Quadratic programs
Cone programs

10-725 Optimization
Geoff Gordon
Ryan Tibshirani

Administrivia

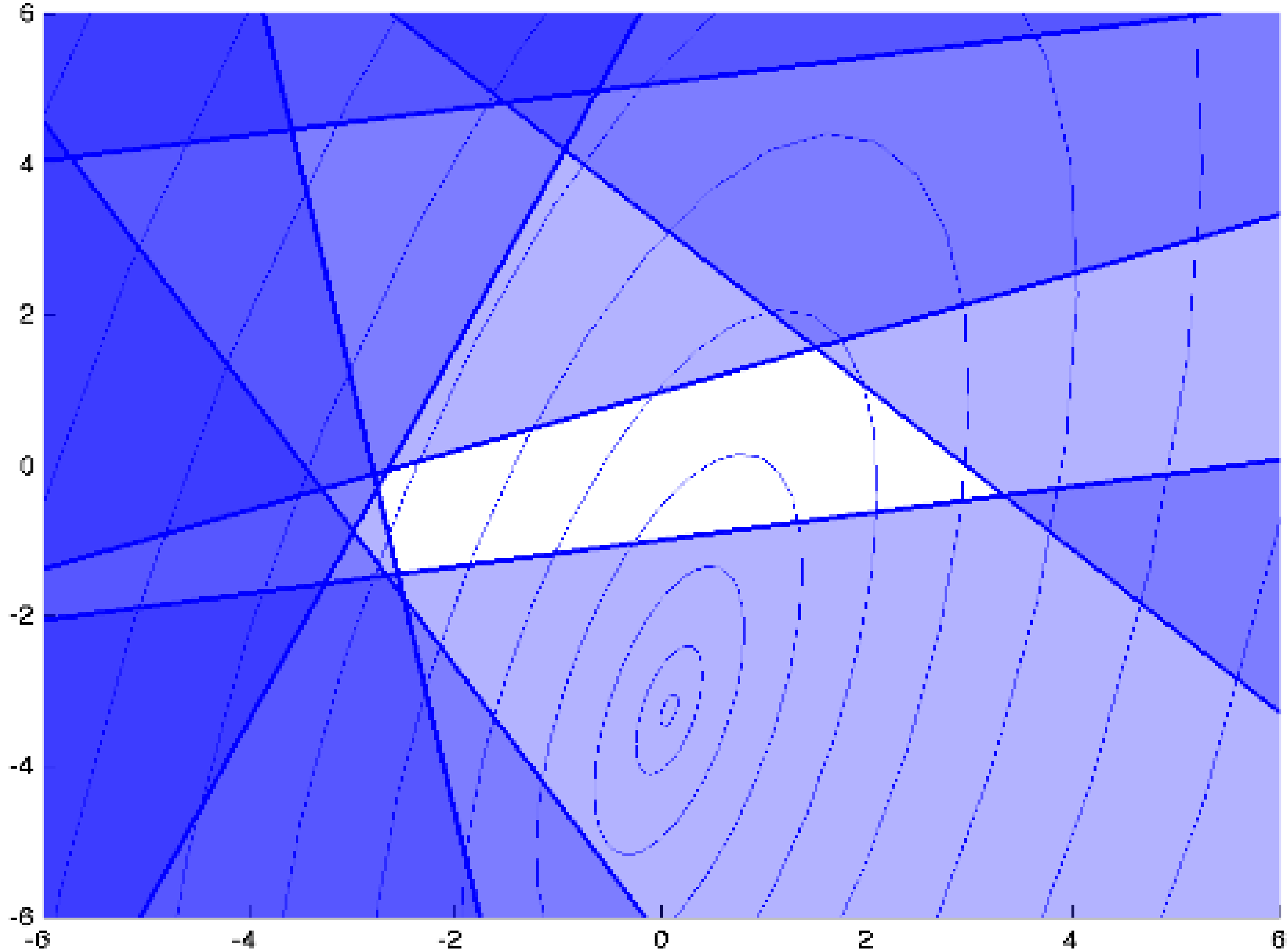
- HW3 back at end of class
- Last day for feedback survey
- All lectures now up on Youtube (and continue to be downloadable from course website)
- Reminder: midterm next Tuesday 11/6!
  - in class, 1 hr 20 min, one sheet (both sides) of notes

Quadratic programs

- m constraints, n vars
  - A: $R^{m \times n}$ b: R^m c: R^n x: R^n H: $R^{n \times n}$
  - [min or max] $x^T H x/2 + c^T x$
  - s.t. $Ax \leq b$ or $Ax = b$ [or some mixture]
  - may have (some elements of) $x \geq 0$
- Convex problem if:
  -

max $2x+x^2+y^2$ s.t.
$x + y \leq 4$
$2x + 5y \leq 12$
$x + 2y \leq 5$
$x, y \geq 0$

For example

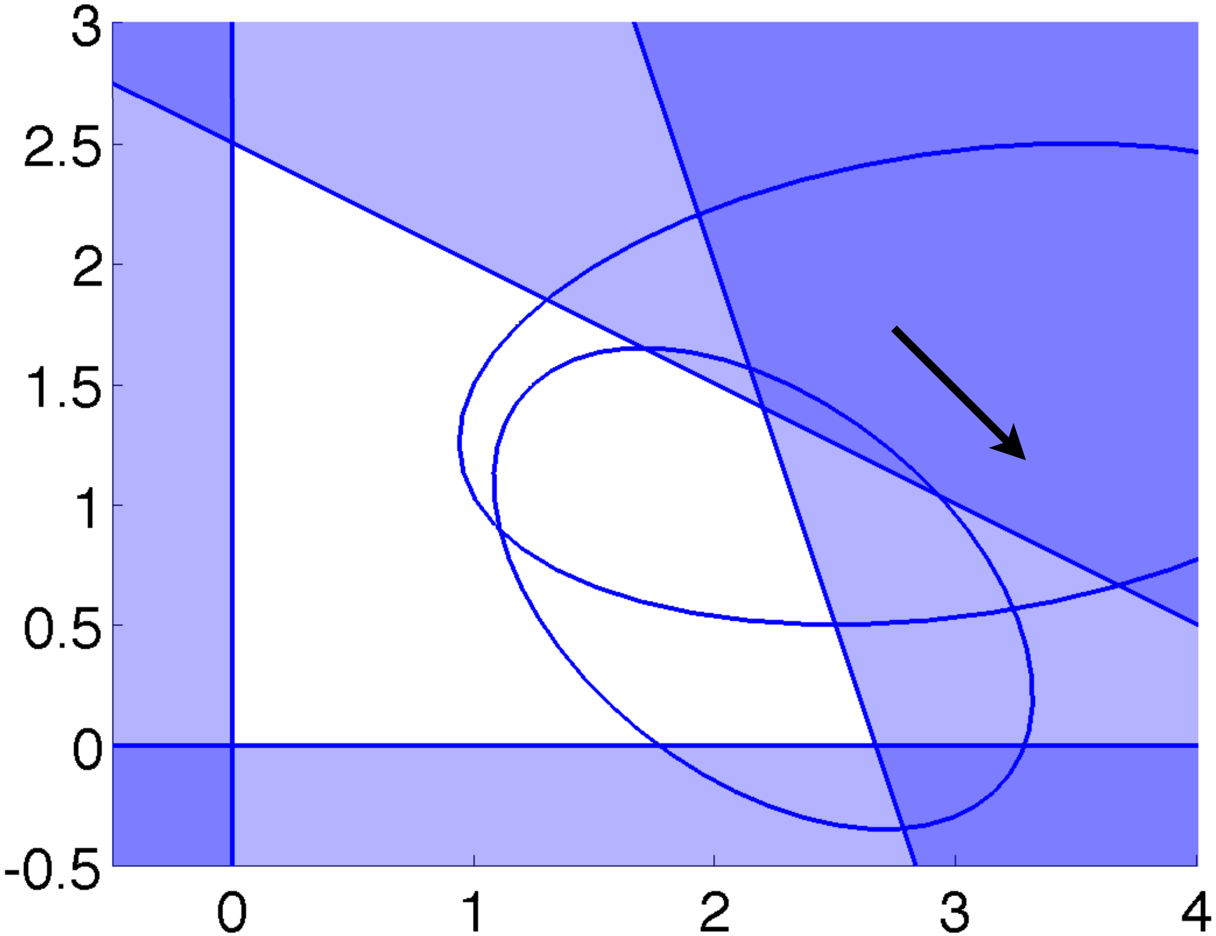


Cone programs

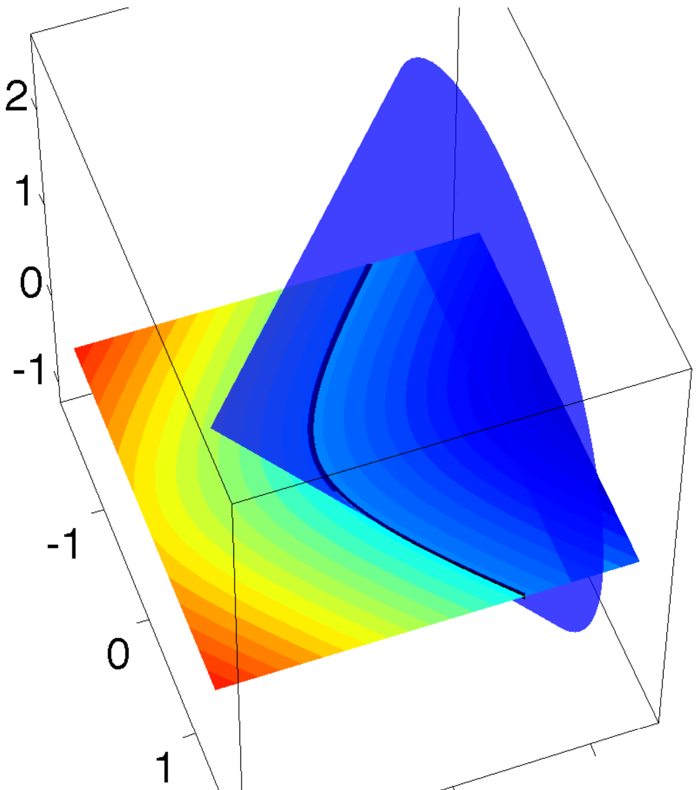
- m constraints, n vars
  - A: $R^{m \times n}$ b: R^m c: R^n x: R^n
  - Cones $K \subseteq R^m$ $L \subseteq R^n$
  - [min or max] $c^T x$ s.t. $Ax + b \in K$ $x \in L$
  - convex if
- E.g., K =
- E.g., L =

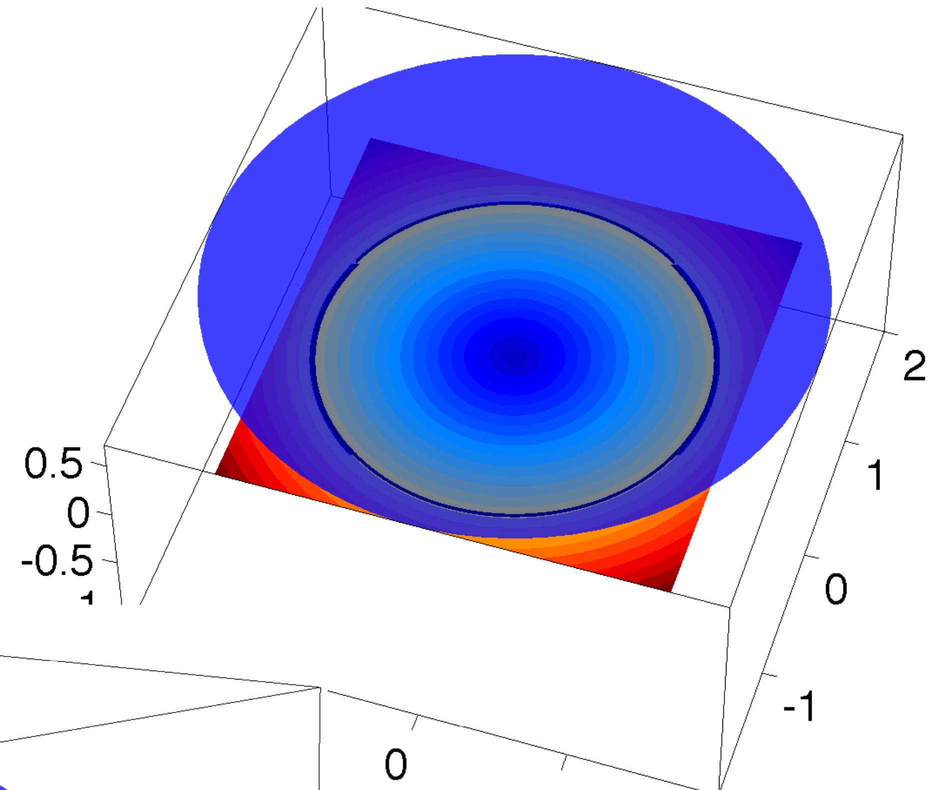
For example: SOCP

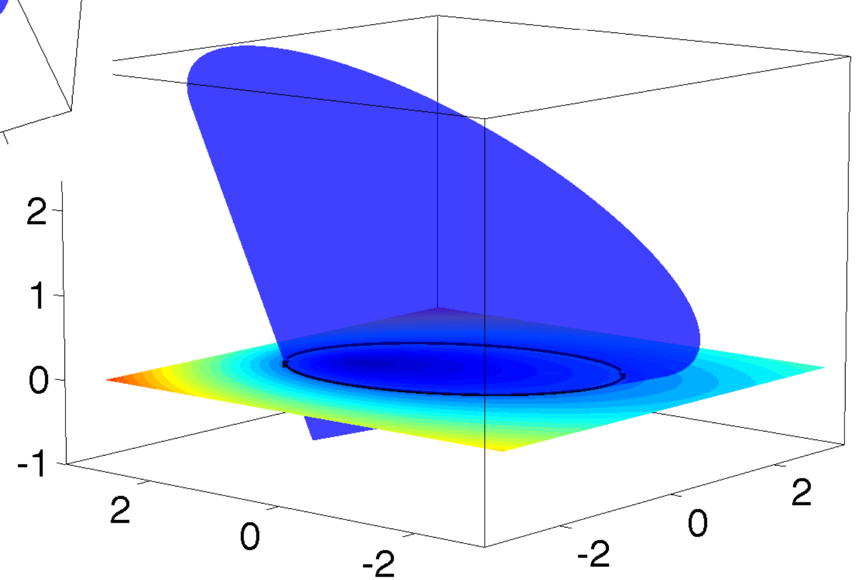
- min $c^T x$ s.t. $A_i x + b_i \in K_i$, $i = 1, 2, \ldots$



Conic sections







QPs are reducible to SOCPs

- min $x^T H x/2 + c^T x$ s.t. …

∃ SOCPs that aren't QPs?

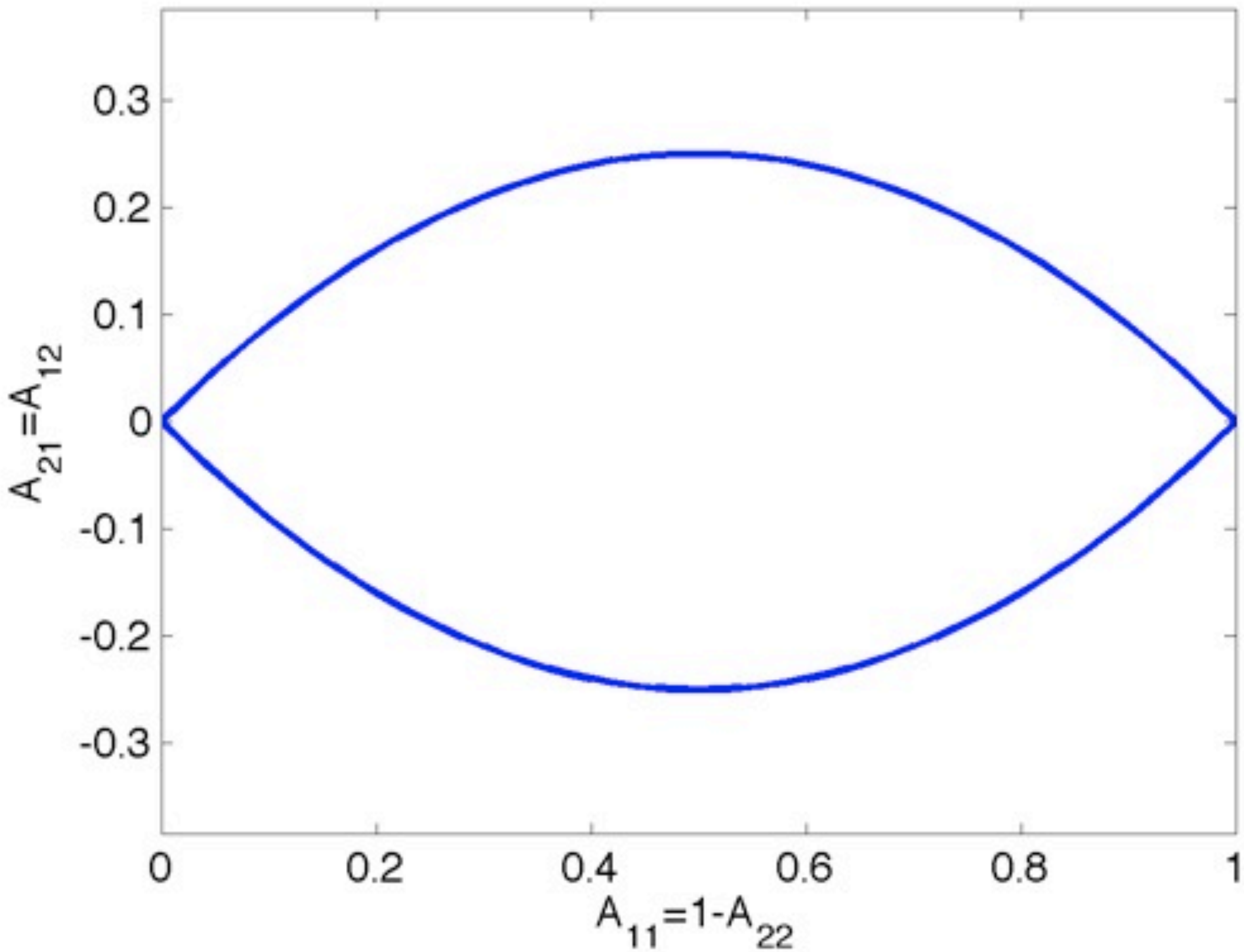
- QCQP: convex quadratic objective & constraints
- minimize $a^2 + b^2$ s.t.
    - $a \geq x^2$, $b \geq y^2$
    - $2x + y = 4$
- Not a QP (nonlinear constraints)
    - but, can rewrite as SOCP

More cone programs: SDP

- Semidefinite constraint:
  - variable $x \in R^n$
  - constant matrices $A_1, A_2, \ldots \in R^{m \times m}$
  - constrain
- Semidefinite program: min $c^T x$ s.t.
  - semidefinite constraints
  - linear equalities
  - linear inequalities

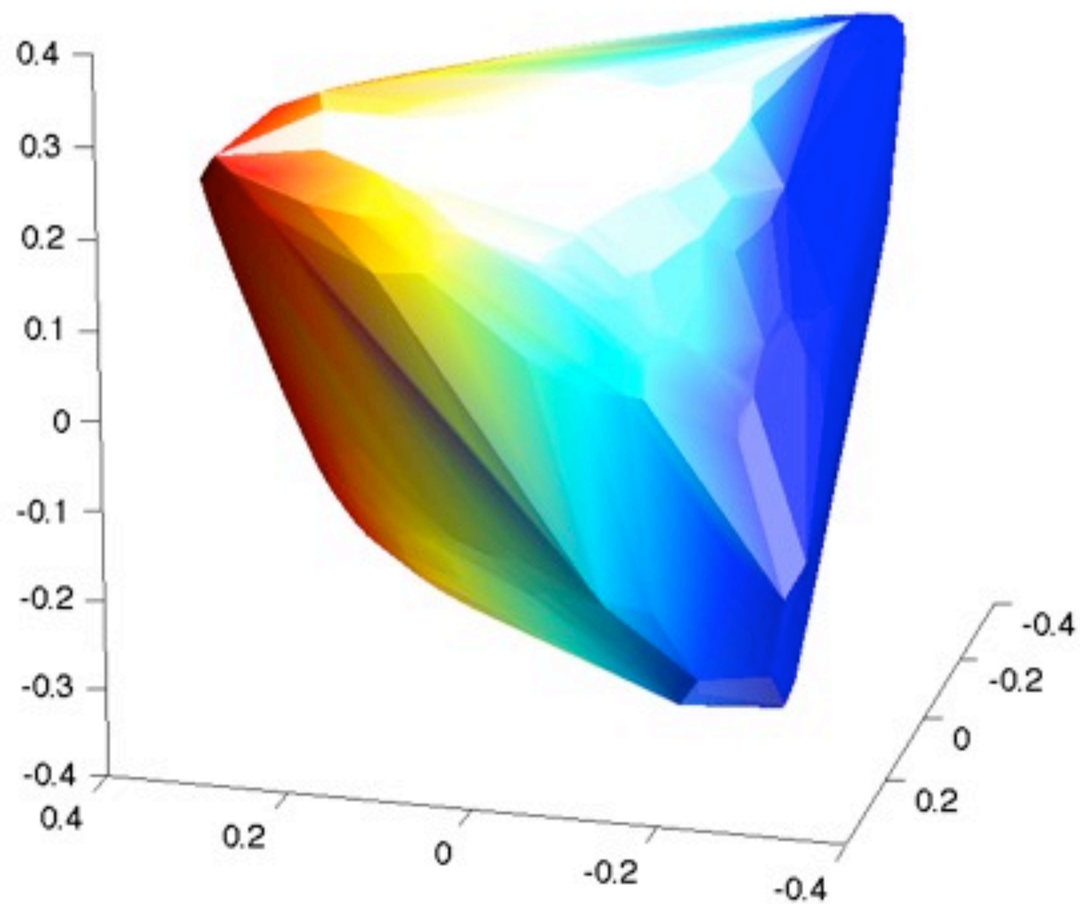
Visualizing S_+

- 2 x 2 symmetric matrices w/ tr(A) = 1

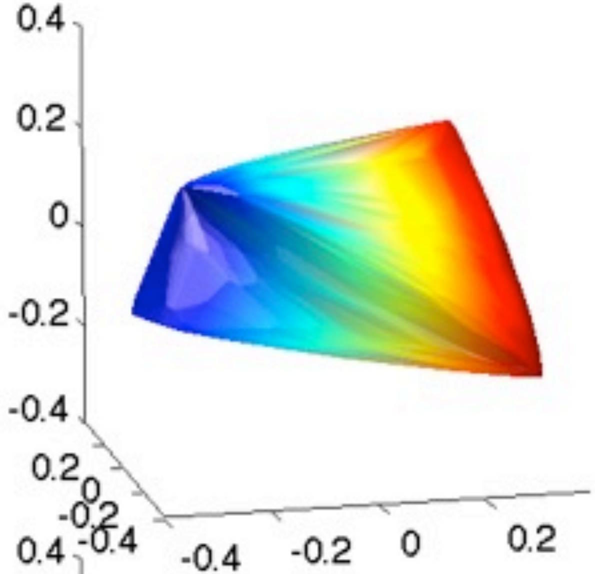


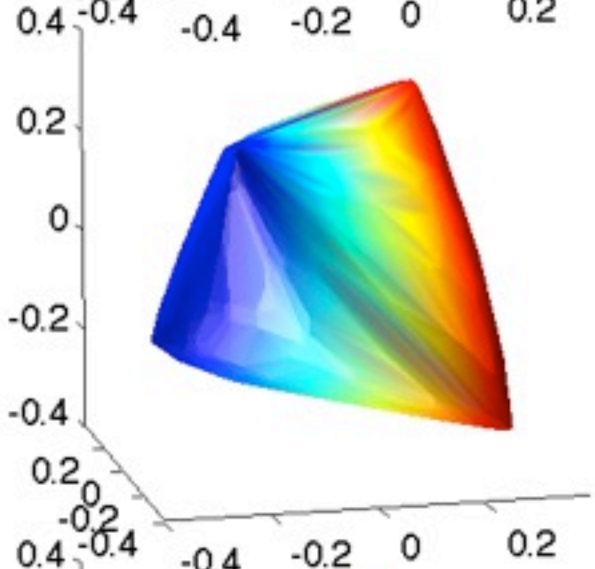
What about 3 x 3?

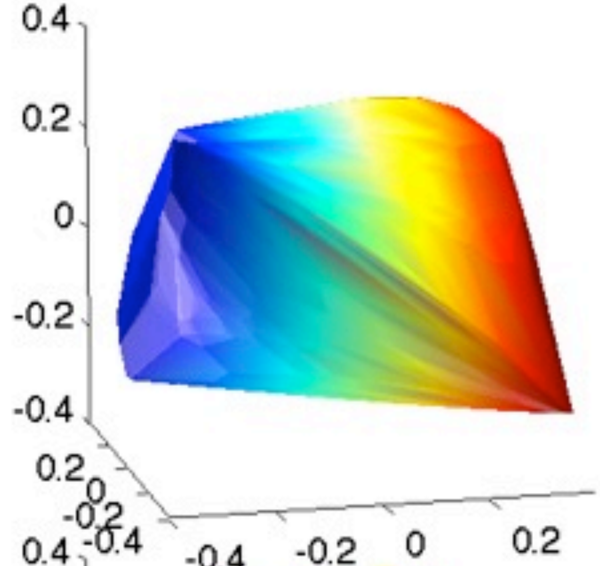
- Try setting entire diagonal to 1/3
  - plot off-diagonal elements (3 of them)

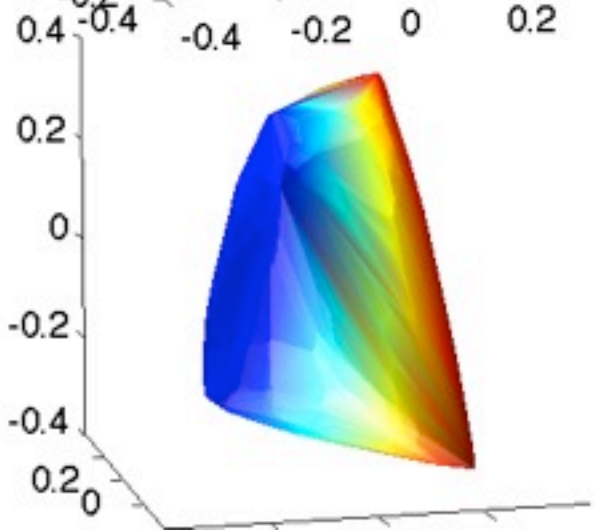


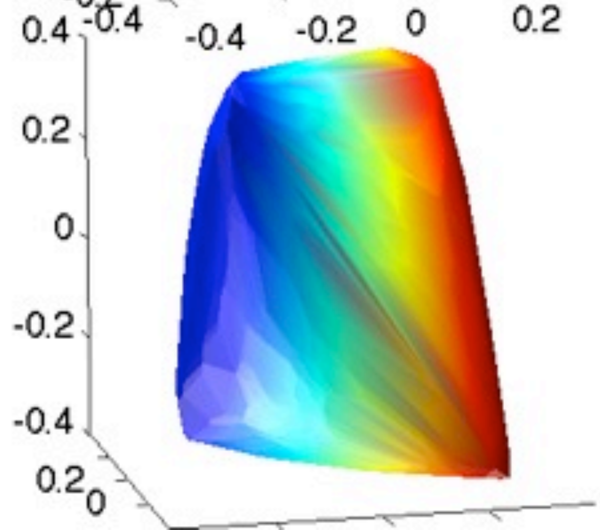
3×3 symmetric psd matrices

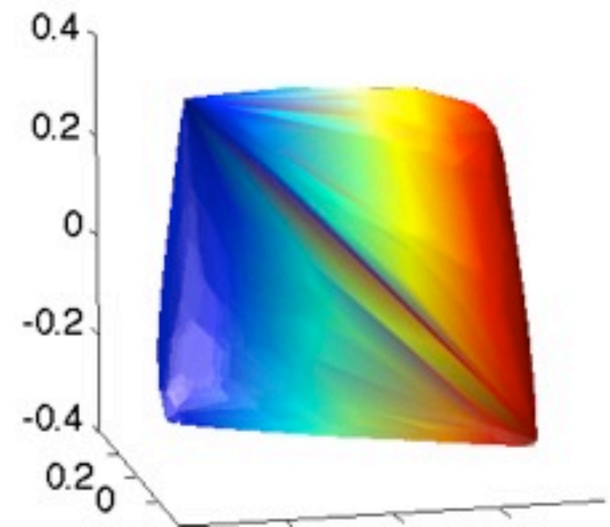












S_+ *is self-dual*

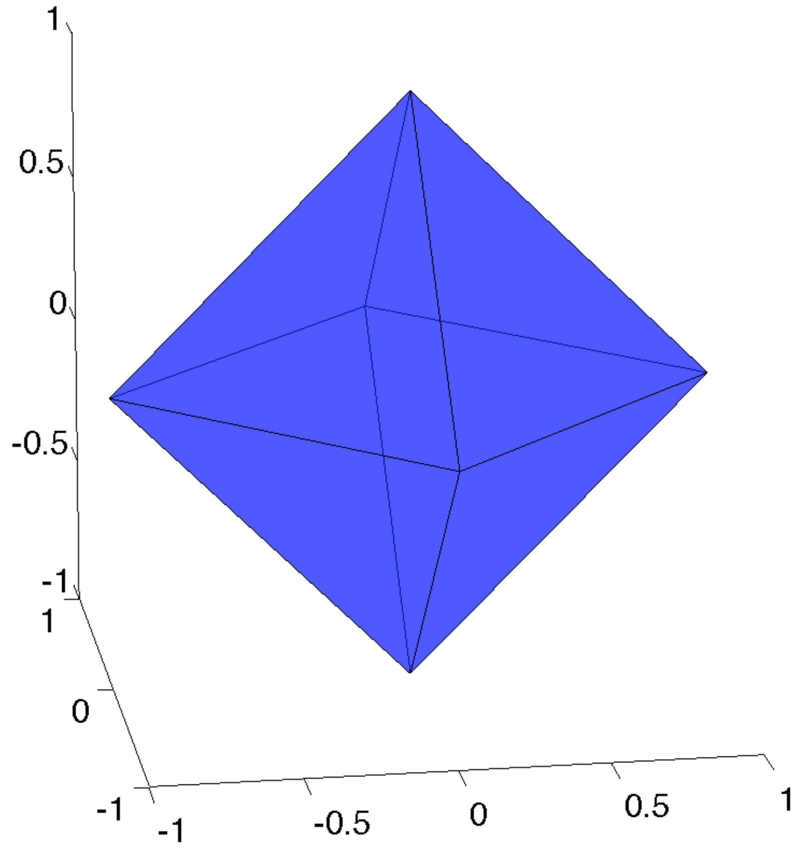
- S_+: { A | $A=A^T$, $x^TAx \geq 0$ for all x }
- [$x^TAx \geq 0$ for all x] ⇔
  [$\mathrm{tr}(B^TA) \geq 0$ for all psd B]

How hard are QPs and CPs?

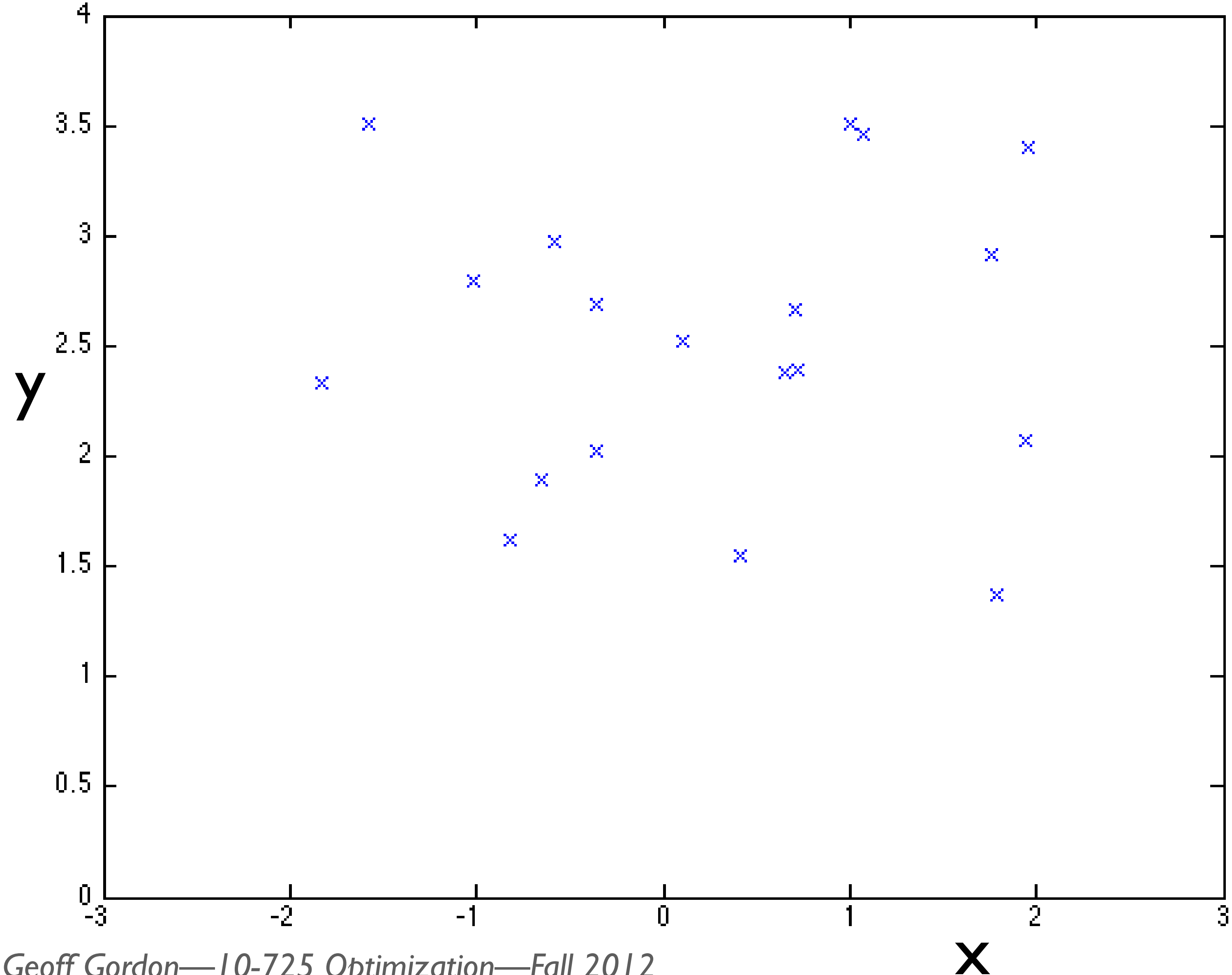
- Convex QP or CP: not much harder than LP!
  - as long as we have an efficient rep'n of the cone
  - poly(L, $1/\epsilon$)   (L = bit length, ϵ = accuracy)
  - can we get strongly polynomial (no $1/\epsilon$)?
    - famous open question, even for LP
- General QP or CP: NP-complete
  - e.g., reduce max cut to QP

QP examples

- Euclidean projection
- LASSO
  - Mahalanobis projection
- Huber regression
- Support vector machine

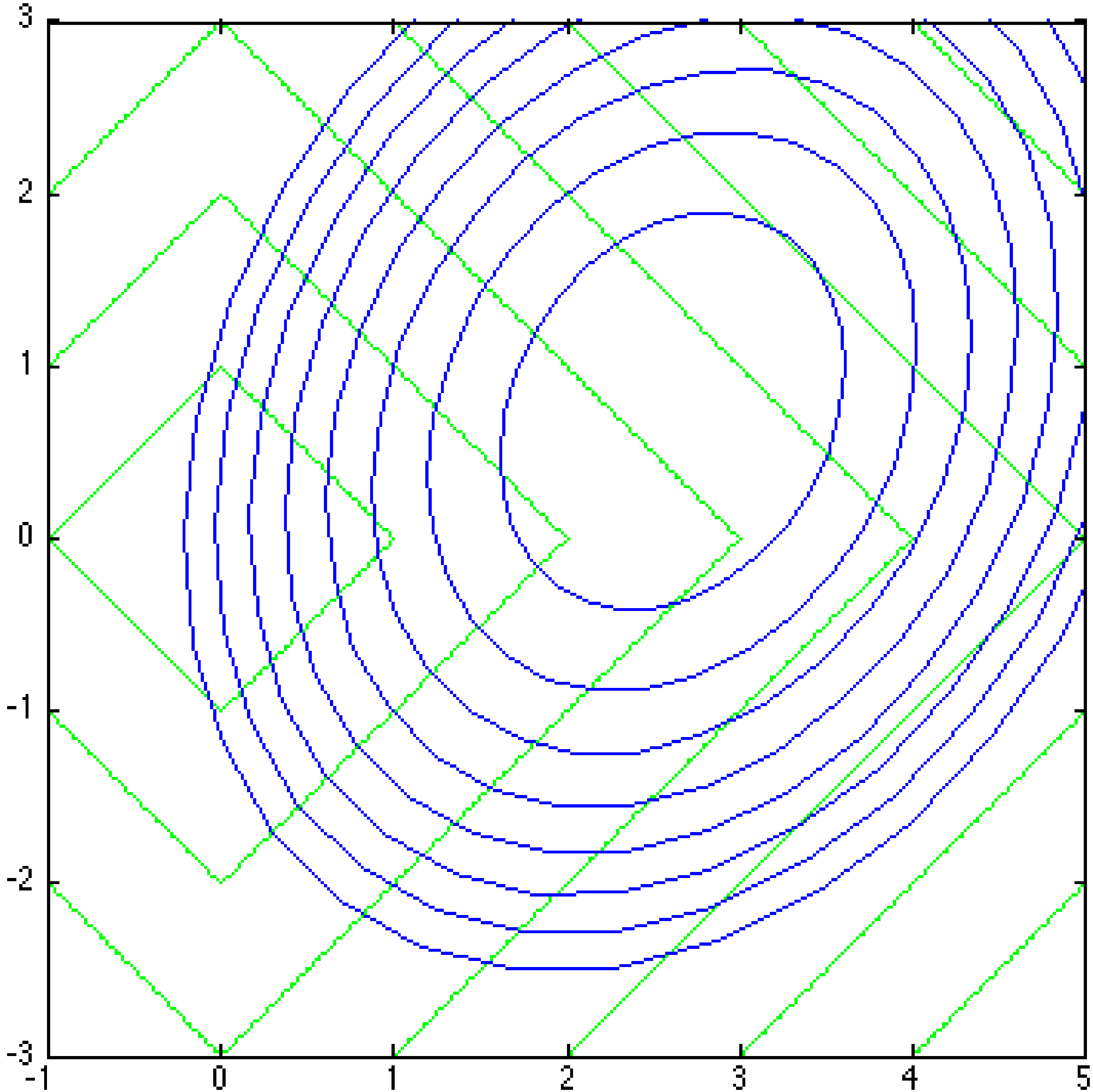


LASSO example



fit y = ax+b
w/ (a,b) sparse

LASSO example



Robust (Huber) regression

- Given points (x_i, y_i)
  - L_2 regression: $\min_w \Sigma_i (y_i - x_i^T w)^2$
- Problem: overfitting!
- Solution: Huber loss
  - $\min_w \Sigma_i \mathrm{Hu}(y_i - x_i^T w)$

$\mathrm{Hu}(z) =$

Huber loss as QP

- $\text{Hu}(z) = \min_{a,b} (z + a - b)^2 + 2a + 2b$
  - s.t. $a, b \geq 0$

Cone program examples

- SOCP
  - (sparse) group lasso
  - discrete MRF relaxation
    - [Kumar, Kolmogorov, Torr, JMLR 2008]
  - min volume covering ellipsoid (nonlinear objective)

Cone program examples

- SDP
  - graphical lasso (nonlinear objective)
  - Markowitz portfolio optimization (see B&V)
  - max-cut relaxation [Goemans, Williamson]
  - matrix completion
  - manifold learning: max variance unfolding

Matrix completion

- Observe A_{ij} for $ij \in E$, write $P_{ij} = \{$
- $\min \|(X-A) \circ P\|_F^2 + \lambda \|X\|_*$

Max-variance unfolding

aka semidefinite embedding

- Goal: given $x_1, \dots x_T \in R^n$
  - find $y_1, \dots, y_T \in R^k$ $(k \ll n)$
  - $||y_i - y_j|| \approx ||x_i - x_j||$ $\forall i,j \in E$
- If x_i were near a k-dim subspace of R^n, PCA!
- Instead, two steps:
  - first look for $z_1, \dots z_T \in R^n$ 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. $\|z_i - z_j\| = \|x_i - x_j\| \quad \forall i,j \in E$

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

Primal-dual pair

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

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: $y^T(Ax+b) = 0 \quad x^T(Hz+c-A^T y) = 0$

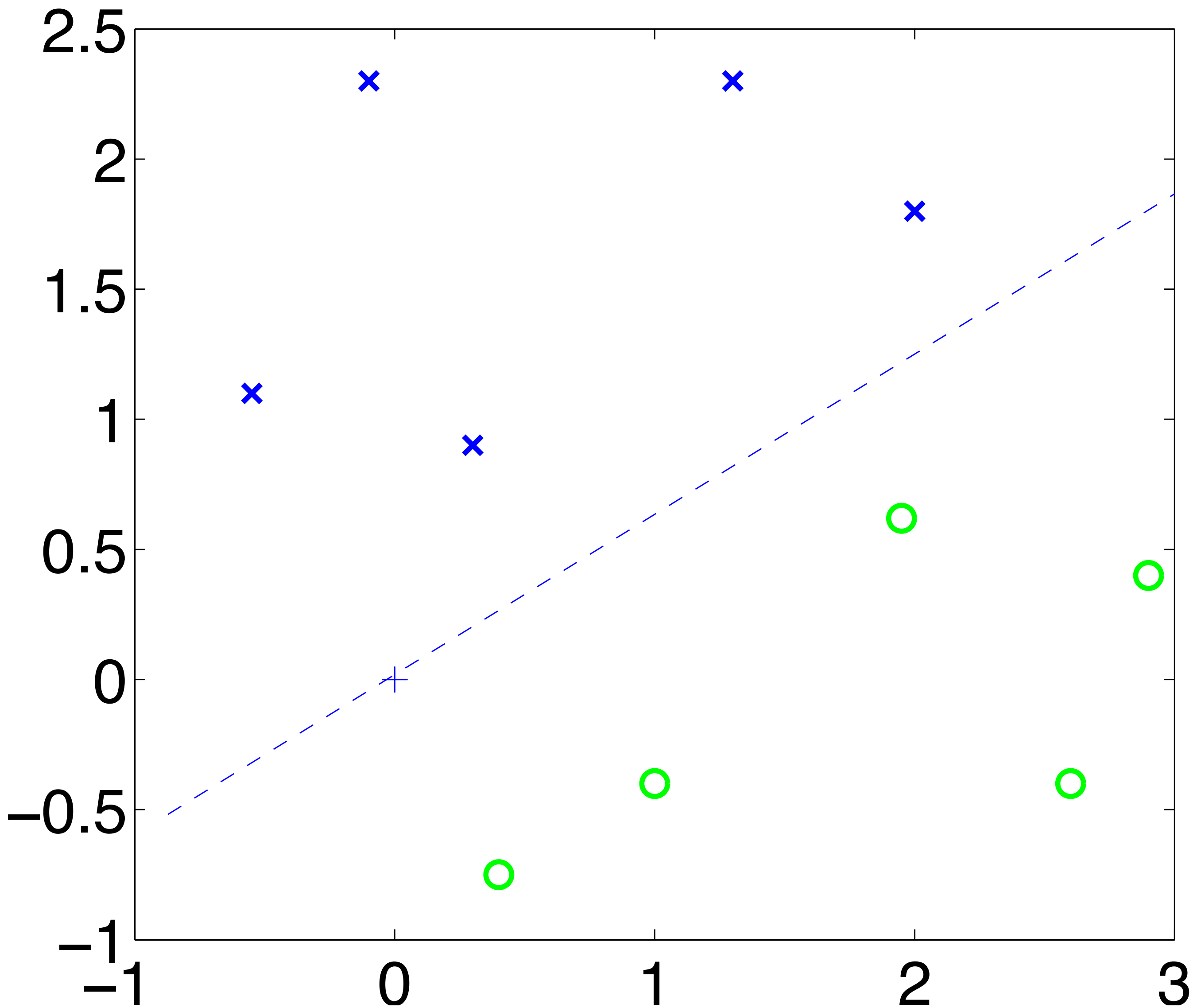
Support vector machines

(separable case)

Maximizing margin

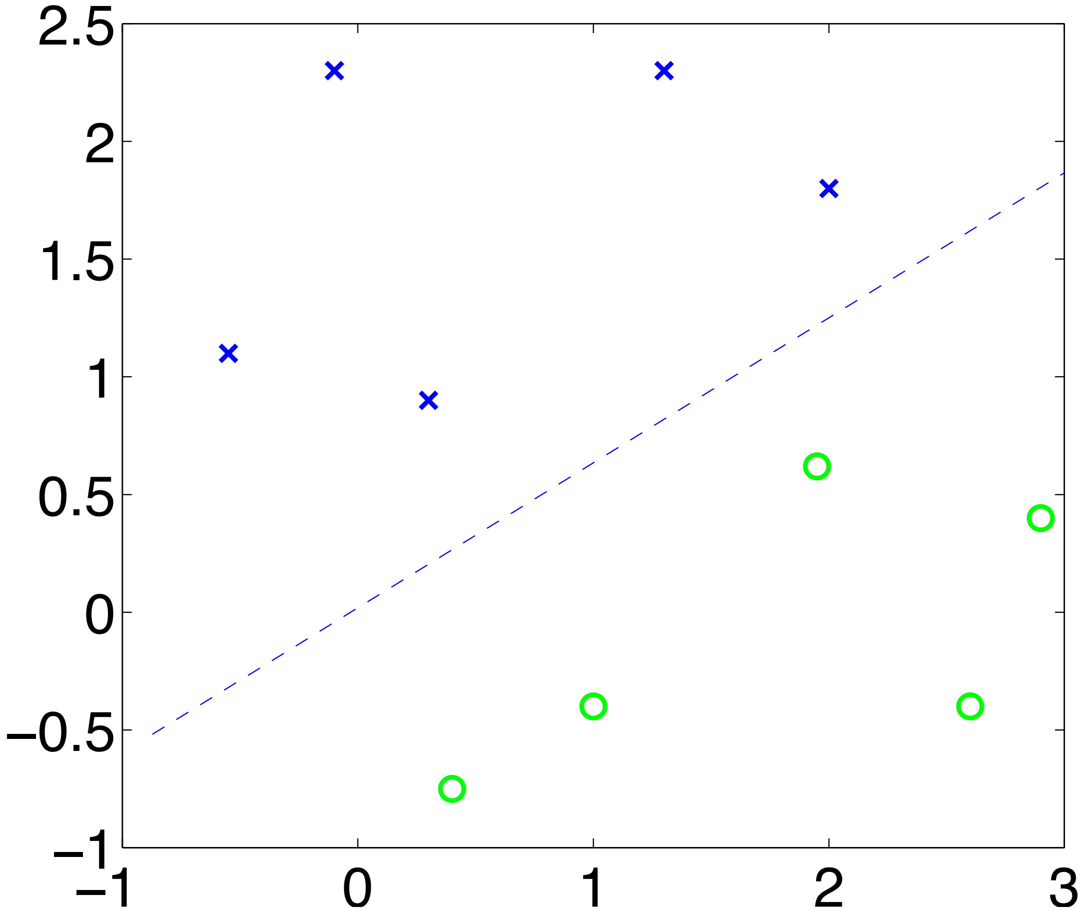
- margin $M = y_i (x_i \cdot \bar{w} - \bar{b})$
- $\max M \text{ s.t. } M \leq y_i (x_i \cdot \bar{w} - \bar{b})$

For example



Slacks

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



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

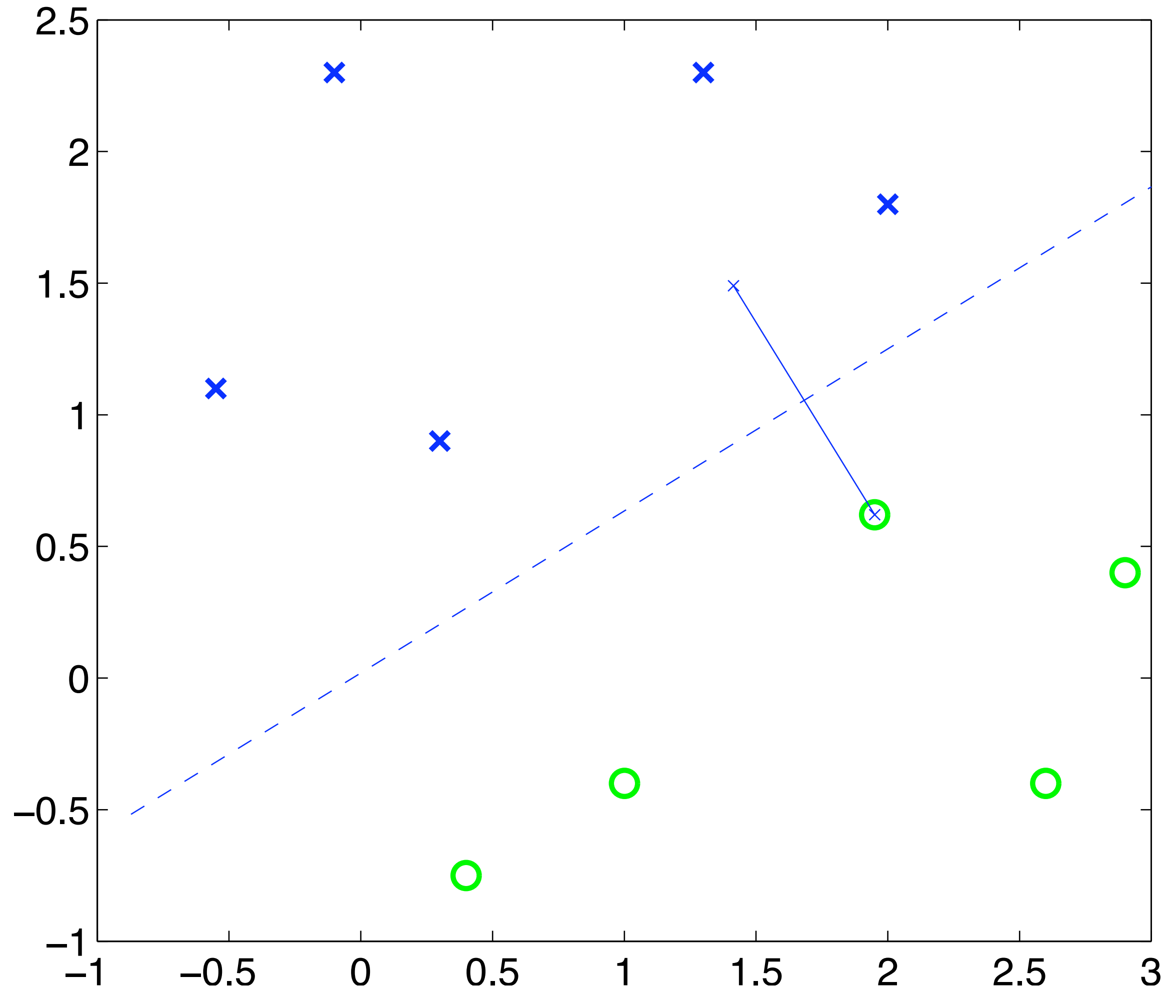
- max $1^T\alpha - \alpha^T K\alpha/2$ s.t. $y^T\alpha = 0 \quad 0 \le \alpha \le 1$

α:

$\alpha > 0$:

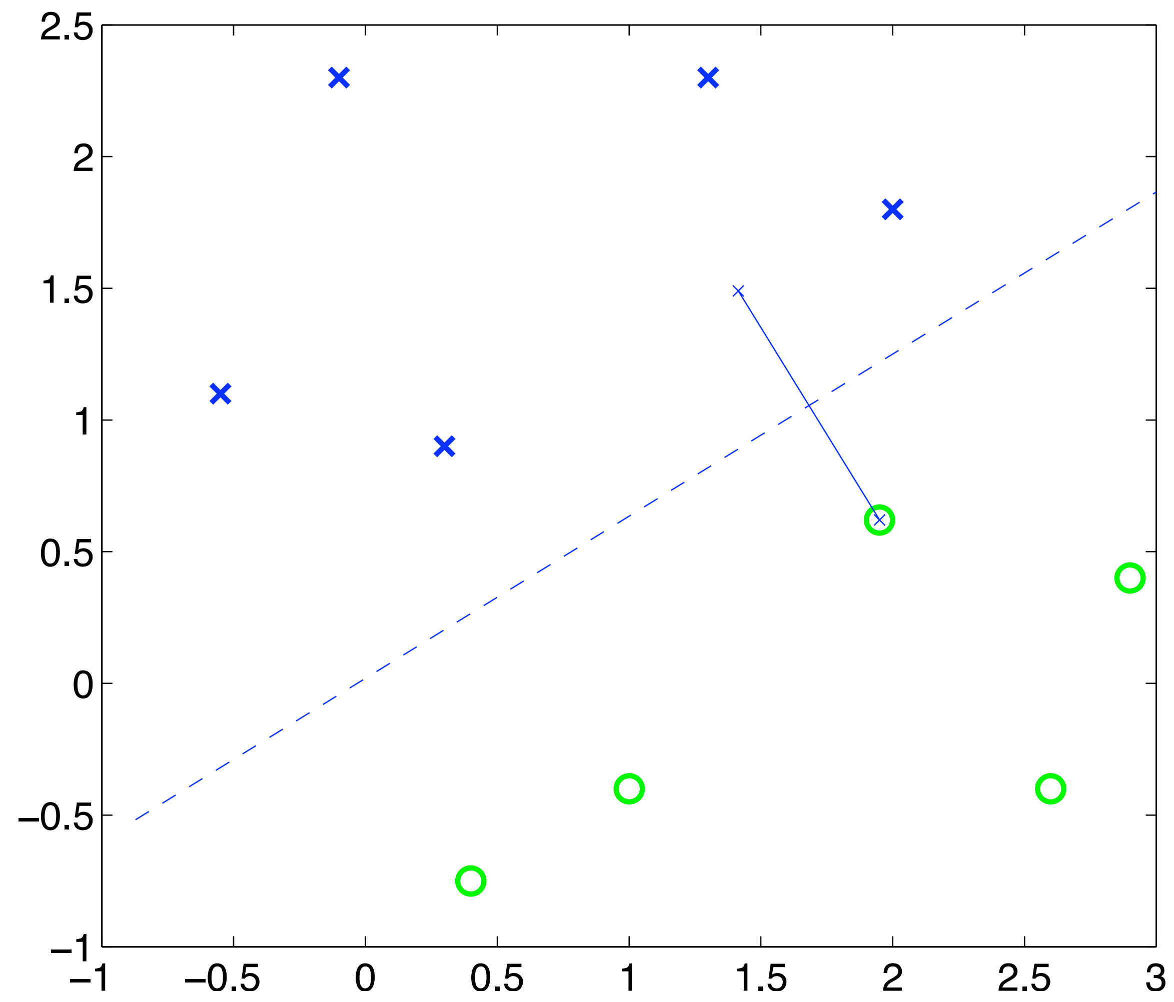
$\alpha < 1$:

$y^T\alpha = 0$:

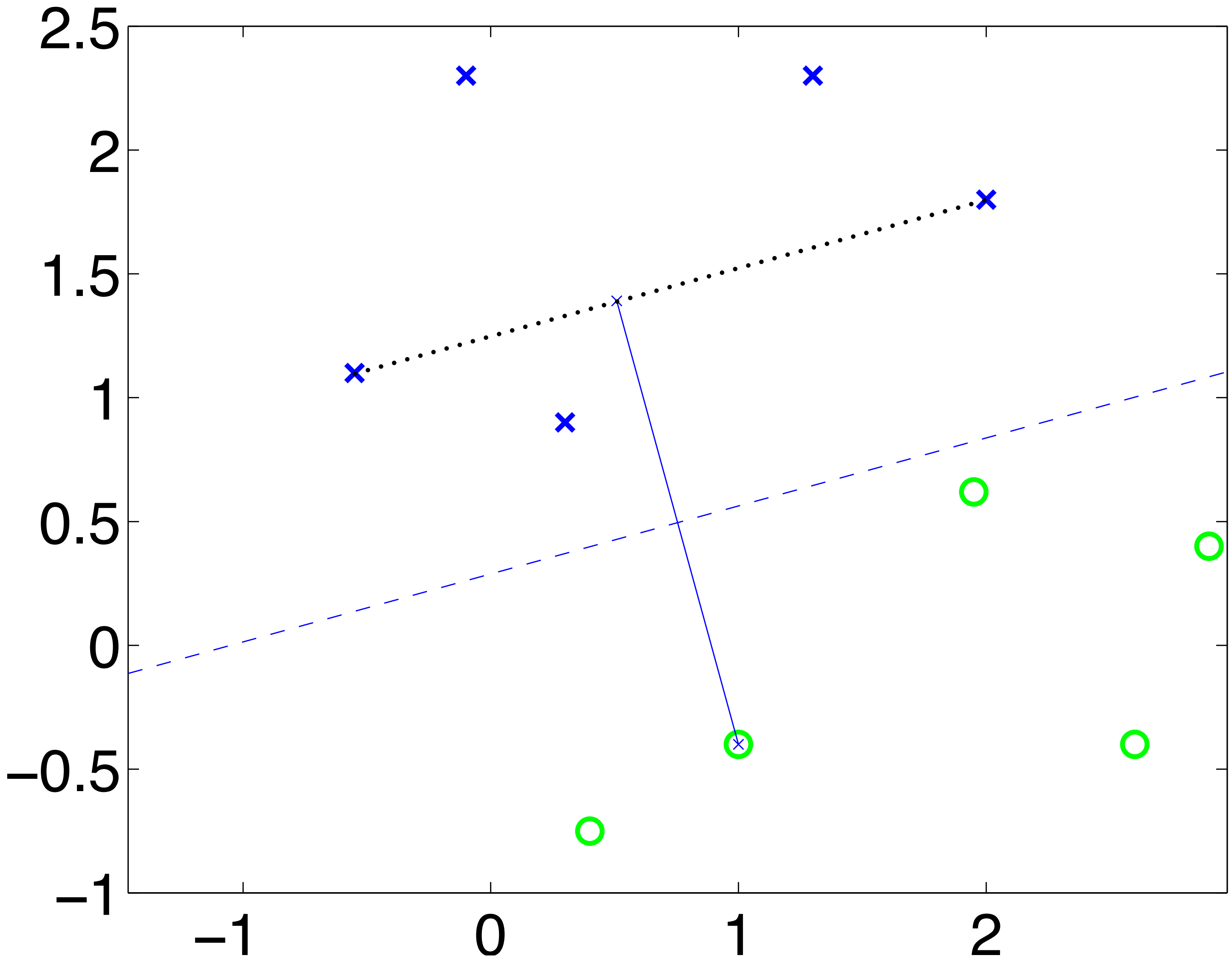


From dual to primal

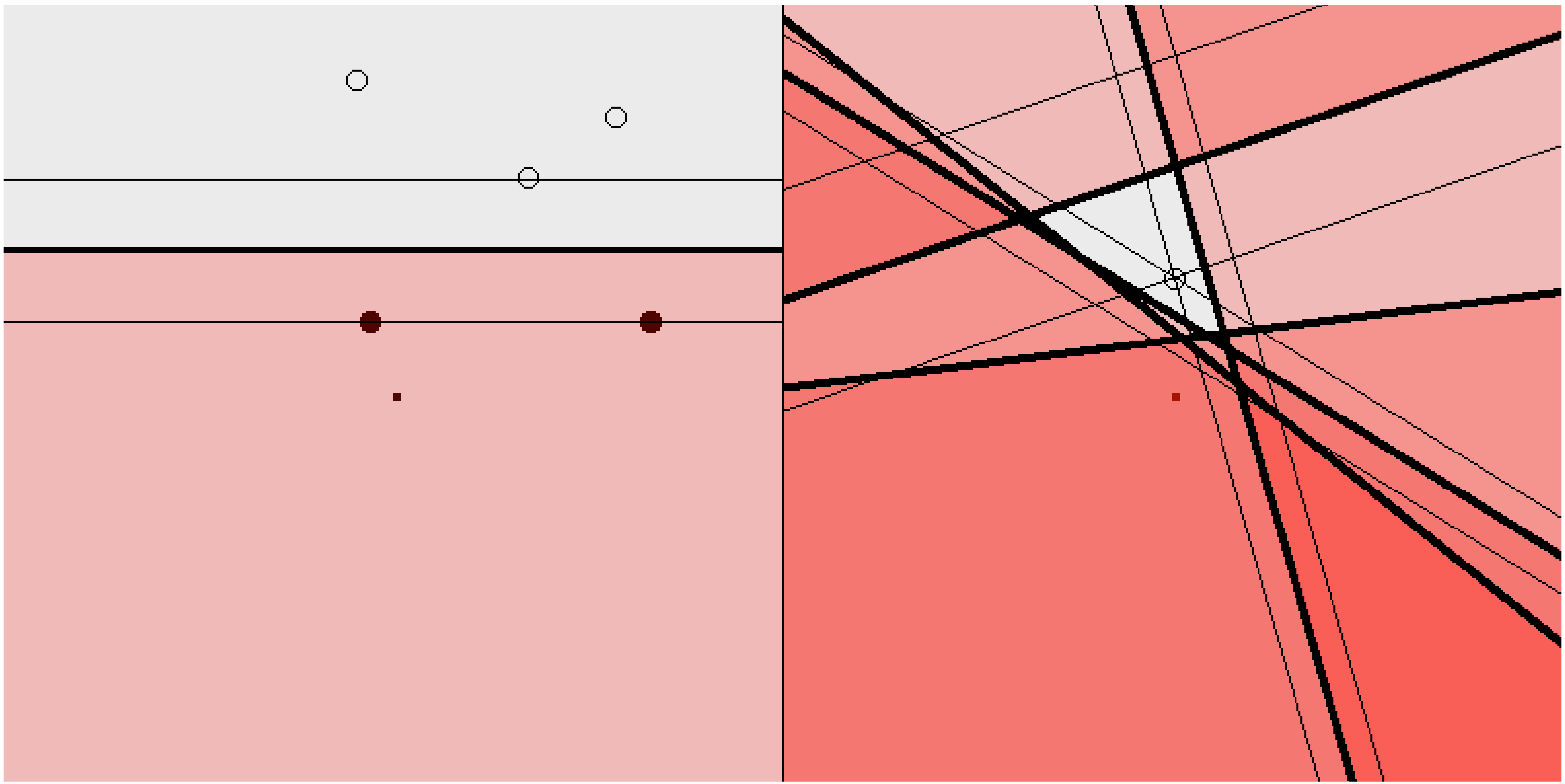
- max $1^T\alpha - \alpha^T K\alpha/2$ s.t. $y^T\alpha = 0 \quad 0 \leq \alpha \leq 1$



A suboptimal support set



SVM duality: the applet



Why is the dual useful?

aka the kernel trick

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

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