
Fall in love
with your next book
Let the booksellers of
Gower Street help you
find your perfect match.

Low-rank matrix completion: optimization on manifolds at work

Nicolas Boumal

Joint work with Pierre-Antoine Absil

Université catholique de Louvain

July 2011

Recommender systems tell you which items you might like

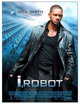
based on a huge database of ratings

$$\begin{array}{cc} & \leftarrow n \text{ cols} \rightarrow \\ X = & \begin{pmatrix} ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? \end{pmatrix} \quad \begin{matrix} \uparrow \\ m \text{ rows} \\ \downarrow \end{matrix} \end{array}$$

We record ratings of movies by the customers

One row per movie, one column per customer

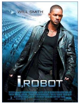
$$X = \begin{array}{cc} & \text{customer } j \\ \begin{pmatrix} ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? \end{pmatrix} & \begin{array}{l} \\ \text{movie } i \\ \\ \\ \\ \end{array} \end{array}$$



We record ratings of movies by the customers

One row per movie, one column per customer

$$
X = \begin{array}{cc}
\begin{matrix} & & & & & & & & & & \text{customer } j & & & & & \end{matrix} & \\
\begin{pmatrix}
? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? \\
? & ? & ? & ? & ? & ? & ? & ? & ? & ? & 4 & ? & ? & ? & ? & ? \\
? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? \\
? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? \\
? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ?
\end{pmatrix} & \begin{matrix} \\ \text{movie } i \\ \\ \\ \\ \end{matrix}
\end{array}
$$



Most ratings are unknown. Our job is to complete X.

$$X = \begin{pmatrix} 1 & ? & 2 & ? & ? & 5 & ? & ? & ? & ? & 5 & ? & ? & ? & 2 & ? \\ 2 & ? & 2 & ? & ? & 4 & ? & ? & ? & 3 & 4 & ? & 5 & ? & ? & ? \\ 1 & ? & 5 & 2 & ? & 4 & ? & 4 & ? & ? & ? & 2 & ? & ? & ? & ? \\ ? & 1 & ? & 3 & ? & ? & ? & 3 & ? & ? & 3 & ? & 2 & ? & 5 & 5 \\ 4 & 4 & ? & ? & ? & ? & 5 & ? & ? & ? & 1 & ? & ? & 1 & ? & 4 \end{pmatrix}$$

We could exploit similarities between customers

to complete the matrix

$$X = \begin{pmatrix} 1 & ? & 2 & ? & ? & 5 & ? & ? & ? & ? & 5 & ? & ? & ? & 2 & ? \\ 2 & ? & 2 & ? & ? & 4 & ? & ? & ? & 3 & 4 & ? & 5 & ? & ? & ? \\ 1 & ? & 5 & 2 & ? & 4 & ? & 4 & ? & ? & ? & 2 & ? & ? & ? & ? \\ ? & 1 & ? & 3 & ? & ? & ? & 3 & ? & ? & 3 & ? & 2 & ? & 5 & 5 \\ 4 & 4 & ? & ? & ? & ? & 5 & ? & ? & ? & 1 & ? & ? & 1 & ? & 4 \end{pmatrix}$$

In a global, automated, scalable way?

We assume that X has low-rank r

Hence, that ratings are inner products in a small space $\mathbb{R}^r$

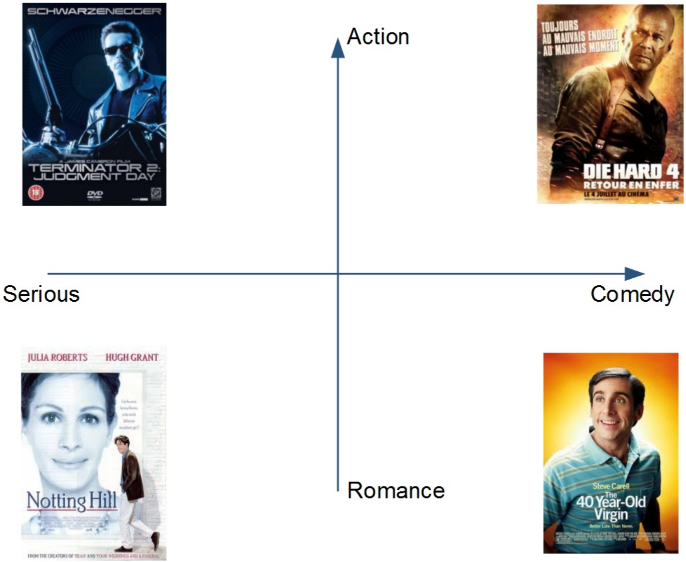
$$X = \begin{pmatrix} 1 & ? & 2 & ? & ? & 5 & ? & ? & ? & ? & 5 & ? & ? & ? & 2 & ? \\ 2 & ? & 2 & ? & ? & 4 & ? & ? & ? & 3 & 4 & ? & 5 & ? & ? & ? \\ 1 & ? & 5 & 2 & ? & 4 & ? & 4 & ? & ? & ? & 2 & ? & ? & ? & ? \\ ? & 1 & ? & 3 & ? & ? & ? & 3 & ? & ? & 3 & ? & 2 & ? & 5 & 5 \\ 4 & 4 & ? & ? & ? & ? & 5 & ? & ? & ? & 1 & ? & ? & 1 & ? & 4 \end{pmatrix}$$

$$\approx \begin{pmatrix} ? & ? \\ ? & ? \\ ? & ? \\ ? & ? \\ ? & ? \end{pmatrix} \cdot \begin{pmatrix} ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? \end{pmatrix}$$

Rationale: only a few factors influence our preferences

We map movies and customers to this r-D space

without any human intervention



Toward a reasonable formulation

The optimal choice UW is an m-by-n matrix of rank r in best agreement with the $|\Omega|$ known entries of X

$$\min_{U \in \mathbb{R}^{m\times r}, W \in \mathbb{R}^{r\times n}} \boxed{\underbrace{\sum_{(i,j)\in\Omega}}_{\text{known entries}} \Big((UW)_{ij} - \overbrace{X_{ij}}^{\text{known ratings}}\Big)^2}$$

This is reasonable, *but*

$$\min_{U \in \mathbb{R}^{m \times r}, W \in \mathbb{R}^{r \times n}} \sum_{(i,j) \in \Omega} \left((UW)_{ij} - X_{ij} \right)^2$$

This objective is invariant under invertible transformations.

This is reasonable, *but*

$$\min_{U \in \mathbb{R}^{m \times r}, W \in \mathbb{R}^{r \times n}} \boxed{\sum_{(i,j) \in \Omega} \big((UW)_{ij} - X_{ij}\big)^2}$$

This objective is invariant under invertible transformations.

The pair (U, W) *is equivalent to* $(UM, M^{-1}W)$,
for any r-by-r invertible matrix M.

We would like to get rid of the invariance. Why?

- The search space $\mathbb{R}^{m\times r} \times \mathbb{R}^{r\times n}$ is bigger than it ought to be;
- Convergence theorems often assume isolated critical points;
- And it may prevent superlinear convergence rates altogether.

We would like to get rid of the invariance. Why?

- The search space $\mathbb{R}^{m\times r} \times \mathbb{R}^{r\times n}$ is bigger than it ought to be;
- Convergence theorems often assume isolated critical points;
- And it may prevent superlinear convergence rates altogether.
- Also, 'cause we can.

Step 1: redefine the objective function

$$\min_{U \in \mathbb{R}^{m \times r}, W \in \mathbb{R}^{r \times n}} \boxed{\sum_{(i,j) \in \Omega} \big((UW)_{ij} - X_{ij}\big)^2}$$

Step 1: redefine the objective function

$$\min_{U\in\mathbb{R}^{m\times r}} \min_{W\in\mathbb{R}^{r\times n}} \boxed{\sum_{(i,j)\in\Omega} \left((UW)_{ij} - X_{ij}\right)^2}$$

Step 1: redefine the objective function

$$\min_{U \in \mathbb{R}^{m \times r}} \min_{W \in \mathbb{R}^{r \times n}} \boxed{\sum_{(i,j) \in \Omega} \left((UW)_{ij} - X_{ij} \right)^2}$$

W is the solution of a simple least squares problem.

Step 1: redefine the objective function

$$\min_{U \in \mathbb{R}^{m \times r}} \boxed{\min_{W \in \mathbb{R}^{r \times n}} \sum_{(i,j) \in \Omega} \left((UW)_{ij} - X_{ij} \right)^2}$$

$f(U)$

W is the solution of a simple least squares problem.

Step 2: identify the right space for the objective function

$$f(U) = \boxed{\min_{W \in \mathbb{R}^{r \times n}} \sum_{(i,j) \in \Omega} \left((UW)_{ij} - X_{ij} \right)^2}$$

For all invertible $M \in \mathbb{R}^{r \times r}$, $\quad f(U) = f(UM)$.

Step 2: identify the right space for the objective function

$$f(U) = \boxed{\min_{W \in \mathbb{R}^{r \times n}} \sum_{(i,j) \in \Omega} \left((UW)_{ij} - X_{ij}\right)^2}$$

For all invertible $M \in \mathbb{R}^{r \times r}$, $\quad f(U) = f(UM)$.

Hence, f is well defined over the set of equivalence classes $[U]$:

$$[U] = \{UM : M \in \mathbb{R}^{r \times r} \text{ is invertible}\}.$$

$$[U] = \{UM : M \in \mathbb{R}^{r \times r} \text{ is invertible}\}.$$

$$[U] = \{UM : M \in \mathbb{R}^{r \times r} \text{ is invertible}\}.$$

They have the same column space.

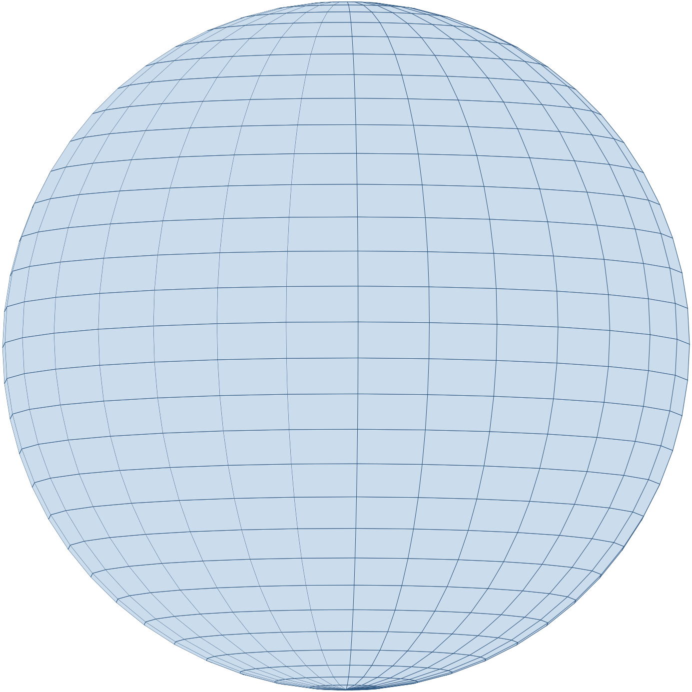
The set of equivalence classes is the Grassmann manifold $\mathrm{Gr}(m, r)$

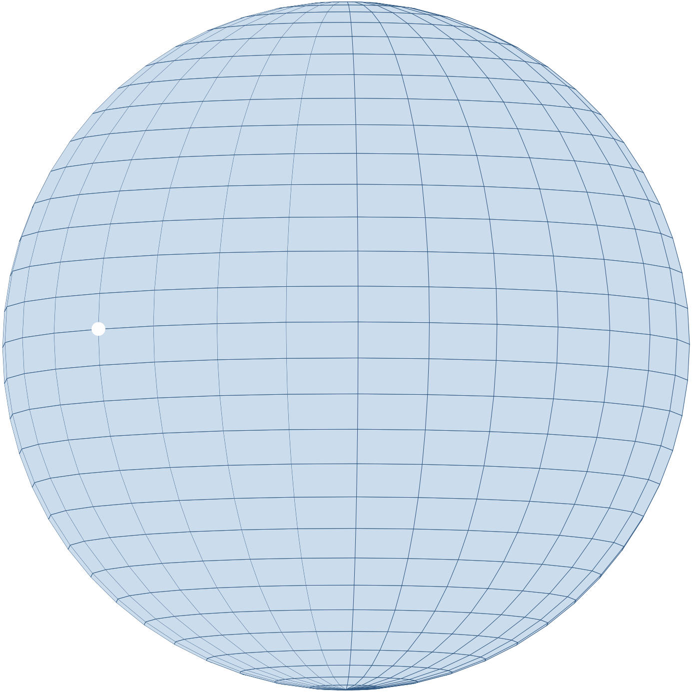
$$[U] = \{UM : M \in \mathbb{R}^{r\times r} \text{ is invertible}\}.$$

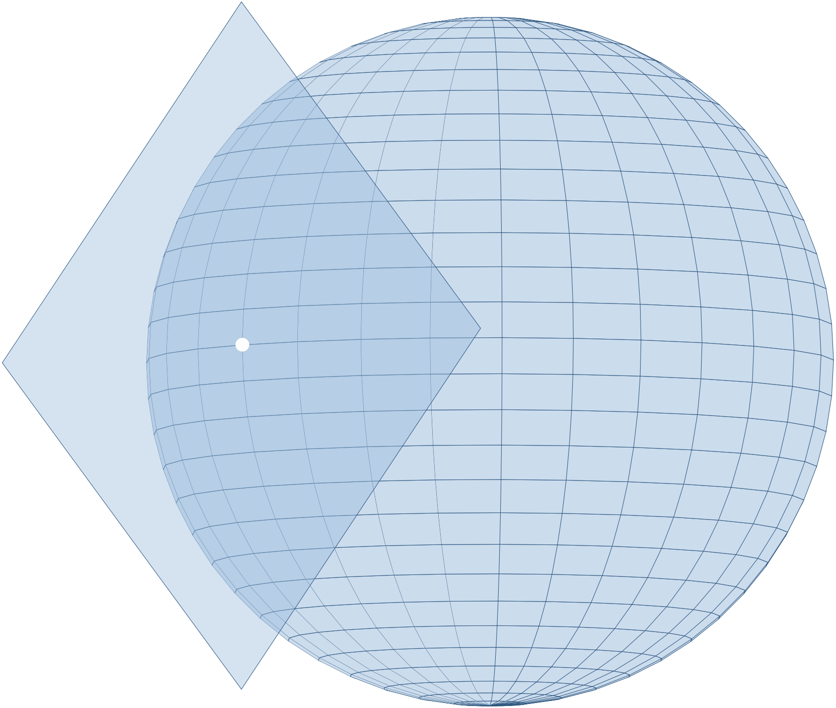
They have the same column space.

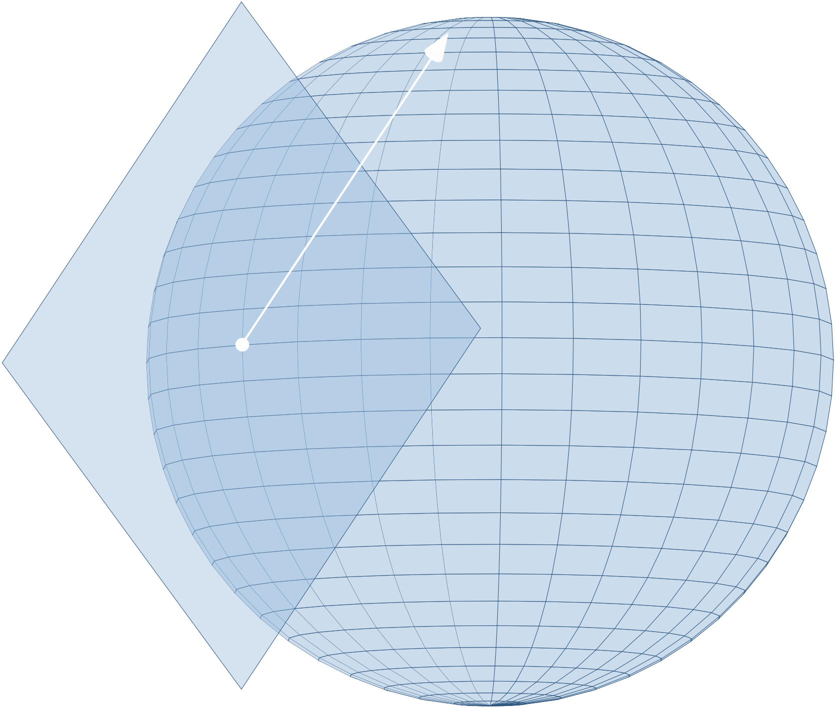
- $\mathrm{Gr}(m, r)$ is the set of r-dimensional subspaces of $\mathbb{R}^m$;
- It has a well studied smooth manifold structure.

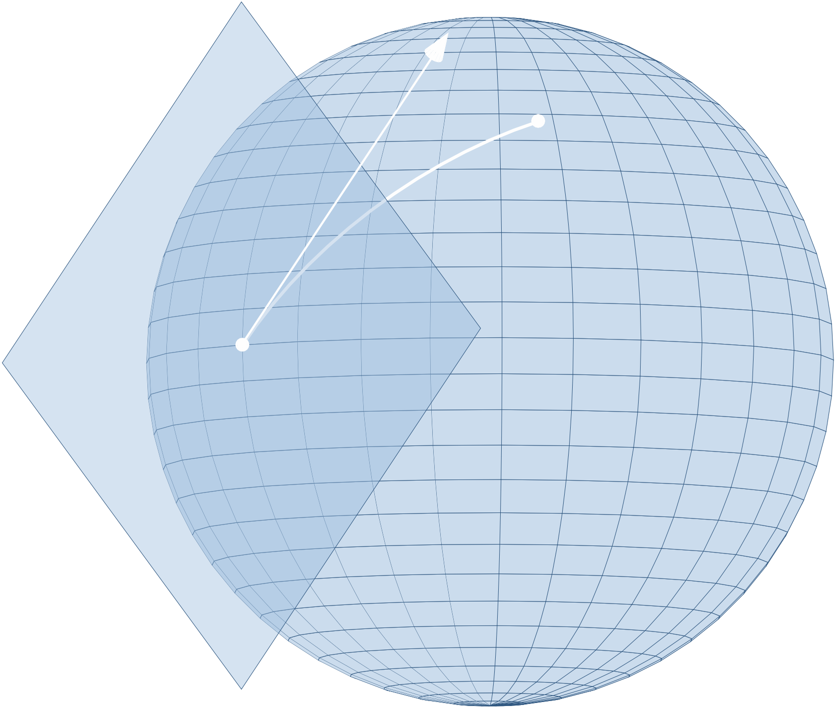
How do you minimize $f([U])$ over the Grassmann manifold?











We use a Riemannian trust-region method: GenRTR

- Absil et al. (2007) generalized trust-region methods to manifolds.

- GenRTR is numerically efficient and comes with proofs.

A few numerical tests

We compare seven algorithms

RTRMC is our method,
with Hessian (2^{nd} order) and without Hessian (1^{st} order);

OptSpace by Keshavan and Oh;

Balanced Factorization by Meyer and Sepulchre;

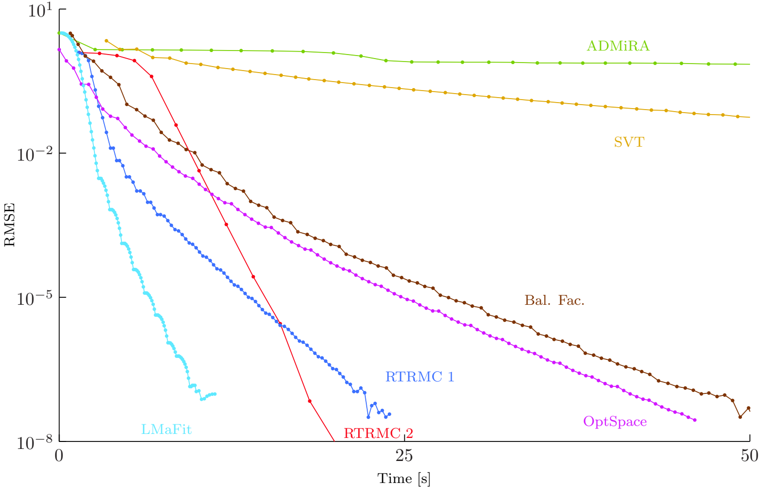
SVT by Candès and Becker;

ADMiRA by Lee and Bresler;

LMaFit by Wen, Yin and Zhang.

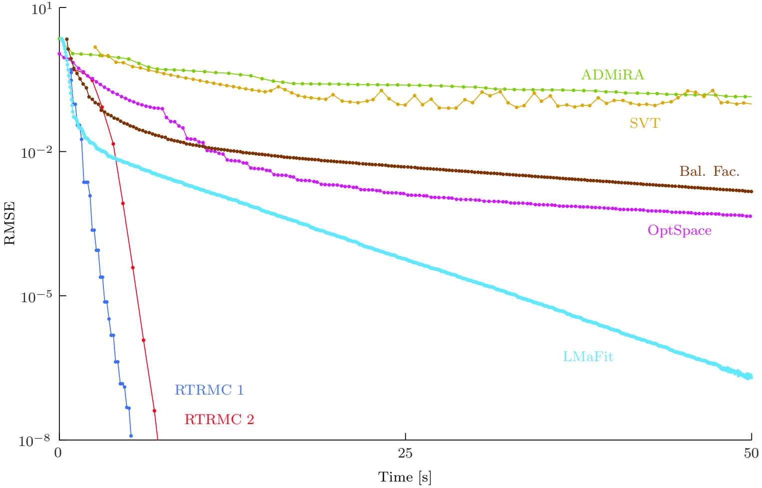
LMaFit is a serious contestant.

$m = n = 10\,000$, $r = 10$, knowledge $= 1\%$



Our method is fastest on rectangular matrices.

$m = 1\,000$, $n = 30\,000$, $r = 5$, knowledge $= 2.6\%$



Final thoughts

Take home message

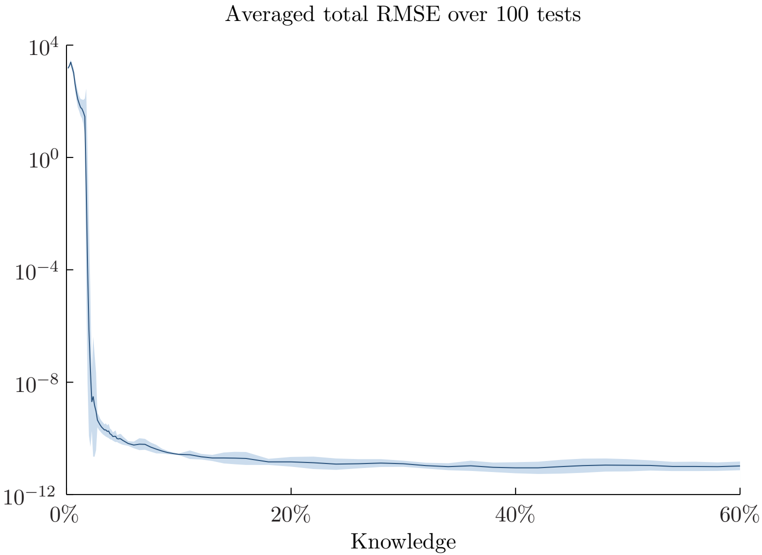
Many practical optimization problems live on a manifold.

Efficient tools are readily available to exploit their geometry.

These tools are backed up by a solid theory.

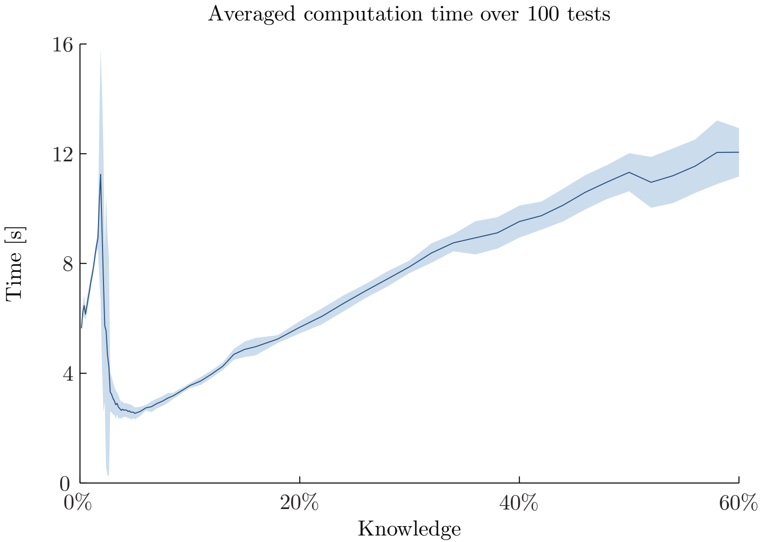
Given enough information, we consistently recover X.

Noiseless, $m = n = 1\,000$, $r = 5$



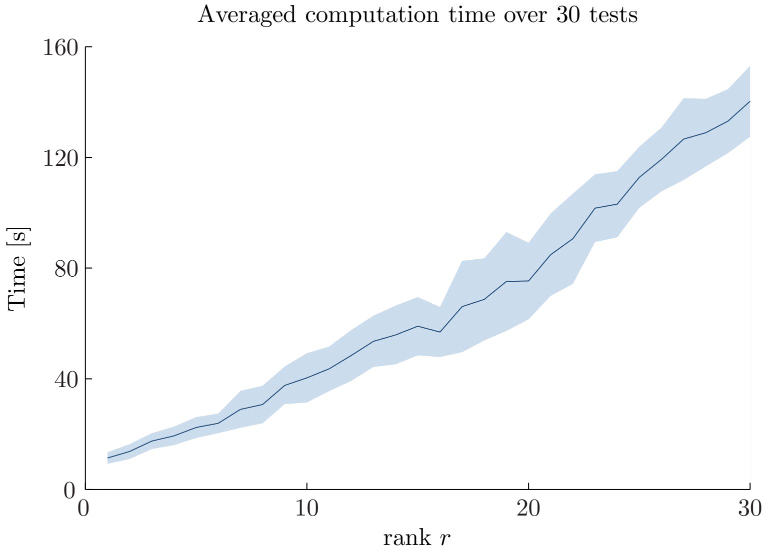
Computation time is proportional to ♯ known entries.

Noiseless, $m = n = 1\,000$, $r = 5$



Computation time scales reasonably with the rank.

Noiseless, $m = n = 1\,000$, knowledge $= 25\%$



In the presence of noise, we are close to optimal.

Noisy, $m = n = 500$, $r = 4$