Constrained optimization

Problem in standard form

$$\text{minimize } f(\mathbf{x})$$

$$\text{subject to } a_i(\mathbf{x}) = 0, \text{ for } i = 1, 2, \cdots p$$

$$c_j(\mathbf{x}) \geq 0 \text{ for } j = 1, 2, \cdots, q$$

$f : \mathbf{R}^n \to \mathbf{R}$ $\quad a_i : \mathbf{R}^n \to \mathbf{R}$ $\quad c_j : \mathbf{R}^n \to \mathbf{R}$

$f(\mathbf{x}^*) = \infty$, if problem is infeasible

$f(\mathbf{x}^*) = -\infty$, if problem is unbounded below

Equality constraints

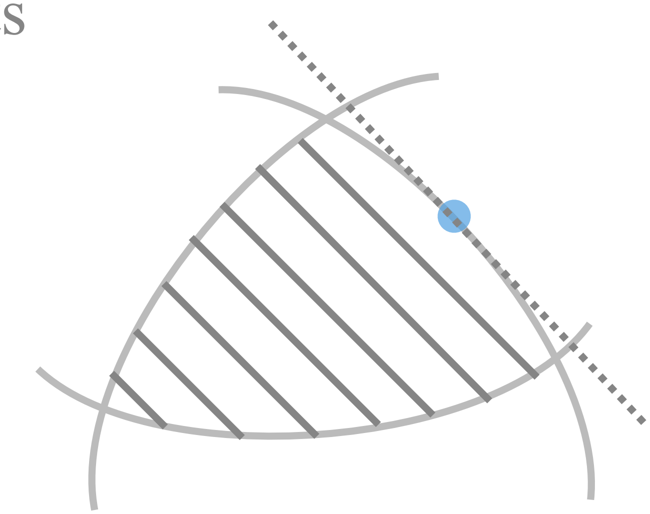
- An *equality constraint* defines a hypersurface where $a_i(\mathbf{x}) = 0$
- A *regular point* is a point in the feasible region and has a full-rank Jacobian
- A tangent plane of the hypersurface determined by the constraint at a regular point x is well defined
- The number of constraints, p, must be less than the dimension of the domain, n

Linear equality constraints

- What is the Jacobian of a linear equality constraint, $\mathbf{Ax} = \mathbf{b}$?
- If rank($\mathbf{A}$) = p, any feasible $\mathbf{x}$ is a regular point
- If rank($\mathbf{A}$) < p, we can test whether contradiction or redundancy exists by checking: rank([$\mathbf{A}$ $\mathbf{b}$])
  - if rank([$\mathbf{A}$ $\mathbf{b}$]) ? rank($\mathbf{A}$), contradiction
  - if rank ([$\mathbf{A}$ $\mathbf{b}$]) ? rank($\mathbf{A}$), redundancy

Inequality constraints

- What is the largest number of inequality constraints in an optimization in R^n?
- Two general approaches to deal with inequality constraints:
  - Divide into active and inactive constraints
  - Convert into equality constraints



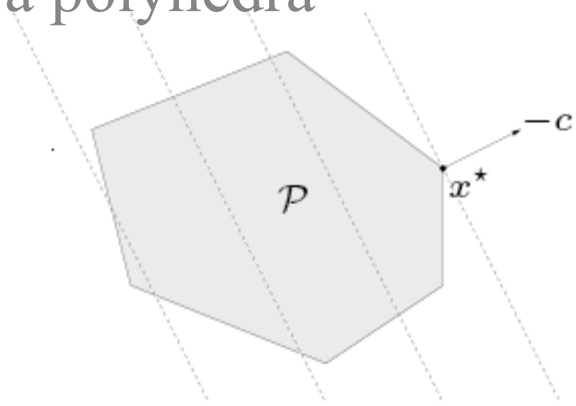
Linear programming

$$\begin{aligned} \text{minimize } & \mathbf{c}^T\mathbf{x} \\ \text{subject to } & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

- Alternative form can be conformed to the standard form

$$\begin{aligned} \text{minimize } & \mathbf{c}^T\mathbf{x} \\ \text{subject to } & \mathbf{A}\mathbf{x} \geq \mathbf{b} \end{aligned}$$

- The feasible set is a polyhedra



Constraint transformations

- We can convert each inequality to equality constraint by introducing slack variable: $\mathbf{y} = \mathbf{Ax} - \mathbf{b}$
  - Inequalities becomes $\mathbf{Ax} - \mathbf{y} = \mathbf{b}$ and $\mathbf{y} \geq 0$
- We can introduce nonnegative bounds on $\mathbf{x}$ by adding two nonnegative vectors $\mathbf{x}^+$ and $\mathbf{x}^-$: $\mathbf{x} = \mathbf{x}^+ - \mathbf{x}^-$
- With new variables, $\hat{\mathbf{x}} = [\mathbf{x}^+ \ \mathbf{x}^- \ \mathbf{y}]$,the problem becomes:

$$\text{minimize } \hat{\mathbf{c}}^T \hat{\mathbf{x}}$$
$$\text{subject to } \hat{\mathbf{A}}\hat{\mathbf{x}} = \hat{\mathbf{b}}$$
$$\hat{\mathbf{x}} \geq \mathbf{0}$$

Examples

diet problem: choose quantities $x_1, \ldots, x_n$ of n foods

- one unit of food j costs c_j, contains amount a_{ij} of nutrient i
- healthy diet requires nutrient i in quantity at least b_i

to find cheapest healthy diet,

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

piecewise-linear minimization

$$\text{minimize} \quad \max_{i=1,\ldots,m}(a_i^T x + b_i)$$

equivalent to an LP

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & a_i^T x + b_i \leq t, \quad i = 1, \ldots, m \end{array}$$

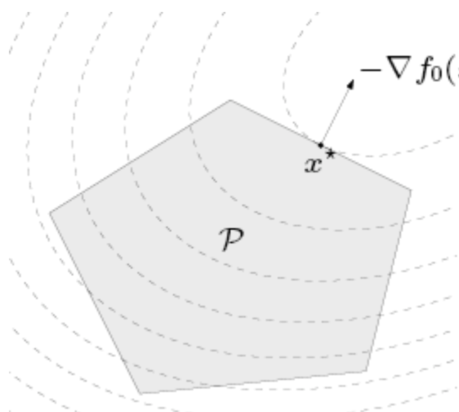
Convex quadratic programming

$$\text{minimize } f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} + \mathbf{x}^T\mathbf{p} + c$$

$$\text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b}$$

$$\mathbf{C}\mathbf{x} \geq \mathbf{d}$$

- If Hessian is positive semidefinite, QP can be regarded as a special class of convex programming
- If Hessisn is indefinite, the problem becomes NP hard



Examples

least-squares

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

- analytical solution $x^\star = A^\dagger b$ ($A^\dagger$ is pseudo-inverse)
- can add linear constraints, *e.g.*, $l \preceq x \preceq u$

linear program with random cost

$$\begin{array}{ll} \text{minimize} & \bar{c}^T x + \gamma x^T \Sigma x = \mathbf{E}\, c^T x + \gamma\, \mathbf{var}(c^T x) \\ \text{subject to} & Gx \preceq h, \quad Ax = b \end{array}$$

- c is random vector with mean $\bar{c}$ and covariance Σ
- hence, $c^T x$ is random variable with mean $\bar{c}^T x$ and variance $x^T \Sigma x$
- $\gamma > 0$ is risk aversion parameter; controls the trade-off between expected cost and variance (risk)

Quadratically constrained QP

$$\text{minimize } f(\mathbf{x}) = \tfrac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} + \mathbf{x}^T\mathbf{p} + c$$

$$\text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b}$$

$$\frac{1}{2}\mathbf{x}^T\mathbf{H}_i\mathbf{x} + \mathbf{x}^T\mathbf{p}_i + c_i, \quad i = 1, \cdots, m$$

- Objective function and constraints are convex objective
- If $\mathbf{H}_i$ are positive definite, the feasible region is the intersection of m ellipsoids and an affine set

Second-order cone programming

$$\text{minimize } \mathbf{b}^T\mathbf{x}$$

[redacted]t to $\|$ [redacted] $\|_2 \leq \;{}_i^T \quad + d_i \quad i = 1, \cdots, q$

- Inequalities are called second-order cone constraints

$$\{\|\mathbf{A}\mathbf{x} + \mathbf{c}\|, \mathbf{b}^T\mathbf{x} + d\} \in \text{second order cone in } R^{n+1}$$

- More general than LP and QCQP
- For $\mathbf{A}_i = \mathbf{0}$ and $\mathbf{c}_i = \mathbf{0}$, reduces to an LP. For $\mathbf{b}_i = \mathbf{0}$, reduces to QCQP

Robust linear programming

the parameters in optimization problems are often uncertain, *e.g.*, in an LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \quad i = 1, \ldots, m, \end{array}$$

there can be uncertainty in c, a_i, b_i

two common approaches to handling uncertainty (in a_i, for simplicity)

- deterministic model: constraints must hold for all $a_i \in \mathcal{E}_i$

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i \text{ for all } a_i \in \mathcal{E}_i, \quad i = 1, \ldots, m, \end{array}$$

- stochastic model: a_i is random variable; constraints must hold with probability η

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \mathbf{prob}(a_i^T x \leq b_i) \geq \eta, \quad i = 1, \ldots, m \end{array}$$

deterministic approach via SOCP

- choose an ellipsoid as $\mathcal{E}_i$:

$$\mathcal{E}_i = \{\bar{a}_i + P_i u \mid \|u\|_2 \leq 1\} \qquad (\bar{a}_i \in \mathbf{R}^n, \quad P_i \in \mathbf{R}^{n \times n})$$

  center is $\bar{a}_i$, semi-axes determined by singular values/vectors of P_i

- robust LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i \quad \forall a_i \in \mathcal{E}_i, \quad i = 1, \ldots, m \end{array}$$

  is equivalent to the SOCP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \bar{a}_i^T x + \|P_i^T x\|_2 \leq b_i, \quad i = 1, \ldots, m \end{array}$$

  (follows from $\sup_{\|u\|_2 \leq 1} (\bar{a}_i + P_i u)^T x = \bar{a}_i^T x + \|P_i^T x\|_2$)

stochastic approach via SOCP

- assume a_i is Gaussian with mean $\bar{a}_i$, covariance Σ_i ($a_i \sim \mathcal{N}(\bar{a}_i, \Sigma_i)$)

- $a_i^T x$ is Gaussian r.v. with mean $\bar{a}_i^T x$, variance $x^T \Sigma_i x$; hence

$$\mathbf{prob}(a_i^T x \leq b_i) = \Phi\left(\frac{b_i - \bar{a}_i^T x}{\|\Sigma_i^{1/2} x\|_2}\right)$$

  where $\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^{x} e^{-t^2/2}\, dt$ is CDF of $\mathcal{N}(0,1)$

- robust LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \mathbf{prob}(a_i^T x \leq b_i) \geq \eta, \quad i = 1, \ldots, m, \end{array}$$

  with $\eta \geq 1/2$, is equivalent to the SOCP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \bar{a}_i^T x + \Phi^{-1}(\eta) \|\Sigma_i^{1/2} x\|_2 \leq b_i, \quad i = 1, \ldots, m \end{array}$$

Semidefinite programming

$$\text{minimize } \mathbf{c}^T\mathbf{x}$$
$$\text{subjecto to } x_1\mathbf{F}_1 + x_2\mathbf{F}_2 + \cdots + x_n\mathbf{F}_n + \mathbf{G} \preceq \mathbf{0}$$
$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad \text{with } \mathbf{F}_i, \mathbf{G} \in \mathbf{S}^n$$

- The inequality constraint is called linear matrix inequality (LMI)

$$x_1\hat{\mathbf{F}}_1 + x_2\hat{\mathbf{F}}_2 + \cdots + x_n\hat{\mathbf{F}}_n + \hat{\mathbf{G}} \preceq \mathbf{0}$$
$$x_1\tilde{\mathbf{F}}_1 + x_2\tilde{\mathbf{F}}_2 + \cdots + x_n\tilde{\mathbf{F}}_n + \tilde{\mathbf{G}} \preceq \mathbf{0}$$

- Multiple LMIs can be represented by one a single LMI

$$x1 \begin{bmatrix} \hat{\mathbf{F}}_1 & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{F}}_1 \end{bmatrix} + \cdots + x_n \begin{bmatrix} \hat{\mathbf{F}}_n & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{F}}_n \end{bmatrix} + \begin{bmatrix} \hat{\mathbf{G}} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{G}} \end{bmatrix} \preceq \mathbf{0}$$

Eigenvalue minimization

$$\text{minimize} \quad \lambda_{\max}(A(x))$$

where $A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n$ (with given $A_i \in \mathbf{S}^k$)

equivalent SDP

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & A(x) \preceq tI \end{array}$$

- variables $x \in \mathbf{R}^n$, $t \in \mathbf{R}$
- follows from

$$\lambda_{\max}(A) \leq t \quad \Longleftrightarrow \quad A \preceq tI$$

Nonconvex problems

- A problem is not convex if one constraint is not convex or the objective function is not convex
- Use SQP or penalty methods (Barrier function methods)

Simple transformation methods

- Introduce equality constraints
- Eliminate equality constraints
- Eliminate nonnegativity bounds
- Eliminate interval-type constraints

Eliminate equality constraints

$$\begin{aligned} &\text{minimize } f(\mathbf{x}) \\ &\quad \text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b} \\ &\qquad\qquad c_i(\mathbf{x}) \geq 0 \quad \text{for } 1 \leq i \leq q \end{aligned}$$

- Use Moore-Penrose pseudo inverse of **A**
  - $\mathbf{A}^{+} = \mathbf{A}^{T}(\mathbf{A}\mathbf{A}^{T})^{-1}$
- $\mathbf{A}^{+}\mathbf{b}$ is a point on the hyperplane
- Introduce a new variable $\boldsymbol{\phi}'$, which is a vector lies on the hyperplane defined by the constraint
- $\boldsymbol{\phi}'$ is in the null space of **A**

Eliminate equality constraints

- Reduce the dimension of variables from n to the null space of **A**
- Apply SVD on $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ to computes the null space of **A**
  - Null space of **A**: $\mathbf{V}_r$, spanned by the last n-m vectors of **V**
- The old variable **x** can be represented by
  - $\mathbf{x} = \mathbf{V}_r\phi + \mathbf{A}^+\mathbf{b}$, where ϕ is an arbitrary vector in R^{n-m}

Eliminate nonnegativity bounds

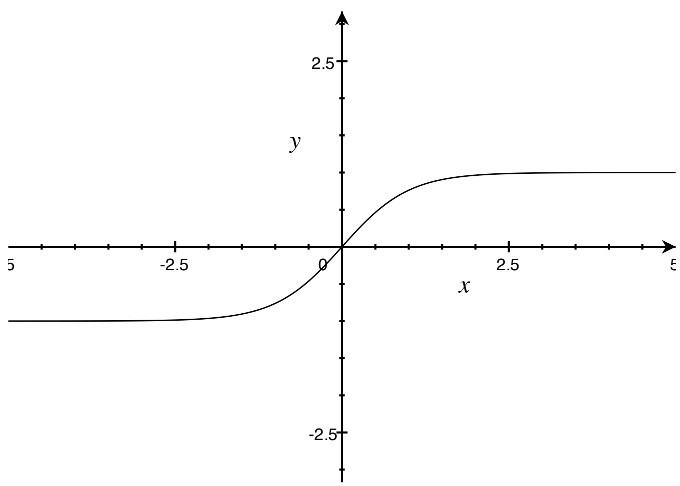
- Nonnegativity bound $x_i \geq 0$ can be eliminated using the variable transformation $x_i = y_i^2$
- Constraint $x_i \geq d$ can be eliminated by the variable transformation $x_i = d + y_i^2$
- What about $x_i \leq d$?

Eliminate interval-type constraints

- Interval constraint $a \leq x \leq b$ can be eliminated by variable transformation

$$x = \frac{b-a}{2}\tanh(z) + \frac{b+a}{2}$$

$$, \text{ where } \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$



References

- A. Antoniou and W.S. Lu, Practical optimization
- S. Boyd, Convex optimization, lecture notes