LECTURE 11: SOLUTION METHODS FOR CONSTRAINED OPTIMIZATION

1. Primal approach
2. Penalty and barrier methods
3. Dual approach
4. Primal-dual approach

Basic approaches

- I. Primal Approach
  - Feasible Direction Method
  - Active Set Method
  - Gradient Projection Method
  - Reduced Gradient Method
  - Variations
- II. Penalty and Barrier Methods
  - Penalty Function Method
  - Barrier Function Method
- III. Dual Approach
  - Augmented Lagrangian Method (Multiplier Method)
  - Cutting Plane Method
- IV. Primal-Dual Approach
  - Lagrange Method

I. Primal approach

Basic concepts:

- A search method that works on the original problem by searching through the feasible domain for optimality.

- Stays feasible at each iteration.

- Decreases the objective function value constantly.

- Given that the original problem has *n* variables with *m* equality constraints being satisfied at each iteration, then the feasible space has *n-m* dimensions for the primal methods to work with.

A. Feasible direction method

- $x^{k+1} = \underbrace{x^k}_{\text{current feasible solution}} + \alpha_k \underbrace{d^k}_{\text{feasible direction}}$
- $\alpha_k > 0$ is a step length that minimizes $f(x^k + \alpha d^k)$ w.r.t. $\alpha > 0$ in a range that $x^k + \alpha d^k$ remains feasible.

Example (Simplified Zoutendijk method)

$$\begin{array}{lll} \text{Minimize} & f(x) & \\ \text{s. t.} & a_1^T x & \leq b_1 \\ & \vdots & \vdots \\ & a_m^T x & \leq b_m \end{array}$$

When x^k is feasible with $I = \{i \mid a_i^T x^k = b_i\}$, d^k is chosen by

$$\begin{array}{ll} \text{Minimize} & \nabla f(x^k) d \\ \text{s. t.} & a_i^T d \leq 0 \ , \quad i \in I \\ & \sum_{i=1}^{n} |d_i| = 1 \end{array}$$

Potential difficulties

- initial feasible solution
- existence of a feasible direction

feasible direction = ?

- convergence

B. Active set method

1. Consider the following problem:

$$(\mathrm{P}) \qquad \begin{array}{ll} \text{Minimize} & f(x) \\ \text{s. t.} & g(x) \leq 0. \end{array}$$

2. The first order necessary condition says that at a local minimum, we have

$$\nabla f(x^*) + \sum_{i \in I(x^*)} \mu_i^* \nabla g_i(x^*) = 0,$$

$$\begin{aligned} g_i(x^*) &= 0, \quad i \in I(x^*) \\ g_i(x^*) &< 0, \quad i \notin I(x^*) \\ \mu_i^* &\geq 0, \quad i \in I(x^*) \\ \mu_i^* &= 0, \quad i \notin I(x^*). \end{aligned}$$

3. When an active set in known, the original problem is reduced to an optimization problem with equality constraints.

4. There are at most 2^m combinations of possible active sets.

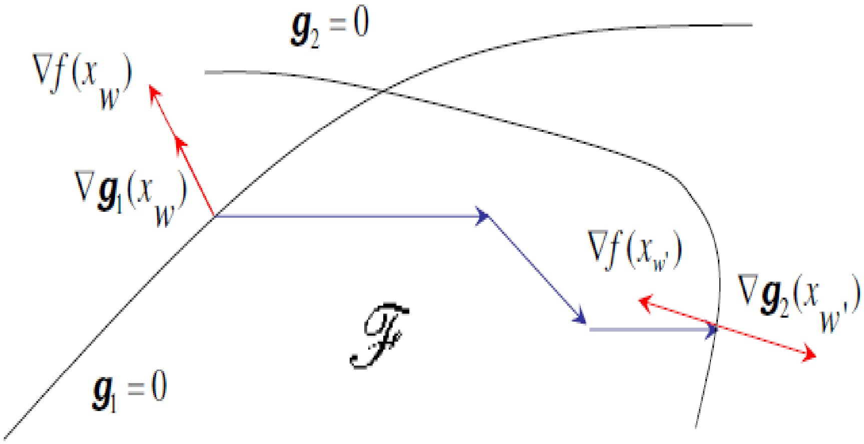
Active set method

5. Given any candidate active set (working set) W, assume that x_w is a solution to

$$(P_w)\quad \begin{array}{ll} \text{Minimize} & f(x) \\ \text{s. t.} & g_i(x) = 0, \quad i \in W. \end{array}$$

If x_w is feasible to (P) and $\mu_i \geq 0$, $\forall\, i \in W$, then x_w is a local optimal solution of (P).

6. If $\exists\, i \in W$ such that $\mu_i < 0$, then dropping constraint i (but staying feasible) will decrease the objective function value due to the sensitivity theorem.



Active set method

7. The surface defined by a working set is called a working surface. By dropping i from W and moving on the new working surface (toward the interior of $\mathscr{F}$), we move to an improved solution.

8. Monitoring the movement to avoid infeasibility until one or more constraints become active, then add them to the working set W. We return to 5. and solve (P_w) again.

9. If we can assure the objective function value is monotonically decreasing, then any working set will not appear twice in the process. Hence the active set method terminates in a finite number of iterations.

10. The solutions to the intermediate problems better be exact solutions to determine the correct sign of $\mu_i's$ and to assure the current working set is not coming back to the iterative process.

Active set theorem

Suppose that for every subset W of the constraint indices, the problem

$$(P_w) \qquad \begin{array}{ll} \text{Minimize} & f(x) \\ \text{s. t.} & g_i(x) = 0, \quad i \in W \end{array}$$

has a unique nondegenerate solution, i.e., $\lambda_i \neq 0$ for $i \in W$. Then the sequence of points generated by the active set method converges to a local solution of (P).

Example:

– Dr. Hao Cheng, SAS Institute, Inc. – "An Active Set Algorithm for Univariate Cubic L_1 Splines."

C. Gradient projection method

- Key Idea: The negative gradient at a current feasible solution is projected onto the working surface for finding the direction of movement.

Example (Linear Constraints)

$$\begin{aligned} \text{Minimize} \quad & f(x) \\ \text{s. t.} \quad & a_i^T x \le b_i \ , \quad i \in I_1 \\ & a_i^T x = b_i \ , \quad i \in I_2 \end{aligned}$$

When x^k is feasible with

$$w(x^k) \triangleq \{i \mid a_i^T x^k = b_i,\ i \in I_1 \cup I_2\},$$

denote $A_q \triangleq [a_i^T]_{i \in w(x^k)}$ and

$$P_k \triangleq I - A_q^T (A_q A_q^T)^{-1} A_q \quad \leftarrow \text{projection map}$$

and $d^k \triangleq -P_k \nabla f(x^k)$.

If $d^k \neq 0$, find

$$\beta_k \triangleq arg \ \max \{\alpha > 0 \mid x^k + \alpha d^k \text{ is feasible}\}$$

and $\alpha_k \triangleq arg \ \min \{f(x^k + \alpha d^k) \mid 0 \le \alpha \le \beta_k\}$.

Gradient projection method

Example (Nonlinear Constraints)

$$\begin{aligned} \text{Minimize} \quad & f(x) \\ \text{s. t.} \quad & h_i(x) \quad = 0, \quad i = 1, 2, \cdots, m \\ & g_{m+j}(x) \quad \leq 0, \quad j = 1, 2, \cdots, p \end{aligned}$$

denote $A_q = \begin{bmatrix} \nabla h_i(x^k) \\ \hline \nabla g_j(x^k) \end{bmatrix}_{i,j \in w(x^k)}$ and

When x^k is feasible with

$$w(x^k) \triangleq \{1, 2, \cdots, m\} \cup \{j \mid g_{m+j}(x^k) = 0\},$$

$P_k \triangleq I - A_q^T (A_q A_q^T)^{-1} A_q$ ← projection mapping.

D. Reduced gradient method

Closely related to the simplex method

$$\begin{aligned} \text{Min} \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

- Under the non-degeneracy assumption, let $A = [B|N]$, then we have

$$\begin{aligned} \text{Min} \quad & c_B^T x_B + c_N^T x_N \\ \text{s.t.} \quad & \\ & \begin{bmatrix} B & N \\ 0 & I \end{bmatrix} \begin{bmatrix} x_B \\ x_N \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix} \\ & x_B \geq 0 \end{aligned}$$

- The reduced cost is

$$r_q \triangleq c_q - c_B^T B^{-1} A_q$$

Notice that $r_q = 0$ if q is basic.

We only need to consider

$$r_N^T = c_N^T - c_B^T B^{-1} N \geq 0 \ \text{ or not.}$$

Reduced gradient method

- Also notice that

$$Bx_B + Nx_N = b$$

If x_N becomes $x_N + \triangle x_N$, x_B has to change to $x_B + \triangle x_B$ in order to keep feasibility.

$$B(x_B + \triangle x_B) + N(x_N + \triangle x_N) = b$$
$$\Rightarrow B \triangle x_B + N \triangle x_N = 0$$
$$\Rightarrow \triangle x_B = -B^{-1} N \triangle x_N.$$

- If there is an $r_q < 0$, then the nonbasic variable x_q enter the basis.

Case 1 – linearly constrained problems

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{s. t.} & Ax = b \\ & x \geq 0 \end{array}$$

Rewrite as

$$\begin{array}{ll} \text{Minimize} & f(y,z) \\ \text{s. t.} & \left[\begin{array}{cc} B & N \end{array}\right] \begin{bmatrix} y \\ z \end{bmatrix} = b \\ & y \geq 0, z \geq 0 \end{array}$$

– Define

$$r_N^T \triangleq \nabla_z f(y,z) - \nabla_y f(y,z) B^{-1} N$$

– The (y,z) satisfies the first-order necessary conditions for optimality if and only if

$$\begin{cases} r_i = 0 & \forall\ z_i > 0 \\ r_i \geq 0 & \forall\ z_i = 0. \end{cases}$$

Linearly constrained case

– One step of the procedure

1. Let $\triangle z_i = \begin{cases} -r_i \ , & \text{if} \quad r_i < 0 \quad \text{or} \quad z_i > 0 \\ 0 \ , & \text{otherwise} \end{cases}$

2. If $\triangle z = 0$, STOP with a solution.
   Otherwise, find

$$\triangle y = -B^{-1}N \triangle z.$$

3. Find

$$\begin{aligned} \beta_1 \quad &\triangleq arg \max\{\alpha \geq 0 \mid y + \alpha \triangle y \geq 0\} \\ \beta_2 \quad &\triangleq arg \max\{\alpha \geq 0 \mid z + \alpha \triangle z \geq 0\} \\ \alpha_k \quad &\triangleq arg \min\{f(x^k + \alpha \triangle x^k) \mid 0 \leq \alpha \leq \beta_1, \\ &\qquad\qquad\qquad\qquad\qquad\qquad 0 \leq \alpha \leq \beta_2\}. \end{aligned}$$

4. Define

$$x^{k+1} = x^k + \alpha_k \triangle x^k.$$

Nonlinearly constrained problems

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{s. t.} & h(x) = 0 \\ & a \leq x \leq b \end{array}$$

where $x \in E^n$ and $h(x)$ is of dimension m.

– Under the non-degeneracy assumption, we let $x = (y, z)$ with $y \in E^m$ and $z \in E^{n-m}$.

The reduced gradient in this case becomes

$$r_N^T \triangleq \nabla_z f(y, z) - \nabla_y f(y, z)[\nabla_y h(y, z)]^{-1} \nabla_z h(y, z)$$

and

$$\triangle\, y \triangleq -[\nabla_y h(y, z)]^{-1} \nabla_z h(y, z) \,\triangle\, z.$$

II. Penalty and barrier methods

- They are procedures for approximating constrained optimization problems by unconstrained problems.

- Penalty methods add to the objective function a term that prescribes a high cost for constraint violation.

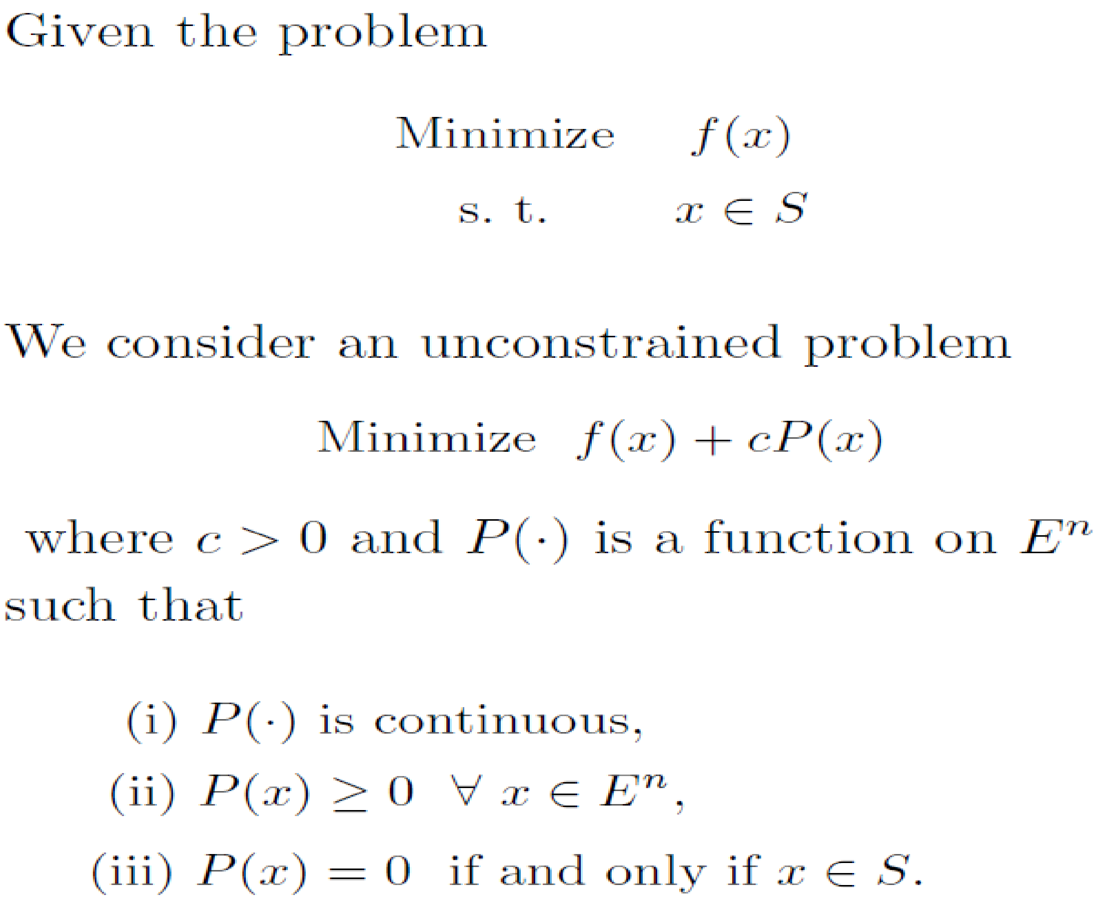
- Barrier methods add a term that favors points interior to the feasible domain over those near the boundary.

- A parameter c is used to control the impact of the additional term.

- They usually work on the *n*-dimensional space of variables directly.

A. Penalty function method

- Basic concept:



Given the problem

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{s. t.} & x \in S \end{array}$$

We consider an unconstrained problem

$$\text{Minimize} \quad f(x) + cP(x)$$

where $c > 0$ and $P(\cdot)$ is a function on E^n such that

(i) $P(\cdot)$ is continuous,

(ii) $P(x) \geq 0 \ \ \forall \ x \in E^n$,

(iii) $P(x) = 0$ if and only if $x \in S$.

Example of penalty functions

- Inequality constraints

Example 1:

$$S = \{x \in E^n \mid g_i(x) \leq 0,\ i = 1, 2, \cdots, p\}$$

- $$P_1(x) \triangleq \frac{1}{2} \sum_{i=1}^{p} [\underbrace{\max\{0,\ g_i(x)\}}_{g_i^+(x)}]^2$$

$$P_2(x) \triangleq \sum_{i=1}^{p} [\underbrace{\max\{0,\ g_i(x)\}}_{g_i^+(x)}]^q \text{ for some } q > 0.$$

Equality constraints

Example 2:

$$S = \{x \in E^n \mid h_i(x) = 0,\ i = 1, 2, \cdots, m\}$$

$$P_3(x) = \tfrac{1}{2}\|h(x)\|^2.$$

General

- Give a problem

$$
\begin{array}{llll}
 & \text{Minimize} & f(x) & \\
(P) & \text{s. t.} & h_i(x) = 0\ , & i = 1, 2, \cdots, m \\
 & & g_j(x) \leq 0\ , & j = 1, 2, \cdots, p
\end{array}
$$

$$P_e(x) \triangleq \sum_{i=1}^{m} |h_i(x)| + \sum_{j=1}^{p} \max\{0,\ g_j(x)\}.$$

Penalty function method

- How it works?

Let $\{c_k \geq 0\} \nearrow +\infty$ with $c_{k+1} > c_k$.

For each k, solve the problem

$$\text{Minimize} \quad q(c_k, x) \triangleq f(x) + c_k P(x)$$

for a solution x^k.

Any good properties?

1. $q(c_k, x^k) \leq q(c_{k+1}, x^{k+1})$

   $P(x^k) \geq P(x^{k+1})$

   $f(x^k) \leq f(x^{k+1})$

2. Let x^* be a solution to the original problem, then

   $$f(x^*) \geq q(c_k, x^k) \geq f(x^k), \quad \forall\, k.$$

3. Let $\{x^k\}$ be a sequence generated by the penalty method. Then any limit point of $\{x^k\}$ is a solution to the original problem.

Question

- Do we always have to solve an infinite sequence of penalty problems to obtain a correct solution to the original problem?

- Answer:

  Exact Penalty Functions are exact in the sense that the solution of the penalty problem yields the exact solution to the original problem for a finite value of the penalty parameter.

Exact penalty theorem

- Give a problem

$$(P)\quad \begin{array}{lll} & \text{Minimize} & f(x) \\ & \text{s. t.} & h_i(x)=0\,, \quad i=1,2,\cdots,m \\ & & g_j(x)\leq 0\,, \quad j=1,2,\cdots,p \end{array}$$

$$P_e(x) \triangleq \sum_{i=1}^{m} |h_i(x)| + \sum_{j=1}^{p} \max\{0,\ g_j(x)\}.$$

- Consider the penalty problem

$$\text{Minimize} \quad f(x)+cP_e(x)$$

- Exact Penalty Theorem

Suppose that x^* satisfies the second-order sufficient conditions for a local minimum of problem (P) and λ^*, μ^* be the corresponding Lagrange multipliers. Then for $c > \max\left\{|\lambda_i^*|,\ \mu_j^* \;\middle|\; i=1,\cdots,m,\ \ j=1,\cdots,p\right\}$, x^* is also a local minimum of the penalty problem with the exact penalty function $P_e(\cdot)$.

- Note that $P_e(x)$ is a non-smooth function.

Related properties – Lagrange multipliers

(a) With $P_1(x)$, corresponding to the sequence $\{x^k\}$ generated by the penalty function method, we define $\mu_k \triangleq c_k\ g^+(x^k) \geq 0$. If $x^k \to x^*$, a solution to the original problem that is a regular point, then $\mu_k \to \mu^*$, the associated Lagrange multiplier vector of the original problem.

(b) When $P_2(x)$ is used, we have

$$\mu_k \triangleq c_k\ q \begin{pmatrix} (g_1^+(x^k))^{q-1} \\ \vdots \\ (g_p^+(x^k))^{q-1} \end{pmatrix} \geq 0\ , \quad \text{for } q > 1$$

such that

$$\mu_k \to \mu^* \quad \text{as} \quad x^k \to x^*.$$

(c) When $P_3(x)$ is used, we have

$$\lambda_k \triangleq c_k\ h(x^k)$$

such that

$$\lambda_k \to \lambda^* \quad \text{as} \quad x^k \to x^*.$$

Related properties – Hessian matrix

- Virtually any choice of penalty function (within the class considered) leads to an ill-conditioned Hessian and to consideration of Hessian of the Lagrangian restricted to subspace that is orthogonal to the subspace spanned by the gradients of the active constraints.

B. Barrier function method

Given the problem

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{s. t.} & x \in S \end{array}$$

where $int(S) \neq \phi$ and any boundary point of S can be approached from the interior (S is robust.)

We consider an unconstrained problem

$$\begin{array}{ll} \text{Minimize} & f(x) + \frac{1}{c} B(x) \\ \text{s. t.} & x \in int(S) \end{array}$$

where $c > 0$ and $B(\cdot)$ is a function defined on $int(S)$ such that

(i) $B(\cdot)$ is continuous,

(ii) $B(x) \geq 0$,

(iii) $B(x) \rightarrow +\infty$ as $x \rightarrow bdry(S)$.

- Example:

$$S = \{x \in E^n \mid g_i(x) \leq 0, \quad i = 1, \cdots, p\} \text{ is robust.}$$

$$B(x) \triangleq -\sum_{i=1}^{p} \frac{1}{g_i(x)}.$$

Barrier function method

- Method: Let $\{c_k \geq 0\} \nearrow +\infty$ with $c_{k+1} > c_k$.

  For each k, solve the problem

  $$\text{Minimize} \quad r(c_k, x) \triangleq f(x) + \frac{1}{c_k} B(x)$$

  for a solution x^k.

- Properties:

  Virtually the same as that of the penalty function method.

III. Dual approach

- Work on the dual problem instead of the original primal problem.

- Study of Lagrange multipliers

- Dual interpretation of cutting planes

Augmented Lagrangian (multiplier) method

- Basic idea:

  - Combination of penalty function and local duality methods. It is one of the most effective algorithms.

Case 1: Equality Constraints:

$$(P) \quad \begin{array}{ll} \text{Minimize} & f(x) \\ \text{s. t.} & h(x) = 0 \end{array}$$

Its "augmented Lagrangian" is the function

$$\ell_c(x, \lambda) \triangleq f(x) + \lambda^T h(x) + \frac{1}{2} c \, \|h(x)\|^2$$

for a positive constant c.

Penalty function viewpoints

1. For a fixed vector λ, $\ell_c(x,\lambda)$ corresponds to

$$(\bar{P}) \quad \begin{array}{ll} \text{Minimize} & f(x) + \lambda^T h(x) \\ \text{s. t.} & h(x) = 0 \end{array}$$

with $P_3(x) = \frac{1}{2} c \, \|h(x)\|^2$ as the penalty function.

2. When the correct Lagrange multiplier vector λ^* is used, then

$$\begin{aligned} \nabla \ell_c(x^*, \lambda^*) &= \underbrace{\nabla f(x^*) + (\lambda^*)^T \nabla h(x^*)} + c \, \underset{0}{\overset{\|}{h(x^*)}} \nabla h(x^*) \\ &= \qquad 0 \qquad + \qquad 0 \\ &= 0 \end{aligned}$$

This means the gradient of $\ell_c(x, \lambda^*)$ would vanish at the solution x^*. In this case, the augmented Lagrangian is seen to be an exact penalty function.

Properties

Property: Assume that the second-order sufficient conditions for a local minimum are satisfied at (x^*, λ^*). Then there exists a $c^* > 0$ such that the augmented Lagrangian $\ell_c(x, \lambda^*)$ has a local minimum at x^* for any $c \geq c^*$.

Note that (P) and $(\bar{P})$ are equivalent. Assume that the Lagrange multiplier vector for (P) is λ^*. If λ_k is chosen for $(\bar{P})$, its Lagrange multiplier becomes $\lambda^* - \lambda_k$. Remember that $c\, h(x^k)$ is approximately equal to the multiplier vector of $(\bar{P})$ when λ_k is used to find x^k that minimizes

$$f(x) + \lambda_k^T h(x) + \frac{1}{2}\|h(x)\|^2$$

in the penalty function method using $P_3(x) \triangleq \frac{1}{2}\|h(x)\|^2$.

Hence

$$c\, h(x^k) \approx \lambda^* - \lambda_k$$

or

$$\lambda^* \approx \lambda_k + c\, h(x^k).$$

General scheme

Step 1: Set $k = 0$ and start with a vector $\lambda_0 \in E^m$.

Step 2: Find $x^k \in E^n$ to minimize

$$f(x) + \lambda_k^T h(x) + \frac{1}{2} c \|h(x)\|^2.$$

Step 3: Update $\lambda_{k+1} = \lambda_k + c\, h(x^k)$.

Step 4: While not converging, increase k by 1 and return to Step 2.

(The major task is to find x^k.)

Duality viewpoints

For a fixed $c > 0$, $\ell_c(x, \lambda)$ corresponds to the Lagrangian for the problem

$$(\bar{\bar{P}}) \quad \begin{array}{ll} \text{Minimize} & f(x) + \frac{1}{2}c\|h(x)\|^2 \\ \text{s. t.} & h(x) = 0 \end{array}$$

1. (P) and $(\bar{\bar{P}})$ are equivalent.

2. The term $\frac{1}{2}c\|h(x)\|^2$ tends to "convexify" the Lagrangian. For sufficiently large c, the Lagrangian will indeed be locally convex.

3. Consider the (local) dual problem of $(\bar{\bar{P}})$.

$$\phi(\lambda) \triangleq \min\left\{f(x) + \frac{1}{2}c\|h(x)\|^2 + \lambda^T h(x)\right\}$$

in a region near (x^*, λ^*).

If $x(\lambda)$ is a solution to the r.h.s., then $h\big(x(\lambda)\big) = \nabla\phi(\lambda)$.

4. To maximize $\phi(\lambda)$, we consider

$$\begin{aligned} \lambda_{k+1} &= \lambda_k + c\nabla\phi(\lambda) \\ &= \lambda_k + c\, h\big(x(\lambda_k)\big) \end{aligned}$$

which is used in Step 3 $\big(\text{with } x^k = x(\lambda_k)\big)$ of a general augmented Lagrangian method.

This is a simple steepest ascent method using a constant step-length c.

Augmented Lagrangian method

- Case 2 : Inequality constraints

$$\begin{aligned} &\text{Minimize} \quad f(x) \\ &\text{s. t.} \quad g(x) \le 0 \end{aligned}$$

We consider an equivalent formulation with equality constraints:

$$\begin{aligned} &\text{Minimize} \quad f(x) \\ &\text{s. t.} \quad g_j(x) + z_j^2 = 0 \ , \quad j = 1, 2, \cdots, p. \end{aligned}$$

Using $P_3(x)$ as the penalty function for augmented Lagrangian,

$$\phi(\mu) \triangleq \min_{x,z} \left\{ f(x) + \sum_{j=1}^{p} \{ \ \mu_j [\ g_j(x) + z_j^2 \ ] + \frac{1}{2} c |g_j(x) + z_j^2|^2 \ \} \right\}.$$

Let $v_j = z_j^2$, then

$$\phi(\mu) \triangleq \min_{v \ge 0, x} \left\{ f(x) + \mu^T [\ g(x) + v \ ] + \tfrac{1}{2} c \|g(x) + v\|^2 \right\}.$$

Taking care of the minimization w.r.t. $v \ge 0$ first, we have

$$\phi(\mu) = \min_{x} \left\{ f(x) + \sum_{j=1}^{p} P_c(\ g_j(x), \mu_j \ ) \right\}$$

where $P_c(t, \mu) \triangleq \frac{1}{2c} \left\{ \left[\max\{0, \mu + ct\} \right]^2 - \mu^2 \right\}.$

Then the general augmented Lagrangian scheme works accordingly.

Cutting plane method

Standard Form:

$$\begin{array}{ll} \text{Minimize} & f(x) = c^T x \\ \text{s. t.} & x \in S \end{array}$$

where $S \subset E^n$ is closed and convex.

– A convex programming problem

$$\begin{array}{ll} \text{Minimize} & g(v) \quad (\text{convex}) \\ \text{s. t.} & v \in V \quad (\text{closed, convex}) \end{array}$$

is equivalent to

$$\begin{array}{ll} \text{Minimize} & r \\ \text{s. t.} & g(v) \leq r \\ & v \in V \end{array}$$

We can define

$$x \triangleq (v, r)$$

$$f(v, r) \triangleq r$$

$$S \triangleq \left\{ (v, r) \mid v \in V, \; g(v) \leq r \right\}$$

to get the standard form.

General scheme

Step 0. Start with a polytope $P_0 \supset S$ and set $k = 0$.

Step 1. Find $x^k \triangleq arg\ \min\{\ c^T x \mid x \in P_k\ \}$.

If $x^k \in S$ output $x^* \triangleq x^k$ and STOP!

Step 2. Find a separating hyperplane H_k with $a_k \in E^n$ and $b_k \in E^1$ such that

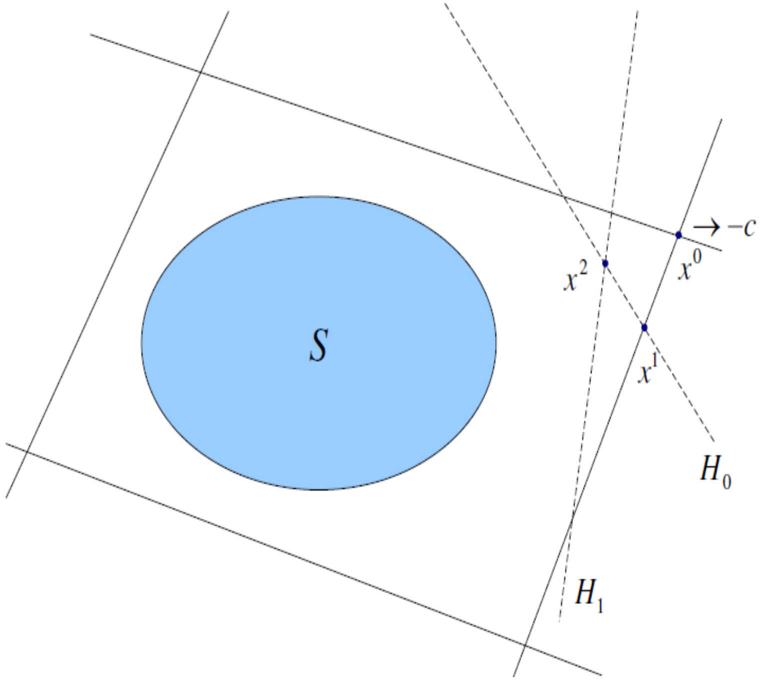
$$S \subset \{\ x \mid a_k^T x \le b_k\ \} \text{ and } x^k \in \{\ x \mid a_k^T x > b_k\ \}.$$

Update $P_{k+1} \leftarrow P_k \cap \{\ x \mid a_k^T x \le b_k\ \}$

and $k \leftarrow k + 1$.

Return to Step 1.

illustration:



Major Task:

Effectively generate "deep" cuts.

Dual interpretation

- Consider the dual problem

$$\begin{array}{ll} \text{Maximize} & \sum_{i \in I} b_i \mu_i \\ \text{s. t.} & \sum_{i \in I} \mu_i a_i = c \\ & \mu_i \geq 0, \; i \in I \end{array}$$

- Minimizing $c^T x$ over a polytope

$$P_k \triangleq \{x \mid a_i^T x \leq b_i, \; i \in I_k\}$$

with $|I_k| < +\infty$ yields x^k with a subset of active constraints $\bar{I}_k$.

The corresponding dual problem with the additional restrictions that $\lambda_i = 0$ for $i \notin \bar{I}_k$ will have a feasible solution, but not necessarily optimal.

Hence, the cutting plane methods work to find a dual optimal solution in each iteration.

Example

- Kelley's Convex Cutting Plane Algorithm:

Consider the following problem

$$\begin{aligned} &\text{Minimize} \quad c^T x \\ &\text{s. t.} \quad g_j(x) \le 0\ , \quad j = 1, 2, \cdots, p \end{aligned}$$

where g_j is convex differentiable.

Let $S \triangleq \{x \mid g(x) \le 0\}$ and $P_0 \supset S$ be an initial polytope such that $c^T x$ is bounded on P_0.

Notice that if $\nabla g_{\bar{j}}(x^k) = 0$, then $S = \phi$.

Apply the general scheme.

If $g(x^k) \le 0$, x^k is optimal.

Otherwise, let $\bar{j}$ be the index maximizing $g_j(x)$.

Then $g_{\bar{j}}(x^k) > 0$.

Notice that for a convex differentiable function $g_{\bar{j}}$,
$g_{\bar{j}}(x) \ge g_{\bar{j}}(x^k) + \nabla g_{\bar{j}}(x^k)\ (x - x^k),\ \forall\ x.$

Hence, we know

$$S \subset \{x \mid g_{\bar{j}}(x^k) + \nabla g_{\bar{j}}(x^k)\ (x - x^k) \le 0\}.$$

In other words,

$$P_{k+1} = P_k \cap \{x \mid g_{\bar{j}}(x^k) + \nabla g_{\bar{j}}(x^k)\ (x - x^k) \le 0\}.$$

Convergence theorem

Let $g_j, \; j = 1, 2, \cdots, p$, be C^1 and convex and Kelley's algorithm generates a sequence of solutions $\{x^k\}$.

Then any limit point of this sequence is a solution to the original problem.

Example

- Supporting Hyperplane Algorithm:

$$\text{Minimize} \quad c^T x$$

$$\text{s. t.} \quad g_j(x) \leq 0 \,, \quad j = 1, 2, \cdots, p$$

Assumptions:

1. $g_j \in C^1$, not necessarily convex.
2. $S \triangleq \{x \mid g_j(x) \leq 0, \ j = 1, 2, \cdots, p\}$ is convex.
3. $\exists \ \bar{x}$ such that $g(\bar{x}) < 0$.
4. $\nabla g_j(x) \neq 0 \ \text{ on } \ \{x \mid g_j(x) = 0\}$.

Apply the general scheme.

1. Let $x^k = arg \ \min\{ \, c^T x \mid x \in P_k \, \}$.

   If $x^k \in S$, STOP!

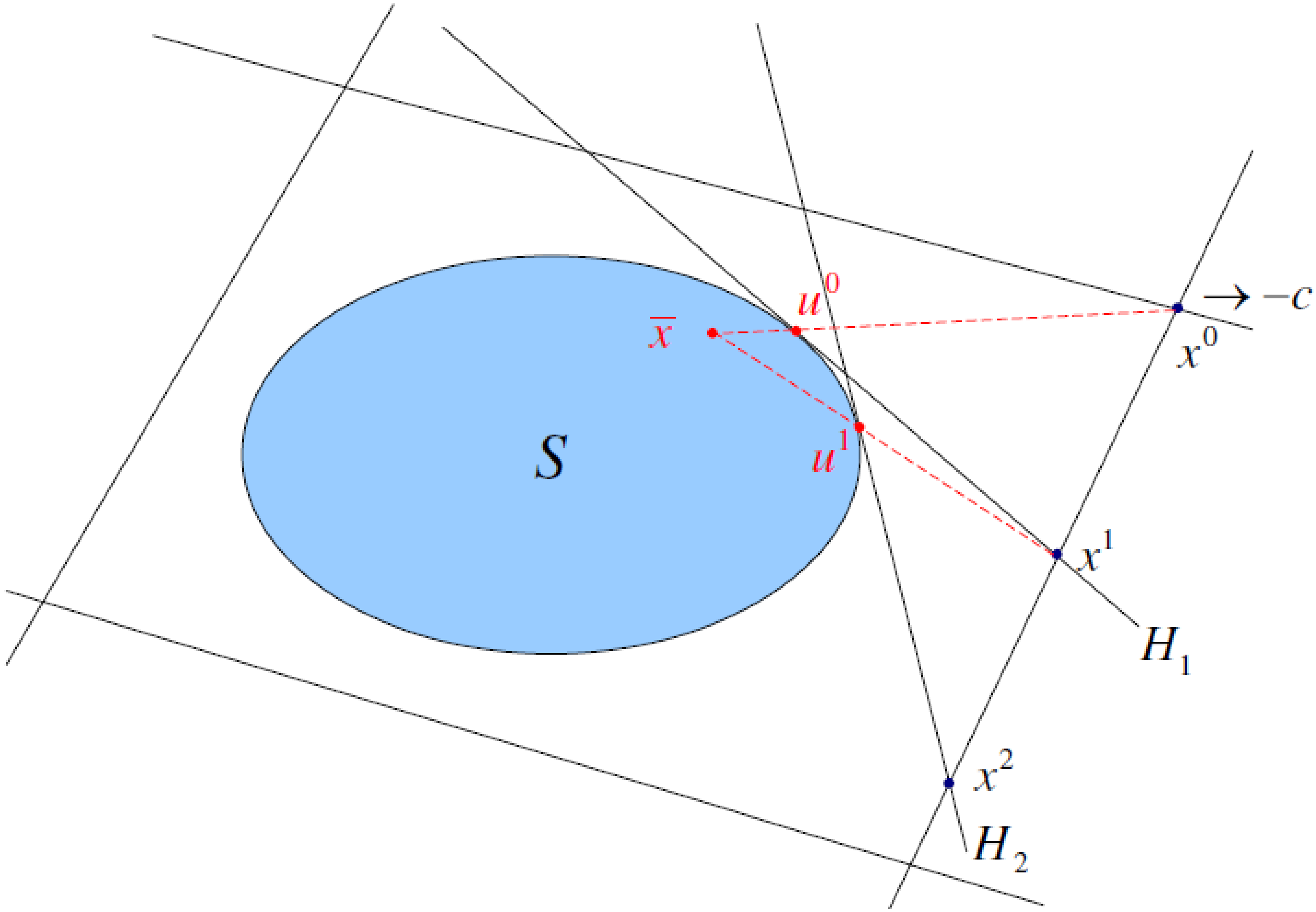
2. Otherwise, find $u^k \in bdry(S) \cap$ the line joining $\bar{x}$ and x^k.

   Let $\bar{j}$ be an index such that $g_{\bar{j}}(u) = 0$.

   Update

$$P_{k+1} \leftarrow P_k \cap \{ \, x \mid \nabla g_{\bar{j}}(u) \, (x - u) \leq 0 \, \}.$$

Illustration



V. Primal-dual approach

- Lagrange method

Basic idea: Directly solve the Lagrange first order necessary condition.

Given

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{s. t.} & h(x) = 0 \end{array}$$

Consider

$n+m$ variables for $n+m$ equations.

$$\begin{cases} \nabla f(x) + \lambda^T \nabla h(x) & = 0 \\ h(x) & = 0 \end{cases}$$

Newton/Modified Newton/Quasi Newton Methods.