The central path

- nonlinear optimization methods for linear optimization
- logarithmic barrier
- central path

Ellipsoid method

ellipsoid algorithm

- a general method for (nonlinear) convex optimization, invented ca. 1972
- Khachiyan (1979): complexity is polynomial when applied to LP

importance

- answered an open question: worst-case complexity of LP is polynomial
- practical performance was disappointing; much slower than simplex
- useful as a very simple algorithm for nonlinear convex optimization
- idea is very different from simplex; motivated research in new directions

Interior-point methods

1950s–1960s: several related methods for nonlinear convex optimization

- sequential unconstrained minimization (Fiacco & McCormick), logarithmic barrier method (Frisch), affine scaling method (Dikin), method of centers (Huard & Lieu)
- no worst-case complexity theory, but often work well in practice

1980s-1990s: interior-point methods for linear optimization

- Karmarkar (1984): new polynomial-time method ('projective algorithm')
- later recognized as related to the earlier methods
- many variations and improvements since 1984
- competitive with simplex; often faster for very large problems

Outline

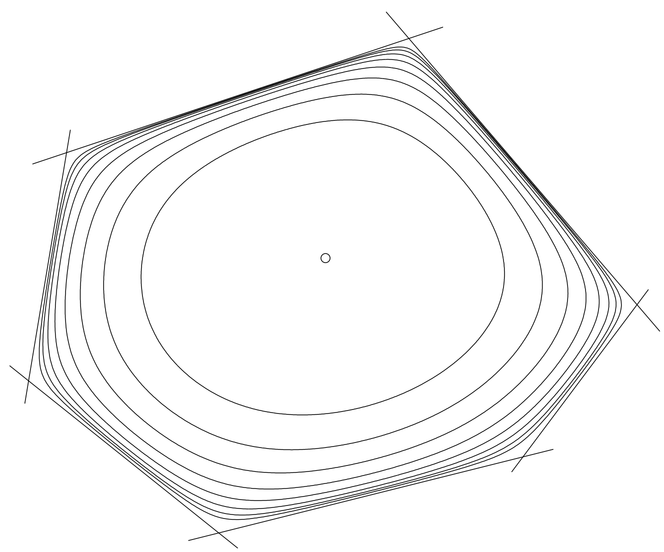
- LP algorithms based on nonlinear optimization
- **logarithmic barrier**
- central path

Logarithmic barrier

- we consider inequalities $Ax \leq b$ with A of size $m \times n$ and with rows a_i^T
- define $P = \{x \mid Ax \leq b\}$ and $P^\circ = \{x \mid Ax < b\}$

logarithmic barrier for the inequalities $Ax \leq b$:

$$\phi(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x) \quad \text{with domain } P^\circ$$



Gradient and Hessian

gradient: $\nabla\phi(x)$ is the n-vector with $\nabla\phi(x)_i = \partial\phi(x)/\partial x_i$

$$\nabla\phi(x) = \sum_{k=1}^{m} \frac{1}{b_k - a_k^T x} a_k = A^T d_x$$

d_x denotes the positive m-vector

$$d_x = \left(\frac{1}{b_1 - a_1^T x}, \; \ldots, \; \frac{1}{b_m - a_m^T x} \right)$$

Hessian: $\nabla^2\phi(x)$ is the $n \times n$-matrix with $\nabla\phi(x)_{ij} = \partial^2\phi(x)/\partial x_i \partial x_j$

$$\nabla^2\phi(x) = \sum_{k=1}^{m} \frac{1}{(b_k - a_k^T x)^2} a_k a_k^T = A^T \, \mathbf{diag}(d_x)^2 A$$

Convexity

second-order condition for convexity of ϕ

- $\nabla^2\phi(x)$ is positive semidefinite for all $x \in P^\circ$:

$$u^T\nabla^2\phi(x)u = u^TA^T\,\mathbf{diag}(d_x)^2Au = \|\,\mathbf{diag}(d_x)Au\|^2 \geq 0 \qquad \forall u$$

- if $\mathrm{rank}(A) = n$, then $\nabla^2\phi(x)$ is positive definite for all $x \in P^\circ$:

$$u^T\nabla^2\phi(x)u = \|\,\mathbf{diag}(d_x)Au\|^2 > 0 \qquad \forall u \neq 0$$

local (semi-)norm: we will use the notation

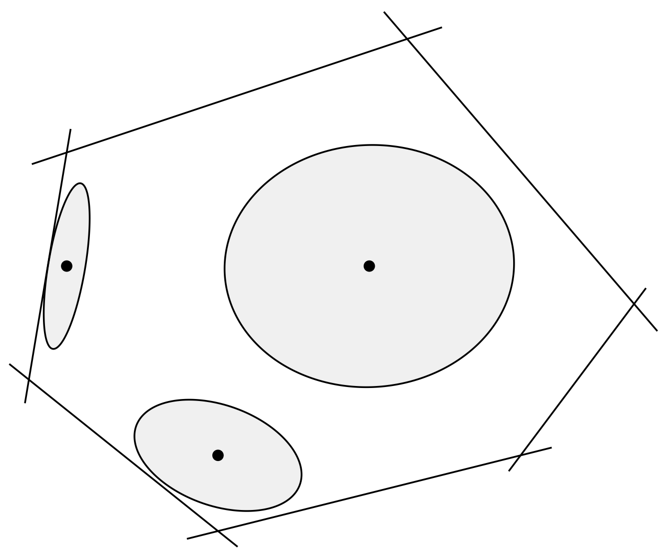
$$\|u\|_x = (u^T\nabla^2\phi(x)u)^{1/2} = \|\,\mathbf{diag}(d_x)Au\|$$

Dikin ellipsoid

definition: the Dikin ellipsoid at $x \in P^\circ$ is the set

$$\begin{aligned}\mathcal{E}_x &= \{y \mid (y-x)^T \nabla^2 \phi(x)(y-x) \leq 1\} \\ &= \{y \mid \|y-x\|_x \leq 1\}\end{aligned}$$

property: Dikin ellipsoid at any $x \in P^\circ$ is contained in P



proof: consider $x \in P^\circ$

- points y in the Dikin ellipsoid at x satisfy

$$\begin{aligned}(y-x)^T \nabla^2 \phi(x)(y-x) &= (y-x)^T A^T \mathbf{diag}(d_x)^2 A(y-x) \\ &= \sum_{i=1}^m \frac{(a_i^T(y-x))^2}{(b_i - a_i^T x)^2} \\ &\leq 1\end{aligned}$$

- therefore each term in the sum is less than or equal to one:

$$-(b - Ax) \leq A(y-x) \leq b - Ax$$

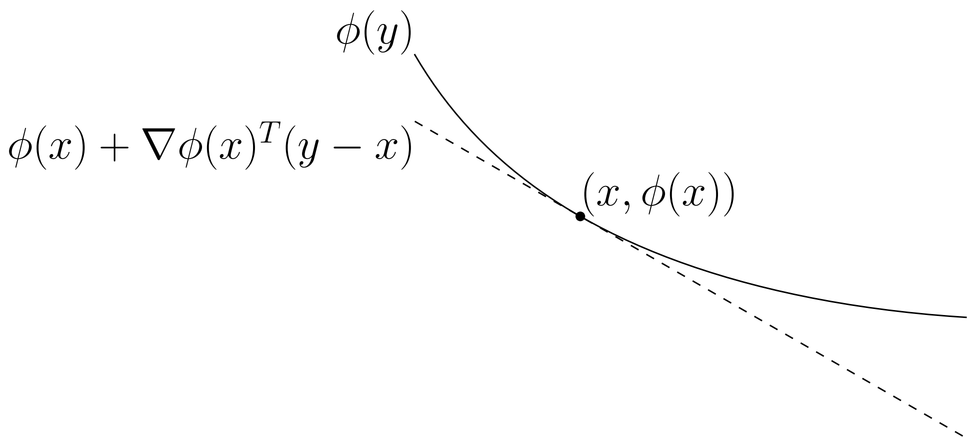
the right-hand side inequality shows that $Ay \leq b$

Convexity: first-order condition

linearization of ϕ at $x \in P^\circ$ gives **lower bound** on ϕ:

$$\phi(y) \geq \phi(x) + \nabla\phi(x)^T(y - x) \qquad \text{for all } x, y \in P^\circ$$

strict inequality holds if $x \neq y$ and $\operatorname{rank}(A) = n$



- x minimizes $\phi(x)$ if and only if $\nabla\phi(x) = 0$
- if $\operatorname{rank}(A) = n$, minimizer of $\phi(x)$ is unique if it exists

proof of lower bound:

$$\begin{aligned}\phi(y) &= -\sum_{i=1}^{m} \log(b_i - a_i^T y) \\ &\geq -\sum_{i=1}^{m} \log(b_i - a_i^T x) + \sum_{i=1}^{m} \frac{a_i^T (y-x)}{b_i - a_i^T x} \\ &= \phi(x) + \nabla \phi(x)^T (y-x)\end{aligned}$$

- inequality follows from $\log u_i \leq u_i - 1$ with $u_i = (b_i - a_i^T y)/(b_i - a_i^T x)$
- equality holds only if $u_i = 1$ for $i = 1, \ldots, m$, *i.e.*, $A(y-x) = 0$

Analytic center

definition: the analytic center of a system of inequalities $Ax \leq b$ is

$$\begin{aligned} x_{\mathrm{ac}} &= \operatorname*{argmin}_x \phi(x) \\ &= \operatorname*{argmin}_x - \sum_{i=1}^{m} \log(b_i - a_i^T x) \end{aligned}$$

- x_{ac} is solution of nonlinear equation

$$\nabla \phi(x) = \sum_{i=1}^{m} \frac{1}{b_i - a_i^T x} a_i = 0$$

- different descriptions $Ax \leq b$ of same polyhedron can have different x_{ac}
- x_{ac} exists and is unique if and only if P° is nonempty and bounded

Outline

- LP algorithms based on nonlinear optimization
- logarithmic barrier
- **central path**

Central path

primal-dual pair of LPs

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b \end{array} \qquad\qquad \begin{array}{ll} \text{maximize} & -b^T z \\ \text{subject to} & A^T z + c = 0 \\ & z \geq 0 \end{array}$$

we assume primal and dual problems are strictly feasible and $\mathrm{rank}(A) = n$

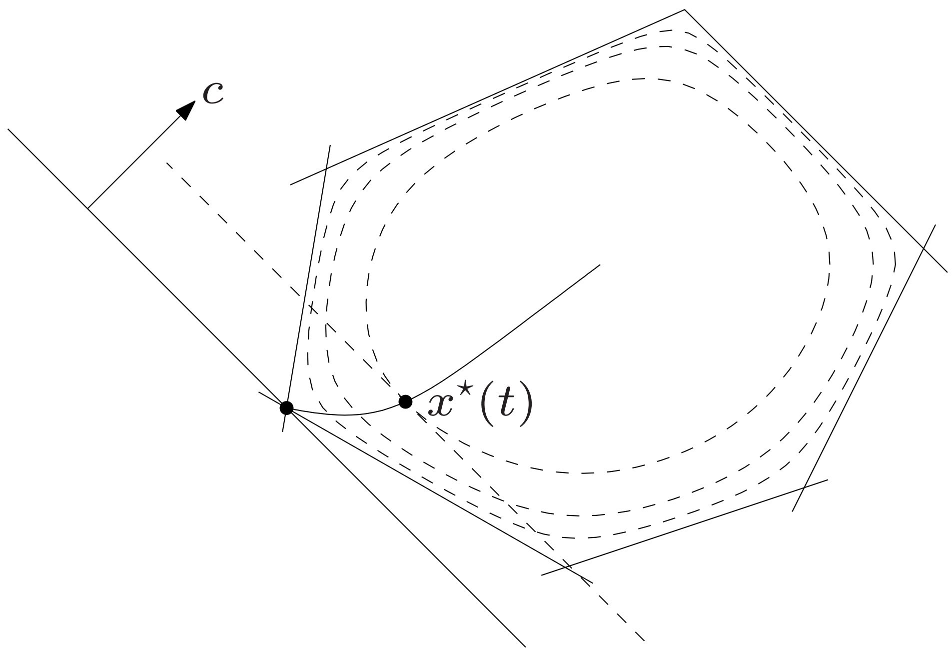
central path: set of points $\{x^\star(t) \mid t > 0\}$ with

$$\begin{aligned} x^\star(t) &= \operatorname*{argmin}_x \left(tc^T x + \phi(x)\right) \\ &= \operatorname*{argmin}_x \left(tc^T x - \sum_{i=1}^m \log(b_i - a_i^T x)\right) \end{aligned}$$

$x^\star(t)$ exists and is unique for all $t > 0$ (constructive proof in next lecture)

Optimality condition

$x^\star(t)$ is solution of $tc + \nabla\phi(x) = 0$



hyperplane $c^T x = c^T x^\star(t)$ is tangent to level curve of ϕ through $x^\star(t)$

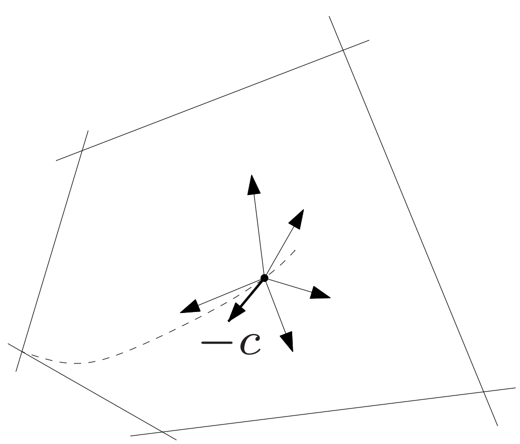
Force field interpretation

- optimality condition can be interpreted as force equilibrium

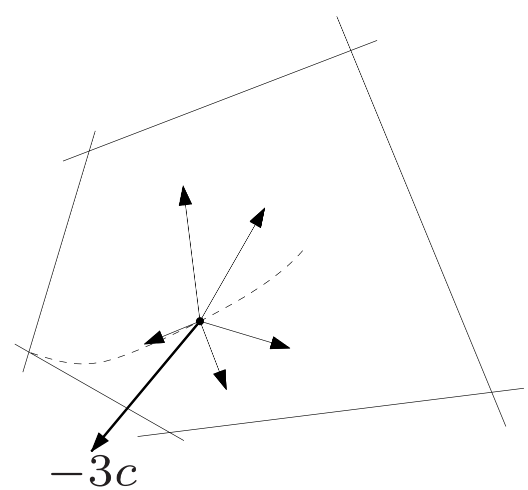
$$-tc + \sum_{i=1}^{m} F_i(x) = 0 \quad \text{with } F_i(x) = \frac{-1}{b_i - a_i^T x} a_i$$

- force $F_i(x)$ decays as inverse distance to $\mathcal{H}_i = \{x \mid a_i^T x = b_i\}$:

$$\|F_i(x)\| = \frac{1}{\mathrm{dist}(x, \mathcal{H}_i)}$$



$t = 1$



$t = 3$

Central path and duality

point $x^\star(t)$ on central path is strictly primal feasible and satisfies

$$c + \sum_{i=1}^{m} z_i^\star(t) a_i = 0 \qquad \text{with} \quad z_i^\star(t) = \frac{1}{t(b_i - a_i^T x^\star(t))}$$

- $z^\star(t)$ is strictly dual feasible: $A^T z^\star(t) + c = 0$ and $z^\star(t) > 0$
- duality gap between $x = x^\star(t)$ and $z = z^\star(t)$ is

$$c^T x + b^T z = (b - Ax)^T z = \frac{m}{t}$$

- gives bound on sub-optimality of $x^\star(t)$

$$c^T x^\star(t) - p^\star \le \frac{m}{t}$$

($p^\star$ is optimal value of LP)

Central path and complementarity

optimality conditions

x, z are primal, dual optimal if and only if

$$s = b - Ax \geq 0, \qquad z \geq 0, \qquad s_i z_i = 0, \quad i = 1, \ldots, m$$

central path equations

$x = x^\star(t)$ and $z = z^\star(t)$ if and only if

$$s = b - Ax > 0, \qquad z > 0, \qquad s_i z_i = \frac{1}{t}, \quad i = 1, \ldots, m$$

Interior-point methods

common characteristics

- follow the central path to find optimal solution
- use Newton's method to follow central path

differences

- algorithms can update primal, dual, or pairs of primal, dual variables
- can keep iterates feasible or allow infeasible iterates (and starting points)
- different techniques for following central path