7. Constraints

- constrained optimization problems
- examples of constraints
- least norm problems
- probabilities and distributions
- functional constraints
- relaxations and sensitivity analysis
- the Lagrangian, Lagrange dual function, and dual problem
- weak and strong duality
- complementary slackness and KKT optimality conditions
- infeasible problems and exact penalty methods

Constrained optimization problems

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

- $f_1, \ldots, f_m : \mathbf{R}^n \to \mathbf{R}$ are inequality constraint functions
- $h_1, \ldots, h_p : \mathbf{R}^n \to \mathbf{R}$ are equality constraint functions
- physical properties, safety requirements, resource limitations, etc.

sometimes written in **abstract form**

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \in \mathcal{X} \end{array}$$

with **constraint set** $\mathcal{X} = \left\{ x \in \mathbf{R}^n \;\middle|\; \begin{array}{l} f_i(x) \leq 0, \; i = 1, \ldots, m \\ h_i(x) = 0, \; i = 1, \ldots, p \end{array} \right\}$

Regularization and constraints

- both represent prior knowledge about the solution
- regularization: soft constraints, penalize violation
- constraints: hard constraints, violation unacceptable

example: solution x should be small in $\|\cdot\|_2$

- Tikhonov regularization:

$$\text{minimize} \quad \|Ax-b\|_2^2 + \lambda\|x\|_2^2$$

- Euclidean ball constraint:

$$\begin{array}{ll} \text{minimize} & \|Ax-b\|_2^2 \\ \text{subject to} & \|x\|_2^2 \leq t \end{array}$$

Examples of constraints

- **nonnegativity**:
  - $x \succeq 0$ for nonnegative least squares
  - componentwise $X_{ij} \geq 0$ for nonnegative matrix factorization
- **box constraints**:

$$\begin{array}{ll} \text{minimize} & \|Ax - b\| \\ \text{subject to} & l \preceq x \preceq u \end{array}$$

  projects b onto the image of the box $\{x \mid l \preceq x \preceq u\}$ under A

  - equivalent to $\|x\|_\infty \leq t$ for equally bounded components
- **probability simplex**: $x \succeq 0$, $\mathbf{1}^T x = 1$ for multinomial distributions
- **linear constraints**: $Cx \preceq d$, $Gx = h$; include, *e.g.*, box constraints by

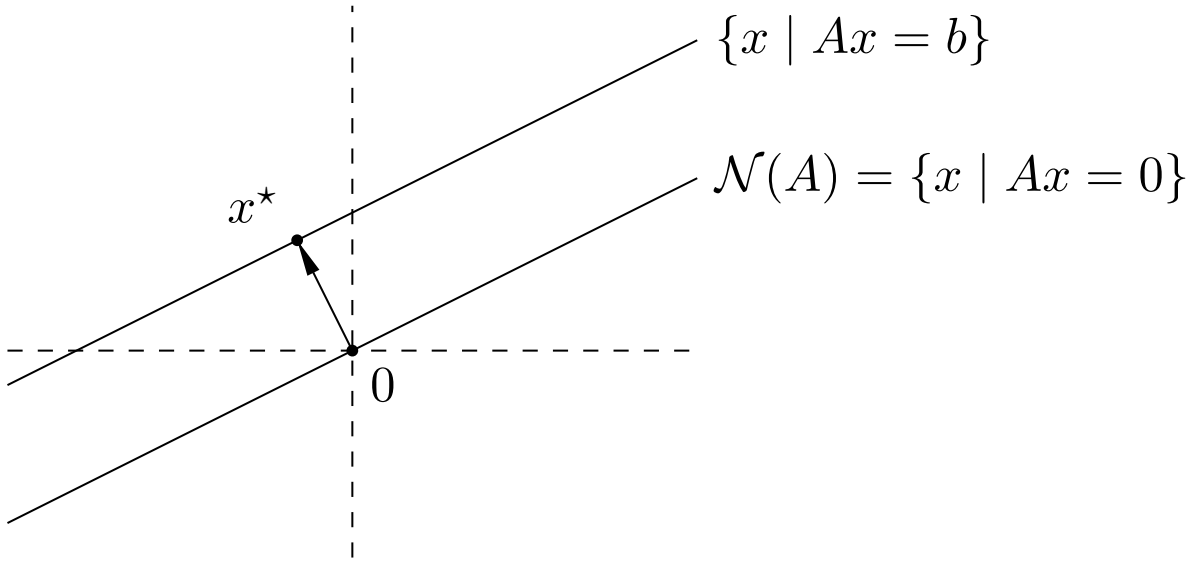
$$C = \begin{bmatrix} I \\ -I \end{bmatrix}, \quad d = \begin{bmatrix} u \\ -l \end{bmatrix}, \quad G = 0, \quad h = 0$$

- **trust region constraints**: $\|x - x_0\| \leq t$ for some $x_0 \in \mathbf{R}^n$

Least norm problems

$$\begin{array}{ll} \text{minimize} & \|x\| \\ \text{subject to} & Ax = b \end{array}$$

- $Ax = b$ underdetermined, *e.g.*, $A \in \mathbf{R}^{m \times n}$ is fat with $m < n$
  - affine solution set $\{x_0 + z \mid z \in \mathcal{N}(A)\}$ (x_0 a particular solution)
- **geometric interpretation**: find point in the solution set closest to zero



Probabilities and distributions

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1 \\ & x \in \mathcal{X} \end{array}$$

- find nonparametric distribution x minimizing some objective $f_0(x)$
- many useful types of constraint set $\mathcal{X}$

moment constraints:

$$\mathbf{E}\, Z^k = \sum_{i=1}^n c_i^k x_i$$

is kth moment of random variable $Z \in \{c_1, \ldots, c_n\}$

- linear equality constraints: $\mathbf{E}\, Z = c^T x = \mu$; $\mathbf{E}\, Z^k = \sum_{i=1}^n c_i^k x_i = \eta$
- linear inequality constraints: $\alpha \leq \sum_{i=1}^n c_i^k x_i \leq \beta$
- generalization: expectation $\mathbf{E}\, g(Z) = \sum_{i=1}^n g(c_i) x_i$ of some function g

probability constraints:

$$\mathbf{prob}(Z \in S) = p^T x, \qquad p_i = \begin{cases} 1, & c_i \in S \\ 0, & \text{otherwise,} \end{cases} \qquad i = 1, \ldots, n$$

- probability of $Z \in \{c_1, \ldots, c_n\}$ in set S is linear function of x
- known or bounded probability: $p^T x = \mu$ or $\alpha \leq p^T x \leq \beta$
- bounded conditional probability:

$$\mathbf{prob}(Z \in A \mid Z \in B) = \frac{\mathbf{prob}(Z \in A \cap B)}{\mathbf{prob}(Z \in B)} = \frac{p^T x}{q^T x}$$

with $p_i = 1$ if $c_i \in A \cap B$, $q_i = 1$ if $c_i \in B$, and 0 otherwise, so

$$\alpha \leq \frac{p^T x}{q^T x} \leq \beta \quad \Longleftrightarrow \quad (p - \alpha q)^T x \geq 0 \quad \text{and} \quad (\beta q - p)^T x \geq 0$$

variance constraints: for random variable $Z \in \{c_1, \ldots, c_n\}$

$$\mathbf{var}\, Z = \mathbf{E}\, Z^2 - (\mathbf{E}\, Z)^2 = \sum_{i=1}^n c_i^2 x_i - \left(\sum_{i=1}^n c_i x_i\right)^2$$

$$= \sum_{i=1}^n c_i^2 x_i - x^T (cc^T) x$$

- difference of linear and convex quadratic functions of x
- lower bound $\mathbf{var}\, Z \geq \mu$ is convex constraint

$$x^T (cc^T) x - \sum_{i=1}^n c_i^2 x_i \leq -\mu$$

- upper bound $\mathbf{var}\, Z \leq \eta$ is DC constraint

$$\sum_{i=1}^n c_i^2 x_i - x^T (cc^T) x \leq \eta$$

entropy constraints:

- distribution x has at least μ level of uncertainty:

$$\sum_{i=1}^{n} x_i \log x_i \leq -\mu$$

- KL-divergence between x and given reference q is at most η:

$$\sum_{i=1}^{n} x_i \log \frac{x_i}{q_i} \leq \eta$$

both are convex constraints in x

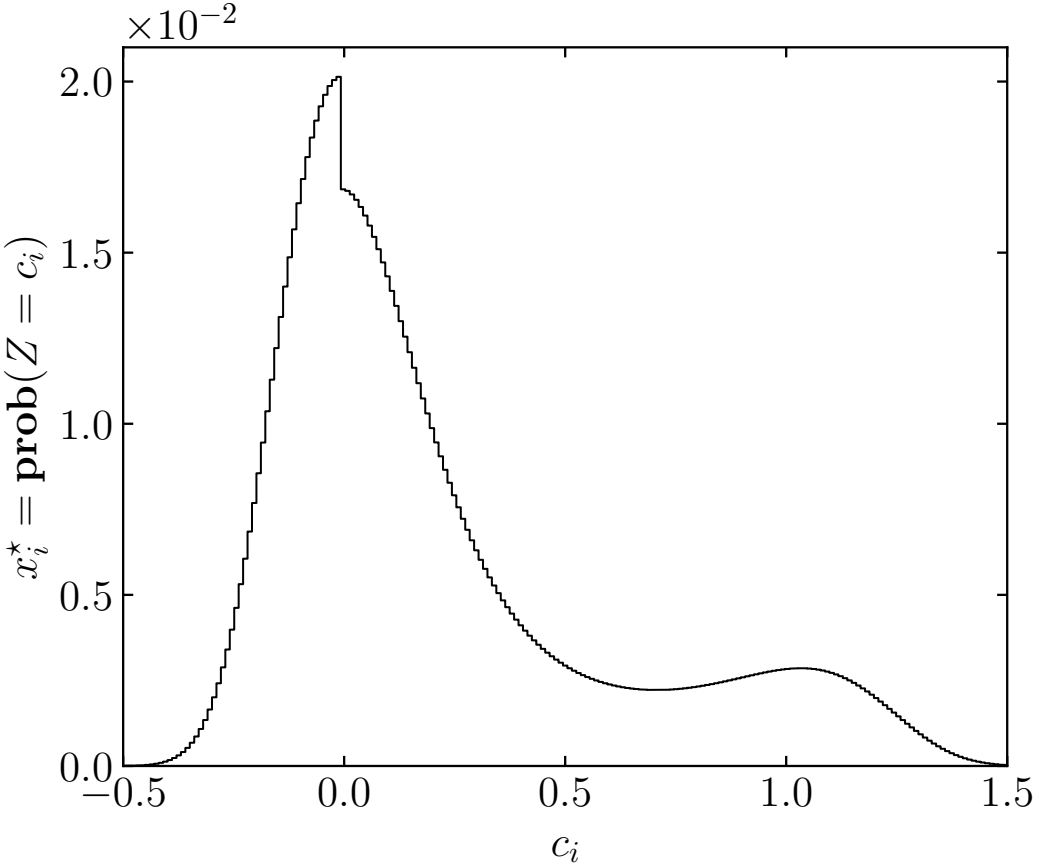
example:

- maximum entropy estimation:

  minimize $\sum_{i=1}^n x_i \log x_i$

  subject to $x \succeq 0,\ \mathbf{1}^T x = 1,\ \mathbf{prob}(Z < 0) \geq 0.35,\ \mathbf{var}\, Z \geq 0.1$

  $-0.2 \leq \mathbf{E}\, Z \leq 0.5,\ \mathbf{E}\, Z^2 \leq 0.2,\ \mathbf{E}(3Z^4 - 7Z^3) \leq -0.7$

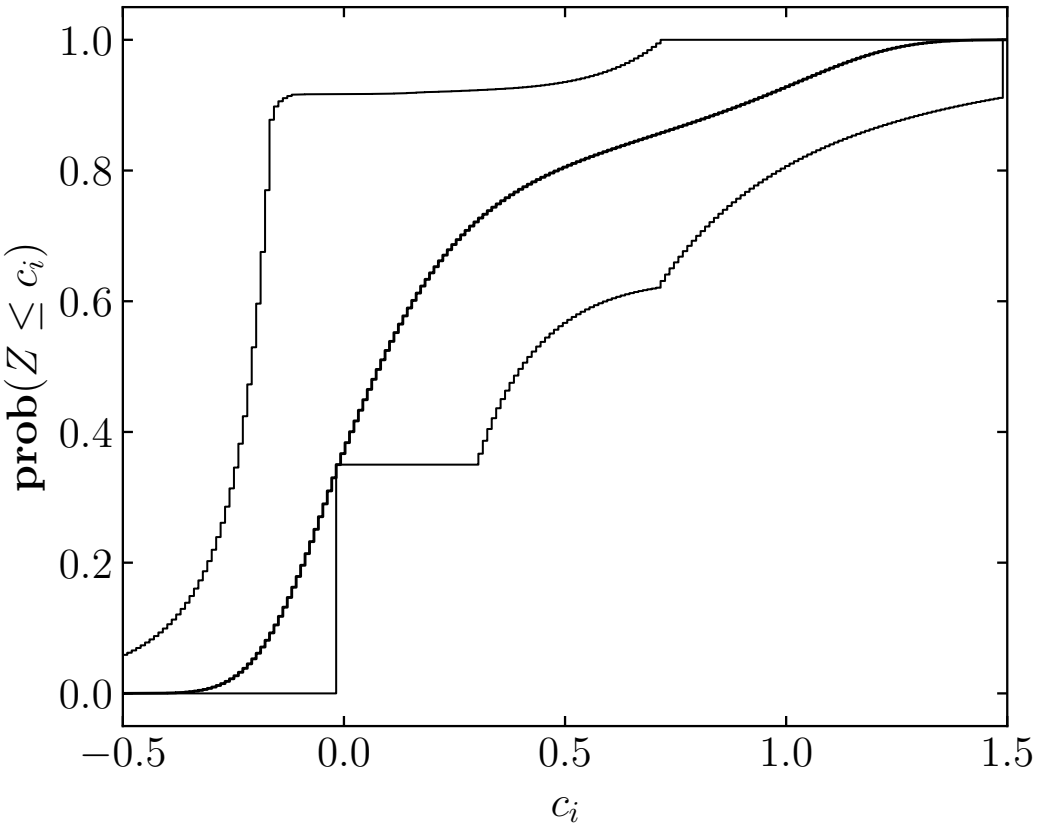


- bounding cumulative density function:

  min./max. $\mathbf{prob}(Z \leq c_i)$

  subject to $x \succeq 0,\ \mathbf{1}^T x = 1,\ \mathbf{prob}(Z < 0) \geq 0.35,\ \mathbf{var}\, Z \geq 0.1$

  $-0.2 \leq \mathbf{E}\, Z \leq 0.5,\ \mathbf{E}\, Z^2 \leq 0.2,\ \mathbf{E}(3Z^4 - 7Z^3) \leq -0.7$



(maximum entropy distribution shown thicker)

Functional constraints

function fitting problems: given $(z_i, y_i) \in \mathbf{R}^k \times \mathbf{R}$, $i = 1, \ldots, m$

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m \phi(f(z_i) - y_i) \\ \text{subject to} & x \in \mathcal{X} \end{array}$$

fits $f : \mathbf{R}^k \to \mathbf{R}$ with $f(z) = x_1 f_1(z) + \cdots + x_n f_n(z)$ to the data points

- $x \in \mathbf{R}^n$ are unknown coefficients of *basis functions* $f_1, \ldots, f_n : \mathbf{R}^k \to \mathbf{R}$
- $\phi : \mathbf{R} \to \mathbf{R}$ is penalty for misfit
- approximation interpretation: $a_i = (f_1(z_i), \ldots, f_n(z_i))$ so $f(z_i) = a_i^T x$
- example: polynomial fitting for $(z_i, y_i) \in \mathbf{R}^2$ with
  - $f(z) = x_1 + x_2 z + \cdots + x_n z^{n-1}$
  - $\phi(u) = u^2$

interpolation constraints: for (s_i, t_i), $i = 1, \ldots, p$, require

$$f(s_i) = t_i, \ i = 1, \ldots, p \quad \Longleftrightarrow \quad Ax = t$$

- $A \in \mathbf{R}^{p \times n}$ with rows $a_i^T = (f_1(s_i), \ldots, f_n(s_i))^T$
- feasible if and only if $t \in \mathcal{R}(A)$

bounding constraints: for (s_i, t_i), $i = 1, \ldots, p$, require

$$f(s_i) \in [\alpha_i, \beta_i], \ i = 1, \ldots, p \quad \Longleftrightarrow \quad \alpha \preceq Ax \preceq \beta$$

- nonnegativity: $f(s_i) = a_i^T x \geq 0$
- monotonicity: $f(s_{i+1}) - f(s_i) = (a_{i+1} - a_i)^T x \geq 0$
- lower bounding data points: $f(s_i) = a_i^T x \leq t_i$

(generalize to full $\mathbf{dom}\, f$ under some assumptions, *e.g.*, f piecewise affine)

Lipschitz continuity constraints: at s_i, $i = 1, \ldots, p$, require

$$|f(s_i) - f(s_j)| \leq L\|s_i - s_j\|, \quad i, j = 1, \ldots, p$$

$$\iff \quad -L\|s_i - s_j\| \leq (a_i - a_j)^T x \leq L\|s_i - s_j\|, \quad i, j = 1, \ldots, p$$

derivative constraints:

- bounding $\nabla f(s) = x_1 \nabla f_1(s) + \cdots + x_n \nabla f_n(s)$:

$$\|\nabla f(s)\| \leq M \quad \iff \quad \|Ax\| \leq M$$

  are convex, where $A = \begin{bmatrix} \nabla f_1(s) & \cdots & \nabla f_n(s) \end{bmatrix}$

- bounding $\nabla^2 f(s) = x_1 \nabla^2 f_1(s) + \cdots + x_n \nabla^2 f_n(s)$:

$$\alpha I \preceq \nabla^2 f(s) \preceq \beta I$$

  are LMIs in x

example: quadratic basis functions

$$f_i(z) = \frac{1}{2} z^T P_i z + q_i^T z + r_i, \quad i = 1, \ldots, n$$

so $f(z) = \frac{1}{2} z^T P z + q^T z + r$ with

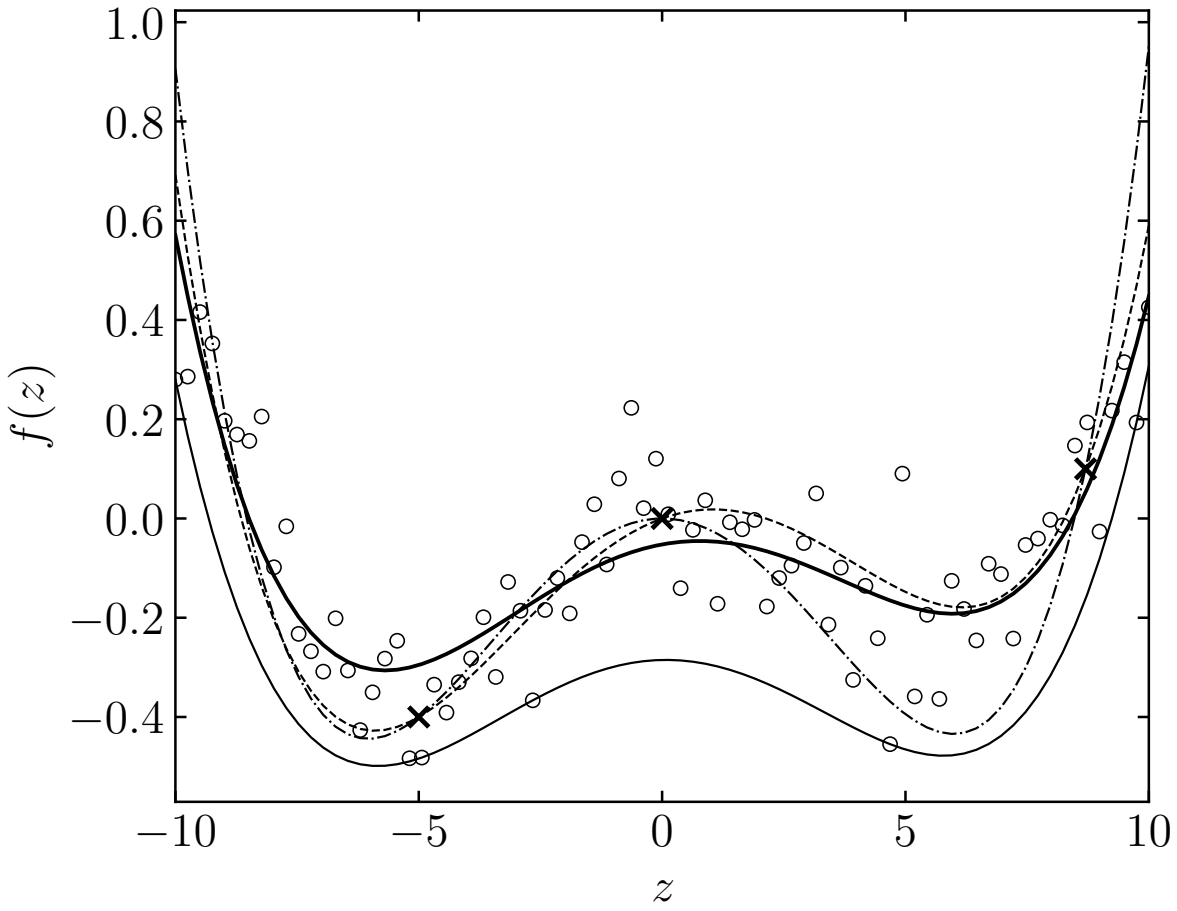
$$P = \sum_{i=1}^{n} x_i P_i, \quad q = \sum_{i=1}^{n} x_i q_i, \quad r = \sum_{i=1}^{n} x_i r_i$$

convexity constraints on f

$$\nabla^2 f(z) = P = \sum_{i=1}^{n} x_i P_i \succeq 0$$

are LMIs in x

numerical example:



- thick: unconstrained
- solid: lower bounding
- dashed: interpolation; dashdotted: interpolation + derivative

Relaxations

for two optimization problems

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \in \mathcal{P} \end{array} \quad \text{and} \quad \begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \in \mathcal{Q} \end{array}$$

call the latter a **relaxation** of the former if $\mathcal{P} \subseteq \mathcal{Q}$

lower bounding property: let $p^\star, q^\star$ be optimal values of each problem,

$$q^\star = \inf_{x \in \mathcal{Q}} f_0(x) \leq \inf_{x \in \mathcal{P}} f_0(x) = p^\star$$

- especially useful if $\mathcal{Q}$ is convex and $\mathcal{P}$ is not
- relaxation infeasible $\implies$ original problem infeasible
- if relaxation optimal $x^\star \in \mathcal{P}$, then $x^\star$ is also original problem optimal

objective relaxation: if $g_0(x) \leq f_0(x)$ for all $x \in \mathcal{P}$, then

$$\text{minimize} \quad g_0(x) \qquad \text{relax} \qquad \text{minimize} \quad f_0(x)$$

- often seen as $g_0 = \mathbf{conv}\, f_0$
- first problem sometimes called *surrogate problem* of the second
- relaxation follows from epigraph formulation
- lower bounding property of optimal values holds
- optimal points relationship depends on shape of g_0

Sensitivity analysis

for (possibly nonconvex) optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \end{array}$$

relax the inequality constraints by $u \in \mathbf{R}^m$:

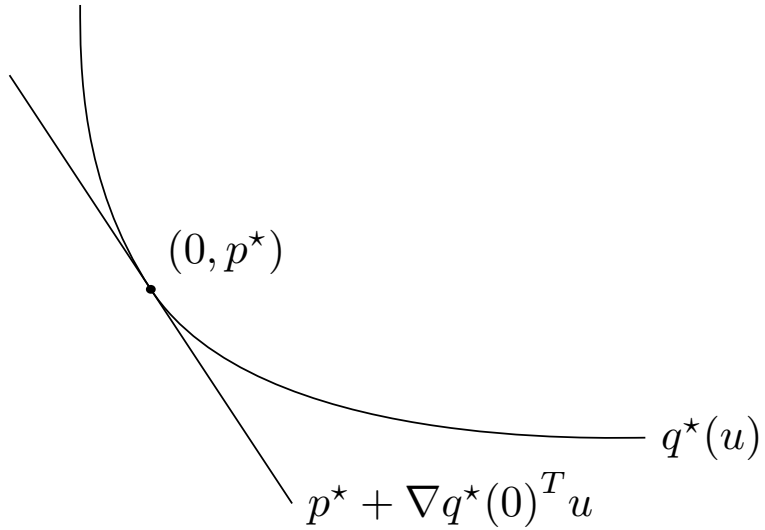
$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq u_i, \quad i = 1, \ldots, m \end{array}$$

- second problem sometimes called *perturbation* of the first
- $u = 0$ is original problem
- $u_i > 0$: the ith constraint is relaxed
- $u_i < 0$: the ith constraint is tightened

let $p^\star$ be optimal value of original problem, define $q^\star\colon \mathbf{R}^m \to \mathbf{R}$

$$q^\star(u) = \inf \left\{ f_0(x) \;\middle|\; \begin{array}{l} x \in \bigcap_{i=0}^m \mathbf{dom}\, f_i \\ f_i(x) \le u_i, \quad i = 1, \ldots, m \end{array} \right\}$$

- gives optimal value of perturbed problem, as function of perturbation u
- for some u, $q^\star(u) = \infty \implies$ perturbation by u is infeasible
- $q^\star$ is convex if original problem is convex
- $q^\star(u) - q^\star(0)$ gives sensitivity to constraint perturbation



Cardinality and rank constraints

- cardinality constrained problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & \mathbf{card}\, x \leq k, \quad \|x\|_\infty \leq 1 \end{array}$$

  has convex relaxation

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & \|x\|_1 \leq k, \quad \|x\|_\infty \leq 1 \end{array}$$

- similarly for rank constrained problem

$$\begin{array}{ll} \text{minimize} & f_0(X) \\ \text{subject to} & \mathbf{rank}\, X \leq k, \quad \|X\|_2 \leq 1 \end{array}$$

  we have convex relaxation

$$\begin{array}{ll} \text{minimize} & f_0(X) \\ \text{subject to} & \|X\|_* \leq k, \quad \|X\|_2 \leq 1 \end{array}$$

Boolean LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & x_i \in \{0,1\}, \quad i = 1, \ldots, n \end{array}$$

is often relaxed to LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & 0 \preceq x \preceq 1 \end{array}$$

- boolean LP is NP-hard, but the relaxation can be solved very efficiently
- if relaxation solution $x^\star$ happens to be boolean, then it is also optimal for the original problem
- similar ideas apply to other problems with integer constraints

Two-way partitioning

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i \in \{-1, 1\}, \quad i = 1, \ldots, n \end{array}$$

- feasible x consists in partitioning n objects into two groups -1 and 1:

$$\{1, \ldots, n\} = \{i \mid x_i = 1\} \cup \{i \mid x_i = -1\}$$

- $W \in \mathbf{S}^n$: weight matrix encoding the pairwise relationships
  - $W_{ij}/-W_{ij}$: cost of putting i and j in the same/different group
- $x_i \in \{-1, 1\} \implies x_i^2 = 1$, so we have equivalent formulation

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(WX) \\ \text{subject to} & X \succeq 0, \quad \mathbf{rank}\, X = 1 \\ & X_{ii} = 1, \quad i = 1, \ldots, n \end{array}$$

where $X = xx^T \in \mathbf{S}_+^n$ is a rank-one matrix

- relaxation: drop rank constraint, get SDP

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(WX) \\ \text{subject to} & X \succeq 0 \\ & X_{ii} = 1, \quad i = 1, \ldots, n \end{array}$$

The Lagrangian

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

- variable $x \in \mathbf{R}^n$; no assumptions on $f_i, h_i \colon \mathbf{R}^n \to \mathbf{R}$

Lagrangian $L \colon \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ of the problem is

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- $\mathbf{dom}\, L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$, where $\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\, h_i$
- λ_i, ν_i are **Lagrange multipliers** for the ith inequality and equality constraints, respectively
- $\lambda \in \mathbf{R}^m$, $\nu \in \mathbf{R}^p$ are **dual variables**

Lagrange dual function

Lagrange dual function (or just *dual function*) $g\colon \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ is

$$g(\lambda,\nu) = \inf_{x\in\mathcal{D}} L(x,\lambda,\nu) = \inf_{x\in\mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

- partial minimization of Lagrangian L over x for fixed λ, ν
- $g(\lambda,\nu) = -\infty$ if $L(x,\lambda,\nu)$ is unbounded below on $\mathcal{D}$

g is **concave** (in λ,ν) even if original problem is not convex:

- for each fixed x, $L(x,\lambda,\nu)$ is affine in (λ,ν)
- pointwise infimum of affine functions is concave

g provides **lower bound** on optimal value $p^\star$ of original problem, *i.e.*,

$$g(\lambda, \nu) \leq p^\star \text{ for all } \lambda \succeq 0,\ \nu$$

- holds trivially if $g(\lambda, \nu) = -\infty$ for some $\lambda \succeq 0,\ \nu$
- informative only when $\lambda \succeq 0$, $(\lambda, \nu) \in \mathbf{dom}\, g$ (*i.e.*, $g(\lambda, \nu) > -\infty$)
  such (λ, ν) are called **dual feasible** points

proof of lower bound:

- let $\tilde{x} \in \mathcal{D}$ be feasible, so

$$f_i(\tilde{x}) \leq 0, \quad i = 1, \ldots, m, \qquad h_i(\tilde{x}) = 0, \quad i = 1, \ldots, p$$

- then for all $\lambda \in \mathbf{R}^m_+$, $\nu \in \mathbf{R}^p$,

$$\sum_{i=1}^m \lambda_i f_i(\tilde{x}) + \sum_{i=1}^p \nu_i h_i(\tilde{x}) \leq 0$$

- hence, we have upper bound on Lagrangian

$$L(\tilde{x}, \lambda, \nu) = f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \leq f_0(\tilde{x})$$

- taking infimum over $x \in \mathcal{D}$ gives

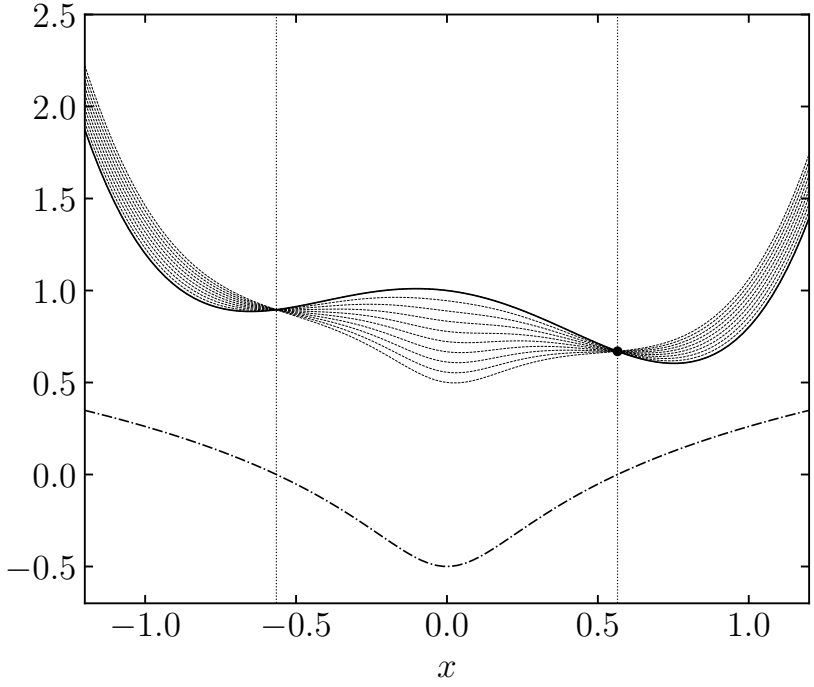
$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x})$$

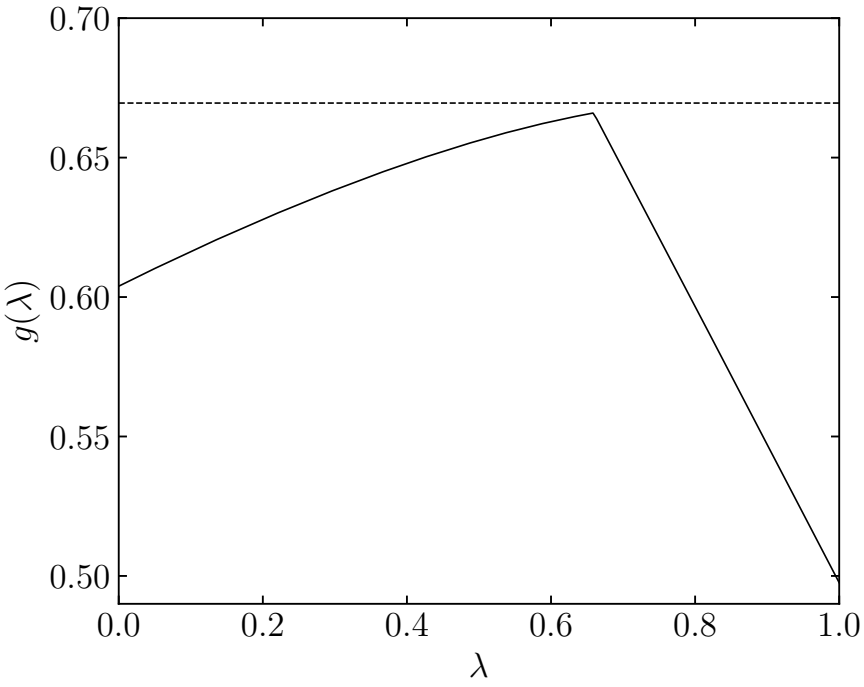
- since the inequality holds for all feasible points $\tilde{x} \in \mathcal{D}$, we have

$$g(\lambda, \nu) \leq \inf_{x \in \mathcal{D}} \left\{ f_0(x) \;\middle|\; \begin{array}{ll} f_i(x) \leq 0, & i = 1, \ldots, m \\ h_i(x) = 0, & i = 1, \ldots, p \end{array} \right\} = p^\star$$

example:

$$\begin{array}{ll} \text{minimize} & f_0(x) = x^4 - x^2 - (1/5)x + 1 \\ \text{subject to} & f_1(x) = (1/4)\log(1 + 20x^2) - 0.5 \leq 0 \end{array}$$





- $f_0(x)$ solid; $f_1(x)$ dashdotted; $L(x, \lambda)$ dashed for different $\lambda > 0$
- dashed line on right shows optimal value $p^\star$; $g(\lambda)$ shown solid

Interpretation of dual function

relaxation interpretation: $g(\lambda, \nu)$ is optimal value of

$$\begin{array}{ll} \text{minimize} & f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \end{array}$$

$$\begin{array}{ll} \text{minimize} & f_0(x) + \lambda^T u + \nu^T v \\ \iff \quad \text{subject to} & f_i(x) \leq u_i, \quad i = 1, \ldots, m \\ & h_i(x) = v_i, \quad i = 1, \ldots, p \end{array}$$

- $x \in \mathbf{R}^n$, $u \in \mathbf{R}^m$, $v \in \mathbf{R}^p$ are variables
- $\lambda \in \mathbf{R}^m$, $\nu \in \mathbf{R}^p$ are fixed parameters
- always feasible for any λ, ν, so original problem is relaxed to an unconstrained problem
- constraint violation is penalized by $\lambda^T u + \nu^T v$

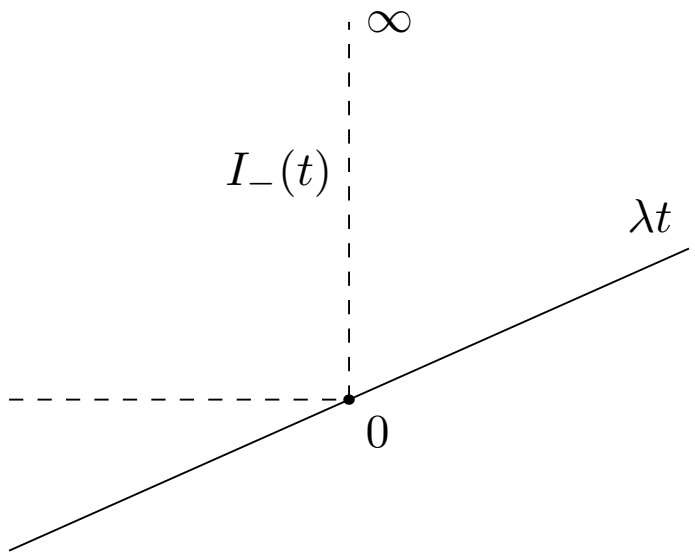
linear approximation interpretation: original problem equals to

$$\text{minimize} \quad f_0(x) + \sum_{i=1}^m I_-(f_i(x)) + \sum_{i=1}^p I_0(h_i(x))$$

where $I_-, I_0 : \mathbf{R} \to \mathbf{R}$ are *indicator functions* of $-\mathbf{R}_+$ and $\{0\}$, *i.e.*,

$$I_-(t) = \begin{cases} 0, & t \leq 0 \\ \infty, & \text{otherwise} \end{cases} \quad \text{and} \quad I_0(t) = \begin{cases} 0, & t = 0 \\ \infty, & \text{otherwise} \end{cases}$$

so $g(\lambda, \nu)$ consists in replacing I_-, I_0 by their linear approximations



- objective relaxation: $\lambda_i f_i(x) \leq I_-(f_i(x))$, $\nu_i h_i(x) \leq I_0(h_i(x))$ for all x

Lagrange dual problem

call original problem **primal problem**

then its associated **(Lagrange) dual problem** is

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- variables: $\lambda \in \mathbf{R}^m$, $\nu \in \mathbf{R}^p$
- find best lower bound on optimal value of primal problem
- dual problem always convex, even if primal is not
- feasible set: $\{(\lambda, \nu) \in \mathbf{dom}\, g \mid \lambda \succeq 0\}$
- optimal value $d^\star$ called **dual optimal** and $p^\star$ is **primal optimal**

Weak duality

for *all* primal and dual problem pairs, we have

$$d^\star \leq p^\star$$

- follows from $g(\lambda, \nu) \leq f_0(x)$ for all feasible x and dual feasible (λ, ν)
- $p^\star - d^\star \geq 0$ is **(optimal) duality gap**
- some useful consequences:
  - dual problem unbounded above $\implies$ primal problem infeasible
  - primal unbounded below $\implies$ dual infeasible

  serves as a *certificate* of problem feasibility and boundedness

example: two-way partitioning ($W \in \mathbf{S}^n$)

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array}$$

$$L(x, \nu) = x^T W x + \sum_{i=1}^{n} \nu_i (x_i^2 - 1) = x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu$$

$$g(\nu) = \inf_x L(x, \nu) = \inf_x x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu$$

$$= \begin{cases} -\mathbf{1}^T \nu, & W + \mathbf{diag}(\nu) \succeq 0 \\ -\infty, & \text{otherwise} \end{cases}$$

hence, dual problem is an SDP (efficiently solved; lower bounding primal)

$$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T \nu \\ \text{subject to} & W + \mathbf{diag}(\nu) \succeq 0 \end{array}$$

Strong duality

for *some* primal and dual problem pairs, we have

$$d^\star = p^\star$$

- condition for strong duality called **constraint qualification**
  - many types exist, depending on problem structure
  - example: for LP, strong duality holds if primal feasible
- *roughly speaking*, strong duality *usually* holds for convex problems

certificate of optimality:

- for $d^\star$ with strong duality and primal feasible $x^\star$:

$$f_0(x^\star) = d^\star \quad \Longrightarrow \quad x^\star \text{ primal optimal}$$

- in general, for primal feasible $\tilde{x}$ and dual feasible $(\tilde{\lambda}, \tilde{\nu})$,

$$p^\star \in [g(\tilde{\lambda}, \tilde{\nu}), f_0(\tilde{x})], \qquad d^\star \in [g(\tilde{\lambda}, \tilde{\nu}), f_0(\tilde{x})]$$

  - $f_0(\tilde{x}) - g(\tilde{\lambda}, \tilde{\nu})$ is **duality gap** and gives suboptimality of $\tilde{x}$
  - if $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$, then $\tilde{x}$ is primal optimal and $(\tilde{\lambda}, \tilde{\nu})$ is dual optimal

sensitivity analysis: perturbed primal optimal value $q^\star : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$,

$$q^\star(u, v) = \inf \left\{ f_0(x) \;\middle|\; \begin{array}{l} x \in \mathcal{D} \\ f_i(x) \le u_i, \;\; i = 1, \ldots, m \\ h_i(x) = v_i, \;\; i = 1, \ldots, p \end{array} \right\}$$

assume strong duality, we have

$$\begin{aligned} & q^\star(0,0) = p^\star = d^\star = g(\lambda^\star, \nu^\star) = \inf_{x \in \mathcal{D}} L(x, \lambda^\star, \nu^\star) \\ \Longrightarrow \quad & q^\star(0,0) \le f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x), \quad x \in \mathcal{D} \end{aligned}$$

since for $x \in \mathcal{D}$, $f_i(x) \leq u_i$, $h_i(x) = v_i$, we have

$$f_0(x) \geq q^\star(0,0) - \lambda^{\star T} u - \nu^{\star T} v$$

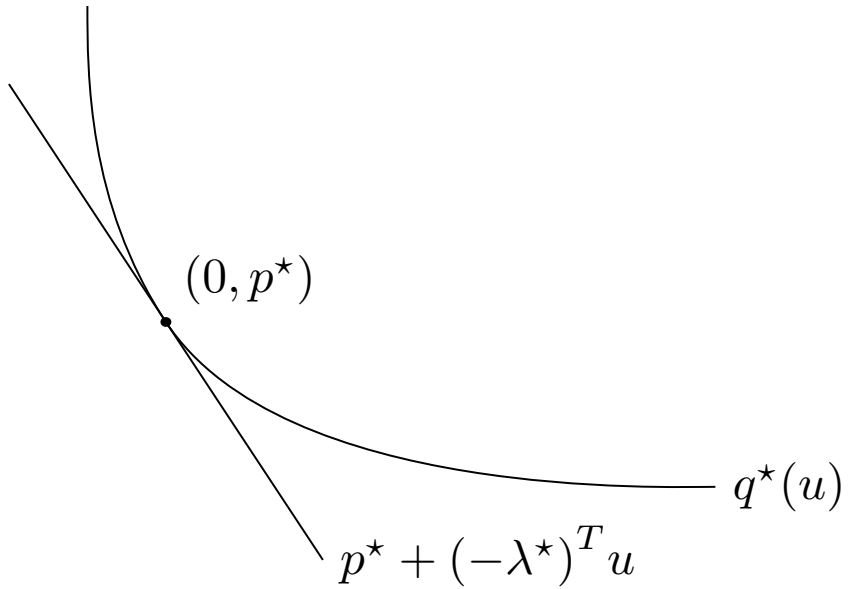
$$\Longrightarrow \quad q^\star(u,v) \geq q^\star(0,0) - \lambda^{\star T} u - \nu^{\star T} v$$

provides *global* linear underestimator of $q^\star$ at primal optimal

for differentiable $q^\star$:

$$\frac{\partial q^\star(0,0)}{\partial u_i} = -\lambda_i^\star, \quad i = 1, \ldots, m$$

$$\frac{\partial q^\star(0,0)}{\partial v_i} = -\nu_i^\star, \quad i = 1, \ldots, p$$



dual optimal $\lambda^\star$ and $\nu^\star$ give local sensitivity of primal optimal value to perturbations in constraints

Complementary slackness

suppose strong duality holds, $x^\star$ primal optimal, $(\lambda^\star, \nu^\star)$ dual optimal,

$$\begin{aligned} f_0(x^\star) = p^\star = d^\star &= g(\lambda^\star, \nu^\star) \\ &= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) \\ &\leq f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star) \\ &\leq f_0(x^\star) \end{aligned}$$

so inequality must hold with equality

therefore we have the following *necessary* conditions:

- $x^\star$ primal optimal $\implies$ $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$ (but not necessarily unique)

- **complementary slackness**:

$$\sum_{i=1}^{m} \lambda_i^\star f_i(x^\star) = 0 \quad \Longrightarrow \quad \lambda_i^\star f_i(x^\star) = 0,\ i = 1, \ldots, m$$

  in other words,

  – if $\lambda_i^\star > 0$, then $f_i(x^\star) = 0$ (constraint is active)

  – if $f_i(x^\star) < 0$, then $\lambda_i^\star = 0$ (constraint inactive)

  so either the ith constraint is active or its dual variable is zero

KKT optimality conditions

assume differentiable objective, strong duality holds,

necessary conditions for $x^\star$ primal optimal, $(\lambda^\star, \nu^\star)$ dual optimal:

$$\begin{aligned}
f_i(x^\star) &\leq 0, \quad i = 1, \ldots, m \\
h_i(x^\star) &= 0, \quad i = 1, \ldots, p \\
\lambda_i^\star &\geq 0, \quad i = 1, \ldots, m \\
\lambda_i^\star f_i(x^\star) &= 0, \quad i = 1, \ldots, m \\
\nabla f_0(x^\star) + \textstyle\sum_{i=1}^m \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^p \nu_i^\star \nabla h_i(x^\star) &= 0
\end{aligned}$$

- first two: primal feasibility; third: dual feasibility
- forth: complementary slackness
- last: stationarity of Lagrangian
- no convexity assumption needed

if primal problem is convex, KKT conditions are **sufficient** for optimality

to see this, suppose $\tilde{x}$ primal feasible, $(\tilde{\lambda}, \tilde{\nu})$ dual feasible, and

$$\begin{array}{rcll} \tilde{\lambda}_i f_i(\tilde{x}) & = & 0, & i = 1, \ldots, m \\ \nabla f_0(\tilde{x}) + \sum_{i=1}^m \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^p \tilde{\nu}_i \nabla h_i(\tilde{x}) & = & 0 & \end{array}$$

for convex primal problem, Lagrangian $L(x, \tilde{\lambda}, \tilde{\nu})$ is convex in x, so

$$g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) = f_0(\tilde{x}) + \sum_{i=1}^m \tilde{\lambda}_i f_i(\tilde{x}) + \sum_{i=1}^p \tilde{\nu}_i h_i(\tilde{x}) = f_0(\tilde{x})$$

- first equality follows from stationarity
- last equality from complementary slackness and feasibility
- implies zero duality gap, so $\tilde{x}$ primal optimal and $(\tilde{\lambda}, \tilde{\nu})$ dual optimal
- no strong duality assumption needed

Infeasible problems

consider *infeasible* feasibility problem

$$\begin{array}{ll} \text{find} & x \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

to find the largest subset of constraints that is feasible,

$$\begin{array}{ll} \text{minimize (over } x, u, v) & \sum_{i=1}^m u_i + \sum_{i=1}^p |v_i| \\ \text{subject to} & u \succeq 0, \quad f_i(x) \leq u_i, \quad i = 1, \ldots, m \\ & h_i(x) = v_i, \quad i = 1, \ldots, p \end{array}$$

- relaxation variables $u \in \mathbf{R}^m$, $v \in \mathbf{R}^p$ measure constraint violations
- objective equivalent to $\|u\|_1 + \|v\|_1$
- so u and v are expected to be sparse

Exact penalty methods

exact penalty method for constrained optimization

$$\begin{array}{ll} \text{minimize} & f_0(x) + \gamma\left(\sum_{i=1}^m u_i + \sum_{i=1}^p |v_i|\right) \\ \text{subject to} & u \succeq 0, \quad f_i(x) \leq u_i, \quad i = 1, \ldots, m \\ & h_i(x) = v_i, \quad i = 1, \ldots, p \end{array}$$

- variable $x \in \mathbf{R}^n$, $u \in \mathbf{R}^m$, $v \in \mathbf{R}^p$; penalty parameter $\gamma > 0$

regularization interpretation: transform constraints into regularization terms in objective; certainty of constraint satisfaction depends on γ

- if original problem feasible, sufficiently large γ guaranteed to give originally feasible solution

- if original problem infeasible, solution gives sparsest constraint violation

example: infeasible least norm problem

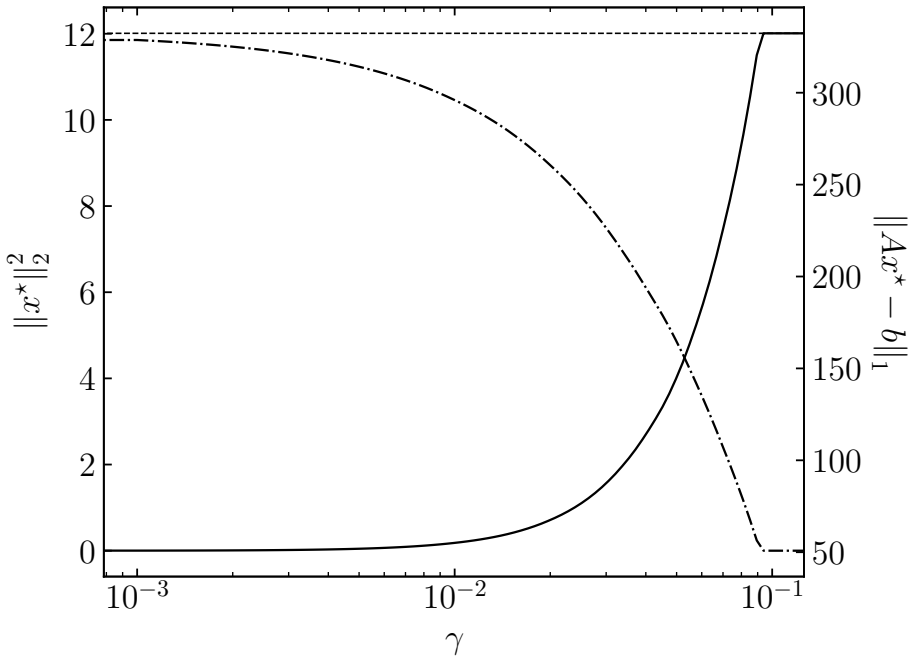
$$\begin{array}{ll} \text{minimize} & \|x\|_2^2 \\ \text{subject to} & Ax = b \end{array}$$

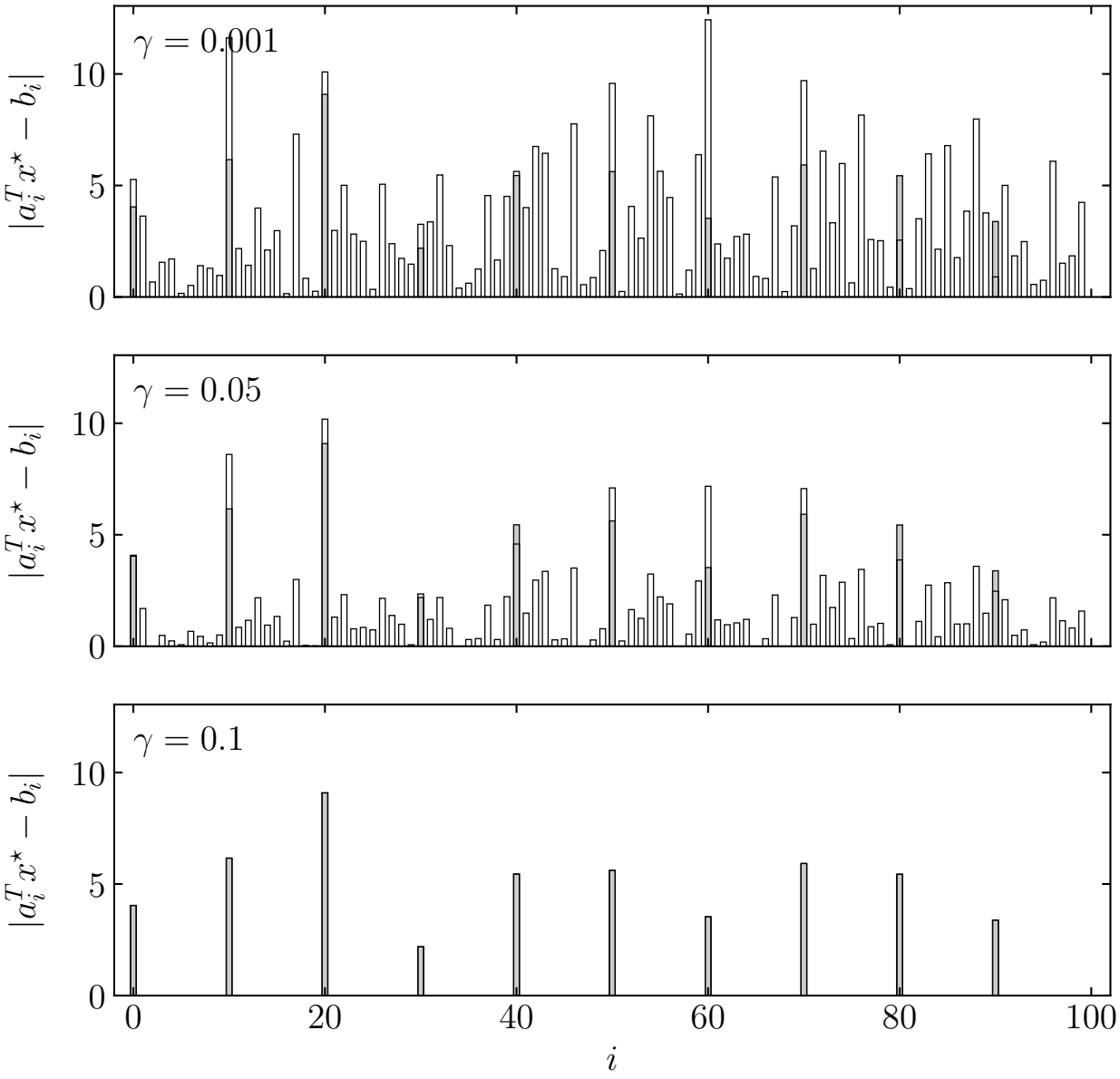
where linear system $Ax = b$ is inconsistent, exact penalty method gives

$$\text{minimize} \quad \|x\|_2^2 + \gamma \|Ax - b\|_1$$

for $A \in \mathbf{R}^{100\times 10}$ with different γ:

- $\|x^\star\|_2^2$ shown solid
- $\|Ax^\star - b\|_1$ dashdotted
- top dashed line shows $\|x^\star\|_2^2$ on largest feasible subset





(violation for solution on largest feasible subset shown darker)