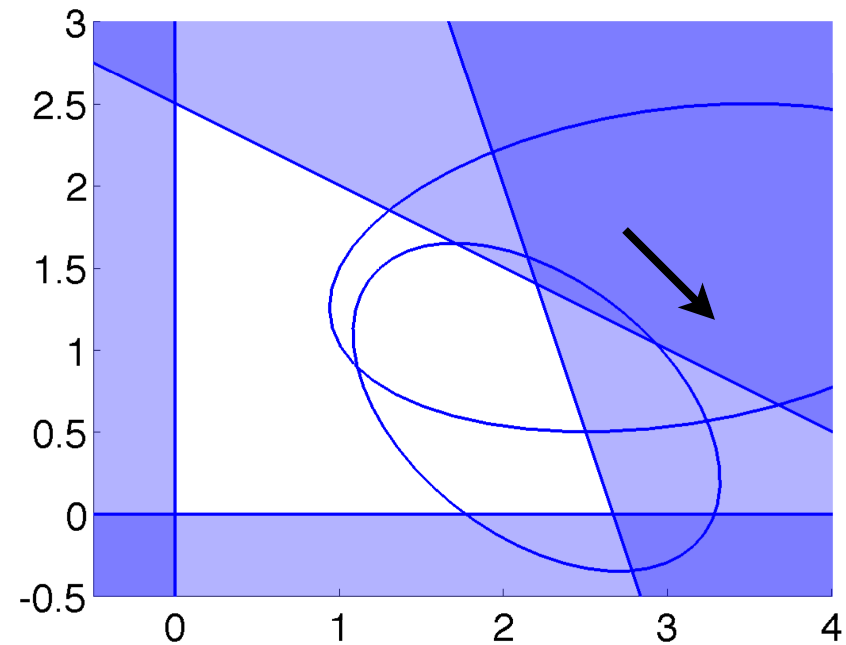
QP & cone program duality
Support vector machines

10-725 Optimization
Geoff Gordon
Ryan Tibshirani

Review

- Quadratic programs
- Cone programs
  - SOCP, SDP
  - QP ⊆ SOCP ⊆ SDP
  - SOC, S_+ are self-dual



- Poly-time algos (but not strongly poly-time, yet)
- Examples: group lasso, Huber regression, matrix completion

Matrix completion

- Observe A_{ij} for $ij \in E$, write $\mathcal{O}_{ij} = \begin{cases} 1 & ij \in E \\ 0 & o/w \end{cases}$
- $\min_X \|(X-A) \circ P\|_F^2 + \lambda \|X\|_*$

$\lambda(\mathrm{tr}(P) + \mathrm{tr}(Q))/2$

$M = \begin{bmatrix} P & X \\ X^T & Q \end{bmatrix} \succeq 0$ s.t.

$X = U\Sigma V^T$

$M \succeq 0 \iff \mathrm{tr}(B^T M) \geq 0 \;\forall B \succeq 0$ take $B = \begin{pmatrix} UU^T & -UV^T \\ -VU^T & VV^T \end{pmatrix}$

$0 \leq \mathrm{tr}(B^T M) = \mathrm{tr}(UU^T P) + \mathrm{tr}(VV^T Q) - 2\mathrm{tr}(VU^T X)$

$\mathrm{tr}(P) + \mathrm{tr}(Q) \geq 2\,\mathrm{tr}(\underbrace{U^T X V}_{\Sigma}) = 2\|X\|_*$

$P = U\Sigma U^T \quad Q = V\Sigma V^T$

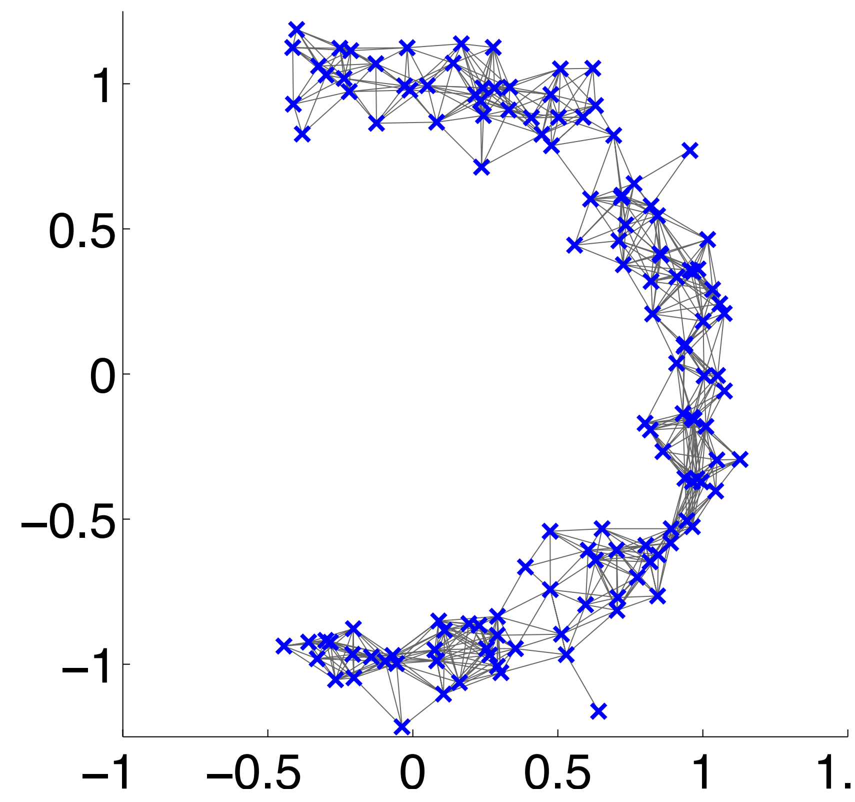
$\mathrm{tr}(P) = \mathrm{tr}(U^T U \Sigma) \quad \mathrm{tr}(P) + \mathrm{tr}(Q) = 2\|X\|_*$

$\begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix}^T \begin{pmatrix} U\Sigma U^T & -U\Sigma V^T \\ -V\Sigma U^T & V\Sigma V^T \end{pmatrix} \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} = \begin{pmatrix} \Sigma & -\Sigma \\ -\Sigma & \Sigma \end{pmatrix} \succeq 0$

Max-variance unfolding

aka semidefinite embedding

- Goal: given $x_1, \ldots x_T \in R^n$
  - find $y_1, \ldots, y_T \in R^k$ $(k \ll n)$
  - $||y_i - y_j|| \approx ||x_i - x_j||$ $\forall i,j \in E$
- If x_i were near a k-dim subspace of R^n, PCA!
- Instead, two steps:
  - first look for $z_1, \ldots z_T \in R^n$ with
    - $||z_i - z_j|| = ||x_i - x_j||$ $\forall i,j \in E$
    - and var(z) as big as possible
  - then use PCA to get y_i from z_i



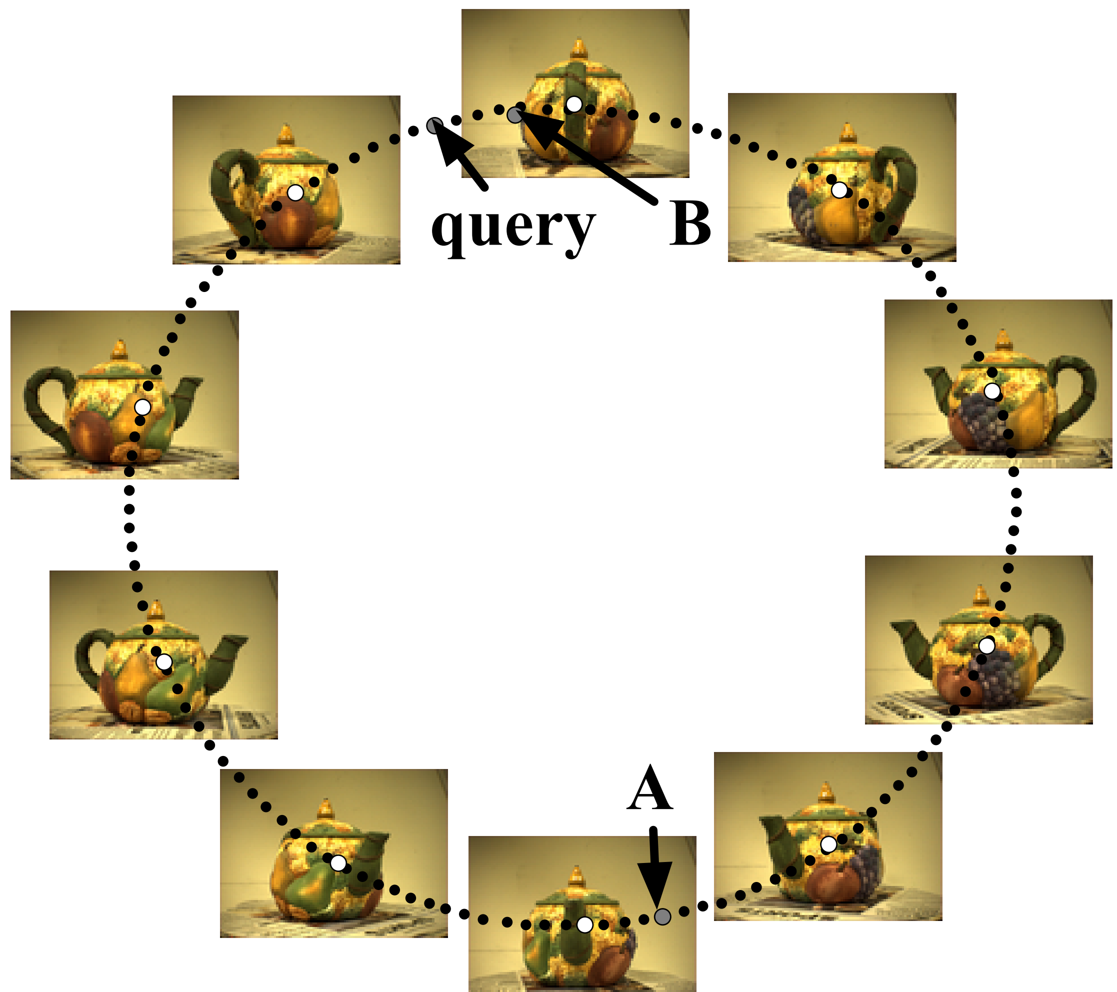
MVU/SDE

- $\max_z \operatorname{tr}(\operatorname{cov}(z))$ s.t. $\|z_i - z_j\| = \|x_i - x_j\| \quad \forall i,j \in E$

Result

- Embed 400 images of a teapot into 2d

Euclidean distance from query to A is smaller; after MVU, distance to B is smaller



[Weinberger & Saul, AAAI, 2006]

Duality for QPs and Cone Ps

- Combined QP/CP:
  - min $c^T x + x^T H x/2$ s.t. $Ax + b \in K$ $x \in L$
  - cones K, L implement any/all of equality, inequality, generalized inequality
  - assume K, L proper (closed, convex, solid, pointed)

Primal-dual pair

- Primal:
  - $\min\ c^T x + x^T H x/2 \quad \text{s.t.} \quad Ax + b \in K \quad x \in L$
- Dual:
  - $\max\ -z^T H z/2 - b^T y \quad \text{s.t.} \quad Hz + c - A^T y \in L^* \quad y \in K^*$

KKT conditions

primal-dual pair

- min $c^Tx + x^THx/2$ s.t. $Ax + b \in K$ $x \in L$
- max $-b^Ty - z^THz/2$ s.t. $Hz + c - A^Ty \in L^*$ $y \in K^*$

KKT conditions

- primal: $Ax+b \in K \quad x \in L$
- dual: $Hz + c - A^T y \in L^* \quad y \in K^*$
- quadratic: $Hx = Hz$
- comp. slack: $y^T(Ax+b) = 0 \quad x^T(Hz+c-A^T y) = 0$

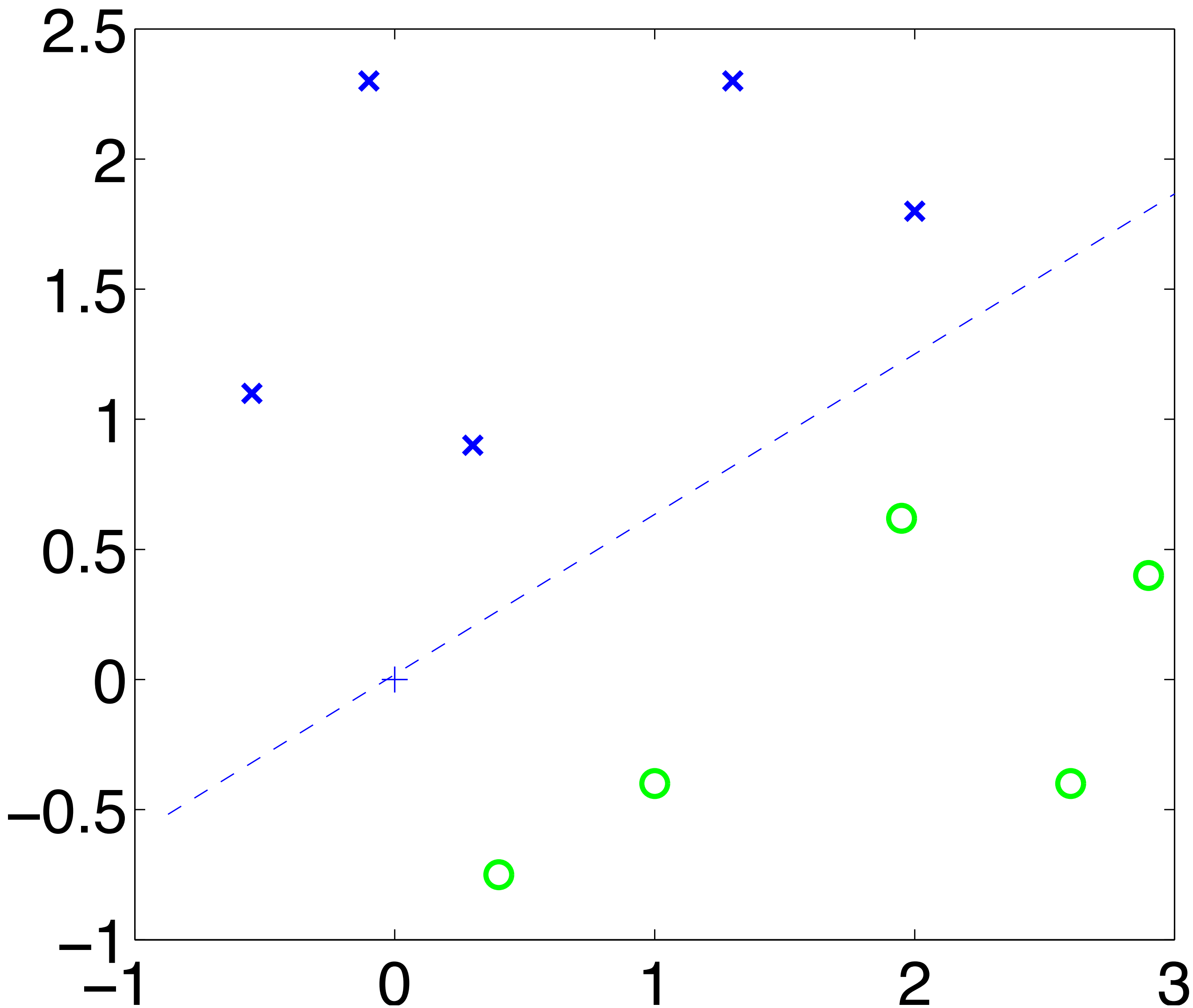
Support vector machines

(separable case)

Maximizing margin

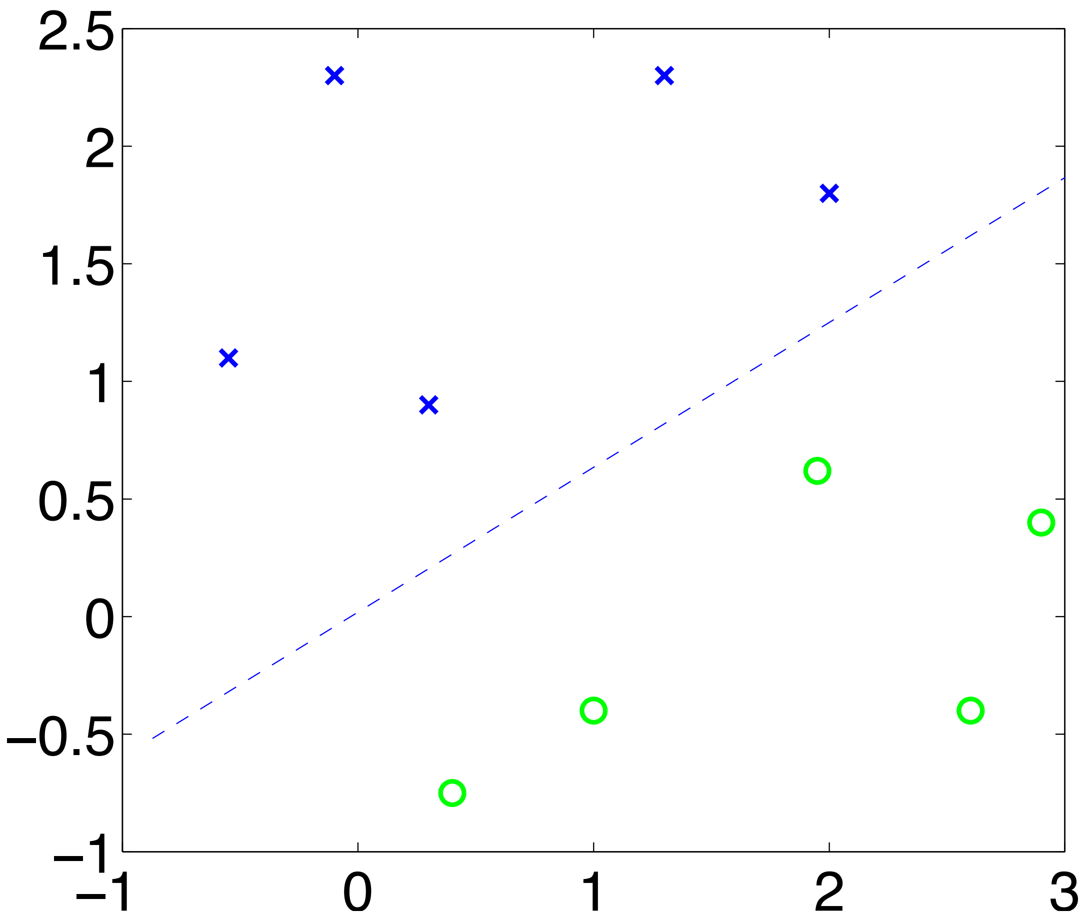
- margin $M = y_i (x_i \cdot \bar{w} - \bar{b})$
- max M s.t. $M \leq y_i (x_i \cdot \bar{w} - \bar{b})$

For example



Slacks

- $\min \|v\|^2/2 \quad \text{s.t.} \quad y_i\,(x_i^T v - d) \geq 1 \quad \forall i$



SVM duality

- min $||v||^2/2 - \Sigma s_i$ s.t. $y_i (x_i^T v - d) \geq 1 - s_i \quad \forall i$
- min $v^T v/2 + 1^T s$ s.t. $Av - yd + s - 1 \geq 0$

Interpreting the dual

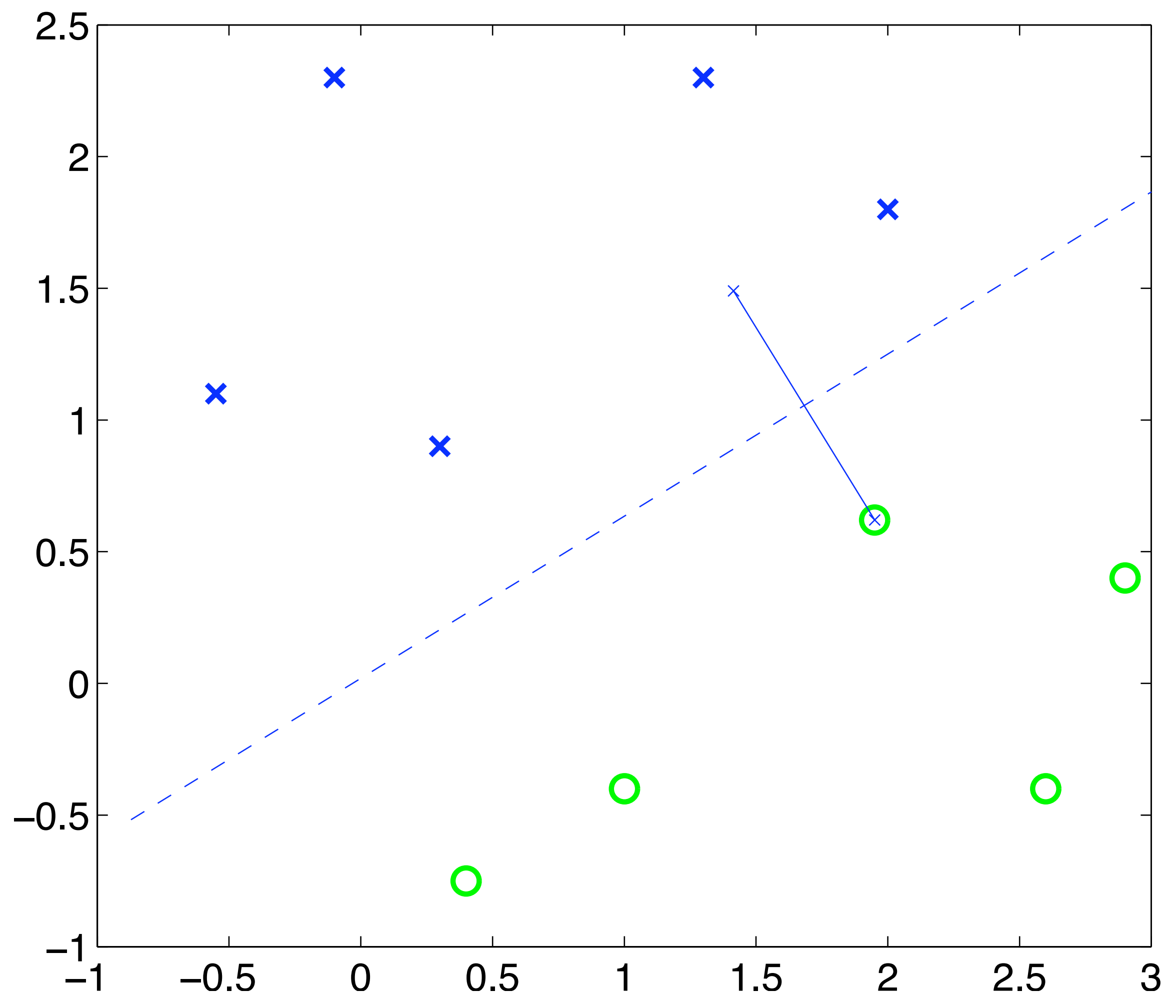
- $\max\ 1^T\alpha - \alpha^T K\alpha/2$ s.t. $y^T\alpha = 0 \quad 0 \le \alpha \le 1$

α:

$\alpha > 0$:

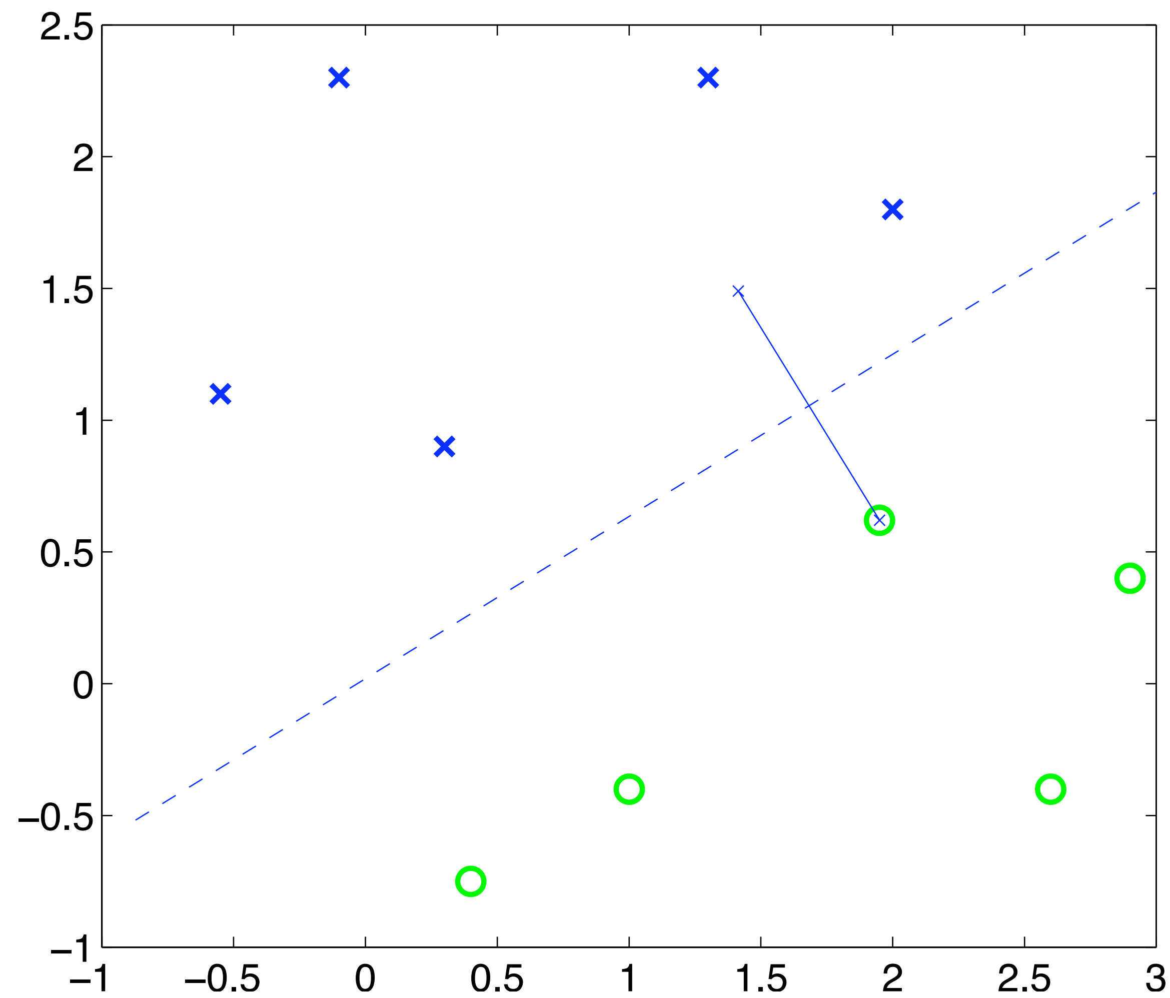
$\alpha < 1$:

$y^T\alpha = 0$:

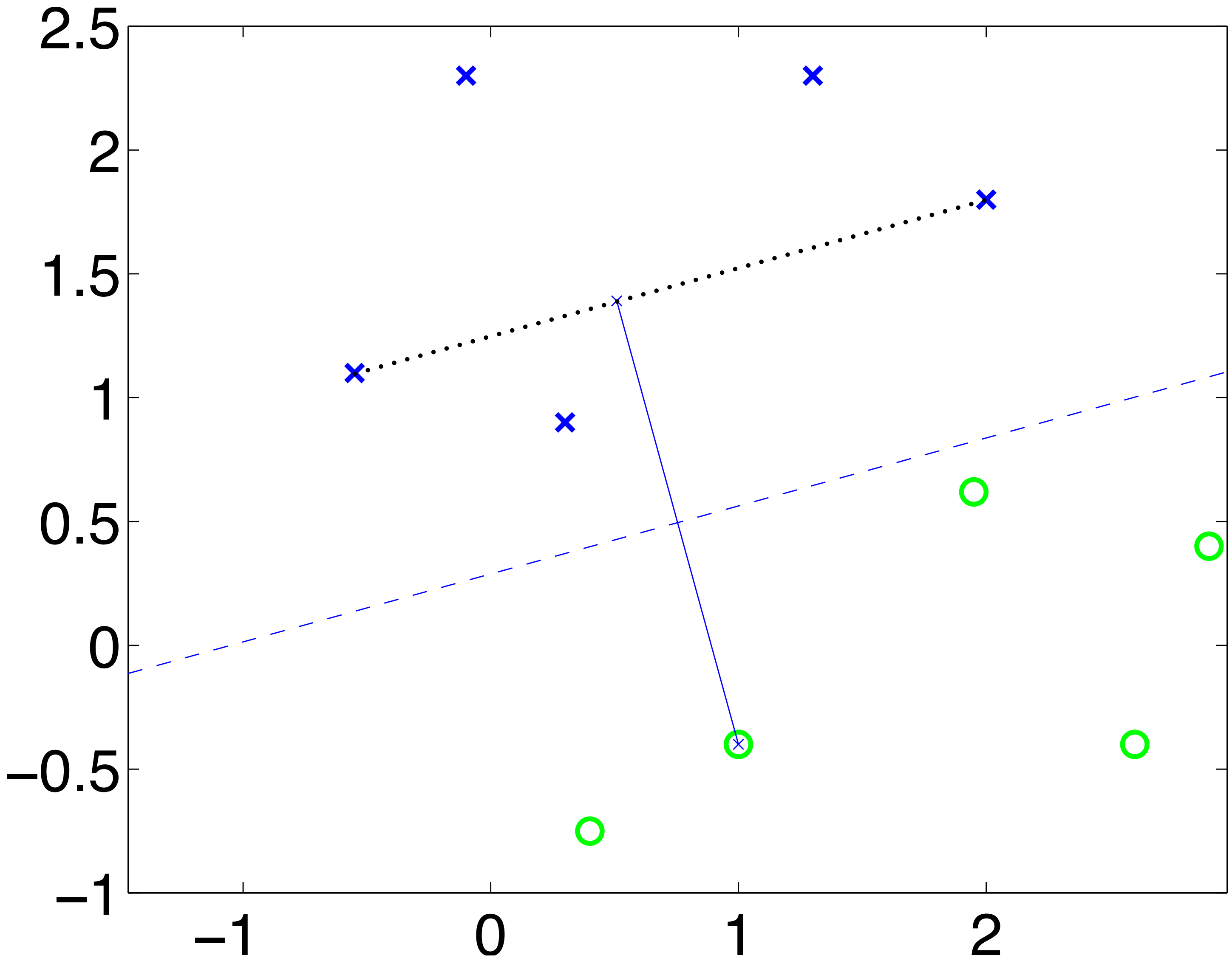


From dual to primal

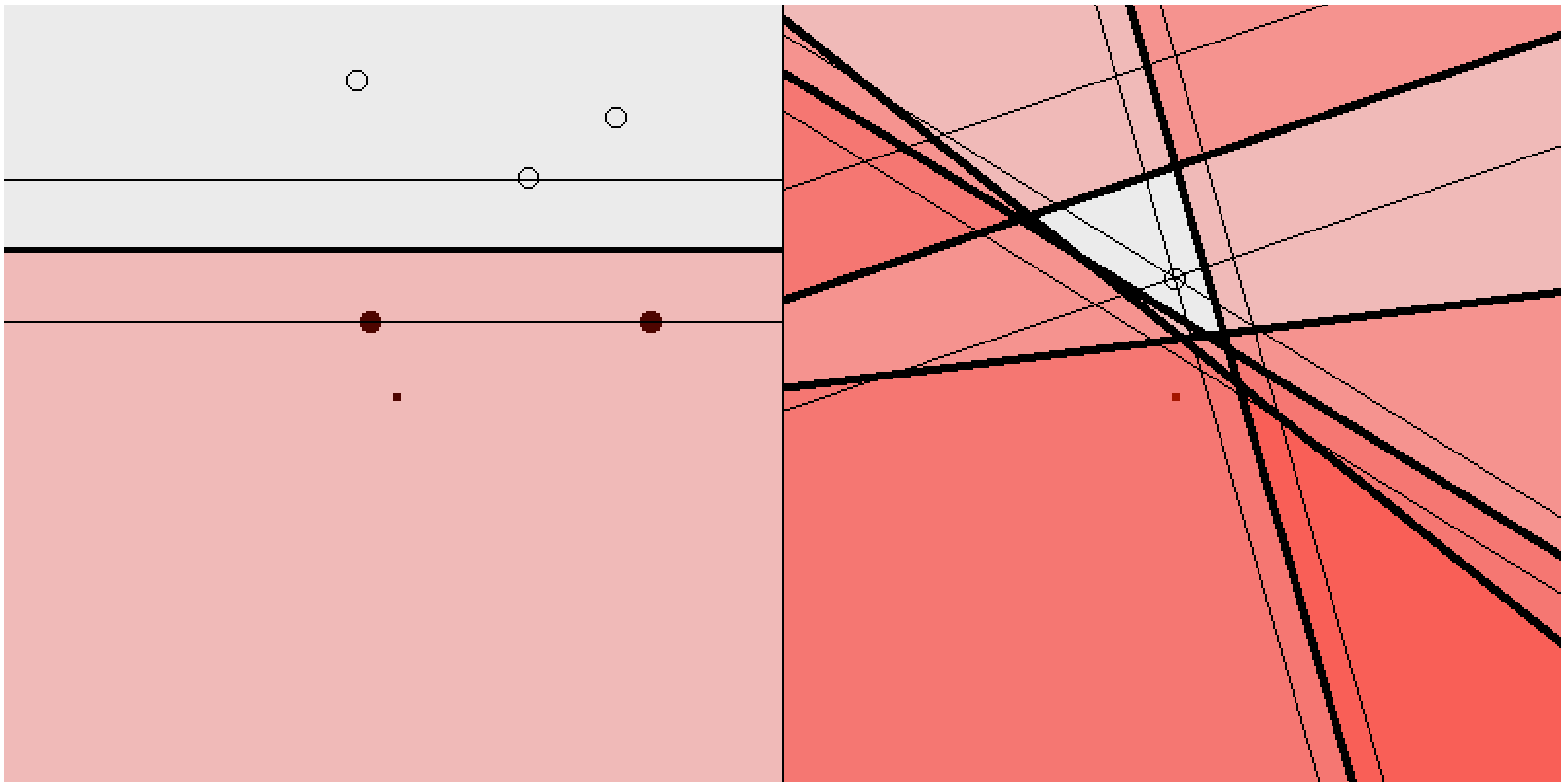
- max $1^T\alpha - \alpha^T K\alpha/2$ s.t. $y^T\alpha = 0 \quad 0 \le \alpha \le 1$



A suboptimal support set



SVM duality: the applet



Why is the dual useful?

aka the kernel trick

$$\max\ 1^T\alpha - \alpha^T A A^T \alpha/2 \quad \text{s.t.}\ \ y^T\alpha = 0 \quad 0 \le \alpha \le 1$$

- SVM: n examples, m features
  - primal:
  - dual:

Interior-point methods

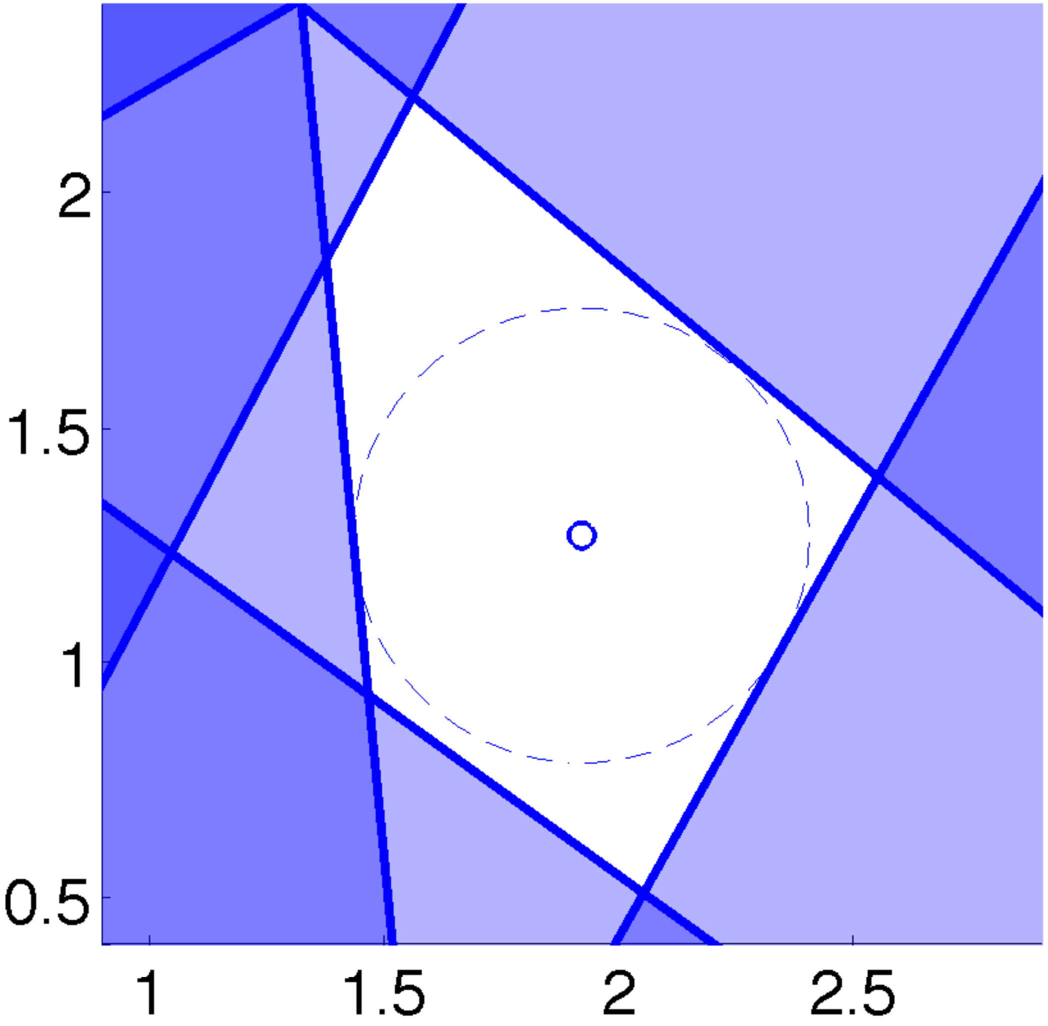
10-725 Optimization

Geoff Gordon
Ryan Tibshirani

Ball center

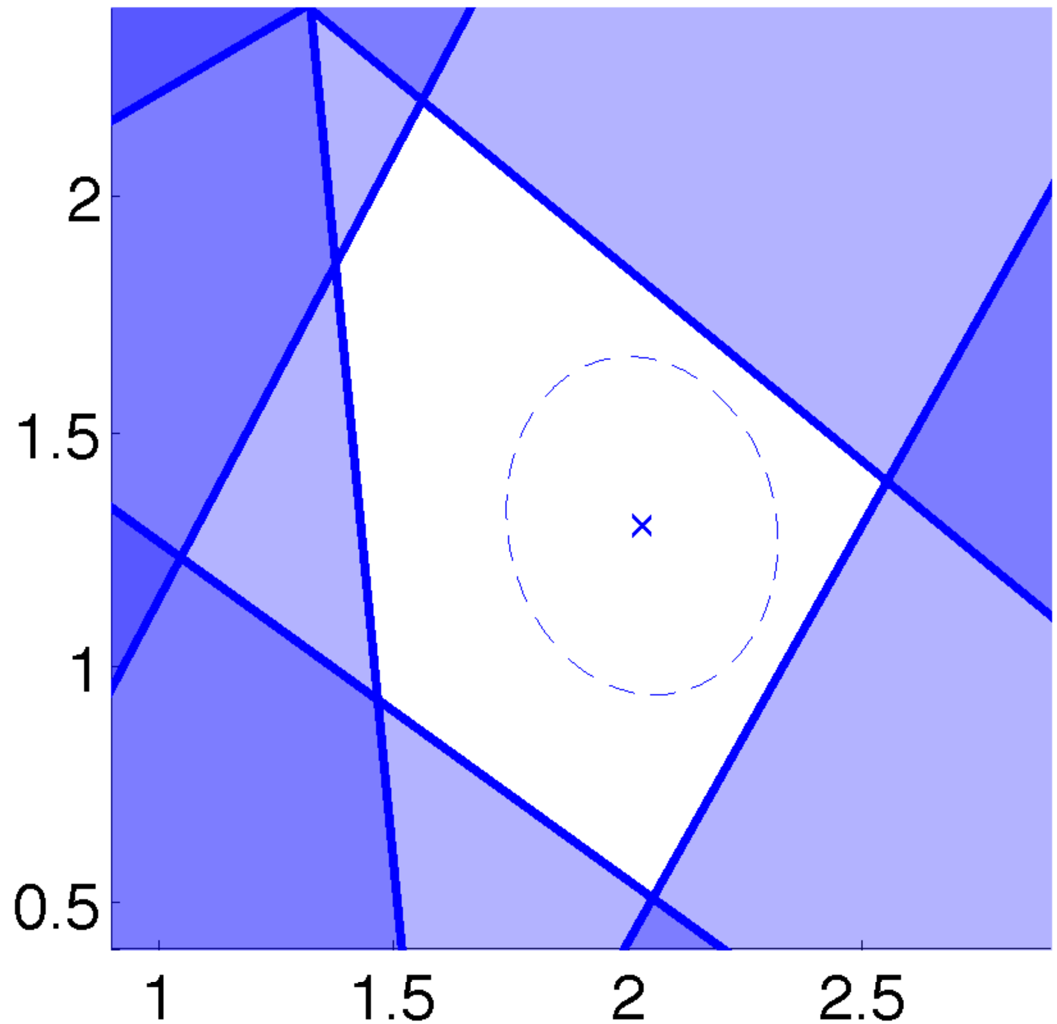
aka Chebyshev center

- X = { x | Ax + b ≥ 0
- Ball center:
  -
  - if $||a_i|| = 1$
  - in general:

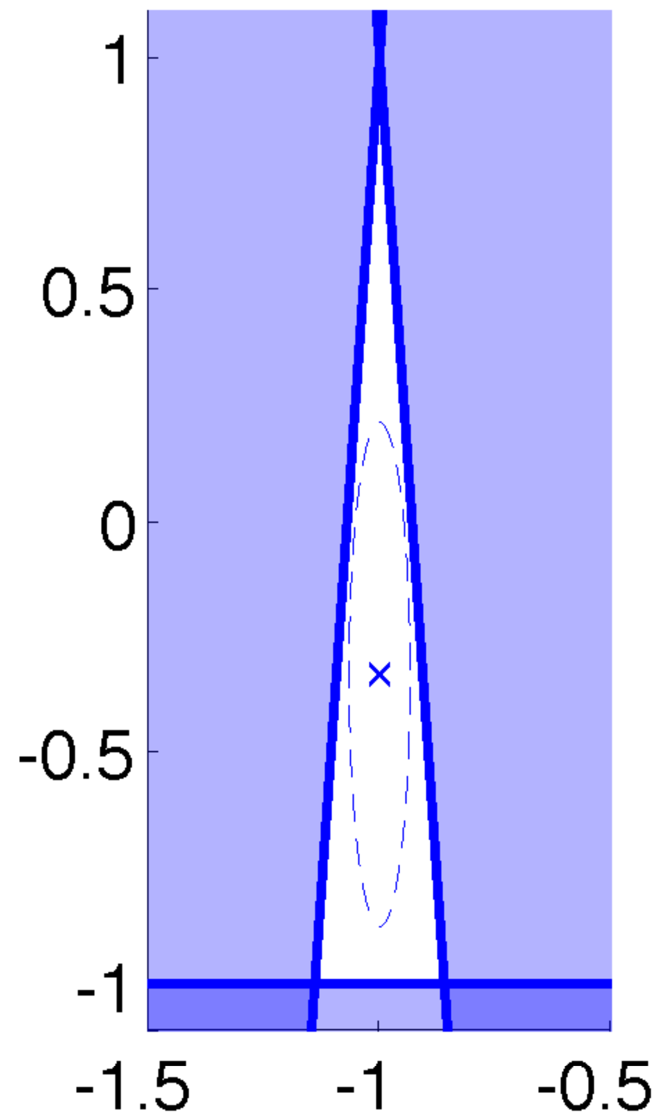


Analytic center

- Let s = Ax + b
- Analytic center:
  -
  -



Bad conditioning? No problem.



Newton for analytic center

- Lagrangian $L(x,s,y) = -\sum \ln s_i + y^T(s - Ax - b)$

Adding an objective

- Analytic center was for $\{ x \mid Ax + b = s \geq 0 \}$
- Now: min $c^T x$ st $Ax + b = s \geq 0$
- Same trick:
  - min $t\, c^T x - \sum \ln s_i$ st $Ax + b = s \geq 0$
    - parameter $t \geq 0$
  - central path =
    - $t \to 0$: $t \to \infty$:
  - $L(x,s,y) =$

Newton for central path

- $L(x,s,y) = t\, c^T x - \sum \ln s_i + y^T(s - Ax - b)$