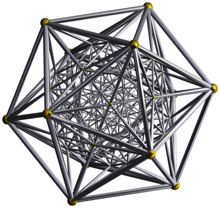
Computational Principles for High-dim Data Analysis

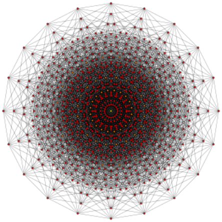
(Lecture Nine)

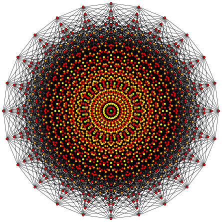
Yi Ma and Jiantao Jiao

EECS Department, UC Berkeley

September 28, 2021







Convex Methods for Low-Rank Matrix Recovery
(Random Measurements)

"Mathematics is the art of giving the same name to different things."
– Henri Poincaré

Problem

Recovering a sparse signal x_o:

$$\underbrace{\boldsymbol{y}}_{\text{observation}} = \boldsymbol{A} \underbrace{\boldsymbol{x}_o}_{\text{unknown}}, \tag{1}$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is a linear map.

Recovering a low-rank matrix $\boldsymbol{X}_o$:

$$\underbrace{\boldsymbol{y}}_{\text{observation}} = \mathcal{A}\left[\underbrace{\boldsymbol{X}_o}_{\text{unknown}}\right], \tag{2}$$

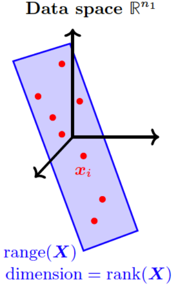
where, $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ is a linear map.

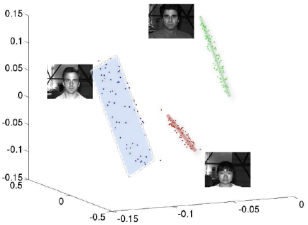
Examples of Low-rank Modeling

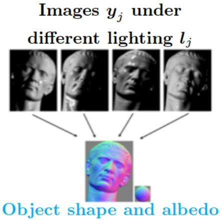
Multiple images of a Lambertian object with different light:

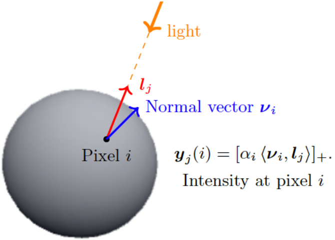
$$\boldsymbol{Y} = \mathcal{P}_\Omega[\boldsymbol{N}\boldsymbol{L}].$$

$\boldsymbol{X} = \boldsymbol{N}\boldsymbol{L}$ has rank 3. (Details in Chapter 14)



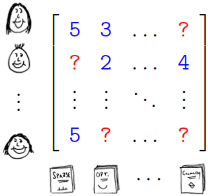






Examples of Low-rank Modeling

Recommendation Ratings:



$$\begin{array}{c} \text{Users} \end{array} \begin{bmatrix} 5 & 3 & \dots & ? \\ ? & 2 & \dots & 4 \\ \vdots & \vdots & \ddots & \vdots \\ 5 & ? & \dots & ? \end{bmatrix} = \mathcal{P}_\Omega \left(\underbrace{\begin{bmatrix} 5 & 3 & \dots & 5 \\ 4 & 2 & \dots & 4 \\ \vdots & \vdots & \ddots & \vdots \\ 5 & 5 & \dots & 3 \end{bmatrix}}_{\text{Complete Ratings } \boldsymbol{X}} \right)$$

Items

Observed (Incomplete) Ratings $\boldsymbol{Y}$

We observe:

$$\underbrace{\boldsymbol{Y}}_{\text{Observed ratings}} = \mathcal{P}_\Omega \Big[\underbrace{\boldsymbol{X}}_{\text{Complete ratings}} \Big],$$

where $\Omega \doteq \big\{(i,j) \mid \text{user } i \text{ has rated product } j\big\}$.

Examples of Low-rank Modeling

Many other examples:

- Euclidean Distance Matrix Embedding
- Latent Semantic Indexing
- ...

Singular Value Decomposition

Theorem (Compact SVD)

Let $\boldsymbol{X} \in \mathbb{R}^{n_1 \times n_2}$ *be a matrix, and* $r = \operatorname{rank}(\boldsymbol{X})$. *Then there exist* $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$ *with numbers* $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ *and matrices* $\boldsymbol{U} \in \mathbb{R}^{n_1 \times r}$, $\boldsymbol{V} \in \mathbb{R}^{n_2 \times r}$, *such that* $\boldsymbol{U}^*\boldsymbol{U} = \boldsymbol{I}$, $\boldsymbol{V}^*\boldsymbol{V} = \boldsymbol{I}$ *and*

$$\boldsymbol{X} \;=\; \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^* \;=\; \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^*. \tag{3}$$

Properties of the SVD:

- The left (or right) singular vectors $\boldsymbol{u}_i$ are the eigenvectors of $\boldsymbol{X}\boldsymbol{X}^*$ (or $\boldsymbol{X}^*\boldsymbol{X}$).
- The nonzero singular values σ_i are the positive square roots of the positive eigenvalues λ_i of $\boldsymbol{X}^*\boldsymbol{X}$ (or $\boldsymbol{X}\boldsymbol{X}^*$).

Singular Vectors via Nonconvex Optimization

To compute singular vector, say $\boldsymbol{u}_1$, consider the optimization problem:

$$\min \varphi(\boldsymbol{q}) \;\equiv\; -\tfrac{1}{2}\boldsymbol{q}^*\boldsymbol{\Gamma}\boldsymbol{q} \quad \text{s.t.} \quad \|\boldsymbol{q}\|_2^2 = 1 \tag{4}$$

with $\boldsymbol{\Gamma} \doteq \boldsymbol{X}\boldsymbol{X}^*$.

$\boldsymbol{q}$ is a critical point of $\varphi(\boldsymbol{q})$ over the sphere $\mathbb{S}^{n-1}$ if and only if (**why?**):

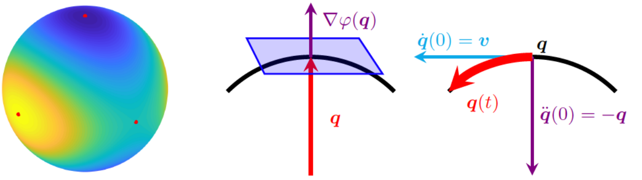
$$\nabla\varphi(\boldsymbol{q}) \propto \boldsymbol{q}. \tag{5}$$

The critical points are precisely the eigenvectors $\pm\boldsymbol{u}_i$ *of* $\boldsymbol{\Gamma}$*:*

$$\boldsymbol{\Gamma}\boldsymbol{q} = \lambda\boldsymbol{q} \quad \text{for some } \lambda. \tag{6}$$

All $\pm\boldsymbol{u}_i$ are unstable critical points of φ over $\mathbb{S}^{n-1}$ except $\pm\boldsymbol{u}_1$!

Singular Vectors via Nonconvex Optimization



Objective: $\varphi(\boldsymbol{q}) = -\frac{1}{2}\boldsymbol{q}^*\boldsymbol{\Gamma}\boldsymbol{q}$

Critical point: $\nabla\varphi(\boldsymbol{q}) = \lambda\boldsymbol{q}$

Curvature:

$$\frac{d^2}{dt^2}\varphi(\boldsymbol{q}(t))\Big|_{t=0} = \boldsymbol{v}^*\nabla^2\varphi(\boldsymbol{q})\boldsymbol{v} + \langle\nabla\varphi(\boldsymbol{q}), -\boldsymbol{q}\rangle\,\|\boldsymbol{v}\|_2^2.$$

A great circle (a *geodesics*) on $\mathbb{S}^{n-1}$: $\boldsymbol{q}(t) = \boldsymbol{q}\cos(t) + \boldsymbol{v}\sin(t)$ with $\boldsymbol{v} \perp \boldsymbol{q}$ and $\|\boldsymbol{v}\|_2 = 1$. The 2nd directional derivative of $\varphi(\boldsymbol{q}(t))$ at $\bar{\boldsymbol{q}} = \pm\boldsymbol{u}_i$:

$$\frac{d^2}{dt^2}\varphi(\boldsymbol{q}(t))\Big|_{t=0} = \underbrace{\boldsymbol{v}^*\nabla^2\varphi(\boldsymbol{q})\boldsymbol{v}}_{\text{Curvature of } \varphi} \; \underbrace{-\,\langle\nabla\varphi(\boldsymbol{q}), \boldsymbol{q}\rangle\boldsymbol{v}^*\boldsymbol{v}}_{\text{Curvature of the sphere}} = \boldsymbol{v}^*\underbrace{\big(-\boldsymbol{\Gamma} + \lambda_i\boldsymbol{I}\big)}_{\text{"Hessian"}}\boldsymbol{v}.$$

Best Low-Rank Matrix Approximation

Theorem (Best Low-rank Approximation)

Let $\boldsymbol{Y} \in \mathbb{R}^{n_1 \times n_2}$*, and consider the following optimization problem*

$$\min \|\boldsymbol{X} - \boldsymbol{Y}\|_F \quad \textit{s.t.} \quad \operatorname{rank}(\boldsymbol{X}) \leq r. \tag{7}$$

The optimal solution $\hat{\boldsymbol{X}}$ *has the form* $\hat{\boldsymbol{X}} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^*$*, where* $\boldsymbol{Y} = \sum_{i=1}^{\min(n_1,n_2)} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^*$ *is the singular value decomposition of* $\boldsymbol{Y}$*.*

The same solution (truncating the SVD) applies to minimizing the rank of the unknown matrix $\boldsymbol{X}$, subject to a data fidelity constraint:

$$\min \operatorname{rank}(\boldsymbol{X}) \quad \text{s.t.} \quad \|\boldsymbol{X} - \boldsymbol{Y}\|_F \leq \epsilon. \tag{8}$$

General Rank Minimization

Problem: recover a low-rank matrix $\boldsymbol{X}$ from linear measurements:

$$\min \operatorname{rank}(\boldsymbol{X}) \quad \text{subject to} \quad \mathcal{A}[\boldsymbol{X}] = \boldsymbol{y} \tag{9}$$

where $\boldsymbol{y} \in \mathbb{R}^m$ is an observation and $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ is a linear map:

$$\mathcal{A}[\boldsymbol{X}] = (\langle \boldsymbol{A}_1, \boldsymbol{X} \rangle, \ldots, \langle \boldsymbol{A}_m, \boldsymbol{X} \rangle), \quad \boldsymbol{A}_i \in \mathbb{R}^{n_1 \times n_2}. \tag{10}$$

Since $\operatorname{rank}(\boldsymbol{X}) = \|\boldsymbol{\sigma}(\boldsymbol{X})\|_0$, the problem is equivalent to the (NP-hard) ℓ^0 minimization:

$$\min \|\boldsymbol{\sigma}(\boldsymbol{X})\|_0 \quad \text{subject to} \quad \mathcal{A}[\boldsymbol{X}] = \boldsymbol{y}. \tag{11}$$

Convex Relaxation of Rank Minimization

Replace the rank, which is the ℓ^0 norm $\boldsymbol{\sigma}(\boldsymbol{X})$ with the ℓ^1 norm of $\boldsymbol{\sigma}(\boldsymbol{X})$:

$$\text{Nuclear norm:} \quad \|\boldsymbol{X}\|_* \doteq \|\boldsymbol{\sigma}(\boldsymbol{X})\|_1 = \sum_i \sigma_i(\boldsymbol{X}). \tag{12}$$

This is also known as the *trace norm* (for symmetric positive semidefinite matrices), the *Schatten* 1*-norm*, or the *Ky-Fan* k*-norm*.

Nuclear norm minimization problem:

$$\min \|\boldsymbol{X}\|_* \quad \text{subject to} \quad \mathcal{A}[\boldsymbol{X}] = \boldsymbol{y}. \tag{13}$$

Nuclear Norm – Convex Envelope of Rank

Why $\|X\|_*$ is a norm (hence convex)?

Theorem

For $M \in \mathbb{R}^{n_1 \times n_2}$, let $\|M\|_* = \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(M)$. Then $\|\cdot\|_*$ is a norm. Moreover, the nuclear norm and the the spectral norm are dual norms*:*

$$\|M\|_* = \sup_{\|N\| \le 1} \langle M, N \rangle, \quad \textit{and} \quad \|M\| = \sup_{\|N\|_* \le 1} \langle M, N \rangle. \tag{14}$$

Why $\|X\|_*$ is tight to approximate rank(X)?

Theorem

$\|M\|_*$ *is the convex envelope of* $\operatorname{rank}(M)$ *over*

$$\mathsf{B}_{op} \doteq \{M \mid \|M\| \le 1\}. \tag{15}$$

Nuclear Norm – Variational Forms

How to compute besides SVD?

$\|\boldsymbol{X}\|_*$ is equivalent to the following variational forms:

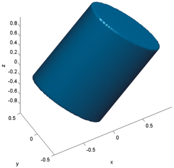
1. $\|\boldsymbol{X}\|_* = \min_{\boldsymbol{U},\boldsymbol{V}} \frac{1}{2}(\|\boldsymbol{U}\|_F^2 + \|\boldsymbol{V}\|_F^2)$, s.t. $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{V}^*$.
2. $\|\boldsymbol{X}\|_* = \min_{\boldsymbol{U},\boldsymbol{V}} \|\boldsymbol{U}\|_F \|\boldsymbol{V}\|_F$, s.t. $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{V}^*$.
3. $\|\boldsymbol{X}\|_* = \min_{\boldsymbol{U},\boldsymbol{V}} \sum_k \|\boldsymbol{u}_k\|_2 \|\boldsymbol{v}_k\|_2$, s.t. $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{V}^* \doteq \sum_k \boldsymbol{u}_k \boldsymbol{v}_k^*$.

These are useful in parameterizing low-rank matrices and finding them numerically, say via optimization.

Success of Nuclear Norm – Geometric Intuition

Nuclear norm ball: consider the set of 2×2 symmetric matrices, parameterized as

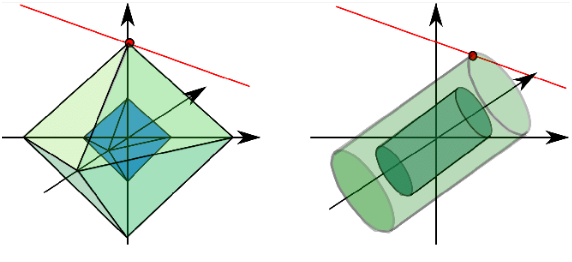
$$\boldsymbol{M} = \left[\begin{smallmatrix} x & y \\ y & z \end{smallmatrix}\right] \in \mathbb{R}^{2\times 2}.$$



The nuclear norm (unit) ball

$$\mathsf{B}_* = \{\boldsymbol{M} \mid \|\boldsymbol{M}\|_* \leq 1\}$$

is a cylinder in $\mathbb{R}^3$. The two circles at both ends of the cylinder correspond to matrices of rank 1.



(a) (b)

Rank-RIP

Definition (Rank-Restricted Isometry Property)

The operator $\mathcal{A}$ has the rank-restricted isometry property of rank r with constant δ, if $\forall\, \boldsymbol{X}$'s that satisfy $\operatorname{rank}(\boldsymbol{X}) \le r$, we have

$$(1-\delta)\|\boldsymbol{X}\|_F^2 \;\le\; \|\mathcal{A}[\boldsymbol{X}]\|_2^2 \;\le\; (1+\delta)\|\boldsymbol{X}\|_F^2. \tag{16}$$

The rank-r restricted isometry constant $\delta_r(\mathcal{A})$ is the smallest δ such that the above property holds.

Theorem (Uniqueness)

If $\boldsymbol{y} = \mathcal{A}[\boldsymbol{X}_o]$, with $r = \operatorname{rank}(\boldsymbol{X}_o)$ and $\delta_{2r}(\mathcal{A}) < 1$, then $\boldsymbol{X}_o$ is the unique optimal solution to the rank minimization problem

$$\min \operatorname{rank}(\boldsymbol{X}) \quad \textit{subject to} \quad \mathcal{A}[\boldsymbol{X}] = \boldsymbol{y}. \tag{17}$$

Rank-RIP for Nuclear Norm Minimization

Theorem (Nuclear Norm Minimization Success)

Suppose that $\boldsymbol{y} = \mathcal{A}[\boldsymbol{X}_o]$ *with* $\operatorname{rank}(\boldsymbol{X}_o) \le r$, *and that* $\delta_{4r}(\mathcal{A}) \le \sqrt{2} - 1$. *Then* $\boldsymbol{X}_o$ *is the unique optimal solution to the nuclear norm minimization problem*

$$\min \|\boldsymbol{X}\|_* \quad \textit{subject to} \quad \mathcal{A}[\boldsymbol{X}] = \boldsymbol{y}. \tag{18}$$

Let $\boldsymbol{X}_o = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*$ be the SVD of the true solution $\boldsymbol{X}_o$.

"Support" of X_o (compared to I of $\boldsymbol{x}_o$):

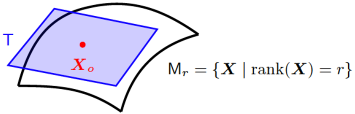
$$\mathsf{T} \doteq \{\boldsymbol{U}\boldsymbol{R}^* + \boldsymbol{Q}\boldsymbol{V}^* \mid \boldsymbol{R} \in \mathbb{R}^{n_2 \times r},\ \boldsymbol{Q} \in \mathbb{R}^{n_1 \times r}\} \subseteq \mathbb{R}^{n_1 \times n_2}. \tag{19}$$

"Sign" of X_o (compared to $\boldsymbol{\sigma}$ of $\boldsymbol{x}_o$):
$\boldsymbol{U}\boldsymbol{V}^*$ plays the role of the "signs" of $\boldsymbol{X}_o$ since $\boldsymbol{U}\boldsymbol{V}^* \in \mathsf{T}$ and

$$\langle \boldsymbol{X}_o, \boldsymbol{U}\boldsymbol{V}^* \rangle = \|\boldsymbol{X}_o\|_*. \tag{20}$$

Rank-RIP for Nuclear Norm Minimization

Proof Ideas: very similar to certifying optimality of $\boldsymbol{x}_o$ for the ℓ^1 minimization.



Let $\hat{\boldsymbol{X}} = \boldsymbol{X}_o + \boldsymbol{H}$ be any optimal solution to our problem (18). Then $\boldsymbol{H} = \hat{\boldsymbol{X}} - \boldsymbol{X}_o \in \operatorname{null}(\mathcal{A})$ must satisfy the following **cone restriction**:

$$\|\mathcal{P}_{\mathsf{T}^\perp}[\boldsymbol{H}]\|_* \leq \|\mathcal{P}_{\mathsf{T}}[\boldsymbol{H}]\|_* . \tag{21}$$

Rank-RIP for Nuclear Norm Minimization

Definition (Matrix Restricted Strong Convexity)

The linear operator $\mathcal{A}$ satisfies the matrix *restricted strong convexity* (RSC) condition of rank r with constant α if for the support T of every matrix of rank r and for all nonzero $\boldsymbol{H}$ satisfying

$$\|\mathcal{P}_{\mathsf{T}^\perp}[\boldsymbol{H}]\|_* \leq \alpha \cdot \|\mathcal{P}_{\mathsf{T}}[\boldsymbol{H}]\|_* . \tag{22}$$

with some constant $\alpha \geq 1$, we have

$$\|\mathcal{A}[\boldsymbol{H}]\|_2^2 > \mu \cdot \|\boldsymbol{H}\|_F^2, \quad \text{for some } \mu > 0. \tag{23}$$

Rank-RIP for Nuclear Norm Minimization

Theorem (Rank-RIP Implies Matrix RSC)

If a linear operator $\mathcal{A}$ satisfies rank-RIP with $\delta_{4r} < \frac{1}{1+\alpha\sqrt{2}}$, then $\mathcal{A}$ satisfies the matrix-RSC condition of rank r with constant α.

Theorem (Rank-RIP of Gaussian Measurements)

If the measurement operator $\mathcal{A}$ is a random Gaussian map with entries i.i.d. $\mathcal{N}(0, \frac{1}{m})$, then $\mathcal{A}$ satisfies the rank-RIP with constant $\delta_r(\mathcal{A}) \leq \delta$ with high probability, provided $m \geq Cr(n_1 + n_2) \times \delta^{-2} \log \delta^{-1}$, where $C > 0$ is a numerical constant.

Proofs precisely emulate those of ℓ^1 minimization for sparsity!

Matrices with Rank-RIP

Theorem

Let us assume $\{\boldsymbol{U}_1, \boldsymbol{U}_2, \ldots, \boldsymbol{U}_{n^2}\} \subset \mathbb{C}^{n\times n}$ *be a unitary basis for the matrix space* $\mathbb{C}^{n\times n}$ *and with* $\|\boldsymbol{U}_i\| \leq \zeta/\sqrt{n}$ *for some constant* ζ*. Let* $\mathcal{A}$ *consists of* m *randomly selected* $\boldsymbol{A}_i = \frac{n}{\sqrt{m}}\boldsymbol{U}_i$*. Then if*

$$m \geq C\zeta^2 \cdot rn\log^6 n, \tag{24}$$

then w.h.p., $\mathcal{A}$ *satisfies the rank-RIP over the set of all rank-*r *matrices.*

Example: Pauli observables of quantum states.
$\boldsymbol{U}_i = \boldsymbol{P}_1 \otimes \cdots \otimes \boldsymbol{P}_k$ where $\otimes$ is the tensor (Kronecker) product and each $\boldsymbol{P}_i = \frac{1}{\sqrt{2}}\boldsymbol{\sigma}$ where $\boldsymbol{\sigma}$ is a 2×2 matrix of the four possibilities:

$$\boldsymbol{\sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{\sigma}_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \boldsymbol{\sigma}_3 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \boldsymbol{\sigma}_4 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Noisy Measurements - Worst Case

Theorem (Stable Low-rank Recovery via BPDN)

Suppose that $\boldsymbol{y} = \mathcal{A}[\boldsymbol{X}_o] + \boldsymbol{z}$, *with* $\|\boldsymbol{z}\|_2 \le \epsilon$, *and let* $\operatorname{rank}(\boldsymbol{X}_o) = r$. *If* $\delta_{4r}(\mathcal{A}) < \sqrt{2} - 1$, *then any solution* $\hat{\boldsymbol{X}}$ *to the optimization problem*

$$\min \|\boldsymbol{X}\|_* \quad \textit{subject to} \quad \|\mathcal{A}[\boldsymbol{X}] - \boldsymbol{y}\|_2 \le \epsilon. \tag{25}$$

satisfies

$$\|\hat{\boldsymbol{X}} - \boldsymbol{X}_o\|_F \le C\epsilon \tag{26}$$

for some numerical constant $C > 0$.

Noisy Measurements - High Probability

Theorem (Stable Low-rank Recovery via Lasso)

Suppose that $\mathcal{A} \sim_{iid} \mathcal{N}(0, \frac{1}{m})$, *and* $\boldsymbol{y} = \mathcal{A}[\boldsymbol{X}_o] + \boldsymbol{z}$, *with* $\boldsymbol{X}_o$ *of rank* r *and* $\boldsymbol{z} \sim_{iid} \mathcal{N}(0, \frac{\sigma^2}{m})$. *Solve the matrix Lasso*

$$\min \tfrac{1}{2} \left\| \boldsymbol{y} - \mathcal{A}[\boldsymbol{X}] \right\|_2^2 + \lambda_m \left\| \boldsymbol{X} \right\|_*, \tag{27}$$

with regularization parameter $\lambda_m = c \cdot 2\sigma \sqrt{\frac{(n_1+n_2)}{m}}$ *for a large enough* c. *Then with high probability,*

$$\|\hat{\boldsymbol{X}} - \boldsymbol{X}_o\|_F \ \leq \ C'\sigma\sqrt{\frac{r(n_1+n_2)}{m}}. \tag{28}$$

Inexact Low Rank

Theorem (Inexact Low-rank Recovery)

Let $\boldsymbol{y} = \mathcal{A}[\boldsymbol{X}_o] + \boldsymbol{z}$, *with* $\|\boldsymbol{z}\|_2 \le \epsilon$. *Let* $\hat{\boldsymbol{X}}$ *solve the denoising problem*

$$\min \|\boldsymbol{X}\|_* \quad \textit{subject to} \quad \|\boldsymbol{y} - \mathcal{A}[\boldsymbol{X}]\|_2 \le \epsilon. \tag{29}$$

Then for any r *such that* $\delta_{4r}(\boldsymbol{A}) < \sqrt{2} - 1$,

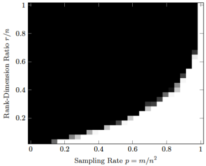
$$\left\| \hat{\boldsymbol{X}} - \boldsymbol{X}_o \right\|_2 \le C \frac{\|\boldsymbol{X}_o - [\boldsymbol{X}_o]_r\|_*}{\sqrt{r}} + C'\epsilon \tag{30}$$

for some constants C *and* C'.

Phase Transition

Let D be the descent cone of the nuclear norm at $\boldsymbol{X}_o \in \mathbb{R}^{n_1 \times n_2}$ of rank r:

$$\mathsf{D} \doteq \{\boldsymbol{H} \mid \|\boldsymbol{X}_o + \boldsymbol{H}\|_* \leq \|\boldsymbol{X}_o\|_*\}. \qquad (31)$$



Theorem (Phase Transition in Low-rank Recovery)

Let $\boldsymbol{G}$ be an $(n_1 - r) \times (n_2 - r)$ matrix with entries i.i.d. $\mathcal{N}(0,1)$. Set

$$\psi(n_1, n_2, r) = \inf_{\tau \geq 0}\{r(n_1 + n_2 - r + \tau^2) + \mathbb{E}_{\boldsymbol{G}}\left[\|\mathcal{D}_\tau[\boldsymbol{G}]\|_F^2\right]\}. \qquad (32)$$

Then there is a phase transition at $m^\star = \delta(\mathsf{D})$:

$$\psi(n_1, n_2, r) - 2\sqrt{n_2/r} \;\leq\; \delta(\mathsf{D}) \;\leq\; \psi(n_1, n_2, r). \qquad (33)$$

Summary

Parallel developments for sparse vectors and low-rank matrices.

Sparse v.s. Low-rank	Sparse Vector	Low-rank Matrix
Low-dimensionality of	individual signal $\boldsymbol{x}$	a set of signals $\boldsymbol{X}$
Compressive sensing	$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$	$\boldsymbol{Y} = \mathcal{A}(\boldsymbol{X})$
Low-dim measure	ℓ^0 norm $\Vert\boldsymbol{x}\Vert_0$	rank$(\boldsymbol{X})$
Convex surrogate	ℓ^1 norm $\Vert\boldsymbol{x}\Vert_1$	nuclear norm $\Vert\boldsymbol{X}\Vert_*$
Success conditions (RIP)	$\delta_{2k}(\boldsymbol{A}) \geq \sqrt{2} - 1$	$\delta_{4r}(\boldsymbol{A}) \geq \sqrt{2} - 1$
Random measurements	$m = O\big(k \log(n/k)\big)$	$m = O(nr)$
Stable/Inexact recovery	$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{z}$	$\boldsymbol{Y} = \mathcal{A}(\boldsymbol{X}) + \boldsymbol{Z}$
Phase transition at	Stat. dim. of descent cone: $m^* = \delta(\mathrm{D})$	

Assignments

- Reading: Sections 4.1-4.3 of Chapter 4.
- Programming Homework # 2.