Regularization Paths with Guarantees for Convex Semidefinite Optimization

Joachim Giesen — FSU Jena, DE
Martin Jaggi — ETH Zurich, CH
Sören Laue — FSU Jena, DE

Pathwise Optimization

Parameterized Convex Optimization:
Minimize a convex function $f_t(x)$
over a compact convex domain $x \in D$.

$$\min_{x \in D} f_t(x)$$

The objective is parameterized by an additional parameter t (e.g. a regularization parameter)

Solution Path:
Maintain an optimal solution $x^*(t)$ along the entire path, as the parameter t changes.

Approximate Solution Path:
Maintain an ε-approximate solution, along the path in t.

Measure of approximation quality: $g_t(x) \le \varepsilon$
Duality gap (quality certificate, easy to compute!)

Path-Following Idea: A Piecewise Constant Path

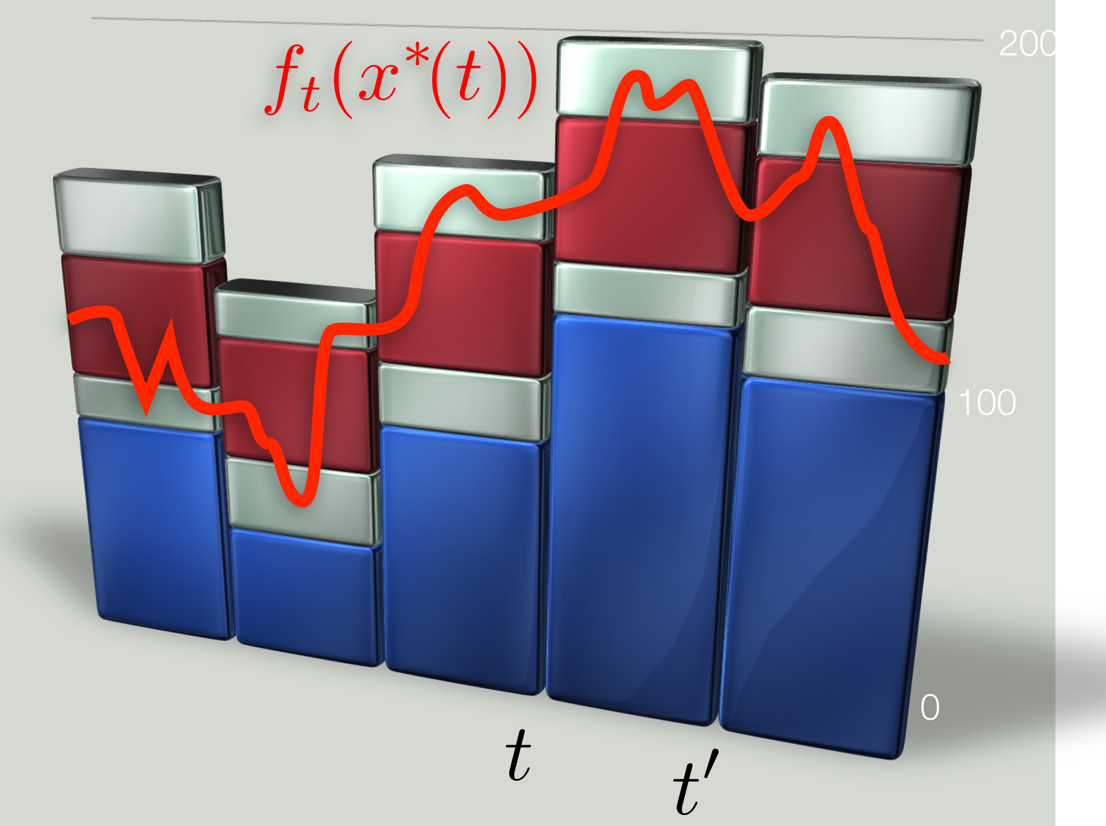
Goal: Guarantee small duality gap $g_t(x) \le \varepsilon$ along the entire path in t

Idea: Keep x constant, change t as far as possible

At the current value t, compute an approximate solution x, of a quality slightly better than necessary $g_t(x) \le \frac{\varepsilon}{2}$

How far can we change the parameter $t \to t'$ such that x is still good enough at t' ? $g_{t'}(x) \le \varepsilon$

Update $t := t'$, and repeat



Stability of Approximate Solutions:
Any t' satisfying $g_{t'}(x) - g_t(x) \le \varepsilon - \frac{\varepsilon}{2}$
will maintain the ε-guarantee for x.
When the duality gap changes continuously in t, we have intervals of size at least $\Omega(\varepsilon)$

Path Complexity

$$\Theta\left(\frac{1}{\varepsilon}\right)$$

Number of intervals of piecewise constant solutions

Theorem 6. *Let f_t be convex and continuously differentiable in X, and let $\nabla f_t(X)$ be Lipschitz continuous in t with Lipschitz constant L, for all feasible X. Then the ε-approximation path complexity of Problem (1) over the parameter range $[t_{\min}, t_{\max}] \subset \mathbb{R}$ is at most*

$$\left\lceil \frac{2L \cdot \gamma}{\gamma - 1} \cdot \frac{t_{\max} - t_{\min}}{\varepsilon} \right\rceil = O\left(\frac{1}{\varepsilon}\right).$$

Applications to Semidefinite Optimization

SDPs
with bounded trace

$$\begin{array}{rl} \min\limits_{X \in \mathbb{R}^{n \times n}} & f_t(X) \\ s.t. & \mathrm{Tr}(X) = 1 \\ & X \succeq 0 \end{array} \qquad (1)$$

Nuclear-Norm Regularized Optimization

arbitrary convex function on matrices

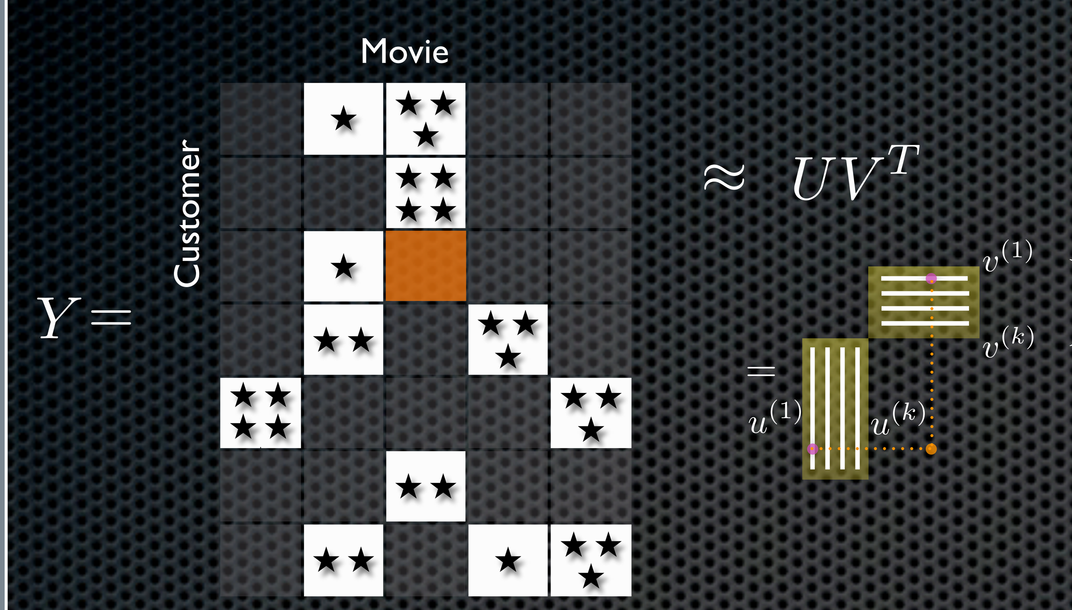
also called „trace norm", sum of singular values

$$\min_{Z \in \mathbb{R}^{m \times n}} f(Z) + \lambda \|Z\|_*$$

Here: constrained variant

$$\begin{array}{rl} \min\limits_{Z} & f(Z) \\ s.t. & \|Z\|_* \le t \end{array}$$

Matrix completion for recommender systems



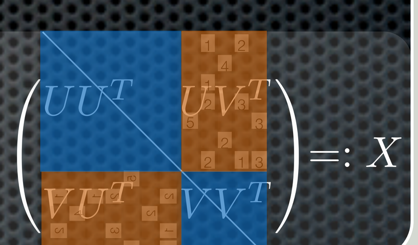
Given a small sample of entries of a matrix, we want to predict all its entries.

Use a **nuclear norm regularization**!

$$\begin{array}{rl} \min\limits_{U,V} & \sum\limits_{(i,j) \in \Omega} (Y_{ij} - (UV^T)_{ij})^2 \\ s.t. & \|U\|_{Fro}^2 + \|V\|_{Fro}^2 \le t \end{array}$$

Is equivalent to:

$$\begin{array}{rl} \min\limits_{X \succeq 0} & f(X) \\ s.t. & Tr(X) \le t \end{array}$$



Experimental Results

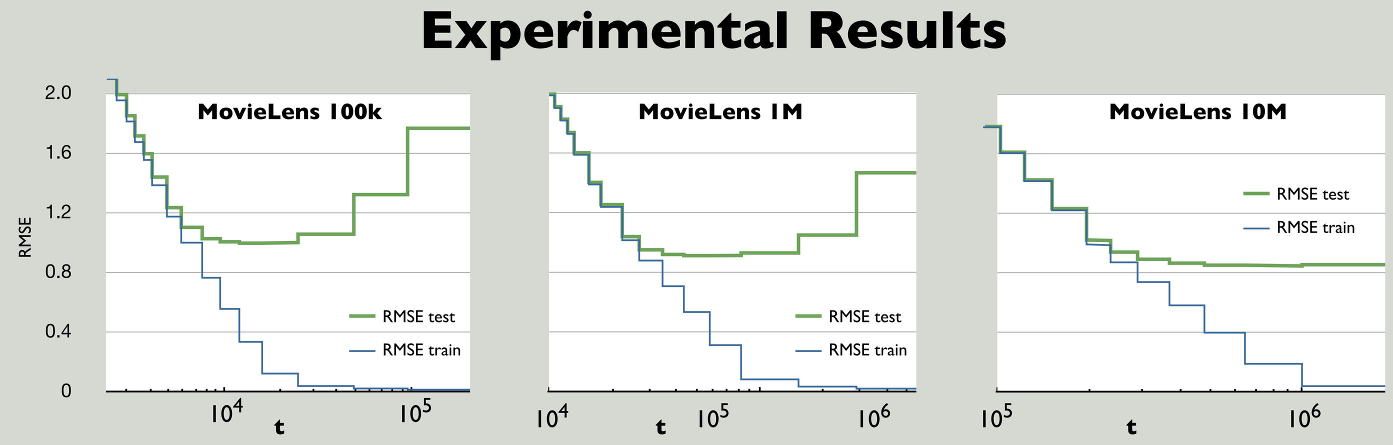


Figure 1: The nuclear norm regularization path for the three MovieLens datasets.

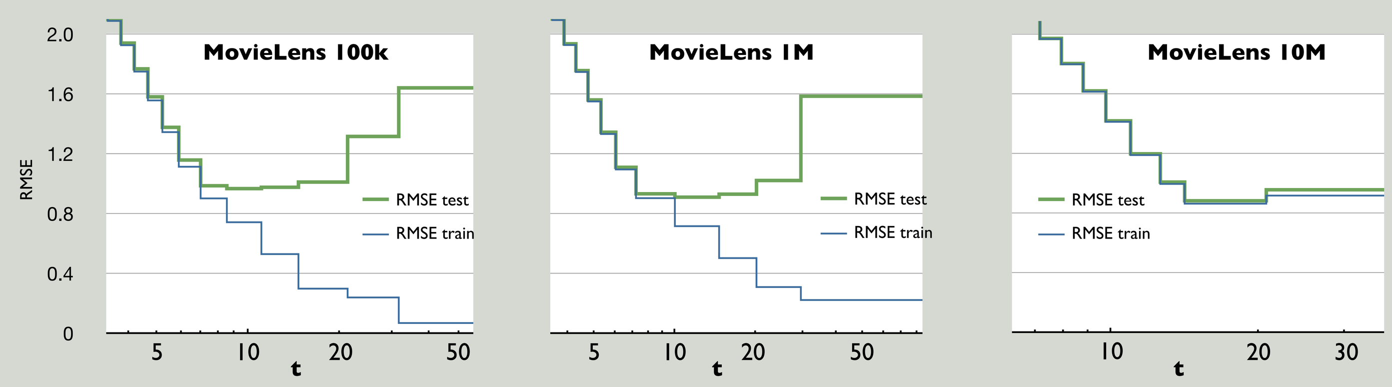


Figure 2: The regularization path for the weighted nuclear norm $\|.\|_{nuc(p,q)}$.

Weighted Nuclear Norm

$$\|Z\|_{nuc(p,q)} := \|PZQ\|_*$$ for P,Q diagonal

Can be reduced to the classical nuclear norm!

weighted nuclear-norm		**classic** nuclear-norm
$\begin{array}{rl} \min\limits_{Z \in \mathbb{R}^{m \times n}} & f(Z) \\ s.t. & \|Z\|_{nuc(p,q)} \le t \end{array}$	$\Leftrightarrow$	$\begin{array}{rl} \min\limits_{\bar{Z} \in \mathbb{R}^{m \times n}} & f(P^{-1}\bar{Z}Q^{-1}) \\ s.t. & \|\bar{Z}\|_* \le t \end{array}$

Guarantees and Algorithms translate to the weighted case

Robust PCA

$$\min_{Z \in \mathbb{R}^{m \times n}} \|Z\|_* + \lambda' \|M - Z\|_1$$

Our regularization path framework applies to the (nuclear norm) constrained variant, if the ℓ_1-loss is smoothened

A Variant of Sparse PCA

We obtain the regularization path for the SDP-relaxation

$$\begin{array}{rl} \min\limits_{X \in \mathbb{S}^{n \times n}} & \rho \cdot \mathbf{e}^T |X| \mathbf{e} - \mathrm{Tr}(MX) \\ s.t. & \mathrm{Tr}(X) = 1 \, , \\ & X \succeq 0 \end{array}$$

Conclusions

Approximate Path	Exact Path
ε-guarantee on gap, continuously along path	exact solution along path
widely applicable and practical	problem-specific, practical only for some problems
low complexity O(1/ε)	complexity can be exponential (in the worst case)
any approx. internal optimizer can be used	exact internal optimizers are necessary