_1, C_2, C_3 as the vector case, when A satisfies the same RIP inequality f.

(X_0^k: best rank-k approximation of X_0)

matrix-vector RIP link: if $\mathcal{A}$ has RIP, all its restrictions to $(U_{n_1 \times n_1}, V_{n_2 \times n_1})$ have RIP

Application to SSP based recovery

Robust recovery: Let $\epsilon = 0$ and X^* be as good as X_0. If $\mathcal{A}$ satisfies Δ-SSP with $\Delta > 4k$,

$$\|X^* - X_0\|_\star \quad \leq \quad C\|X_0 - X_0^k\|_\star$$

where $C = \frac{2}{1-2\sqrt{k/\Delta}}$.

- improves sufficient condition of $k < \frac{\Delta}{6}$ [Dvijotham,F.'10] to $k < \frac{\Delta}{4}$
- simplifies analysis
- matches sufficient condition for sparse vector recovery [Zhang'08]

NSP based robust recovery

Nuclear norm robustness: Let $\epsilon = 0$; for any X_0 and any X^* as good as X_0,

$$\|X_0 - X^*\|_\star < 2C\|X_0 - X_0^k\|_\star$$

iff for all $W \in \mathcal{N}(\mathcal{A})$,

$$\sum_{i=1}^{k} \sigma_i(W) < \frac{C-1}{C+1} \sum_{i=k+1}^{n} \sigma_i(W).$$

Matrix noise robustness: For any X_0 with $\mathrm{rank}(X_0) \leq k$, any $\|\mathbf{z}\|_2 \leq \epsilon$ and any X^* as good as X_0,

$$\|X_0 - X^*\|_F < C\epsilon,$$

iff for any W with $\sum_{i=1}^{k} \sigma_i(W) \geq \sum_{i=k+1}^{n} \sigma_i(W)$,

$$\|W\|_F < \frac{C}{2}\|\mathcal{A}(W)\|_2.$$

Application to recovery via Schatten-p quasinorm

nonconvex surrogate for rank

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n_1} \sigma_i^p(X) \\ \text{subject to} & \mathcal{A}(X) = \mathbf{y} \end{array}$$

- motivated iterative algorithms, e.g. iterative reweighted least squares
- empirically observed to outperform nuclear norm recovery e.g. [Chartrand'08,Foucart,Lai'09]

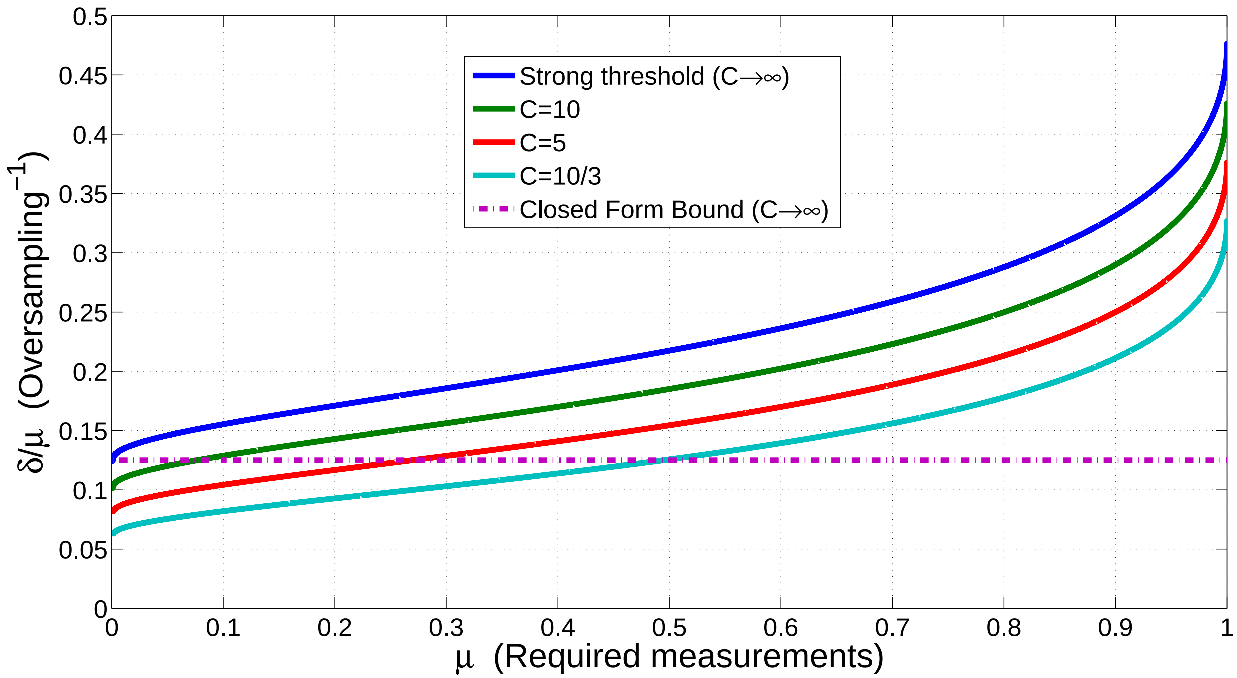
Lemma. if property $\mathbf{S}$ on matrix A implies recovery of all $\mathbf{x}_0$ with sparsity $2k$ via ℓ_p minimization $(0 < p < 1)$, then (extended) $\mathbf{S}$ implies recovery of all matrices with $\mathrm{rank}(X_0) \leq k$ via Schatten-p quasinorm minimization.

- inequality $\sum_{i=1}^{n}(\sigma_i^p(X) - \sigma_i^p(Y)) \leq \sum_{i=1}^{n} \sigma_i^p(X - Y)$ is used. there is gap between necessary and sufficient parts.

- a stronger inequality (with abs values on RHS) will allow bridging the gap. this seems to hold empirically, though no proof yet. . .

Improved thresholds

strong thresholds for Gaussian measurements, using new NSP (plus probabilistic analysis using Gordon's lemma) [Oymak,Hassibi'10]



degrees of freedom per measurement vs number of measurements, for $n \times n$ matrix, constant k/n.

improves known thresholds (e.g. [Recht,Xu,Hassibi'10])

Conclusions

- extend strong recovery conditions from vectors to matrices with no loss
- match best vector RIP ([Cai,et al'10;Lai'10]) and nullspace recovery conditions (e.g., $\delta_k < 0.309$ guarantees rank-k recovery)
- robust recovery: best error bounds via NSP
- some results for nonconvex Schatten-p quasinorm minimization (but not yet tight)
- also help obtain improved thresholds for Gaussian measurements

ref: Oymak, Mohan, Fazel, Hassibi, "A simplified approach to recovery conditions for low rank matrices", http://arxiv.org/abs/1103.1178