Convex Optimization with a Parameter

&

Regularization Paths for Matrix Factorizations

Mittagseminar

2011 / 03 / 03

Martin Jaggi

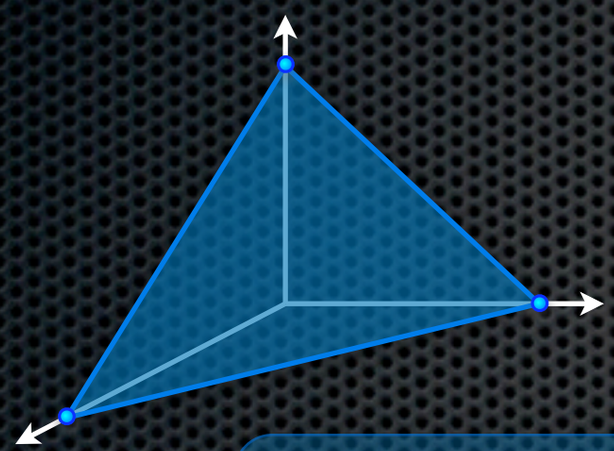
Joint work with Joachim Giesen and Soeren Laue

- Approximating Parameterized Convex Optimization Problems
  ESA 2010
- Regularization Paths with Guarantees for Nuclear Norm Regularized Matrix Factorizations
  Submitted

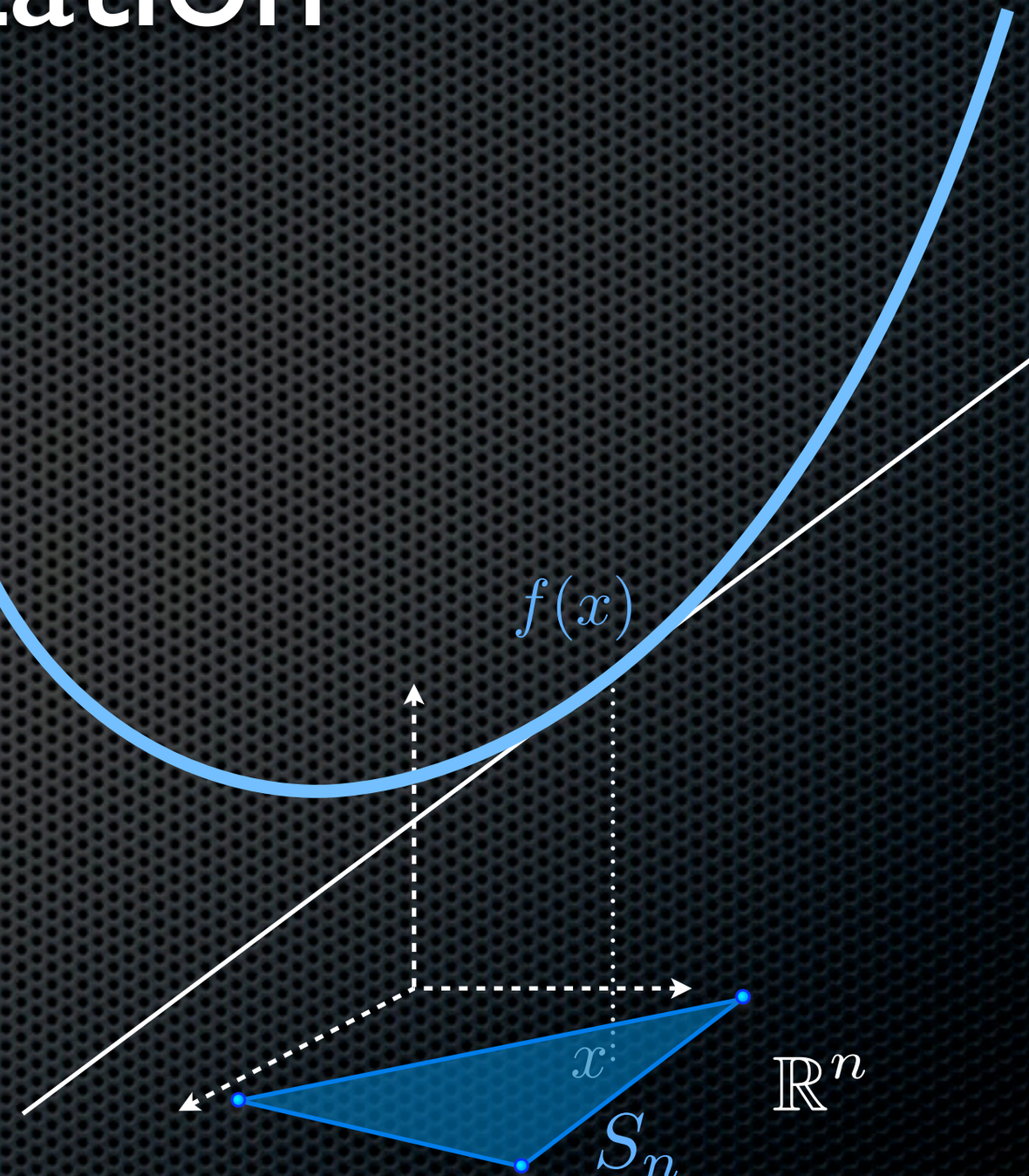
Convex Optimization

The Problem

$$\min_{x \in S_n} f(x)$$



$$S_n := \begin{array}{l} x \in \mathbb{R}^n \\ x \geq 0 \\ \mathbf{1}^T x = 1 \end{array}$$



f( ) convex, differentiable $\forall x, y$

$$f(y) \geq f(x) + (y - x)^T \nabla f(x)$$

Convex Optimization

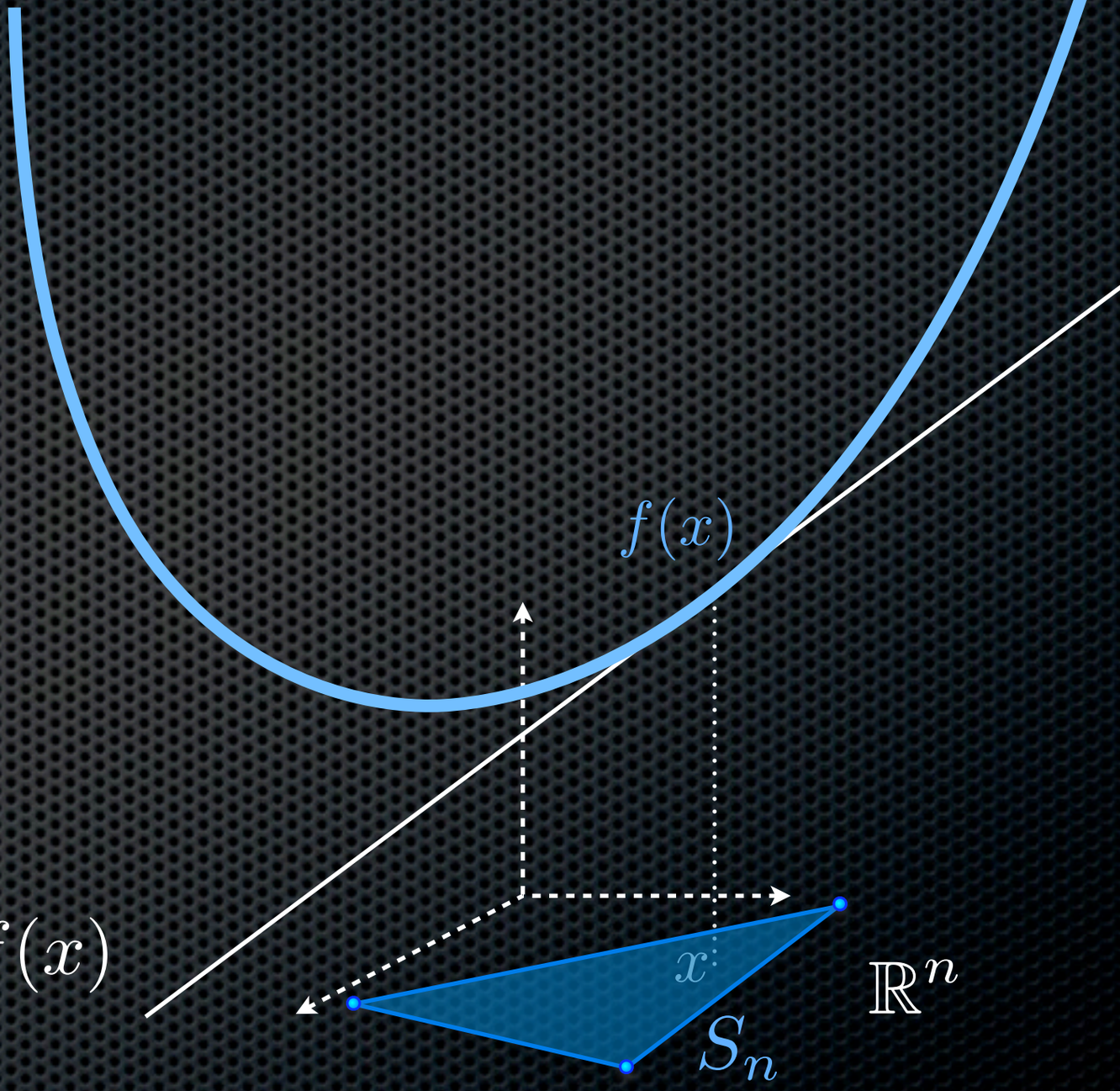
The Problem

$$\min_{x \in S_n} f(x)$$

The Dual

$$\omega(x) :=$$

$$\min_{y \in S_n} f(x) + (y - x)^T \nabla f(x)$$



f( ) convex, differentiable $\forall x, y$

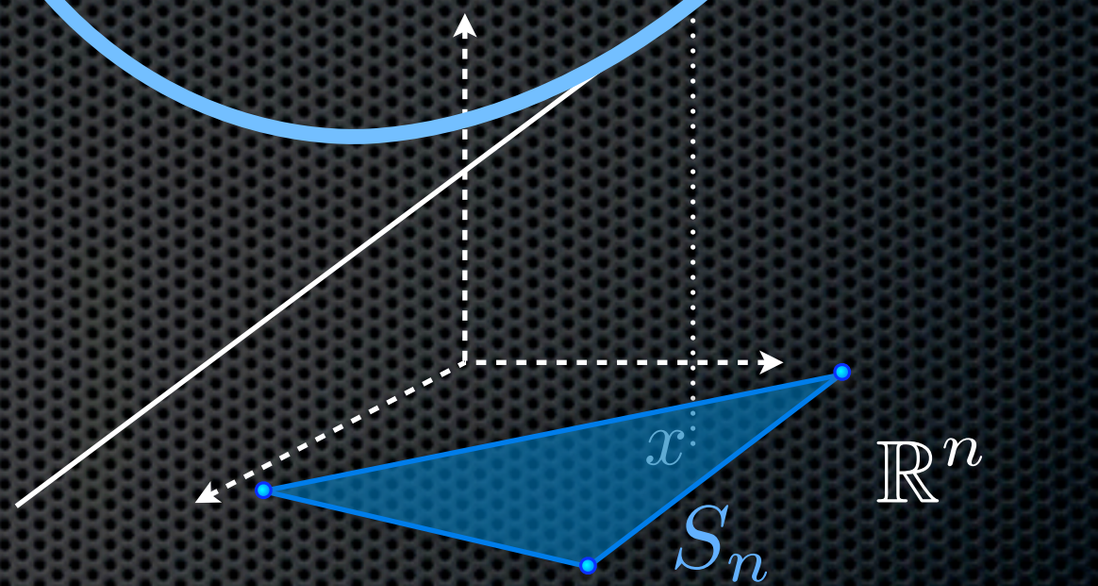
$$f(y) \geq f(x) + (y - x)^T \nabla f(x)$$

Mind the duality gap...

$$g(x) := f(x) - \omega(x)$$
$$= x^T \nabla f(x) - \min_{y \in S_n} y^T \nabla f(x)$$
$$= x^T \nabla f(x) - \min_i (\nabla f(x))_i$$

f( ) convex, differentiable $\forall x, y$

$$f(y) \geq f(x) + (y - x)^T \nabla f(x)$$



...as a measure of approximation

Lemma

$$f(x) - f(x^*) \leq g(x) \quad \forall x$$

Definition

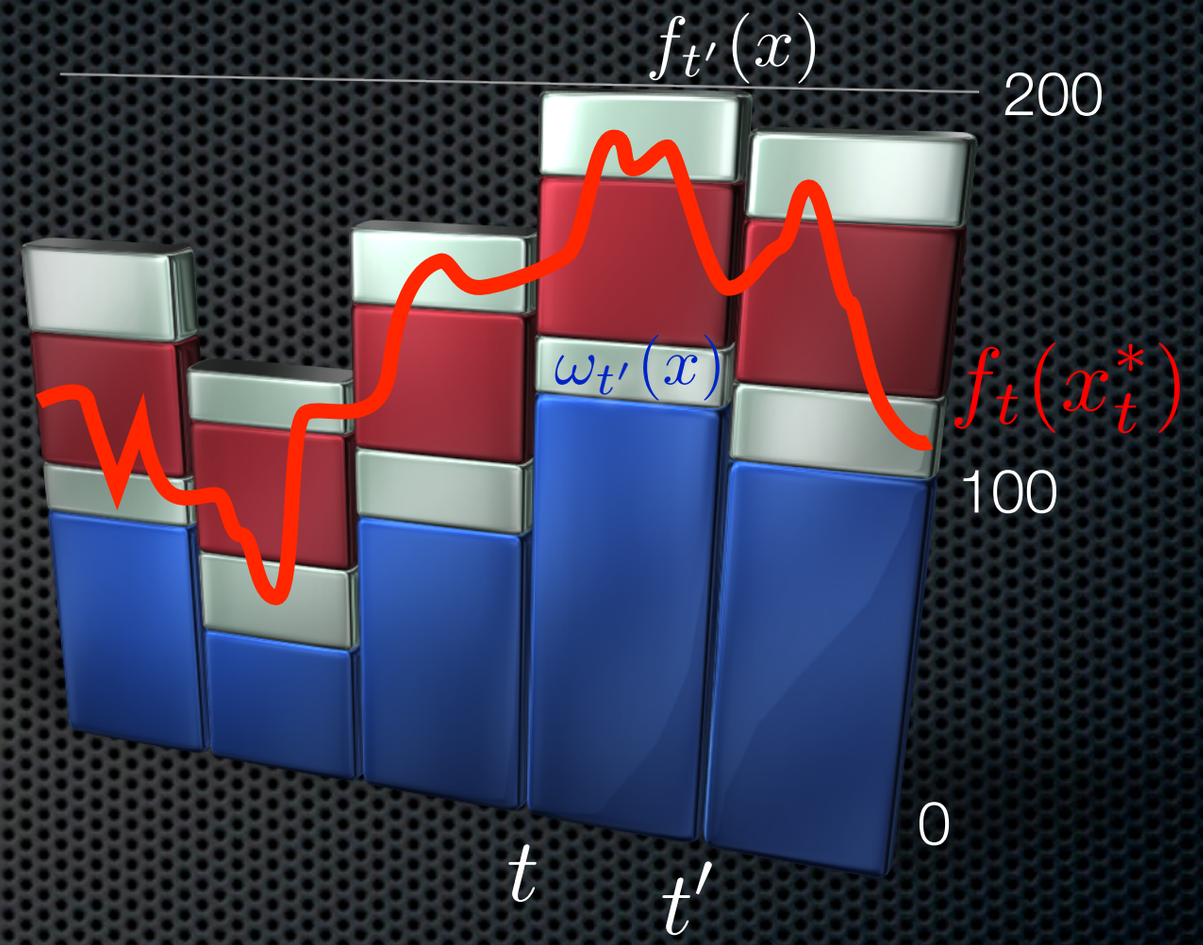
x is an ε-approximation

iff $g(x) \leq \varepsilon$

Pathwise Optimization

The Parameterized Problem

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



"Better than necessary"

$$g_t(x) \leq \tfrac{\varepsilon}{2}$$

The difference

$$g_{t'}(x) - g_t(x) \leq \varepsilon\left(1 - \tfrac{1}{2}\right) \quad \Leftarrow \quad |t' - t| \leq \varepsilon \cdot P_f$$

"Continuity in the parameter"

"Still good enough"

$$g_{t'}(x) \leq \varepsilon$$

Theorem

There are $O\left(\frac{1}{\varepsilon}\right)$ many intervals

of piecewise constant ε-approx. solutions.

Pathwise Optimization

The Parameterized Problem

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

"Better than necessary"

$$g_t(x) \leq \tfrac{\varepsilon}{2}$$

Example:

$$f(x) = x^T(A + t\hat{A})x + (b + t\hat{b})^T x$$

$$\nabla f_{t'}(x) - \nabla f_t(x)$$

$$= (t' - t)(x^T \hat{A} + \hat{b}^T)$$

The difference

$$g_{t'}(x) - g_t(x) \leq \varepsilon\left(1 - \tfrac{1}{2}\right)$$

$$\Leftarrow \quad |t' - t| \leq \varepsilon \cdot P_f$$

"Continuity in the parameter"

"Still good enough"

$$g_{t'}(x) \leq \varepsilon$$

Theorem

There are $O\left(\frac{1}{\varepsilon}\right)$ many intervals

of piecewise constant ε-approx. solutions.

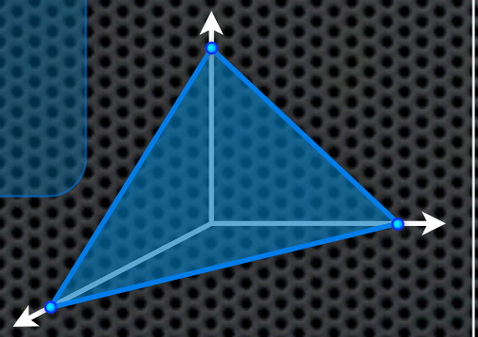
Matrix Problems / Semidefinite Optimization

f( ) convex, differentiable

Vector Problems

$$\min f(x)$$

$$x \in \mathbb{R}^n$$
$$x \geq 0$$
$$\mathbf{1}^T x = 1$$



Matrix Problems

$$\min f(X)$$

$$X \in Sym^{n \times n}$$
$$X \succeq 0$$
$$Tr(X) = 1$$

The Gap

$$g(x) = x^T \nabla f(x) - \min_{y \in S_n} y^T \nabla f(x)$$
$$= x^T \nabla f(x) - \min_i (\nabla f(x))_i$$

$$g(X) = \langle X, \nabla f(X) \rangle - \min_{Y \in \mathbf{S}^n} \langle Y, \nabla f(X) \rangle$$
$$= \langle X, \nabla f(X) \rangle - \lambda_{\min}(\nabla f(X))$$

Finding ε-approximations

[Clarkson SODA '08] Sparsity $O\left(\frac{1}{\varepsilon}\right)$

Rank $O\left(\frac{1}{\varepsilon}\right)$ [Hazan LATIN '08]

Pathwise Optimization

Vector Problems

The Gap

$$g_t(x) = x^T \nabla f_t(x) - \min_i (\nabla f_t(x))_i$$

"Better than necessary"

$$g_t(x) \leq \tfrac{\varepsilon}{2}$$

"Still good enough"

$$g_{t'}(x) \leq \varepsilon$$

Matrix Problems

$$g_t(X) = \langle X, \nabla f_t(X) \rangle - \lambda_{\min}(\nabla f_t(X))$$

Can use Weyl's Theorem
to control the Eigenvalue

"The difference"

$$g_{t'}(x) - g_t(x) \leq \varepsilon \left(1 - \tfrac{1}{2}\right)$$

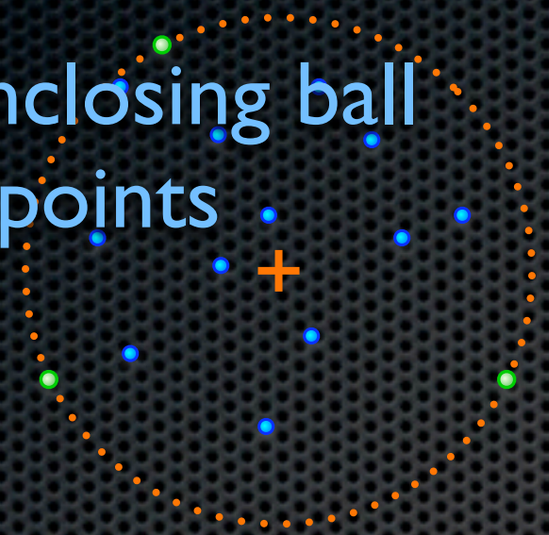
Theorem

There are $O\left(\frac{1}{\varepsilon}\right)$ many intervals

of piecewise constant ε-approx. solutions.

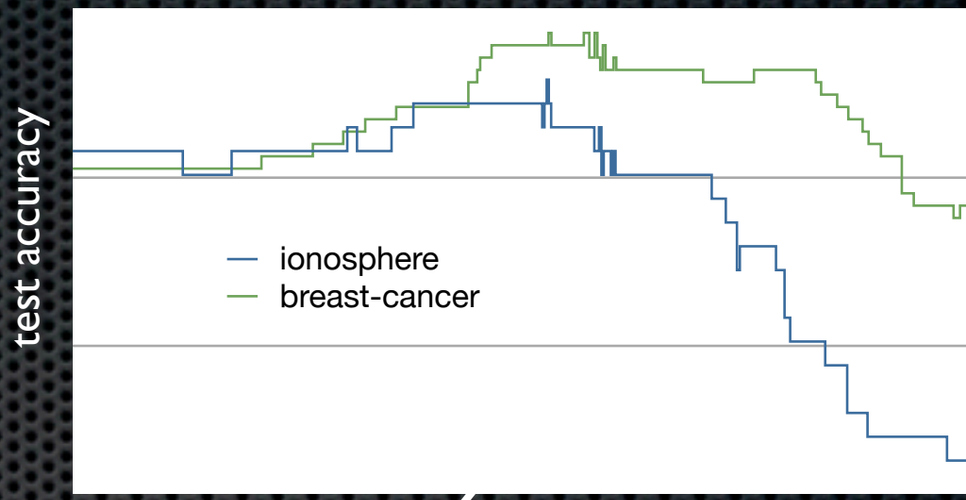
Applications

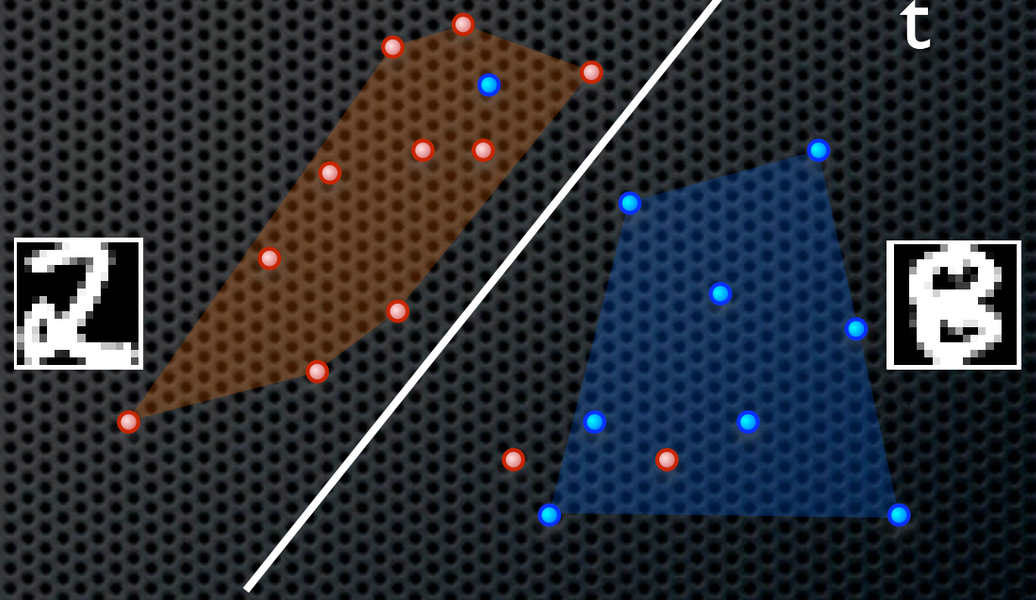
- Smallest enclosing ball of moving points



- SVMs

$$\min_{x \in S_n} x^T (A^T A + t\mathbf{1}) x$$





- Model Predictive Control

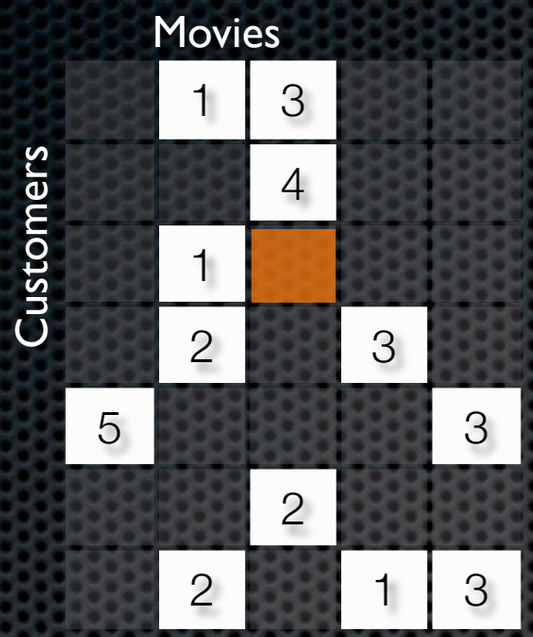
- Mean Variance Portfolio Optimization

$$\min_{x \in S_n} x^T A^T A x - t \cdot b^T x$$

Matrix Factorizations

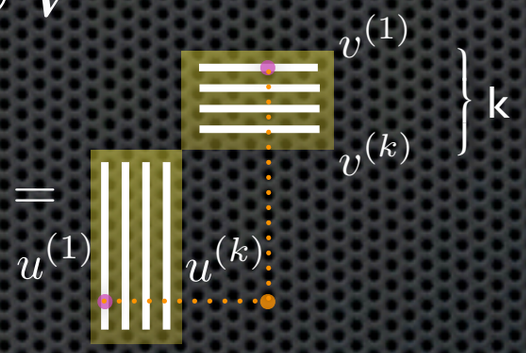
for recommender systems

$Y =$



$\approx UV^T$

$= \sum_{k}$



The Netflix challenge:
17k Movies
500k Customers
100M Observed Entries
(≈ 1%)

$$\min_{U,V} \sum_{(i,j)\in\Omega} (Y_{ij} - (UV^T)_{ij})^2$$

$$s.t.\ \|U\|_{Fro}^2 + \|V\|_{Fro}^2 = t$$

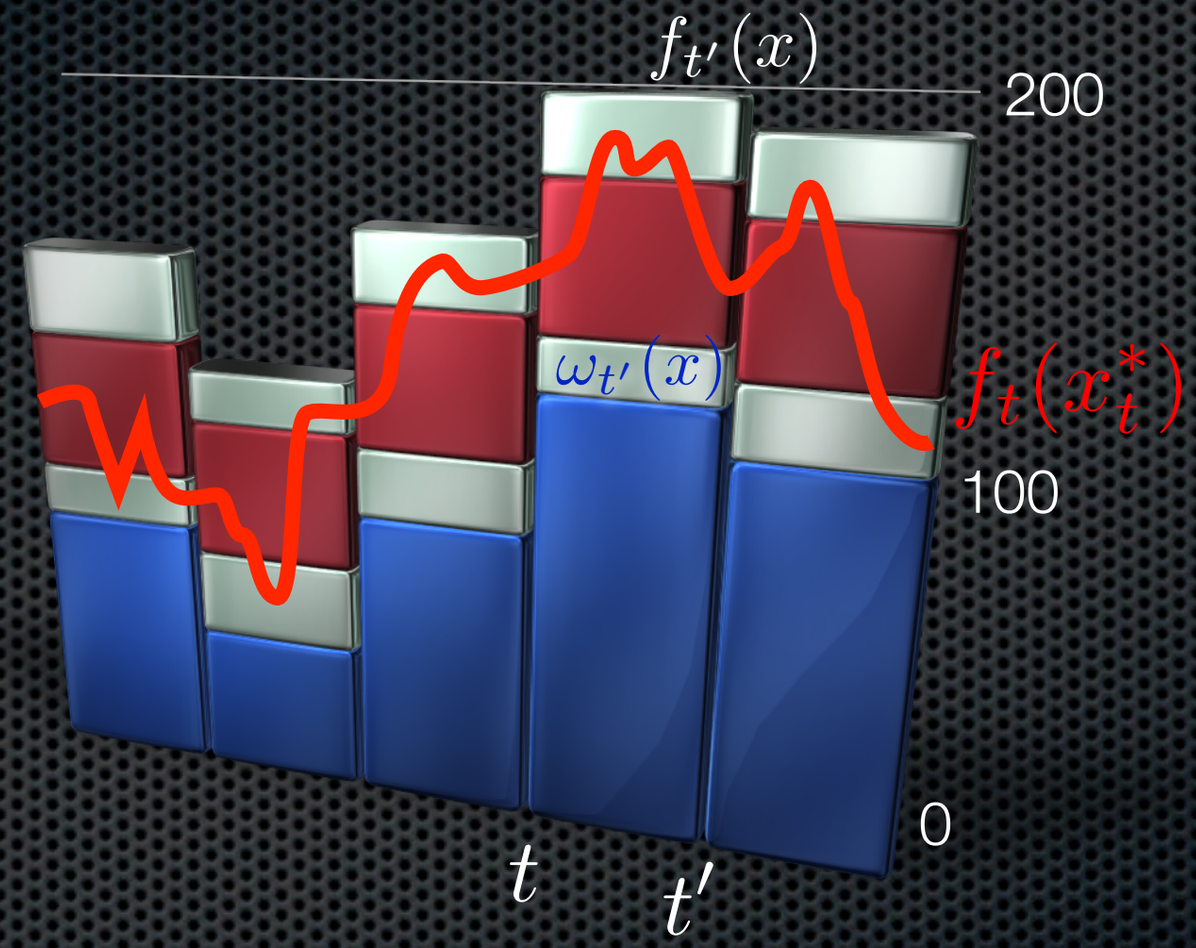
Is equivalent to:

$$\min_{X \succeq 0} f(X)$$

$$s.t.\ Tr(X) = t$$

$$\begin{pmatrix} UU^T & UV^T \\ VU^T & VV^T \end{pmatrix} =: X$$

The end



Thanks