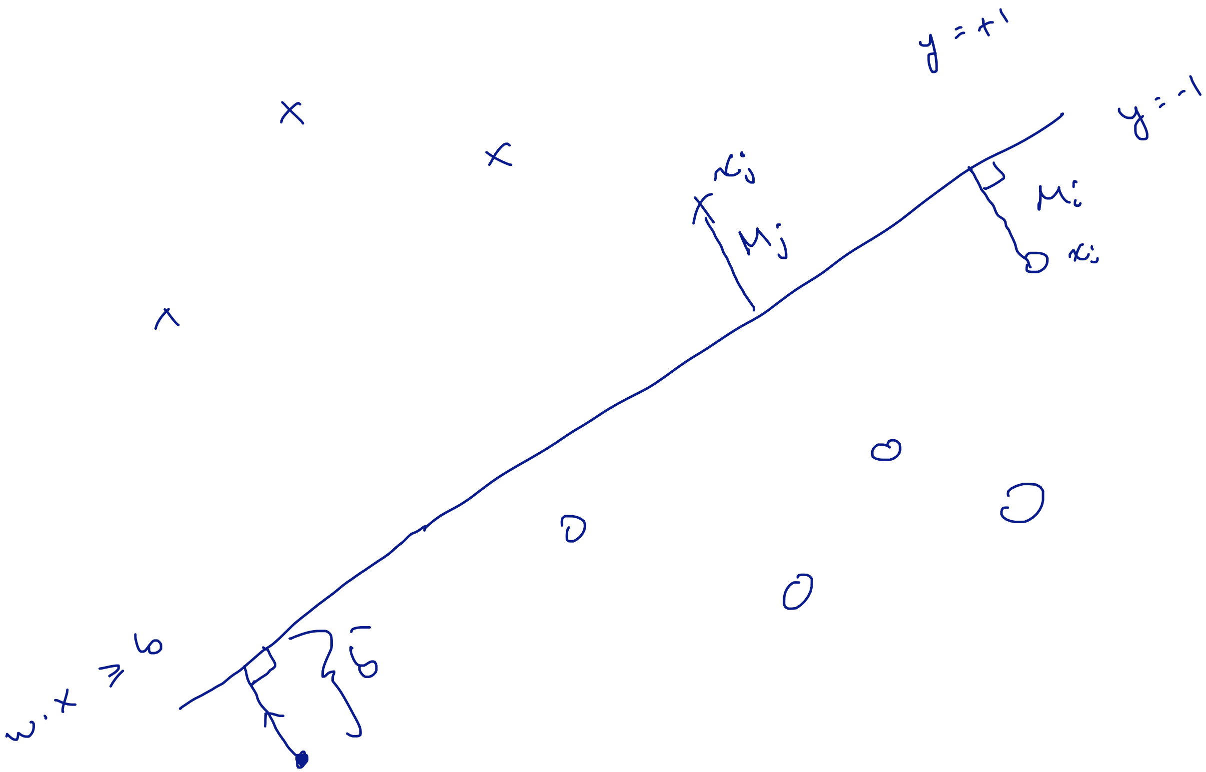
Administrivia

- HW4 out
  - based on feedback survey,
    - fewer questions: 4, but only do 3
    - range of problem types: focus on those that help your understanding
    - split out "spoilers" for Q2
- Midterm
  - mean 65 (out of 95), std dev 11.3
  - back at end of class

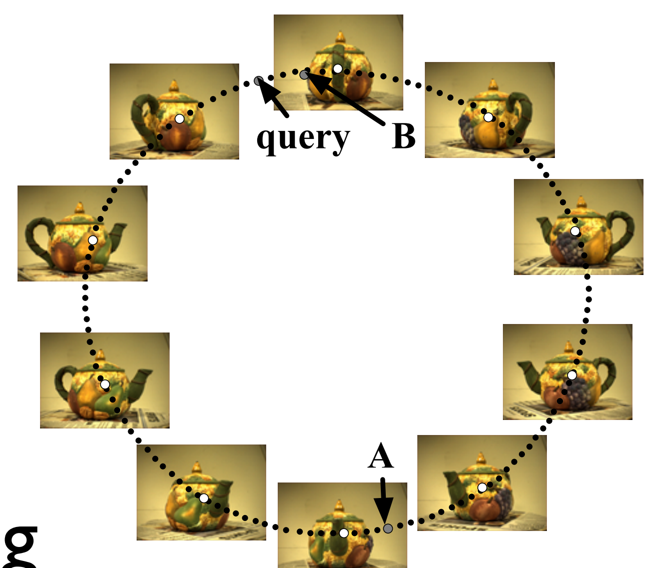
Review

- Cone & QP duality
  - min $c^Tx + x^THx/2$ s.t. $Ax + b \in K$ $x \in L$
  - max $-z^THz/2 - b^Ty$ s.t. $Hz + c - A^Ty \in L^*$ $y \in K^*$
- KKT conditions
  - primal: $Ax+b \in K$ $x \in L$
  - dual: $Hz + c - A^Ty \in L^*$ $y \in K^*$
  - quadratic: $Hx = Hz$
  - comp. slack: $y^T(Ax+b) = 0$ $x^T(Hz+c-A^Ty) = 0$

Review



Support vector machines



Maximum-variance unfolding

Support vector machines

10-725 Optimization
Geoff Gordon
Ryan Tibshirani

SVM duality

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

Interpreting the dual

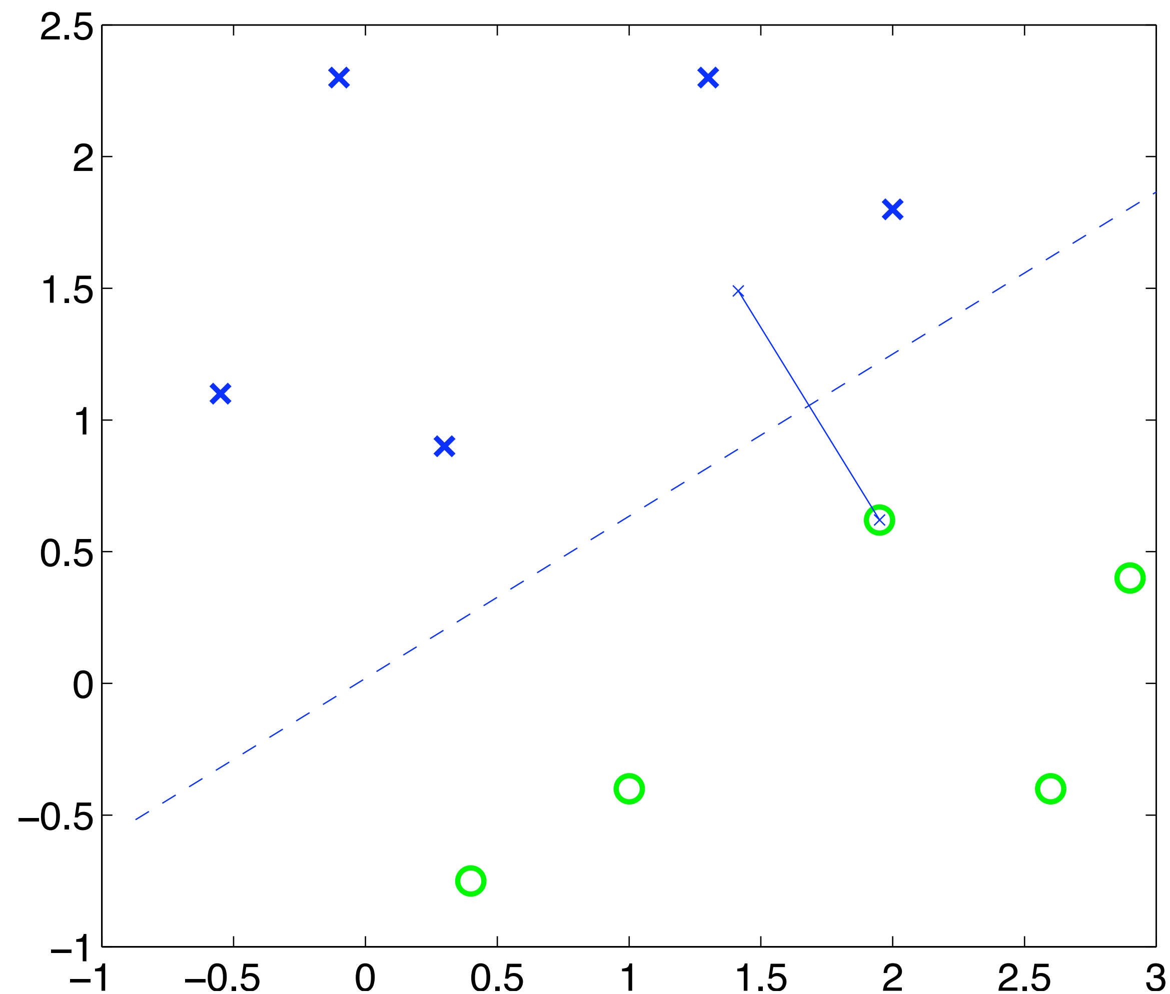
- max $1^T\alpha - \alpha^T K\alpha/2$ s.t. $y^T\alpha = 0$ $0 \le \alpha \le 1$

α:

$\alpha>0$:

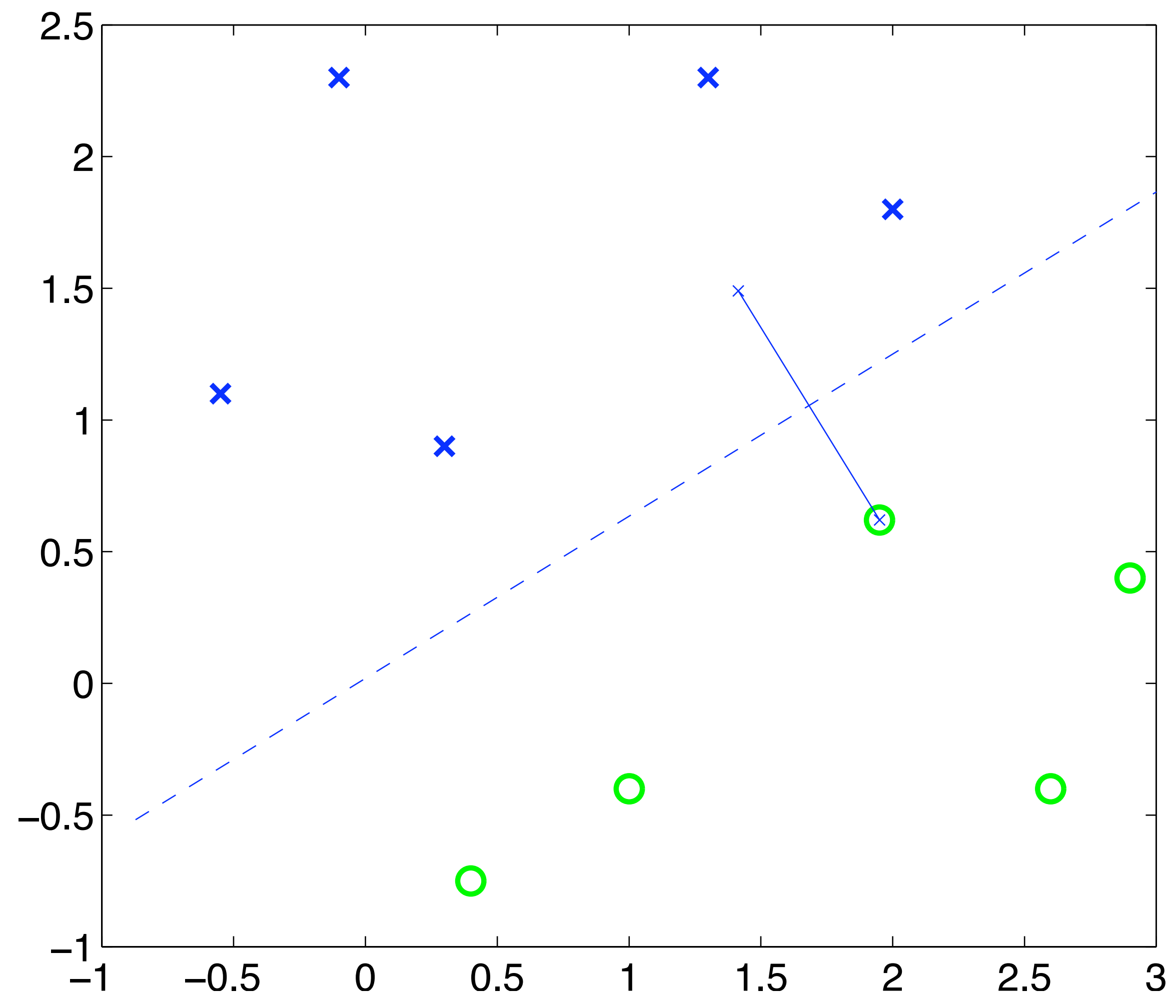
$\alpha<1$:

$y^T\alpha=0$:

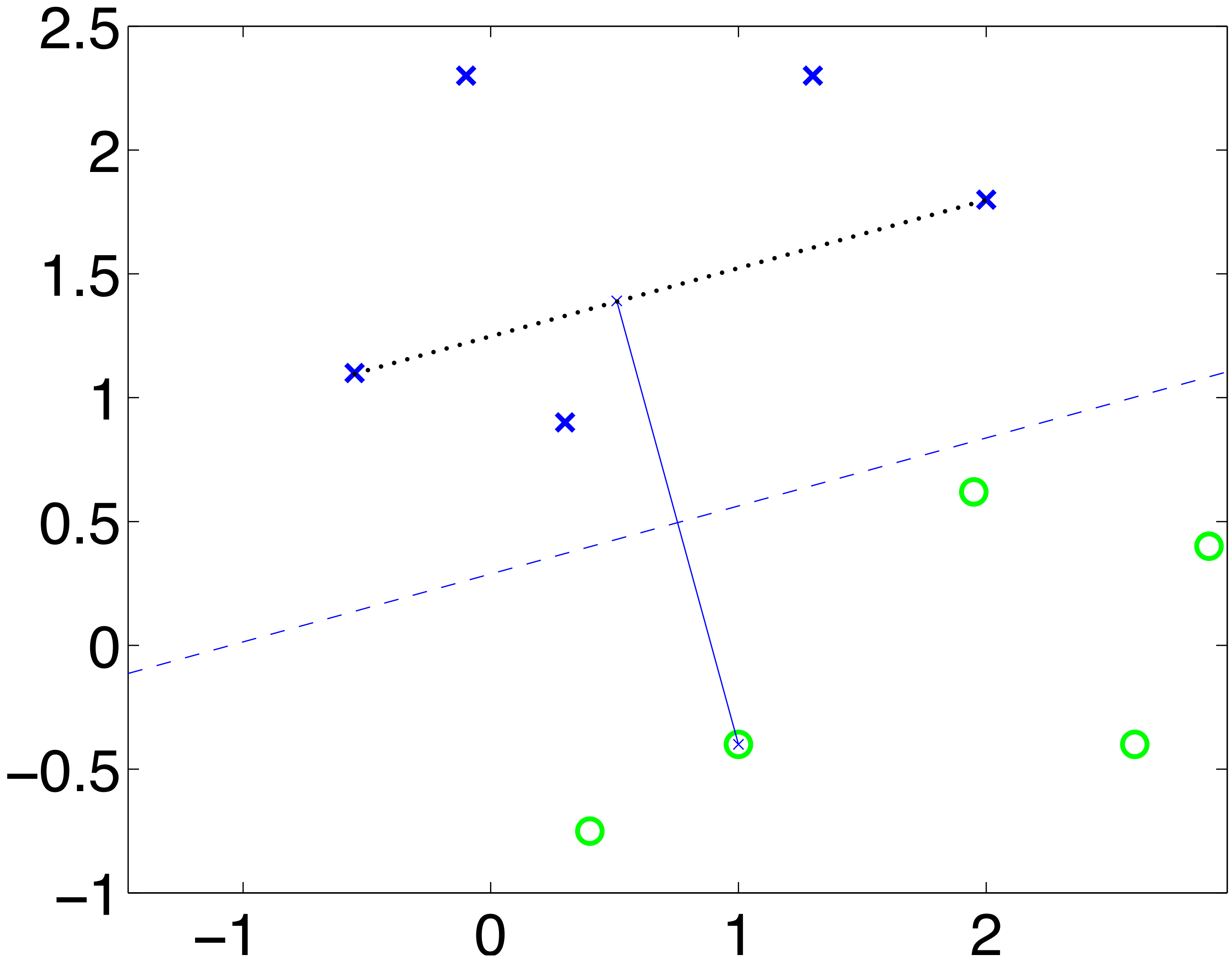


From dual to primal

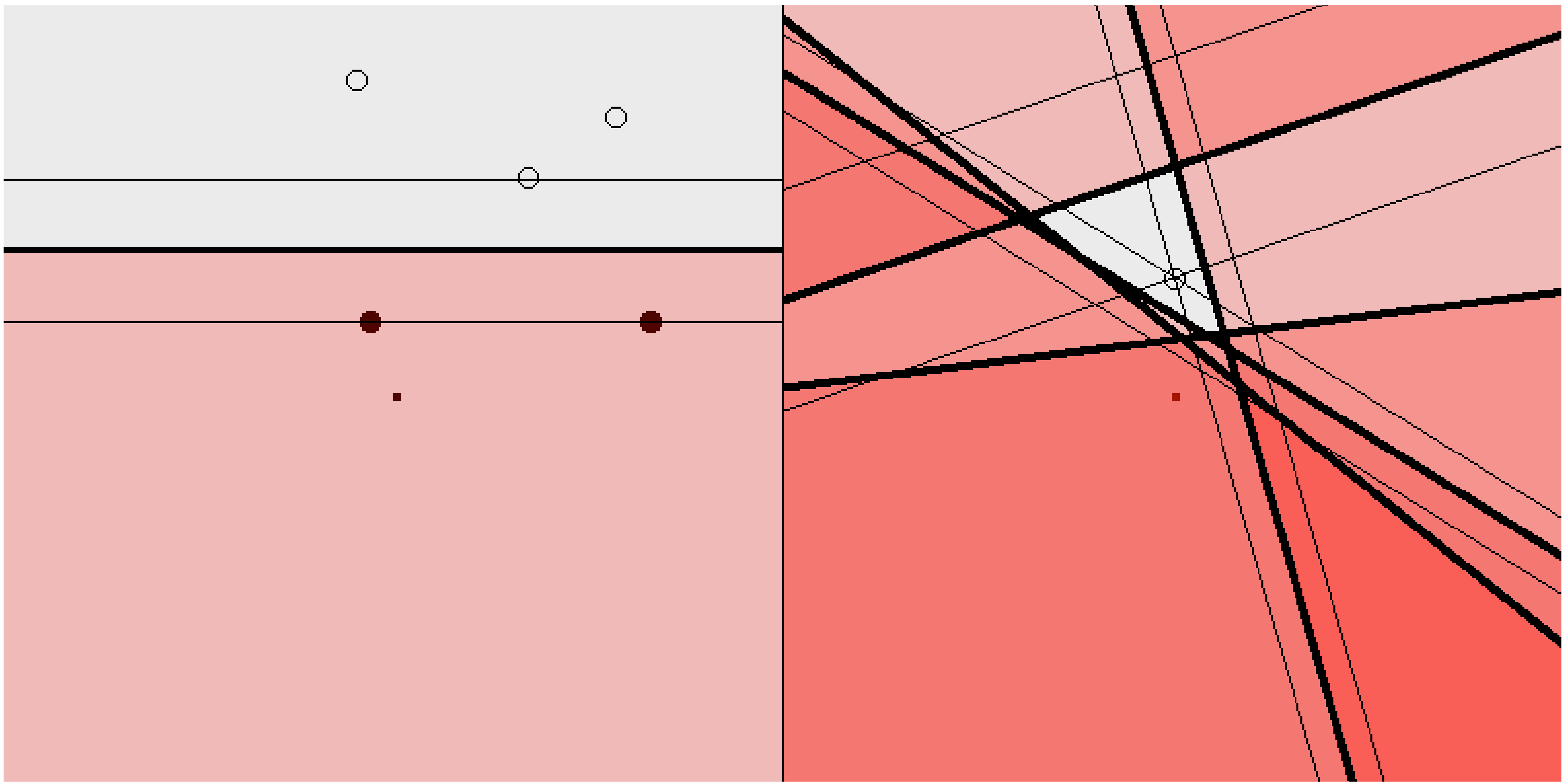
- max $1^T\alpha - \alpha^T K\alpha/2$ s.t. $y^T\alpha = 0 \quad 0 \le \alpha \le 1$



A suboptimal support set



SVM duality: the applet



Why is the dual useful?

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

- SVM: n examples, m features: $x_i = \phi(u_i) \in R^m$
  - primal:
  - dual:

The kernel trick

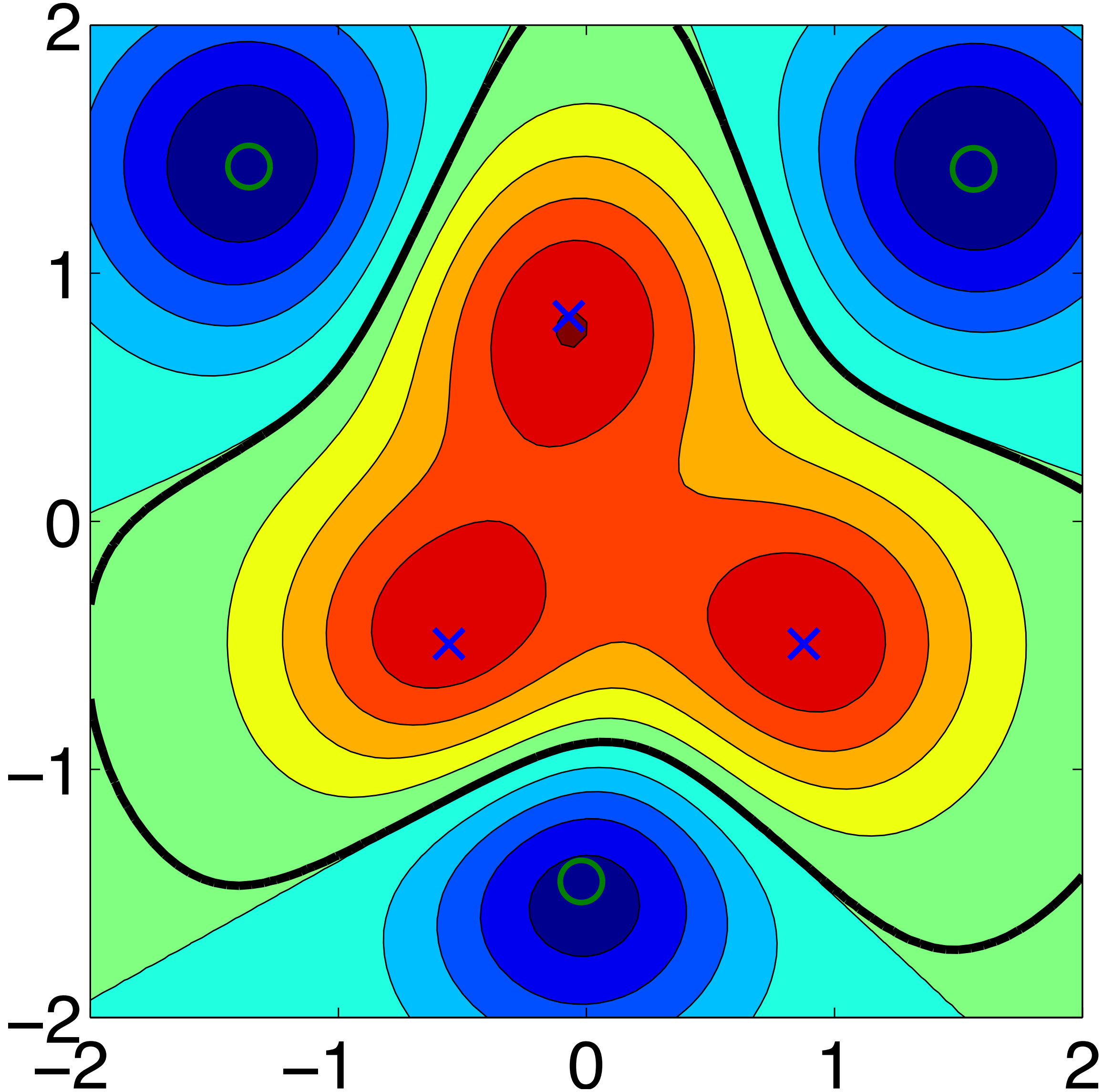
- Don't even need to know features $x_i = \phi(u_i)$, as long as we can compute dot products $x_i^T x_j$
- Matrix of dot products:
  - $K_{ij} =$
  - only need subroutine for k (don't care about ϕ)
  - how do we know k works?
    -
    - this is a "positive definite function," aka "Mercer kernel"—∃ many examples

Examples of kernels

- $K(u_i, u_j) = (1 + u_i^T u_j)^d$
  - can represent any degree-d polynomial
  - i.e., decision surface is $p(u) = b$ for degree-d poly p
- $K(u_i, u_j) = (u_i^T u_j)^d$
  - polynomial where all terms have degree exactly d
  - d=1 reduces to original (linear) SVM
- $K(u_i, u_j) = \exp(-||u_i - u_j||^2 / 2\sigma^2)$
  - Gaussian radial basis functions of width σ

Gaussian kernel

$\sigma = 0.5$



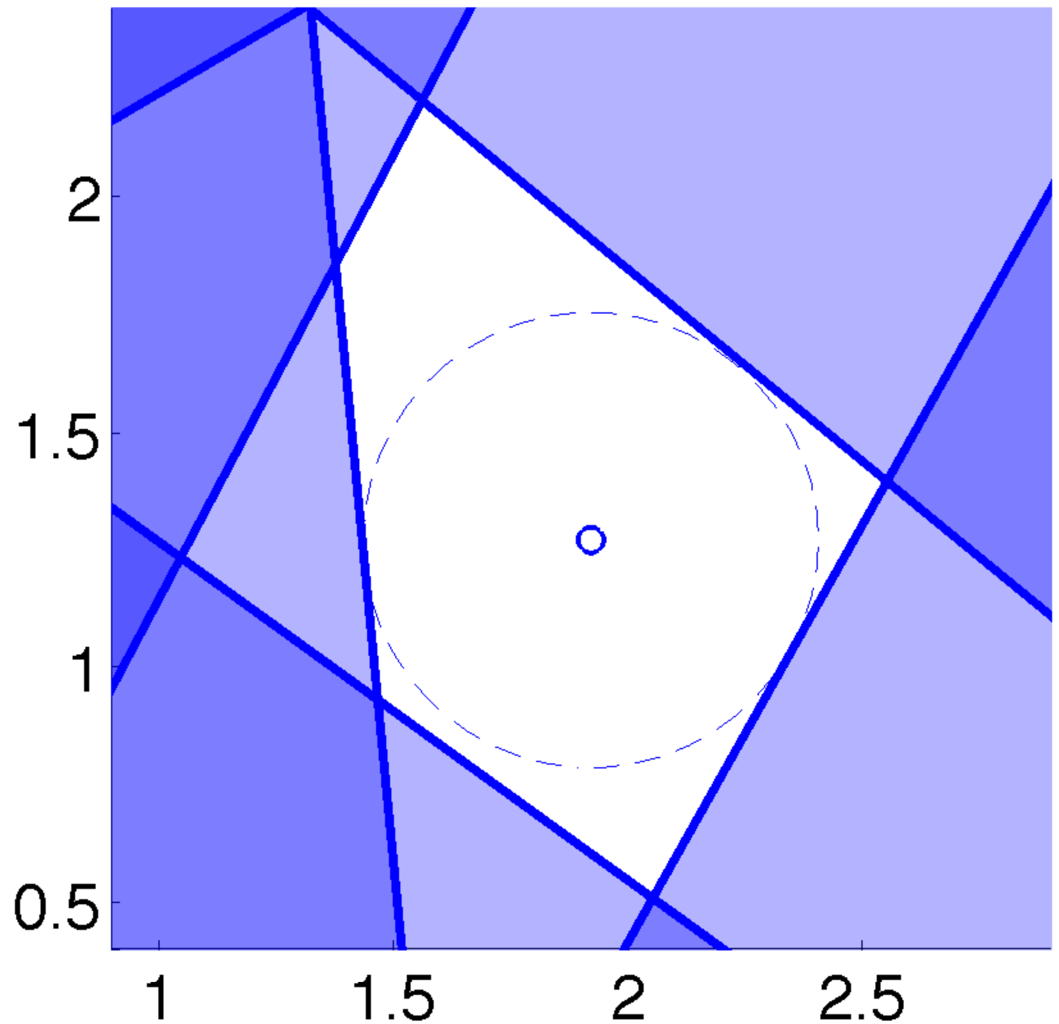
Interior-point methods

10-725 Optimization
Geoff Gordon
Ryan Tibshirani

Ball center

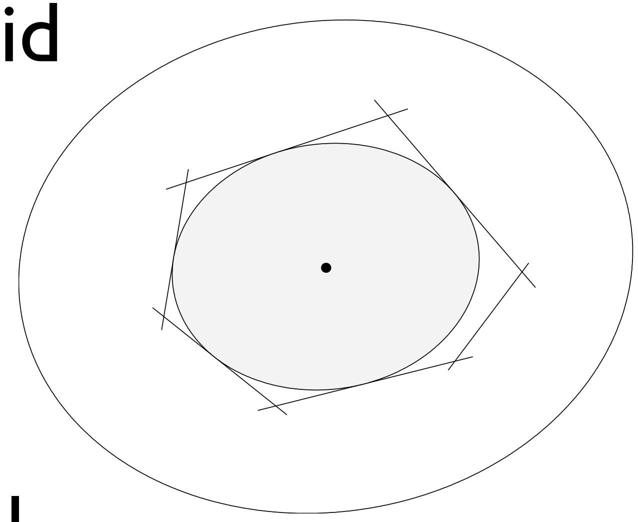
aka Chebyshev center

- $X = \{ x \mid Ax + b \geq 0 \}$
- Ball center:
  -
  - if $||a_i|| = 1$
  - in general:



Ellipsoid center

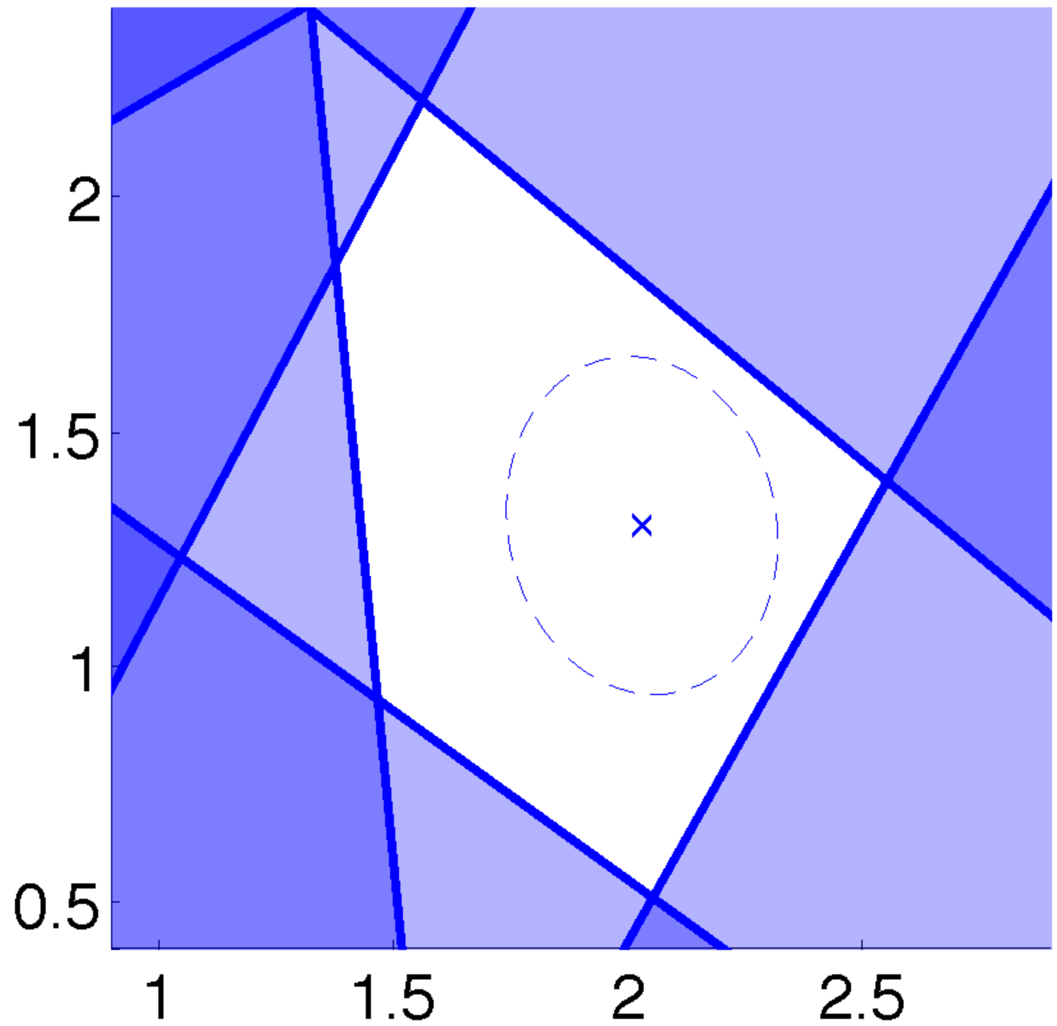
aka max-volume inscribed ellipsoid



- Center d of largest inscribed ellipsoid
  - $E = \{ Bu + d \mid \|u\|_2 \leq 1 \}$
  - $\mathrm{vol}(E) \geq \mathrm{vol}(X)/n$ in R^n
- min log det B^{-1} s.t.
  - $a_i^T(Bu+d) + b_i \geq 0 \quad \forall i \quad \forall u$ with $\|u\| \leq 1$
  - $B \succcurlyeq 0$
- Convex optimization, but relatively expensive:
  - convex objective, semidefinite constraint
  - each (u, a_i, b_i) yields a linear constraint on B, d

Analytic center

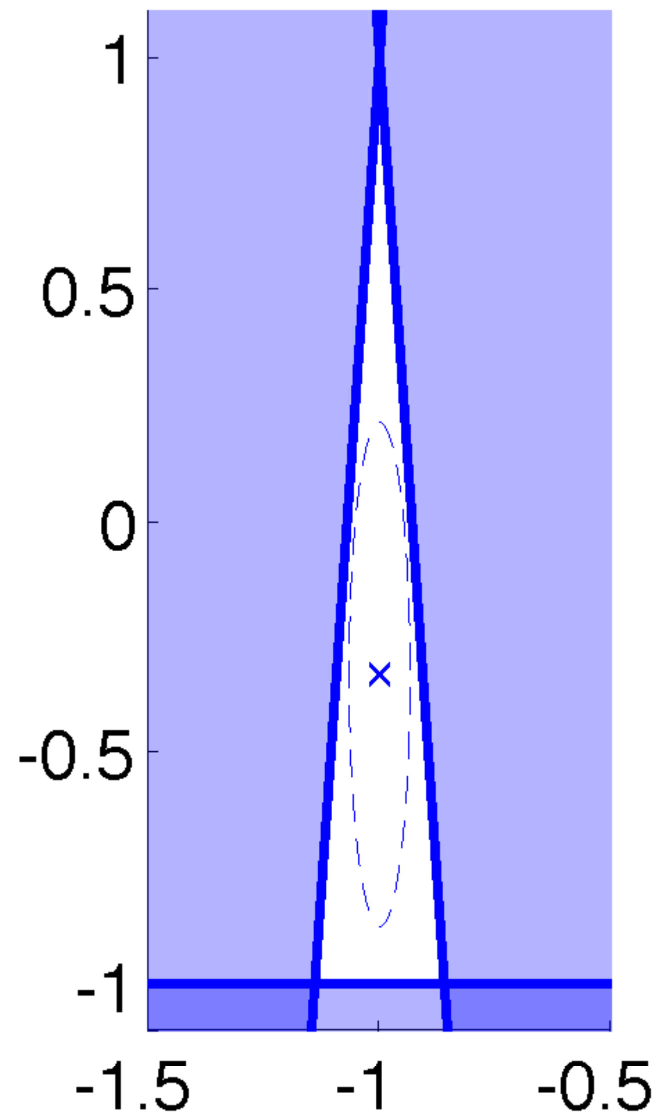
- Let s = Ax + b
- Analytic center:
  -
  -



Bad conditioning? No problem.

$a_i^T x + b_i \geq 0 \quad \min -\sum \ln(a_i^T x + b_i)$

$y = Mx + q$



Newton for analytic center

- $f(x) = -\sum \ln(a_i^T x + b_i)$
  - $df/dx = -\sum a_i / (a_i^T x + b_i)$
  - $d^2f/df^2 =$

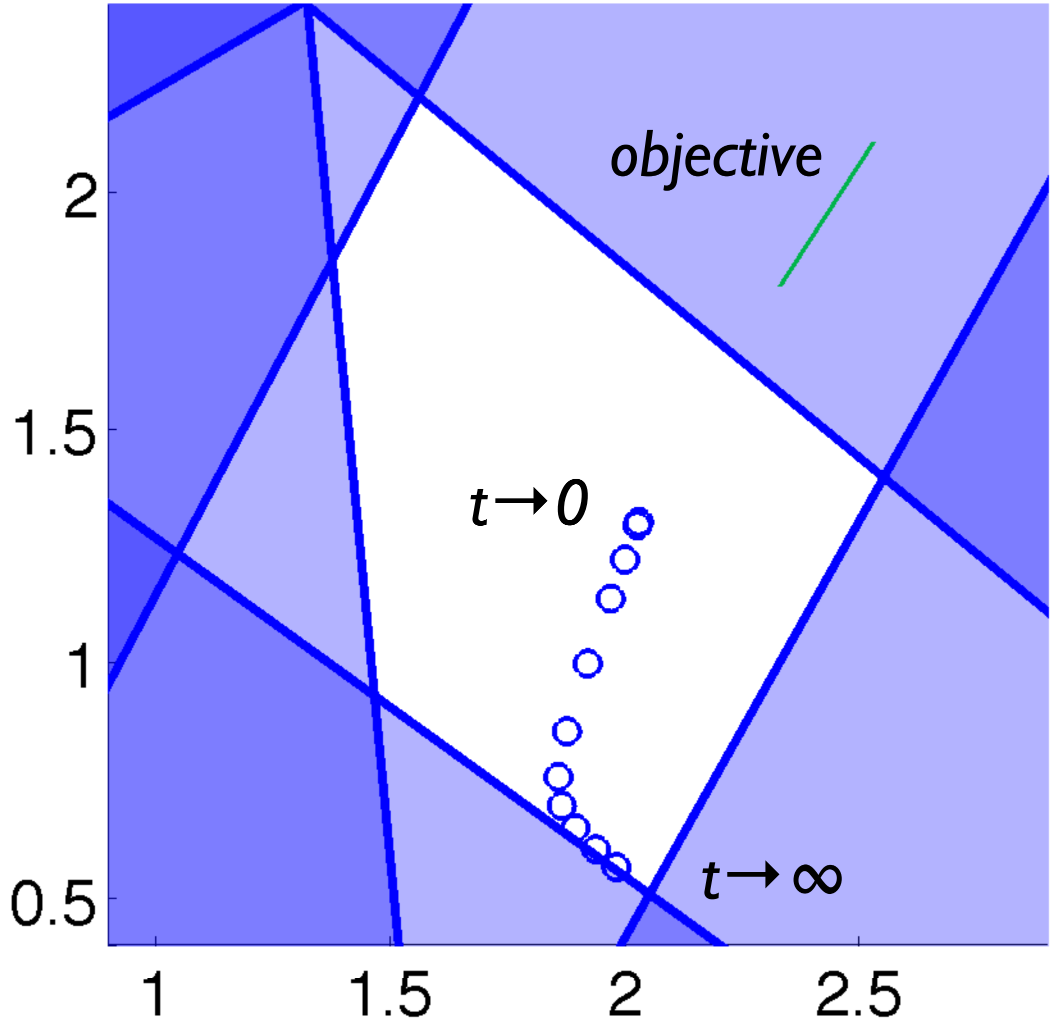
Adding an objective

- Analytic center was for: find x st $Ax + b \geq 0$
- Now: min $c^T x$ st $Ax + b \geq 0$
- Same trick:
  - min $f_t(x) = c^T x - (1/t) \sum \ln(a_i^T x + b_i)$
    - parameter $t > 0$
  - central path =
    - $t \to 0$: $t \to \infty$:

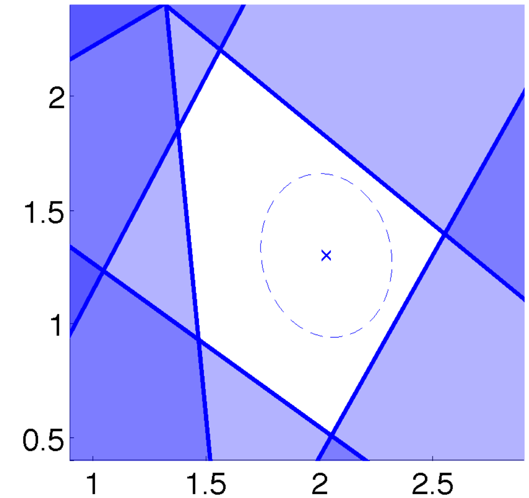
Newton for central path

- min $f_t(x) = c^T x - (1/t) \sum \ln(a_i^T x + b_i)$
  - $df/dx =$
  - $d^2f/dx^2 =$

Central path example



Dikin ellipsoid



- $E(x_0) = \{ x \mid (x-x_0)^T H(x-x_0) \le 1 \}$
  - H = Hessian of log barrier at x_0
  - unit ball of Hessian norm at x_0
- $E(x) \subseteq X$ for any strictly feasible x
  - affine constraints can be just feasible
  - E(x): as above, but intersected w/ affine constraints
- $\mathrm{vol}(E(x_{ac})) \ge \mathrm{vol}(X)/m$
  - weaker than ellipsoid center, but still very useful

$E(x_0) \subseteq X$

- $E(x_0) = \{ x \mid (x-x_0)^T H (x-x_0) \leq 1 \}$
  - $H = A^T S^{-2} A$
  - $S = \mathrm{diag}(s) = \mathrm{diag}(Ax_0 + b)$

Constraint form of central path

- $\min -\sum \ln s_i$ st $Ax + b \geq 0 \quad c^T x \leq \lambda$
- $\exists$ a 1-1 mapping $\lambda(t)$ w/ $x(\lambda(t)) = x(t)\ \forall t>0$
  - but this form is slightly less convenient since we don't know minimal feasible value of λ or maximal nontrivial value of λ

Dual of central path

- min $c^T x - (1/t) \sum \ln s_i$ st $Ax + b = s \geq 0$
    - $\min_{x,s} \max_y L(x,s,y) = c^T x - (1/t) \sum \ln s_i + y^T(s - Ax - b)$

Primal-dual correspondence

- Primal and dual for central path:
  - min $c^T x - (1/t) \sum \ln s_i$ st $Ax + b = s \geq 0$
  - max $(m \ln t)/t + m/t + (1/t) \sum \ln y_i - y^T b$ st $A^T y = c \quad y \geq 0$
- $L(x,s,y) = c^T x - (1/t) \sum \ln s_i + y^T(s - Ax - b)$
  - grad wrt s:
  - to get x:

Duality gap

- At optimum:
  - primal value $c^T x - (1/t) \sum \ln s_i$ =
    dual value $(m \ln t)/t + m/t + (1/t) \sum \ln y_i - y^T b$
  - $s \circ y = te$

Primal-dual constraint form

- Primal-dual pair:
  - min $c^T x$ st $Ax + b \geq 0$
  - max $-b^T y$ st $A^T y = c \quad y \geq 0$
- KKT:
  - $Ax + b \geq 0$ (primal feasibility)
  - $y \geq 0 \quad A^T y = c$ (dual feasibility)
  - $c^T x + b^T y \leq 0$ (strong duality)
  - …or, $c^T x + b^T y \leq \lambda$ (relaxed strong duality)

Analytic center of relaxed KKT

- Relaxed KKT conditions:
  - $Ax + b \geq 0$
  - $y \geq 0$
  - $A^T y = c$
  - $c^T x + b^T y \leq \lambda$
- Central path = {analytic centers of relaxed KKT}