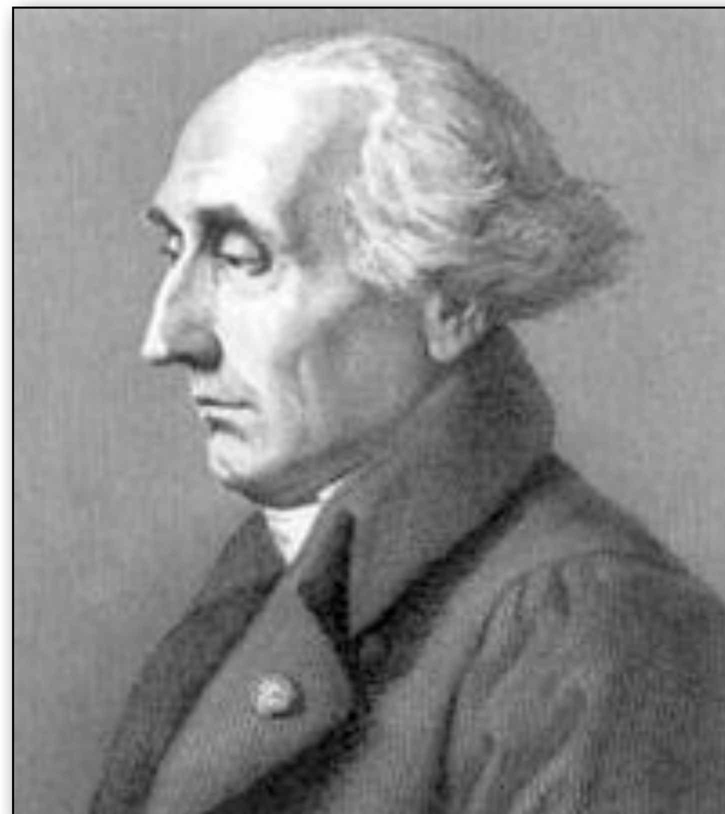
Lagrange multipliers

- Technique for turning constrained optimization problems into unconstrained ones
  - for intuition or for algorithm
- Useful in general
  - but in particular, leads to a famous ML method: the support vector machine



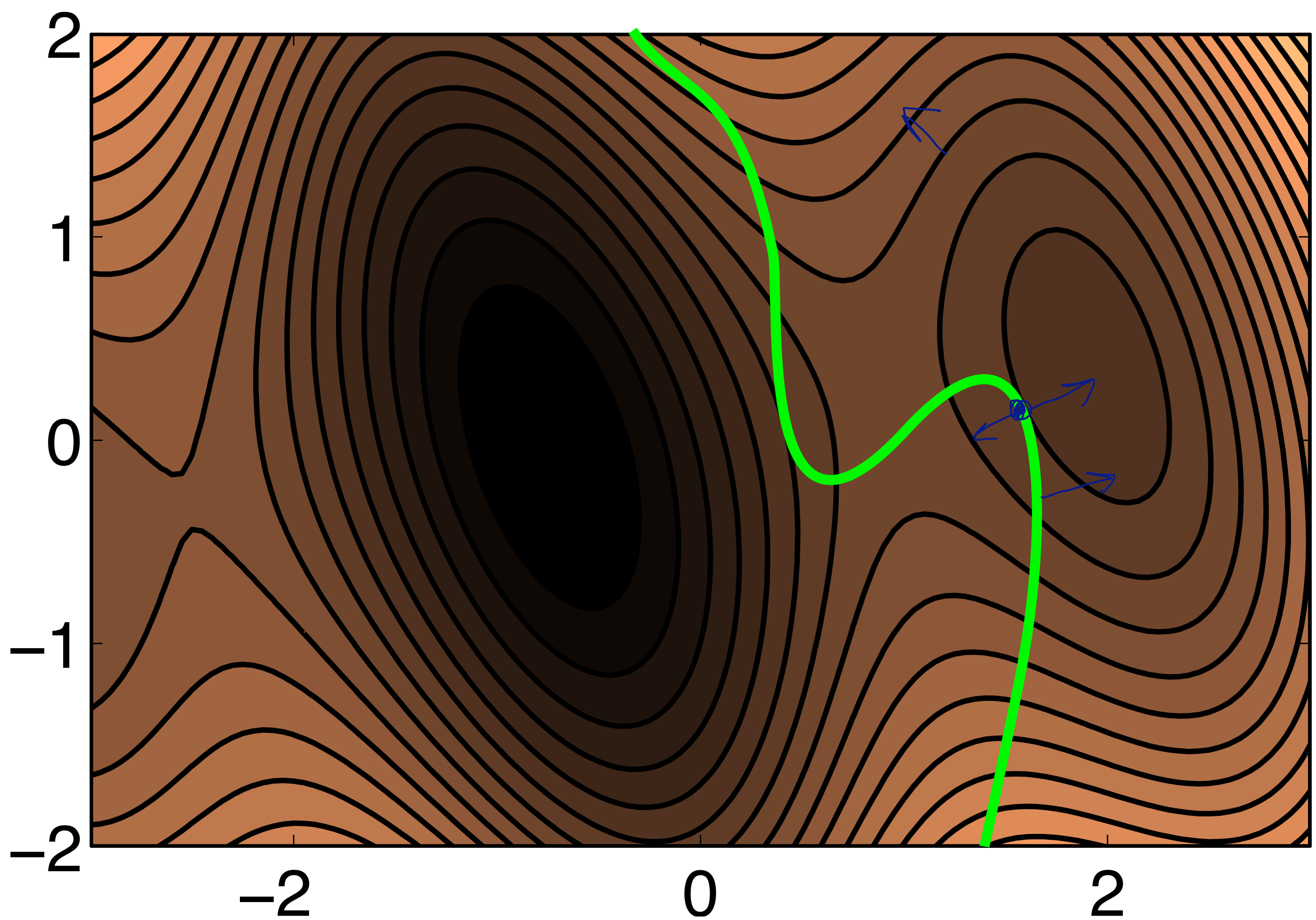
Recall: Newton's method

- $\min_x f(x) \quad \rightarrow \quad H(x)\Delta x + g(x) = 0$
  - $f: R^d \rightarrow R$
- Why: $\quad f(x+\Delta x) \sim f + g^T\Delta x + \Delta x^T H \Delta x/2$
  - $f = f(x)$, $g = g(x)$, $H = H(x)$
  - set derivative wrt Δx to 0

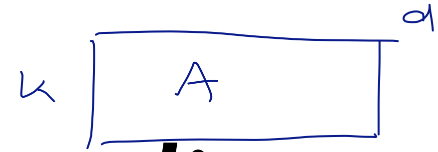
Equality constraints

- min f(x) s.t. p(x) = 0

$\nabla p(x) = \lambda \nabla f(x)$



$\lambda \in \mathbb{R}^k$



Optimality w/ equality

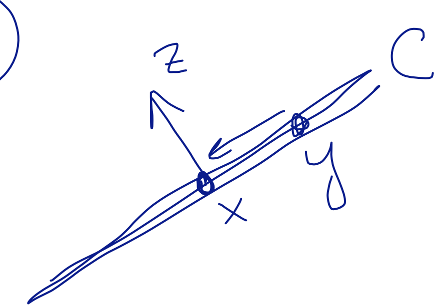
- min f(x) s.t. p(x) = 0
  - f: $R^d \rightarrow R$ p: $R^d \rightarrow R^k$ $(k \leq d)$
  - g: $R^d \rightarrow R^d$ (g = df/dx gradient of f)
- Useful special case: min f(x) s.t. Ax = b

$p(x) = Ax - b$

$C = \{x \mid Ax = b\}$

$z \perp C$ iff $z^T(x-y) = 0$ $\forall x, y \in C$

$\langle z, x-y \rangle$



$\underline{z = A^T\lambda}$

$\lambda^T A(x-y) = \lambda^T(b-b) = 0$

$$\boxed{\begin{array}{l} g(x) = A^T\lambda \\ p(x) = 0 \end{array}}$$

$\min_x f(x)$ st $p(x) = 0$

$p = Ax - b$

$J = A$

More generally

- f: $R^d \to R$ p: $R^d \to R^k$ J: $R^d \to R^{k \times d}$ J = dp/dx

$$\boxed{\begin{aligned} p(x) &= 0 \\ g(x) &= J(x)^T \lambda \end{aligned}}$$

KKT

$$\begin{array}{cc} & \\ k & \boxed{\quad J(x) \quad} \\ & d \end{array}$$

$c = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ $a = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$

z x y

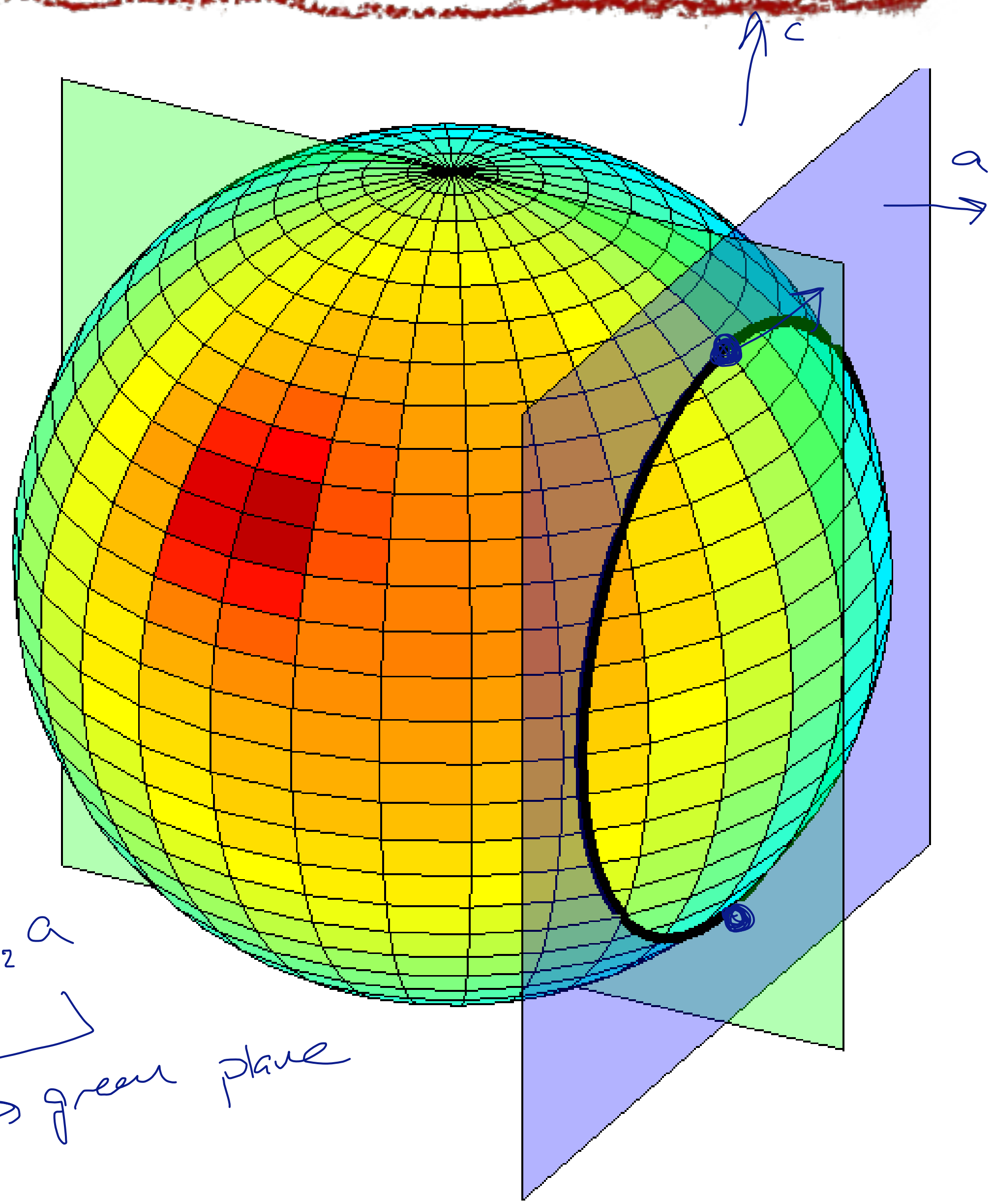
Picture

$$\max_{x,y,z} c^\top \begin{bmatrix} x \\ y \\ z \end{bmatrix} \text{ s.t.}$$

$$x^2 + y^2 + z^2 = 1$$

$$a^\top x = b$$

$\begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix}$ a



feas $\left[\begin{array}{l} x^2 + y^2 + z^2 = 1 \\ a^\top \begin{pmatrix} x \\ y \\ z \end{pmatrix} = b \end{array} \right.$

1st order opt $\left\{ c = \lambda_1 \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix} + \lambda_2 a \right.$

→ green plane

Newton w/ equality

KKT

- min f(x) s.t. p(x) = 0
  - f: $R^d \to R$, p: $R^d \to R^k$

$$df/dx = g(x) \qquad dp/dx = J(x)$$

$$\boxed{\begin{aligned} p(x) &= 0 \\ g(x) &= J(x)^T \lambda \end{aligned}}$$

- Now suppose:
  - dg/dx = $H(x)$ $d \times d$

$$p(x+\Delta x) \approx p(x) + J(x)\Delta x$$

$$g(x+\Delta x) \approx g(x) + H(x)\Delta x$$

- Newton step:

$$p(x) + J(x)\Delta x = 0$$

$$g(x) + H(x)\Delta x = J(x)^T \lambda$$

$$\begin{pmatrix} H & -J^T \\ J & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \lambda \end{pmatrix} = \begin{pmatrix} -g(x) \\ -p(x) \end{pmatrix}$$

Useful special case

- Exact for quadratic:
  min $x^THx/2 + c^Tx$ s.t. $Ax + b = 0$
  - $g(x) = Hx+c$, $J(x) = A$
  - so: $\begin{pmatrix} H & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} -c \\ -b \end{pmatrix}$

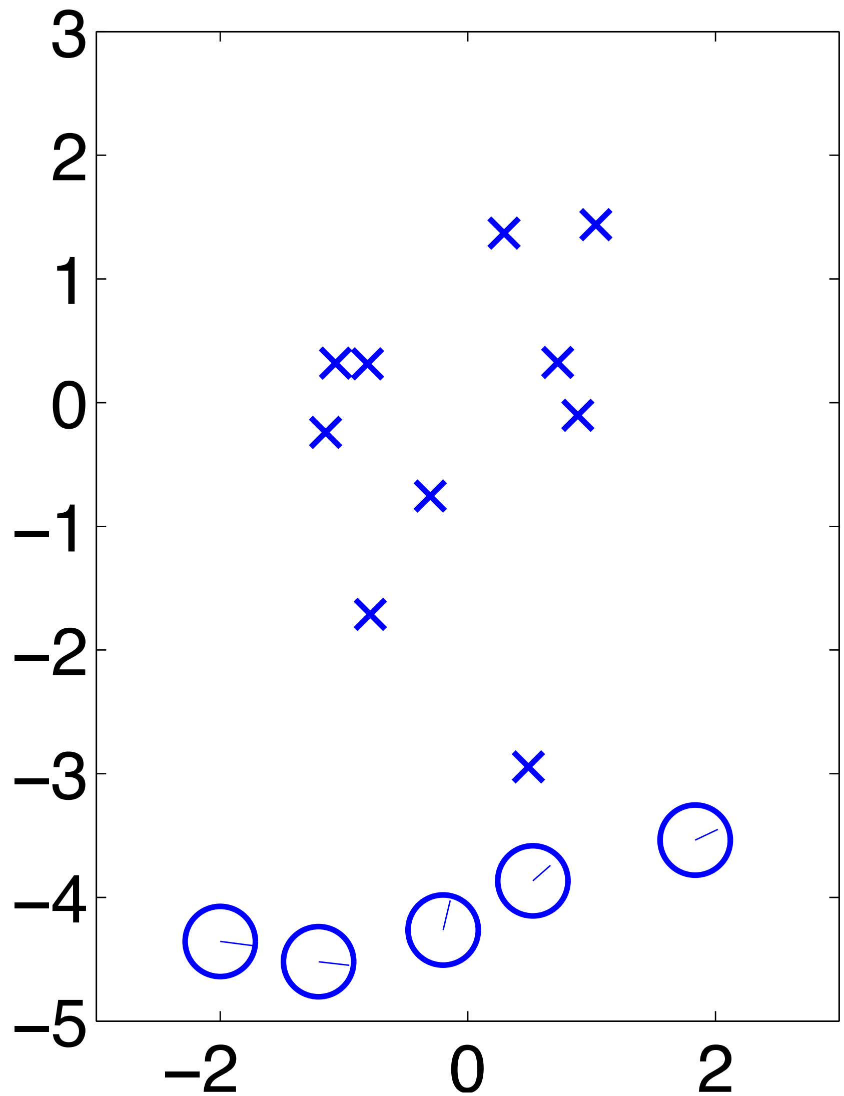
$$p(x) = 0$$
$$g(x) = J(x)^T\lambda$$

$$\begin{pmatrix} H & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \lambda \end{pmatrix} = \begin{pmatrix} -(Hx+c) \\ -(Ax+b) \end{pmatrix}$$

Ex: bundle adjustment for SLAM

- Solve for:
  - Robot positions x_t, θ_t — $x_t \to \mathbb{R}^2$, $\theta_t \to [-\pi, \pi]$
  - Landmark positions y_k → $\mathbb{R}^2$
- Given: odom., radar, vision, …
- Constraints:
  - observations consistent w/ inferred landmark/robot positions

$$d_{tk}^2 = \|x_t - y_k\|^2 + \text{noise} \rightarrow \min \text{noise}^2$$



Lagrange for inequalities

- Even more useful than for equalities
- Review LPs/QPs first

Linear or quadratic programs

- n variables: $x = (x_1, x_2, \ldots, x_n)^T$
  - ranges: $[l_i, u_i]$

    l_i or u_i

    $-\infty$ $+\infty$

- Objective: min or max $\text{sum}_i\ c_i\ x_i$ → $c^T x$
  - optionally: ... + $\text{sum}_{ij}\ q_{ij}\ x_i\ x_j$ /2 → $x^T Q x / 2$

- m constraints
  - for j=1..m, $\text{sum}_i\ a_{ij}\ x_i = b_j$ (or ≥ or ≤)

    $Ax = b$

    ≥ ≤

- Example:

max 2x+3y s.t.

$x + y \leq 4$

$2x + 5y \leq 12$

$x + 2y \leq 5$

$x, y \geq 0$

Sketching an LP

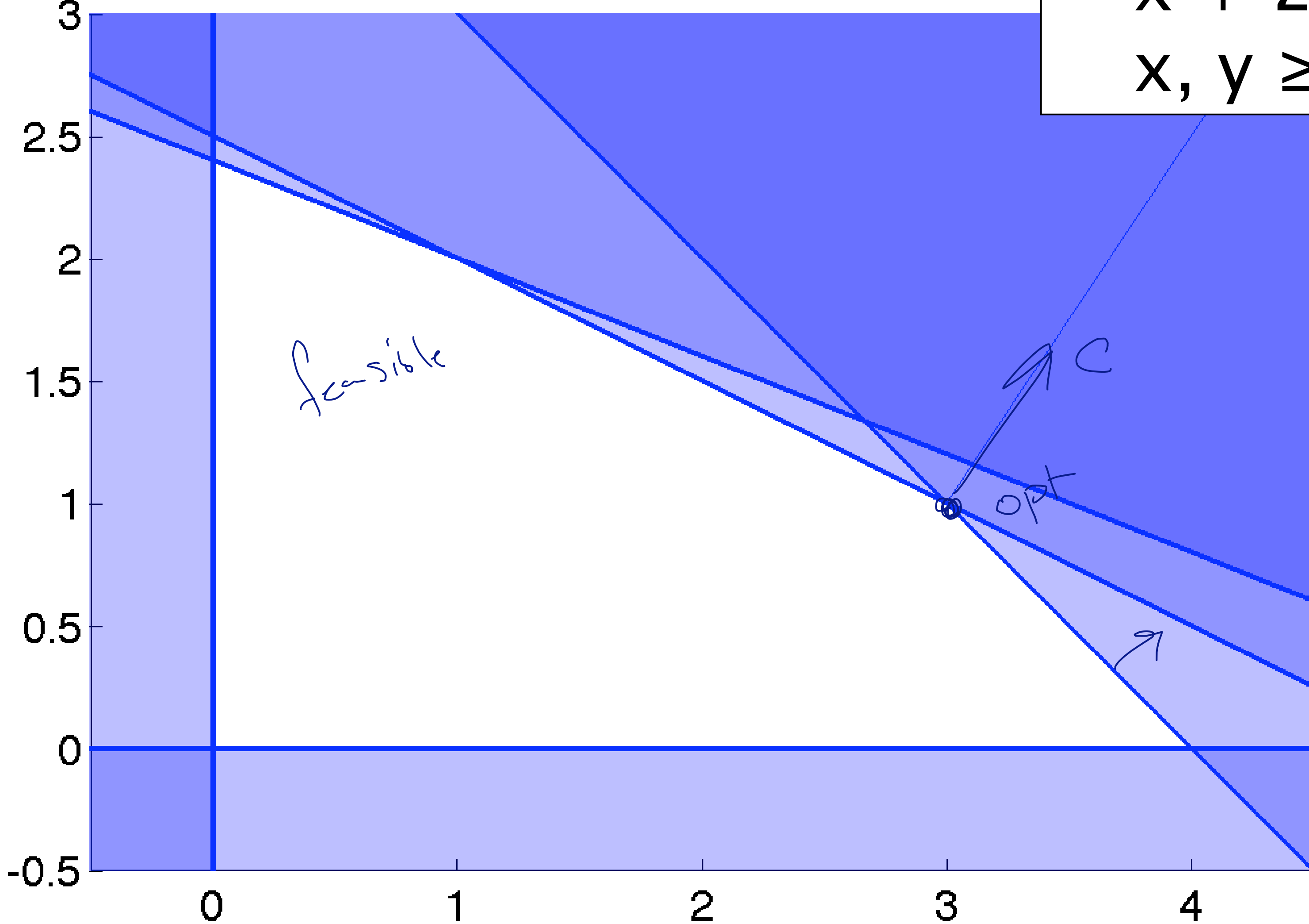
max 2x+3y s.t.

$x + y \leq 4$

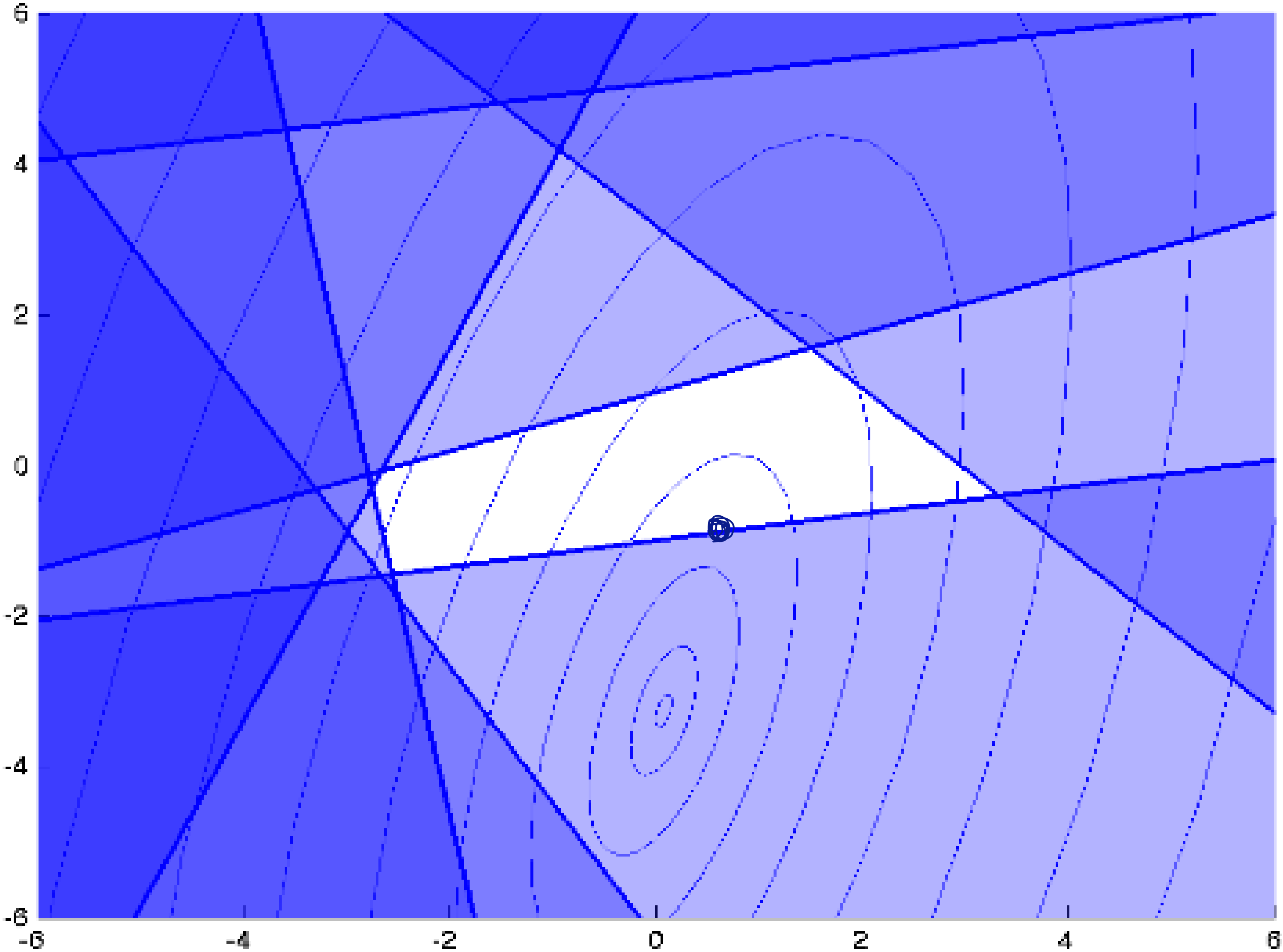
$2x + 5y \leq 12$

$x + 2y \leq 5$

$x, y \geq 0$



Sketching a QP



Matrix notation

- For a vector of variables v and a constant matrix A and vector b,
  - $Av \le b$ [componentwise]
- Objective: $c^T v + v^T Q v/2$
- E.g.:

max $2x+3y$ s.t.
$x + y \le 4$
$2x + 5y \le 12$
$x + 2y \le 5$
$x, y \ge 0$

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 5 \\ 1 & 2 \\ -1 & 0 \\ 0 & -1 \end{pmatrix} \quad b = \begin{pmatrix} 4 \\ 12 \\ 5 \\ 0 \\ 0 \end{pmatrix} \quad c = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

Transforming linear constraints

- Getting rid of inequalities (except variable bounds)

$$x + y \leq 4$$

$$x + y + s \cancel{\leq} = 4 \qquad s \geq 0$$

- Getting rid of equalities

$$x + 2y = 4$$

$$x + 2y \leq 4$$

$$x + 2y \geq 4$$

Transforming linear constraints

- Getting rid of free vars

$$\max x + y \text{ s.t.}$$
$$2x + y \leq 3$$
$$y \geq 0$$

$x_1 \geq 0 \quad x_2 \geq 0$

$\rightarrow \quad 2(x_1 - x_2) + y \leq 3$

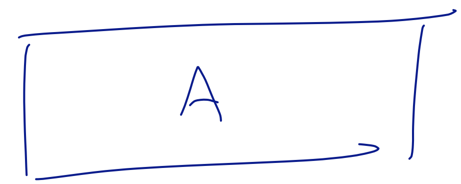
- Getting rid of bounded vars

$$x \in [2, 5]$$

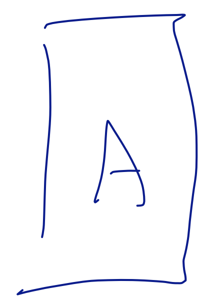
$\rightarrow \quad x \leq 5 \qquad x \geq 2$

Normalizing LPs or QPs

- Standard form:
  - all variables are nonnegative
  - all constraints are equalities
  - max or min $c^T x$ s.t. $Ax = b, x \geq 0$
  - (optionally: A has full row rank)



- Inequality form:
  - all variables are free
  - all constraints are inequalities
  - max or min $c^T x$ s.t. $Ax \geq b$
  - (optionally: A has full column rank)



Now back to Lagrange

- Develop Lagrange multiplier technique for LP/QP
- Leads to an important idea: ***duality***
  - transform one optimization problem (“primal”) to an equivalent but often-very-different-looking one (“dual”)
  - taking dual twice gets back to primal
- Why do we care?
  - dual can be much easier (or harder) to solve than primal
  - if easier: profit!
  - upcoming example: SVMs

Suppose we're lazy

$(2,2) \to 4$

- A "hard" LP (in inequality form):
  - $\min\ x + y \ \text{ s.t. } \ x + y \geq 2 \quad x \geq 0 \quad y \geq 0$

OK, we got lucky

$a, b, c \geq 0$

- What if it were:
  - min $x + 3y$ s.t. $a(x + y \geq 2) + b(x \geq 0) + c(y \geq 0)$

$a = 1 \quad c = 2$

$x + y + 2y \geq 2$

$x + 3y \geq 2$

How general is this?

$a, b, c \geq 0$

- What if it were:
  - $\min\ px + qy \ \text{ s.t. } \ \overset{a}{(x + y \geq 2)} \ \overset{+b}{(x \geq 0)} \ \overset{+c}{(y \geq 0)}$

$a+b = p \qquad a+c = q$

$$(a+b)x + (a+c)y \geq 2a$$

$$px + qy \geq 2a$$

$$\boxed{\begin{array}{l} \max\ 2a \\ a, b, c \geq 0 \\ a+b = p \\ a+c = q \end{array}}$$ dual

Equality constraints

$a, b, c \geq 0 \quad d$ free

- Note = constraint
  - $\min\ x - 2y$ s.t. $a(x + y \geq 2) + b(x \geq 0) + c(y \geq 0) + d(3x + y = 2)$

Quadratic program

primal

- min $c^T x + x^T H x/2$ s.t. $Ax \geq b$

x

$\lambda \geq 0$

$\lambda^T(b - Ax) \leq 0$

$A^T \lambda = c + Hx$

$\lambda^T b - (c + Hx)^T x \leq 0$

$\underbrace{\lambda^T b - \frac{1}{2} x^T H x}_{\text{lower bound}} \leq \underbrace{c^T x + \frac{1}{2} x^T H x}_{\text{obj}}$

$$\boxed{\begin{aligned} &\max_{\lambda, x} \ \lambda^T b - x^T H x/2 \\ &\text{s.t. } A^T \lambda = c + Hx \\ &\qquad \lambda \geq 0 \end{aligned}}$$

dual QP

[could rename $x \to z$ for clarity]