MS-E2139 Nonlinear Programming

Lecture slides

Spring 2017

December 28, 2016

Kimmo Berg

Contents

1 Introduction

Practicalities

- teaching: 4h lectures, 4h exercises per week
- exam (24/30p), assignments (2x4p)
- extra points from homework (3p) and exercises (2p)
- (voluntary programming assignment)
- textbook

History and Applications of optimization
(see the course website)

- Dido, Kepler, Newton, Gauss, Dantzig, Stigler, Karmarkar
- logistics, routing, shape and antenna design, pricing, scheduling
- Markowitz portfolio optimization, diet and Goddard rocket problem

Classes of optimization problems

Nonlinear optimization problem (NLP):

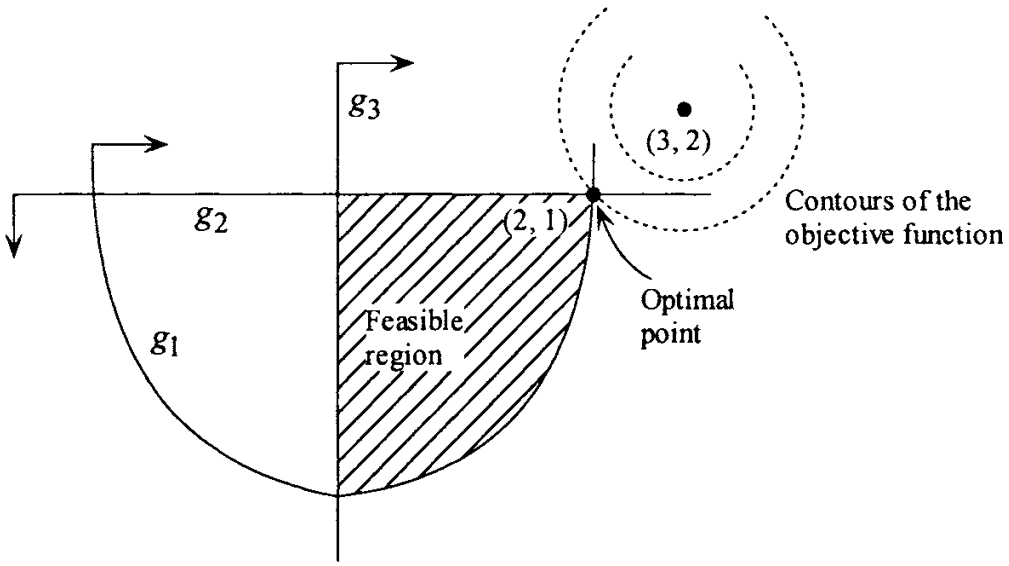
$$\min_{x \in X} \quad f(x)$$

- decision variables $x \in \mathbf{R}^n$
- objective function $f : X \mapsto \mathbf{R}$ or $\mathbf{R}^n \mapsto \mathbf{R}$ (functional), usually continuous and differentiable
- feasible set $X \subset \mathbf{R}^n$
- it can be defined by the constraints: $g_i(x) \leq 0,\ i = 1, \ldots, m$, $h_i(x) = 0,\ i = 1, \ldots, l$, or in matrix form $g(x) \leq \bar{0}$, $g : \mathbf{R}^n \mapsto \mathbf{R}^m$
- if $X = \mathbf{R}^n$ then **unconstrained problem**

Example.

$$\begin{aligned} \min_{x_1, x_2} \quad & (x_1 - 3)^2 + (x_2 - 2)^2 = f(x_1, x_2) = f(x) \\ s.t. \quad & x_1^2 - x_2 - 3 \leq 0, \\ & x_2 - 1 \leq 0, \\ & -x_1 \leq 0. \end{aligned}$$

- *e.g.* $g_1(x) = x_1^2 - x_2 - 3$
- *draw the figure, feasible set, contours/level sets* $\{(x_1, x_2) : f(x_1, x_2) = c \in \mathbf{R}\}$



Linear optimization problem (LP):

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

- linear objective function
- feasible set X polyhedron, g, h linear, $A \in \mathbf{R}^{m \times n}$
- MS-E2140 Linear programming
- MS-E2143 Network optimization (usually LP, transportation)

Definition 1.1. *Function* $f : \mathbf{R}^n \mapsto \mathbf{R}$ *is* **linear** *if*
$f(ax + by) = af(x) + bf(y)$, $f = a^T x$, *or* $f = Ax$.
Function f *is* **affine** *if* $f(x) = L(x) + b$, *where* L *is linear.*
Function f *is* **additive** *if* $f(x + y) = f(x) + f(y)$.
Function f *is* **homogenous of degree k** *if* $f(ax) = a^k f(x)$, $\forall x, a \neq 0$.

Convex optimization problem:

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & g(x) \leq \bar{0}, \quad Ax = b. \end{aligned}$$

- objective $f(x)$ and feasible set $g(x)$ convex
- MS-E2144 Optimization theory
- convexity will be defined shortly

Quadratic programming problem (QP):

$$\begin{aligned} \min \quad & \frac{1}{2}x^T Ax + b^T x \\ s.t. \quad & c^T x \leq d, \quad e^T x = f. \end{aligned}$$

- objective quadratic, feasible set polyhedron (constraints linear)
- if A positive semidefinite then convex QP
- Markowitz portfolio optimization

Definition 1.2. *A matrix Q is* **positive semidefinite** *if $x^T Qx \geq 0$, $\forall x$.*
A matrix S is **positive definite** *if $x^T Sx > 0$, $\forall x \neq \bar{0}$.*
(all eigenvalues are (strictly) positive)

More classes of optimization:

- **stochastic or robust optimization** if f,g,h are not exactly known
- **integer programming** if X discrete (MS-E2146 Integer optimization)
- **dynamic optimization** if $dim\ X = \infty$ (MS-E2148 Dynamic optimization)
- **multicriteria optimization** if multiple objectives (MS-E2153 Multiobjective optimization)

Example

Resource allocation, portfolio, diet problem:

$$\begin{aligned} \max \quad & c^T x \\ s.t. \quad & Ax \leq b, \quad x \geq 0. \end{aligned}$$

- LP, $A \in \mathbf{R}^{m \times n}$, $x \in \mathbf{R}^n$, $b \in \mathbf{R}^m$
- m resources, n activities, x_i level of activity i

- $c_i x_i$ utility from activity i, $f(x) = c^T x = \sum_{i=1}^n c_i x_i$
- activity j with level x_j uses resource i by $a_{ij} x_j$
- total usage given by $Ax = \begin{bmatrix} \sum_{j=1}^n a_{1j} x_j \\ \dots \\ \sum_{j=1}^n a_{mj} x_j \end{bmatrix}$, available resources b

Stochastic problem if c_i stochastic variable, c stochastic vector with expected value $\bar{c}$ and covariance $V = V_{ij} = E\left[(c_i - \bar{c}_i)(c_j - \bar{c}_j)\right]$.
we get portfolio problem: invest b so that $Ax \leq b$ and multiple objectives

$$\begin{aligned} \max \quad & \bar{c}^T x, \text{ expected profit} \\ \min \quad & x^T V x, \text{ variance (risk)} \end{aligned}$$

Assume that the decision maker has utility function for the profit z, $u(z) = 1 - e^{-kz}$, where k is the risk aversion parameter. Also, assume that the profit $z = \bar{c}^T x$ is normally distributed with variance $\sigma^2 = x^T V x$, then $\max E\left[u(z)\right]$ is equivalent to (under monotonic transformation)

$$\begin{aligned} \max \quad & \bar{c}^T x - \frac{1}{2} k x^T V x \\ s.t. \quad & Ax \leq b, \quad x \geq 0. \end{aligned}$$

QP problem, Markowitz
with different k different solutions (draw a figure)
Pareto efficient solutions

What optimization studies?

1. Modeling: assumptions, simplifications, choices for functions
2. Optimization theory: existence, uniqueness, characterization with optimality conditions (local, global, necessary, sufficient, geometric), duality
3. Computation: methods, complexity

Optimality conditions

Definition 1.3. *$x^* \in S$ is a* **global minimum** *if $f(x) \geq f(x^*)$, $\forall x \in S$. (strict if $f(x) > f(x^*)$, $\forall x \in S$, $x \neq x^*$)*

Definition 1.4. *$x^* \in S$ is a* **local minimum** *if $\exists \epsilon > 0$ s.t. $f(x) \geq f(x^*)$, $\forall x \in N_\epsilon(x^*) \cap S$.*

Definition 1.5.

- *int* $S = \{x \mid \exists \epsilon > 0, N_\epsilon(x) \subset S\}$,
- $N_\epsilon(x) = \{y \in \mathbf{R}^n \mid \ \|x - y\| < \epsilon\}, \ \epsilon > 0$,
- *cl* $S = \{x \mid \forall \epsilon > 0, S \cap N_\epsilon(x) \neq \emptyset\}$,
- $\partial S = \{x \mid S \cap N_\epsilon(x) \neq \emptyset, S^C \cap N_\epsilon(x) \neq \emptyset, \forall \epsilon > 0\}$, *where S^C is the* **complement** *of S,*

Note: S is **open** if $S = int\ S$ and S is **closed** if $S = cl\ S$.
Directional derivatives and differentiability

Definition 1.6. *Let $S \subset \mathbf{R}^n$, $S \neq \emptyset$, $f : S \mapsto \mathbf{R}$, $x_0 \in S$ and direction $d \neq \bar{0}$ s.t. $x_0 + \lambda d \in S$, $\forall \lambda \in [0, \lambda_0]$ for some $\lambda_0 > 0$.* **Gateaux derivative** *of f at x_0 in direction d is (when the limit exists)*

$$f'(x_0; d) = \lim_{\lambda \to 0^+} \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}.$$

Definition 1.7. *Function $f : S \mapsto \mathbf{R}$, $S \subset \mathbf{R}^n$ is* **Frechet differentiable** *at $x_0 \in int\ S \neq \emptyset$ if $\exists \nabla f(x_0) \in \mathbf{R}^n$ (***gradient***) and a function $\alpha : \mathbf{R}^n \mapsto \mathbf{R}$ s.t.*

$$f(x) = f(x_0) + \nabla f(x_0)^T (x - x_0) + \|x - x_0\| \alpha(x_0; x - x_0), \ \forall x \in S,$$

where $\alpha(x_0; x - x_0) \to 0$ when $x \to x_0$.

If a function if Frechet differentiable it has all Gateaux derivatives and they equal ($f'(x_0; d) = \nabla f(x)^t d$). If a function is differentiable it is also continuous. The gradient $\nabla f(x)$ is unique and $\nabla f(x) = [\partial f(x)/\partial x_1, \ldots, \partial f(x)/\partial x_n]$. If $f : \mathbf{R}^n \mapsto \mathbf{R}^l$, $f(x) = (f_1, \ldots, f_l)^T$ then the **Jacobian** is
$\nabla f(x) = \left[\nabla f_1(x)^T, \ldots, \nabla f_l(x)^T\right]$.

Example. $\nabla(x^T A x)$.

Definition 1.8. *Function $f : S \subset \mathbf{R}^n, \neq \emptyset \mapsto \mathbf{R}$ is* **twice differentiable** *at $x_0 \in int\ S$ if $\exists \nabla f(x_0) \in \mathbf{R}^n$, a symmetric $n \times n$* **Hessian** *matrix $\nabla^2 f(x_0) \in \mathbf{R}^{n \times n}$ and $\alpha : \mathbf{R}^n \mapsto \mathbf{R}$ s.t. $\alpha(x_0; x - x_0) \to 0$ when $x \to x_0$ and $\forall x \in S$*

$f(x) = f(x_0) + \nabla f(x_0)^T (x - x_0) + 1/2 (x - x_0)^T \nabla^2 f(x_0)(x - x_0) + \|x - x_0\|^2 \alpha(x_0; x - x_0)$.

Note that $(\nabla^2 f(x_0))_{ij} = \frac{\partial^2 f(x_0)}{\partial x_i \partial x_j}$.

Unconstrained optimization

Definition 1.9. *$d \in \mathbf{R}^n$ is a* **descent direction** *of f at x' if $\exists \delta > 0$ s.t. $f(x' + \lambda d) < f(x')$, $\forall \lambda \in (0, \delta)$. The cone of descent directions is $d \in F$.*

Theorem (4.1.2)**.** *If $f : \mathbf{R}^n \mapsto \mathbf{R}$ differentiable at x' then*
$F_0 = \{d, \nabla f(x')^T d < 0\} \subset F$.

Proof. Diff.: $(f(x' + \lambda d) - f(x'))/\lambda = \nabla f(x')^T d + \|d\| \alpha(x'; \lambda d) \Rightarrow \exists \delta > 0$ s.t. $f(x' + \lambda d) - f(x') < 0$, $\forall \lambda (0, \delta)$ since $\alpha(x'; \lambda d) \to 0$ when taking the limit $\lambda \to 0$. So $d \in F$. □

Theorem (Fermat 1646, first order necessary)**.** *Let $f : \mathbf{R}^n \mapsto \mathbf{R}$ diff. ($S \neq \emptyset$ open or $x^* \in$ int S). If x^* is a local optimum then $\nabla f(x^*) = \bar{0}$.*

Proof. Assume $\nabla f(x^*) \neq \bar{0}$. Choose $d = -\nabla f(x^*) \Rightarrow -\|\nabla f(x^*)\|^2 < 0$, i.e., $d \in F_0$. From Theorem 4.1.2., $\exists \lambda > 0$ and $d \in F$ s.t. $f(x^* + \lambda d) < f(x^*)$, which is a contradiction of local optimality. There cannot be descent directions at local minima. □

Theorem (4.1.3, second order necessary)**.** *Let $f : \mathbf{R}^n \mapsto \mathbf{R}$ twice diff. If x^* is a local minimum then $\nabla^2 f(x^*)$ is positive semidefinite.*

Proof. As before but use the second order Taylor explansion instead of the first. □

Theorem (4.1.4, sufficient)**.** *Let $f : \mathbf{R}^n \mapsto \mathbf{R}$ twice diff. If $\nabla f(x^*) = \bar{0}$ and $\nabla^2 f(x^*)$ positive definite then x^* is a strict local minimum.*

2 Convex sets

Definition 2.1. *A set $S \in \mathbf{R}^n$ is* **convex** *if for all $x_1, x_2 \in S$ holds that*
$\lambda x_1 + (1 - \lambda) x_2 \in S$, $\forall \lambda \in (0, 1)$.

Example. *The following sets are convex:*

- *hyperplanes $S = \{x \in \mathbf{R}^n \mid p^T x = a\}$,*
- *open and closed half-spaces $S = \{x \in \mathbf{R}^n \mid p^T x < a\}$ $(\leq)$,*

- *polyhedra* $P = \{x \mid Ax \leq b, Cx = d\}$,
- *norm balls* $B = \{x \mid \|x - x_c\| \leq r\}$, *where* $\|x\| = \sqrt{x \cdot x} = \sqrt{< x, x >} = \sqrt{x^T x} = \sqrt{\sum_{j=1}^{n} x_i^2}$,
- *norm cones* $C = \{(x, t) \mid \|x\| \leq t\} \in \mathbf{R}^{n+1}$,
- *ellipsoids* $E = \{x \mid x^T Q x + p^T x + q \leq 0,\ Q\ p.s.d.\}$

Definition 2.2. *Set C is a* **cone** *from origin if $x \in C \Rightarrow \lambda x \in C$, $\forall \lambda \geq 0$.*
The **dual cone** *of C is $C^* = \{y \mid y^T x \geq 0$ for all $x \in C\}$. The* **polar cone** *of C is $C^0 = \{y \mid y^T x \leq 0$ for all $x \in C\}$.*

Definition 2.3. *The weighted averages $\sum_{j=1}^{k} \lambda_j x_j$ of points $x_1, \ldots, x_k$ are called:*

- **linear combinations** *when $\lambda_j \in \mathbf{R}$,*
- **affine combinations** *when $\sum_{j=1}^{k} \lambda_j = 1$,*
- **conical combinations** *when $\lambda_j \geq 0$,*
- **convex combinations** *when $\sum_{j=1}^{k} \lambda_j = 1$, $\lambda_j \geq 0$.*
- *$conv(S)$ is the set of its convex combinations*
  (shown in exercises: $conv(S) = \bigcap\{C \subset X : C\ conv., S \subset C\}$)
- *Note S is convex if $S = conv(S)$.*

Theorem (2.1.6, Caratheodory)**.** *If $S \subset \mathbf{R}^n$ and $x \in conv(S)$ then $x \in conv(x_1, \ldots, x_{n+1})$.*

Convexity preserving operations for sets

- Theorem 2.1.2: $\bigcap$, $\pm$
  where $A \pm B = \{x \pm y \mid x \in A, y \in B\}$,
- affine functions $f = Ax + b$: scaling, translation, (image and inverse image)
- cartesian product $\times$: $S_1 \times S_2 = \{(x_1, x_2) \mid x_1 \in S_1, x_2 \in S_2\}$,
- perspective functions $P(x, t) = \frac{x}{t}$, $t > 0$,
- linear-fractional functions $g(x) = \frac{Ax+b}{c^T x+d}$, dom $g = \{x \mid c^T x + d > 0\}$,

- interior *int*, closure *cl*, convex hull *conv*.

How to examine if a set is convex?

- proof based on the definition
- using earlier results and convexity preserving operations
- draw a figure
- simulation: numeric testing by choosing points in random and test convexity (proving the set is not convex)

Existence

Definition 2.4. Infimum $\alpha = \inf_{x \in S} f(x)$ *if* $\alpha \leq f(x)\ \forall x \in S$ *and* $\nexists \alpha_0 > \alpha$ *s.t.* $\alpha_0 \leq f(x)\ \forall x \in S$, *and* **minimum** $\alpha = \min_{x \in S} f(x)$ *if* $\exists x^* \in S$ *s.t.* $\alpha = f(x^*) \leq f(x)\ \forall x \in S$.

Note the axiom of real numbers: if $A \neq \emptyset \subset \mathbf{R}$ and $\exists M$ s.t. $x \leq M\ \forall x \in A$ then $\exists \sup A$.

Theorem (2.3.1, Weierstrass)**.** *If* $S \neq \emptyset \subset \mathbf{R}^n$ *compact (closed and bounded) and* $f : S \mapsto \mathbf{R}$ *(lower semi)continuous then* $\exists x^* \in S$ *s.t.* $f(x^*) = \min_{x \in S} f(x) = \inf_{x \in S} f(x)$.

Minimum distance from a convex set

Theorem (2.4.1)**.** *If* $S \neq \emptyset \subset \mathbf{R}^n$ *closed convex and* $y \notin S$ *then* $\exists! x^* \in S$ *s.t.*

$$\|x^* - y\| = \min_{x \in S} \|x - y\| = \inf_{x \in S} \|x - y\|.$$

and x^* *is the minimum* $\Leftrightarrow (y - x^*)^T (x - x^*) \leq 0,\ \forall x \in S$.

Proof. Existence. $S \neq \emptyset \Rightarrow \exists x' \in S$, $S_0 = S \cap \{x, \|x - y\| \leq \|y - x'\|\}$ is compact. $f : S_0 \mapsto \mathbf{R}$ continuous, Weierstrass.

Uniqueness. Assume $\exists x' \in S$ s.t. $\|y - x^*\| = \|y - x'\| = \gamma$. Since S is convex, $1/2(x^* + x') \in S$. Now,

$$\|y - 1/2(x^* + x')\| = \|1/2(y - x^*) + 1/2(y - x')\| \leq 1/2\|y - x^*\| + 1/2\|y - x'\| = \gamma$$

by triangle inequality, and it is a contradiction.

If part. $\|y-x\|^2 = \|y-x^*+x^*-x\|^2 = \|y-x^*\|+\|x^*-x\|^2+2(x^*-x)^T(y-x^*) \geq \|y-x^*\|^2$.

Only if part. x^* is a minimum, i.e., $\|y-x\|^2 \geq \|y-x^*\|^2\ \forall x \in S$. Since S convex, $x \in S \Rightarrow x^* + \lambda(x-x^*) \in S,\ \forall \lambda\,[0,1]$. $\|y-x^*-\lambda(x-x^*)\|^2 \geq \|y-x^*\|^2,\ \forall \lambda\,[0,1]$. $\|y-x^*-\lambda(x-x^*)\|^2 = \|y-x^*\|^2 + \lambda^2\|x-x^*\|^2 - 2\lambda(y-x^*)^T(x-x^*)$. Thus, $\lambda^2\|x-x^*\|^2 \geq 2\lambda(y-x^*)^T(x-x^*),\ \forall \lambda\,[0,1]$. Assume $\lambda > 0 \Rightarrow \lambda\|x-x^*\|^2 \to 0$ when $\lambda \to 0$. When $\lambda \to 0$ then $2(y-x^*)^T(x-x^*) \leq 0$. □

Application: Let $x_1, \ldots, x_m \in \mathbf{R}^n$ linearly independent and $y \in \mathbf{R}^n$.

$$\min_{\alpha_i \geq 0} \|y - \sum_{i=1}^{m} \alpha_i x_i\|,$$

where $S = \{x \mid x = \sum_{i=1}^m \alpha_i x_i,\ \alpha_i \geq 0\}$ is closed and convex cone. Thus, there is a unique minimizer and $(y - \sum \alpha_i^* x_i)^T(\sum x_i(\alpha_i - \alpha_i^*)) \leq 0$. This implies $(y - \sum \alpha_i^* x_i)^T x_i \leq 0$ and $=$ when $\alpha_i^* > 0$. We get so called normal equations

$$\begin{aligned} A\alpha^* - b &= z, \\ z &\geq \bar{0}, \\ \alpha^{*T} z &= 0. \end{aligned}$$

where $b_i = y^T x_i$ and Gram matrix

$$A = \begin{bmatrix} x_1^T x_1 & \ldots & x_1^T x_n \\ \vdots & & \vdots \\ x_n^T x_1 & \ldots & x_n^T x_n \end{bmatrix}.$$

Definition 2.5. *Let $S_1, S_2 \neq \emptyset \in \mathbf{R}^n$. A hyperplane H* **separates** *sets S_1 and S_2 if $S_1 \subset H^+ = \{x \mid p^T x \geq \alpha\}$ and $S_2 \subset H^- = \{x \mid p^T x \leq \alpha\}$.*
If also $S_1 \cup S_2 \notin H$ then H separates **properly**.
Separation is **strict** *if $S_1 \subset int\ H^+$ and $S_2 \subset int\ H^-$. If the sets are open then the separation is strict.*
Separation is **strong** *if $\exists \epsilon > 0$ s.t. $S_1 \subset \{x \mid p^T x \geq \alpha + \epsilon\}$ and $S_2 \subset H^-$.*

Theorem (2.4.4, point and set)**.** *If $S \neq \emptyset \mathbf{R}^n$ closed, convex and $y \notin S$ then $\exists p \in \mathbf{R}^n$, $p \neq \bar{0}$ and $\alpha \in \mathbf{R}$ s.t. $p^T y > \alpha$, $p^T x \leq \alpha$, $\forall x \in S$.*

Proof. From Theorem 2.4.1, $\exists! x^* \in S$ s.t. $(x-x^*)^T(y-x) \leq 0$, $\forall x \in S$. $0 < \|y-x^*\|^2 = y^T(y-x^*) - x^{*T}(y-x^*) = p^T y - \alpha$, where $p = y - x^* \neq \bar{0}$ and $p^T x^* = \alpha$. Substituting p, $p^T(x-x^*) \leq 0 \Leftrightarrow p^T x \leq \alpha$. □

Note the connection to Hahn-Banach separation theorem.

Corollary. If $S \in \mathbf{R}^n$ closed, convex then $S = \bigcap_{S \subset H^-} H^-$, where H^- half-spaces.

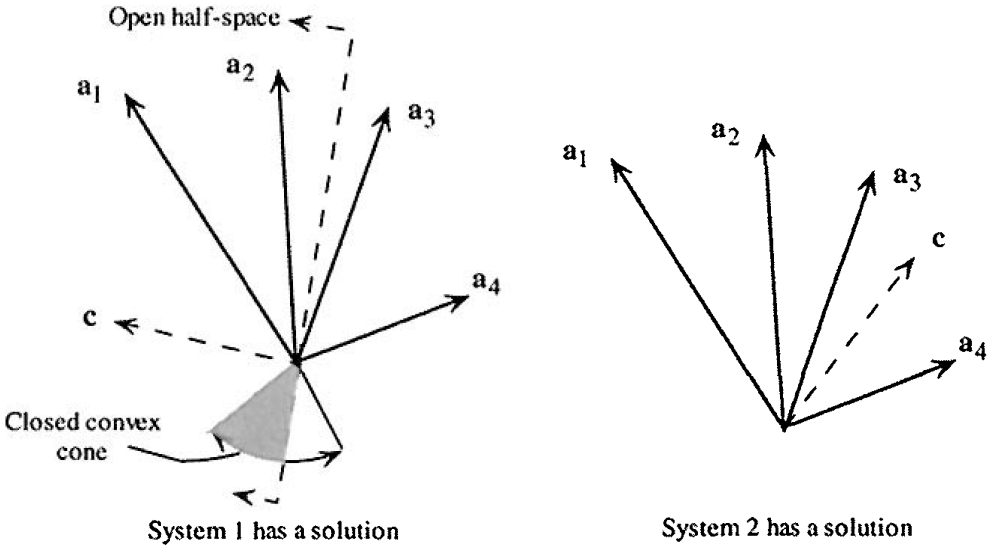
Theorem (2.4.5, Farkas)**.** *Let $A \in \mathbf{R}^{m\times n}$ and $c \in \mathbf{R}^n$. Exactly one system has a solution:*

(1) $Ax \leq \bar{0}$, $c^T x > 0$, *for some* $x \in \mathbf{R}^n$,

(2) $A^T y = c$, *for some* $y \geq \bar{0}$, $y \in \mathbf{R}^m$.

Proof. Assume (2) has a solution. Assume $Ax \leq \bar{0} \Rightarrow c^T x = x^T A^T y = (Ax)^T y \leq \bar{0}$, so (1) does not have a solution.

Assume (2) does not have a solution. Let $S = \{x' = A^T y, y \geq \bar{0}\}$ closed, convex and $c \notin S$. By Theorem 2.4.4, $\exists p \in \mathbf{R}^n$ s.t. $p^T c > \alpha$, $p^T x \leq \alpha$, $\forall x \in S$. Especially, $\bar{0} \in S \Rightarrow \alpha \geq 0 \Rightarrow p^T c > 0$. When $x \in S$, $p^T x = p^T (A^T y) = y^T (Ap) \leq \alpha$, $\forall y \geq \bar{0}$. y can be chosen arbitrarily large $\Rightarrow Ap \leq \bar{0}$. So p solves (1). □



Theorem (2.4.9, Gordan)**.** *Let $A \in \mathbf{R}^{m\times n}$. Exactly one system has a solution:*

(1) $Ax < \bar{0}$, *for some* $x \in \mathbf{R}^n$,

(2) $A^T y = \bar{0}$, *for some* $y \geq \bar{0}$, $y \neq \bar{0} \in \mathbf{R}^m$.

Proof. $Ax < \bar{0} \Leftrightarrow Ax + es \leq \bar{0}$. Choose in Farkas $A' = \begin{bmatrix} A \\ e \end{bmatrix}$, where $e = [1 \dots 1]^T \in \mathbf{R}^n$ and $s > 0$. Farkas and Gordan systems are equal by choosing

$x' = [x\ s]^T$ and $c' = [0 \dots 0\ 1]$:

$$[A\ \ e] \begin{bmatrix} x \\ s \end{bmatrix} \leq \bar{0} \quad , \quad [0 \dots 0\ 1] \begin{bmatrix} x \\ s \end{bmatrix} > 0,$$

$$\begin{bmatrix} A^T \\ e^T \end{bmatrix} y = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},\ y \geq \bar{0} \quad \Leftrightarrow \quad A^T y = \bar{0}, e^T y = 1, \text{ for some } y \geq 0$$

$$\Leftrightarrow \quad A^T y = \bar{0}, y \geq \bar{0}, y \neq \bar{0}.$$

□

Theorem (Motzkin)**.** *Let $A_1 \in \mathbf{R}^{m \times n}$ and $A_2 \in \mathbf{R}^{l \times n}$. Exactly one system has a solution:*

(1) $A_1 d < \bar{0}$, $A_2 d = \bar{0}$, *for some* $d \in \mathbf{R}^n$, $(A_3 d \leq \bar{0})$,

(2) $A_1^T y_1 + A_2^T y_2 (+A_3 y_3) = \bar{0}$, $y_1 \geq \bar{0}$, $y_1 \neq \bar{0} \in \mathbf{R}^m$, $y_2 \in \mathbf{R}^l$. $(y_3 \geq \bar{0})$

3 Convex functions and subgradients

Definition 3.1. *A function $f : S \mapsto \mathbf{R}$, $S \subseteq \mathbf{R}^n$, $S \neq \emptyset$ convex set, is (strictly)* **convex** *in set S if for all $x_1, x_2 \in S$, $\lambda \in (0, 1)$ holds that*
$f(\lambda x_1 + (1 - \lambda) x_2) \leq \lambda f(x_1) + (1 - \lambda) f(x_2)$ $(<$ *when* $x_1 \neq x_2)$.

Example. *The following functions are convex:*

- *affine* $f(x) = p^T x + a$
- *norms* $\|x\|$
- *pos.sem.def. quadratic functions* $f(x) = x^T A x + b^T x + c$, *A p.s.d*
- $\exp(ax)$, $a \in \mathbf{R}$
- x^a, $x > 0$, $a \geq 1$ *or* $a \leq 0$
- $-x^a$, $x > 0$, $0 \leq a \leq 1$
- $-\log(x)$ *or* $x \log(x)$, $x > 0$,

Theorem (one dimensional property)**.** *f convex $\Leftrightarrow$ $g(t) = f(x+tv)$ convex $\forall x \in$ dom f, $v \in \mathbf{R}^n$.*

Convexity preserving operations for functions

- non-negative weighted sum $g = w_1 f_1 + \ldots + w_m f_m$, $w_i \geq 0$
- affine scaling $g(x) = f(Ax + b)$, $\operatorname{dom} g = \{x \mid Ax + b \in \operatorname{dom} f\}$
- pointwise maximum $g(x) = \max\{f_1(x), \ldots, f_m(x)\}$, $\operatorname{dom} g = \bigcap \operatorname{dom} f_i$
- over infinite set $g(x) = \sup_{y \in A} f(x, y)$
- composition $g(x) = f(h(x))$ if f convex, non-decreasing (non-increasing) and h convex (concave)
- minimization $g(x) = \inf_{y \in C} f(x, y)$, $C \neq \emptyset$ convex

Example. *These operations can for example be applied in*

- $f(x) = -\sum_{i=1}^m \log(b_i - a_i^T x)$*, when* $(b_i - a_i^T x > 0)$ *(sum, log, affine)*
- $f(x) = x_{[1]} + \ldots + x_{[k]}$ *(sum of k largets components)*
  $=\max\{x_{i_1} + \ldots + x_{i_k} \mid 1 \leq i_1 < \ldots < i_k \leq n\}$ $(n!/(k!(n-k)!$ *combinations, max of linear*)
- $f(X) = \sup\{y^T X y,\ \|y\|_2 = 1\}$ *(maximum eigenvalue, sup of linear)*

Jensen inequality

f convex $\Leftrightarrow$ $f(\sum_{i=1}^k \lambda_i x_i) \leq \sum_{i=1}^k \lambda_i f(x_i)$, $\sum_{i=1}^k \lambda_i = 1$, $\lambda_i \geq 0$, $x_i \in S$

Example. *geometric mean is smaller than arithmetic mean:*

$$(a_1 \cdot \ldots \cdot a_n)^{1/n} \leq (a_1 + \ldots + a_n)/n, \quad a_i > 0$$

Example. *If f convex then $f(Ex) \leq Ef(x)$ (expectation of random variable)*

You can derive other inequalities like Hölder's inequality by applying Jensen inequality to some appropriate functions.

Connection between convex sets and functions

Definition 3.2. Epigraph *(hypograph, hyp) of a function is* *epi* $f = \{(x, y) \mid x \in S, y \geq f(x)\} \subset \mathbf{R}^{n+1}$ $(\leq)$.

Theorem (3.2.2)**.** *If* $S \subseteq \mathbf{R}^n$, $S \neq \emptyset$ *convex,* $f : S \mapsto \mathbf{R}$ *then* f *convex* $\Leftrightarrow$ *epi* f *convex (set)*

Properties of convex functions

Definition 3.3. Lower-level-set *(upper)* $S_\alpha = lev_\alpha f = \{x \in S,\ f(x) \leq \alpha\}$, $\alpha \in \mathbf{R}$ $(\geq)$.

Theorem (3.1.2)**.** *If* $S \subset \mathbf{R}^n$, $S \neq \emptyset$, $f : S \mapsto \mathbf{R}$ *convex then* $lev_\alpha f$ *is convex for all* $\alpha \in \mathbf{R}$.

Note that a function whose all lower-level-sets are convex need not be convex.

Definition 3.4. *Function* f *is* **quasiconvex** *if* $f(\lambda x_1 + (1-\lambda)x_2) \leq \max\{f(x_1), f(x_2)\}$, *for all* $x_1, x_2 \in S$, $\lambda \in (0, 1)$.

Quasiconvexity is **strict** if $(<)$ $\forall f(x_1) \neq f(x_2)$ and **strong** if $(<)$ $\forall x_1 \neq x_2$.

Theorem (3.5.2)**.** f *is quasiconvex* $\Leftrightarrow$ $lev_\alpha f$ *convex* $\forall \alpha \in R$.

Definition 3.5. *Function* f *is* **pseudoconvex** *if* $\forall x_1, x_2 \in S$, $\nabla f(x_1)^T(x_2 - x_1) \geq \bar{0} \Rightarrow f(x_2) \geq f(x_1)$. **Strict** *if* $f(x_2) > f(x_1)$ *when* $x_1 \neq x_2$.

Continuity of convex functions

Definition 3.6. *Limit* $x_n \to \bar{x}$ *means* $\forall \delta > 0, \exists N$ *s.t.* $\forall n > N$, $\|x_n - \bar{x}\| < \epsilon$.

Definition 3.7. *Function* f *is* **continuous** *in* $\bar{x}$ *if* $\forall \epsilon > 0$, $\exists \delta > 0$ *s.t.* $\|x - \bar{x}\| \leq \delta \Rightarrow |f(x) - f(\bar{x})| \leq \epsilon$. $(\forall x_n \to \bar{x} \Rightarrow f(x_n) \to f(x))$

Theorem (3.1.3)**.** *If* $f : S \mapsto \mathbf{R}$ *convex then* f *continuous in int* S.

Legendre-Fenchel conjugate function

Definition 3.8. Convex hull $conv(f) = \sup\{g : S \mapsto \mathbf{R}$ *convex,* $g \leq f\}$.

Definition 3.9. Conjugate function $f^*(y) = \sup_x\{y^T x - f(x)\}$ *(convex, sup of affine)*

Definition 3.10. *Biconjugate* $f^{**} = conv(f)$.

Directional derivatives of convex functions

Theorem (3.1.5)**.** *Let* $S \subset \mathbf{R}^n$, $S \neq \emptyset$ *convex,* $f : S \mapsto \mathbf{R}$ *convex,* $x_0 \in S$ *and* $d \neq \bar{0}$ *s.t.* $x_0 + \lambda d \in S$, $\forall \lambda \in [0, \lambda_0]$ *for some* $\lambda_0 > 0$ *then*
$\exists f'(x_0; d)$ *(possibly* $\pm\infty$*),*
if $x_0 \in int\ S$*, then* $|f'(x_0; d)| < \infty$.

The gradient is a global underestimator with local information and the gradient is monotone.

Theorem (3.3.3 and 3.3.4)**.** *If* $S \neq \emptyset$ *open, convex,* f *differentiable then*

$$f \text{ convex } \Leftrightarrow \begin{array}{l} i)\ f(x) \geq f(x_0) + \nabla f(x_0)^T(x - x_0), \quad \forall x \in S\ (> \textit{strictly}) \\ ii)\ (\nabla f(x_2) - \nabla f(x_1))^T(x_2 - x_1) \geq 0, \quad \forall x_1, x_2 \in S \end{array}$$

Proof. Let us show i): Only if part. Let $x, y \in S$. Since f is convex there is $0 < \lambda \leq 1$

$$\begin{aligned} \frac{f(x + \lambda(y - x)) - f(x)}{\lambda} &\leq f(y) - f(x), \\ \|y - x\| \frac{f(x + \lambda(y - x)) - f(x) - \nabla f(x)^T \lambda(y - x)}{\|\lambda(y - x)\|} + \nabla f(x)^T(y - x) &\leq f(y) - f(x), \\ \Leftrightarrow \nabla f(x)^T(y - x) &\leq f(y) - f(x), \end{aligned}$$

where the first part $\to 0$ when $\lambda \to 0$.

If part. Let $x', y' \in S$, $0 \leq \lambda \leq 1$ and $x = \lambda x' + (1 - \lambda)y'$. From the assumption we get

$$\begin{aligned} f(x') &\geq f(x) + \nabla f(x)^T(x' - x), \\ f(y') &\geq f(x) + \nabla f(x)^T(y' - x). \end{aligned}$$

Multiplying the first by λ and second by $(1 - \lambda)$ and summing

$$\lambda f(x') + (1 - \lambda) f(y') \geq f(x) + \nabla f(x)^T(\lambda x' + (1 - \lambda)y' - (\lambda x' + (1 - \lambda)y')) = f(x).$$

□

Theorem (3.3.7)**.** *Let* $S \neq \emptyset$ *open convex,* $f : S \mapsto \mathbf{R}$ *twice differentiable on* S*. Function* f *is convex if and only if Hessian is positive semidefinite at each point in* S*.*

Theorem (3.3.8)**.** *Let $S \neq \emptyset$ open convex, $f : S \mapsto \mathbf{R}$ twice differentiable on S. If Hessian is positive definite in S then f is strictly convex. If f is strictly convex then Hessian is positive semidefinite in S. (p.s. if quadratic)*

Note. Positive definite Hessian is sufficient for strictly convexity but not necessary. $f(x) = x^4$ is strictly convex even though $f''(0) = 0$ (p.s.d).

How to prove that a function is convex?

- use convex functions and convexity preserving operations
- convexity is one dimensional property
- f' monotonic and non-decreasing
- f'' non-negative
- if f twice differentiable, $\nabla^2 f$ p.s.d. in int S

Supporting hyperplanes

Definition 3.11. *Let $S \neq \emptyset \subset \mathbf{R}^n$ and $x' \in \partial S$.*
H is a **supporting hyperplane** *of S at x' if either $S \subset H^+$ or $S \subset H^-$. If also $S \not\subset H$ then H is* **proper support**.

Note. H supports $S \Leftrightarrow p^T x' = \inf_{x \in S} p^T x$ or $p^T x' = \sup_{x \in S} p^T x$.

Theorem (2.4.7)**.** *If $S \neq \emptyset \subset \mathbf{R}^n$ convex and $x' \in \partial S$ then*
$\exists p \neq \bar{0}$ *s.t.* $p^T(x - x') \leq 0$, $\forall x \in cl\ S$.

Proof. Let us separate the points in closure from the points in interior. When $x \in \partial S \Rightarrow \exists$ sequence y_k, $y_k \in cl\ S$ s.t. $y_k \to x$. Theorem 2.4.4 implies $\forall y_k\ \exists p_k$ s.t. $p_k^T y_k > p_k^T x$, $\forall x \in cl\ S$. Since p_k is bounded, there is subsequence p_{k_i} s.t. $p_{k_i} \to p$, when $i \to \infty$, $\|p\| = 1$. This implies $p^T x' \geq p^T x$, $\forall x \in cl\ S$. (= when $x = x' \in cl\ S$) □

Corollary. S convex, $x' \notin int\ S \Rightarrow \exists p \neq \bar{0}$ s.t. $p^T(x - x') \leq 0$, $\forall x \in cl\ S$. Proof. if $x \notin cl\ S$ Theorem 2.4.4 and if $x \in cl\ S$ Theorem 2.4.7.

Theorem (2.4.8, proper separation)**.** *If $S_1, S_2 \neq \emptyset$ convex, $S_1 \cap S_2 = \emptyset$ then*

$$\exists p \neq \bar{0}\ s.t.\ \inf_{x \in S_1} p^T x \geq \sup_{x \in S_2} p^T x$$

Proof. Let $S = S_1 - S_2$, which is convex. $S_1 \cap S_2 = \emptyset \Rightarrow \bar{0} \notin S$. Let us separate $\bar{0}$ and S by Theorem 2.4.7: $\exists p \neq \bar{0}$ s.t. $p^T x \geq 0$, $\forall x \in S \Leftrightarrow p^T x_1 \geq p^T x_2$, $\forall x_1 \in S_1,\ x_2 \in S_2$. □

Theorem (2.4.10, strong separation)**.** *If S_1, S_2 closed convex, S_1 bounded, $S_1 \cap S_2 = \emptyset$ then*

$$\exists p \neq \bar{0}, \epsilon > 0 \ s.t. \ \inf_{x \in S_1} p^T x \geq \epsilon + \sup_{x \in S_2} p^T x.$$

Proof. Let $S = S_1 - S_2$, which is closed and convex. Use Theorem 2.4.4. □

Subgradients

Definition 3.12. *A vector $\xi \in \mathbf{R}^n$ is a* **subgradient** *of function f at $x' \in S$ if*

$$f(x) \geq f(x') + \xi^T (x - x'), \quad \forall x \in S.$$

$\xi \in \partial f(x')$ denotes the set of subgradients, i.e., the **subdifferential***, at x'.*

It is shown in the exercises that the subdifferential is a convex and closed set.

Example. $f(x) = |x|$, $\partial f(0) = \{\xi, -1 \leq \xi \leq 1\}$. *(unit square in* $\mathbf{R}$*)*

Note that $f = \max_{i=1,\ldots,k} f_i(x)$ typically has solution at a corner.

Theorem. *If f convex, $x_0 \in int \operatorname{dom} f$ then for all $d \in \mathbf{R}^n$*

$$f'(x_0; d) = \sup_{\xi \in \partial f(x_0)} \xi^T d.$$

Theorem (3.2.5)**.** *If $f : S \mapsto \mathbf{R}$ convex, $x' \in int\ S \neq \emptyset$, then $\partial f(x') \neq \emptyset$.*

Proof. From Theorem 3.2.2, *epi* f is convex. From Theorem 2.4.7, $\exists (\xi_0, \mu) \neq (\bar{0}, 0)$, $\xi \in \mathbf{R}^n$, $\mu \in \mathbf{R}$ s.t.

$$\xi_0^t (x - x') + \mu(y - f(x')) \leq 0, \quad \forall (x, y) \in epi\ f,$$

where y can be arbitrarily large, and thus $\mu \leq 0$. If $\mu = 0$ then $\xi_0^T (x - x') \leq 0$, $\forall x \in S$. If $x' \in int\ S$ then $\exists \lambda > 0$ s.t. $x' + \lambda \xi_0 \in S$, $\lambda \xi_0^T \xi_0 \leq 0 \Rightarrow \xi_0 = \bar{0}$. This means that $(\xi_0, \mu) = (\bar{0}, 0)$ which is a contradiction and it should be that $\mu < 0$. Now, we can denote $\xi = -\xi_0/\mu$ and we get

$$\xi^T (x - x') - y + f(x') \leq 0, \quad \forall (x, y) \in epi\ f.$$

So $(-1, \xi)$ is a supporting hyperplane for *epi* f and the above equation means $\xi \in \partial f(x')$ when $y = f(x)$. □

Theorem (3.2.6). *If $f : S \neq \emptyset \mapsto \mathbf{R}$, $\partial f(x) \neq \emptyset$, $\forall x \in int\ S$ then $f : int\ S \mapsto \mathbf{R}$ convex.*

Proof. Let $x_1, x_2 \in int\ S$. Then $y = \lambda x_1 + (1-\lambda)x_2 \in int\ S$, $\lambda \in (0,1)$. Especially,

$$\begin{aligned} f(x_1) &\geq f(y) + (1-\lambda)\xi^T(x_1 - x_2), \\ f(x_2) &\geq f(y) + \lambda\xi^T(x_2 - x_1), \\ \Rightarrow \lambda f(x_1) &+ (1-\lambda)f(x_2) \geq f(y), \end{aligned}$$

where the third equation is a sum of the first equation multiplied by λ and the second by $(1-\lambda)$. □

Theorem (3.3.2). *if $f : S \neq \emptyset \mapsto \mathbf{R}$ convex and differentiable at $x' \in int\ S$ then $\partial f(x') = \{\nabla f(x')\}$.*

Proof. From Theorem 3.2.5, $\exists \xi \in \partial f(x')$. Let $d \neq \bar{0} \in \mathbf{R}^n$ and $\exists \lambda > 0$ s.t. $x' + \lambda d \in S$. From the definition of ξ and differentiability

$$\begin{aligned} f(x' + \lambda d) &\geq f(x') + \lambda\xi^T d, \\ f(x' + \lambda d) &= f(x') + \lambda\nabla f(x')^T + \nabla\|d\|\alpha(x'; \lambda d), \\ \Rightarrow 0 &\geq \lambda(\xi - \nabla f(x'))^T d - \lambda\|d\|\alpha(x'; \lambda d). \end{aligned}$$

Dividing by λ and taking the limit $\lambda \to 0$, we get $(\xi - \nabla f(x'))^T d \leq 0$. By choosing $d = \xi - \nabla f(x')$, we get $\xi = \nabla f(x')$. □

Theorem (Dubovitsky-Milyutin). *If $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ then $\partial f(x) = conv\{\bigcup \partial f_i(x), f_i(x) = f(x)\}$, $x \in \bigcap\ int\ dom\ f_i$.*

Example. *$f = \max\{f_1(x), f_2(x)\}$, where $f_1(x), f_2(x)$ convex and differentiable.*

Example. *Subdifferentials for norms: $f(x) = \|x\|_1 = \sum_{i=1}^n |x_i|$, $f(x) = \|x\|_2$, $f(x) = \|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ and $f(x) = \|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$.*

Example. *No subgradients at zero even though convex function:*

$$f(x) = \begin{cases} -\sqrt{x}, & x \geq 0, \\ \infty, & x < 0. \end{cases}$$

Theorem (3.4.3, corollary). *Uncostrained optimization revisited: x^* global minimum $\Leftrightarrow \bar{0} \in \partial f(x^*)$.*

4 Optimality conditions

Theorem (3.4.2)**.** $\min_{x\in S} f(x)$, *S convex, x^* local minimum*

i) if f convex then x^* is a global minimum,

ii) if f strictly convex then x^* is the unique global optimum.

Proof. i) Assume x^* is not a global minimum, which means that there is $x_0 \in S$ s.t. $f(x_0) < f(x^*)$. Since S is convex, we have $f(\lambda x_0 + (1-\lambda)x^*) \leq \lambda f(x_0) + (1-\lambda)f(x^*) < \lambda f(x^*) + (1-\lambda)f(x^*) = f(x^*)$. $\lambda x_0 + (1-\lambda)x^* \in N_\epsilon(x^*) \cap S$ when λ small and thus it is a contradiction to local optimality.
ii) Assume $\exists x_0 \in S$, $x_0 \neq x^*$ s.t. $f(x_0) = f(x^*)$. Pick the middle point which belongs to S. Due to strict convexity $f((x_0 + x^*)/2) < (f(x_0) + f(x^*))/2 = f(x^*)$, which is a contradiction of optimality. □

Corollary. Let f be convex and diff. x^* is global optimum iff $\nabla f(x^*) = \bar{0}$.

Corollary2. Let f twice differentiable and $\nabla^2 f(x)$ p.s.d $\forall x$. x^* is global optimum iff $\nabla f(x^*) = \bar{0}$.

Theorem (4.1.5)**.** *Let f be pseudoconvex. x^* is global optimum iff* $\nabla f(x^*) = \bar{0}$.

Convex optimization

"... in fact, the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity."
by R. Tyrell Rockafellar 1993

Theorem (3.4.3, necessary and sufficient)**.** *Let $f : \mathbf{R}^n \mapsto \mathbf{R}$ convex, $S \neq \emptyset \subset \mathbf{R}^n$ convex,* $\min_{x\in S} f(x)$.

$$x^* \in S \text{ global optimum} \Leftrightarrow \ \xi^T(x - x^*) \geq 0, \ \forall x \in S, \text{ for some } \xi \in \partial f(x^*).$$

Proof. If part. $f(x) \geq f(x^*) + \xi^T(x - x^*) \geq f(x^*)$, $\forall x \in S$, so x^* optimum.
Only if part. Let us separate the following two sets:

$$\begin{aligned} S_1 &= \{(x - x^*, y), x \in \mathbf{R}^n, y > f(x) - f(x^*) \geq 0\} \subset \mathbf{R}^{n+1}, \\ S_2 &= \{(x - x^*, y), x \in S, y \leq 0\} \subset \mathbf{R}^{n+1}, \end{aligned}$$

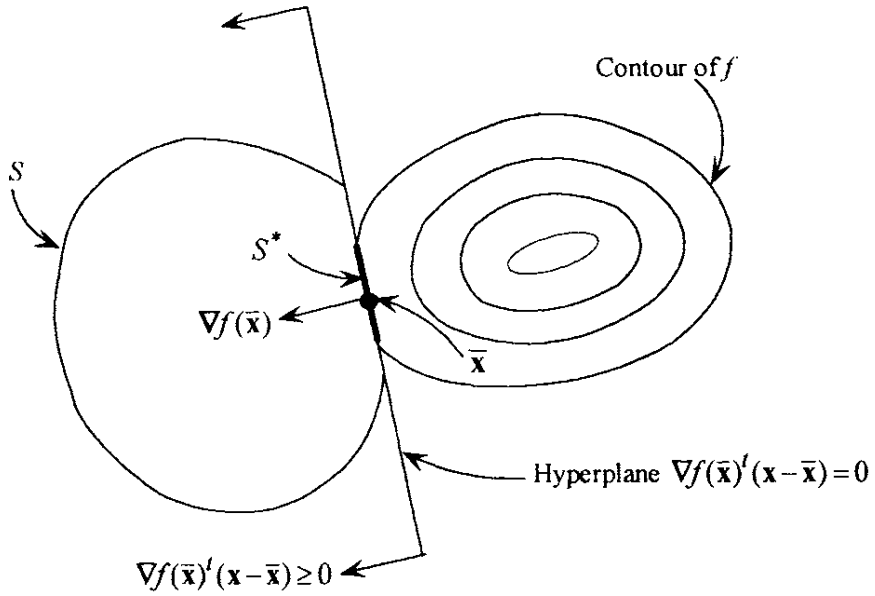
where $(x^*, 0) \in S_2$. S_1, S_2 are convex and $S_1 \cap S_2 = \emptyset$. From Thorem 2.4.8 $\exists(\xi_0, \mu) \neq (\bar{0}, 0)$ and $\alpha \in \mathbf{R}$ s.t.

$$\begin{aligned} \xi_0^T(x - x^*) + \mu y &\leq \alpha, x \in \mathbf{R}^n, y < f(x) - f(x^*), \\ \xi_0^T(x - x^*) + \mu y &\geq \alpha, x \in S, y \leq 0. \end{aligned}$$

Especially, $(x^*, 0) \in S_2 \Rightarrow \alpha \leq 0$ from the second equation. Also, $\forall \epsilon > 0$, $(x^*, \epsilon) \in S_1$ and from the first equation $\mu\epsilon \leq \alpha \leq 0 \Rightarrow \mu \leq 0$. When ϵ is arbitrarily small then $\alpha \geq 0$ and thus $\alpha = 0$.
Assume $\mu = 0$. From the first equation $\xi_0^T(x - x^*) \leq 0$, $\forall x \in \mathbf{R}^n$ and especially with $x = x^* + \xi_0$ $\|\xi_0\|^2 \leq 0$ and thus $\xi_0 = \bar{0}$ is a contradiction. So it must be that $\mu < 0$ and we can define $\xi = -\xi_0/\mu$.
From the first equation: $f(x) \geq f(x^*) + \xi^T(x - x^*)$, $\forall x \in \mathbf{R}^n$, i.e. $\xi \in \partial f(x^*)$.
From the second equation: $\xi^T(x - x^*) - y \geq 0$, $x \in S$, $y \leq 0$. When $y = 0$ $\xi^T(x - x^*) \geq 0$, $\forall x \in S$. □

Corollary. With same assumptions and S open, then x^* global optimum iff $\bar{0} \in \partial f(x^*)$. Proof. Since S is open $x = x^* - \lambda\xi \in S$ for some $\lambda > 0$, $\forall \xi \in \partial f(x^*)$. Thus, $-\lambda\|\xi\|^2 \geq 0 \Rightarrow \xi = \bar{0}$.

Corollary2. With same assumptions and f differentiable, then x^* global optimum iff $\nabla f(x^*)^T(x - x^*) \geq 0$, $\forall x \in S$.



Note. In general **variational inequality problem**:
find x_0 s.t. $f : S \mapsto \mathbf{R}^n$

$$f(x_0)^T(x - x_0) \geq 0, \ \forall x \in S.$$

This problem class includes e.g. the **complementary problem**:
find $x_0 \geq 0$ s.t.

$$\nabla f(x_0) \geq 0, \quad \nabla f(x_0)^T x_0 = 0.$$

For example, finding a Nash equilibrium in game theory.

The result allows a simple numerical method to find a minimum. At nonoptimal point x' where $\nabla f(x')^T(x - x') < 0$ for some $x \in S$, it is easy to find an improving solution. Direction $d = x - x'$ can be used and the step size can be solved using some one-dimensional line search method. The update can be written as $x_{k+1} = x_k + \lambda_k(x - x_k) \in S$, where λ_k is the step size. It can be repeated until $\nabla f(x_k)^T(x - x_k) \geq 0$, $\forall x \in S$. This is called the **method of feasible direction**.

The result could also be derived the following way with more strict assumptions.

Theorem (Moreau-Rockafellar)**.** *If f, g are convex then $\partial f + \partial g \subset \partial(f + g)$ and if int dom $f \cap$ dom $g \neq \emptyset$ then $\partial(f + g) \subset \partial f + \partial g$.*

Definition 4.1. *The* **indicator function** *of set S is* $\chi_S = \begin{cases} 0, & x \in S, \\ \infty, & x \notin S. \end{cases}$

Definition 4.2. *The* **extension** *of $f : S \mapsto \mathbf{R}$ is $\bar{f} = f + \chi_S$, $\bar{f} : \mathbf{R}^n \mapsto \bar{\mathbf{R}}$, where $\bar{\mathbf{R}} = \mathbf{R} \cup \infty$.*

Let $S \neq \emptyset$ convex, f convex. $\inf_{x \in S} f(x) \Leftrightarrow \inf_{x \in \mathbf{R}^n} f(x) + \chi_S$. x^* is a global minimum iff

$$
\begin{aligned}
&\bar{0} \in \partial(f + \chi_S)(x^*) = \partial f(x^*) + \partial \chi_S(x^*), \\
&\Leftrightarrow \bar{0} = \xi + \xi', \ \xi \in \partial f(x^*), \ \xi' \in \partial \chi_S(x^*), \\
&\Leftrightarrow -\xi \in \partial \chi_S(x^*) \Leftrightarrow \chi_S(x) \geq \chi_S(x^*) + (-\xi)^T(x - x^*), \ \forall x \in \mathbf{R}^n, \\
&\Leftrightarrow \xi^T(x - x^*) \geq 0, \ \forall x \in S.
\end{aligned}
$$

We need to assume either *int dom $f \cap S \neq \emptyset$* or *dom $f \cap$ int $S \neq \emptyset$*.

Maximizing convex function

Theorem (3.4.6)**.** $\max_{x \in S} f(x)$, *f, S convex. If x' is a local maximum then*

$$\xi^T(x - x') \leq 0, \ \forall x \in S, \ \forall \xi \in \partial f(x').$$

Note that it is not a sufficient condition.

Example. $f(x) = x^2$, $S = \{x, -1 \leq x \leq 2\}$, $x^* = 2$, $f(x^*) = 4$,
$f'(2)(x - 2) \leq 0$, $\forall x \in S$ *but also* $f'(0)(x - 0) = 0 \leq 0$, $\forall x \in S$.

Theorem (3.4.7)**.** *If S is polyhedron then x^* is an extreme point of S.*

Applications: Risk management in portfolio optimization

$$\begin{aligned} \max \quad & r^T x \\ s.t. \quad & 1/2 x^T Q x \leq V, \\ & Ax \leq b,\ e^T x = 1,\ x \geq 0, \end{aligned}$$

where $r^T x$ expected profit, $1/2 x^T Q x$ variance, $e = (1, \ldots, 1)$. Covariance matrix Q is always symmetric positive semidefinite and thus it is a convex problem. The variance, however, measures both downside and upside risks, when typically downside risk should be considered.

$$\begin{aligned} \max \quad & r^T x \\ s.t. \quad & RM(x) \leq \gamma, \\ & Ax \leq b,\ e^T x = 1,\ x \geq 0, \end{aligned}$$

where $RM(x)$ is a risk measure, e.g., value at risk

$$VaR_\alpha(\xi) = \min\ \gamma,\ \text{s.t. } P(\xi\gamma) \geq \alpha,$$

where α is the confidence level (e.g. 95%). The measure tells that the loss is at most VaR_α with probability α. This measure is a popular measure in finance industry, even though it is not convex nor coherent (sub-additive). These are properties that good risk measures should satisfy. VaR has many local minima and finding the best solution can be difficult.

The following measure is convex and coherent

$$CVaR_\alpha(\xi) = E(\xi, \xi \geq VaR_\alpha(\xi)).$$

See slides in the course website and Uryasev and Rockafellar. The constraints can be linearized, which allows very large problems to be solved with fast and stable algorithms. This shows that the modeling part may have great effect on how difficult optimization problem needs to be solved.

Robust optimization
(slides from the course website)

$$\begin{aligned} \min \quad & c^T x + d \\ s.t. \quad & Ax \leq b, \end{aligned}$$

where c, d, A, b are in uncertainty set U due to data uncertainty, which can be from forecasts, prediction, measurement and implementation errors.

5 Optimality for inequality constrained problem

$$\min_{x \in S} f(x), \quad S = \{x \in X, g_i(x) \leq 0, 1 \leq i \leq m\},$$

where $g_i : \mathbf{R}^n \mapsto \mathbf{R}$, $X \subset \mathbf{R}^n$ open, $g = \begin{bmatrix} g_1 \\ \vdots \\ g_m \end{bmatrix}$.

Definition 5.1. *$d \in \mathbf{R}^n$ is a* **descent direction** *of f at x' if $\exists \delta > 0$ s.t. $f(x' + \lambda d) < f(x')$, $\forall \lambda \in (0, \delta)$. The cone of descent directions is $d \in F$.*

Definition 5.2. *Let $S \subset \mathbf{R}^n$, $x' \in cl\ S$. The cone of* **feasible directions** *of S at x' is $D = \{d \in \mathbf{R}^n, d \neq \bar{0}, x' + \lambda d \in S,\ \forall \lambda \in (0, \delta),\ \text{for some } \delta > 0\}$.*

Theorem (geometric optimality)**.** *x^* is a local minimum iff there are no feasible descent directions $D \cap F = \emptyset$.*

Theorem (4.2.2)**.** *Let f diff. at $x^* \in S$. If x^* is a local minimum then $F_0 \cap D = \emptyset$.*

Proof. $F_0 \subset F \Rightarrow F_0 \cap D = \emptyset$ by geometric optimality. □

Note. the condition is sufficient if f pseudoconvex and $\forall x \in S \cap N_\epsilon(x^*) \Rightarrow x - x^* \in D$.

Definition 5.3. *The* **index set of active constraints** *at x' is denoted by $I = \{i, g_i(x') = 0\}$ and the corresponding cone*

$$G_0 = \{d, \nabla g_i(x')^T d < 0,\ \forall i \in I\}.$$

Theorem (4.2.4)**.** *If g_i, $i \notin I$, continuous at x' and g_i, $i \in I$, differentiable at x' then $G_0 \subseteq D$.*

Proof. Since $x' \in X$ open, $\exists \delta_1 > 0$ s.t. $x' + \lambda d \in X$, $\forall \lambda \in (0, \delta_1)$. Since $g_i(x') < 0$, $i \notin I$, are continuous, $g_i(x' + \lambda d) < 0$, $i \notin I$, $\forall \lambda \in (0, \delta_2)$. If $d \in G_0$ then $\nabla g_i(x')^T d < 0$, $i \in I$. By Theorem 4.1.2 $g_i(x' + \lambda d) < g_i(x') = 0$, $\forall \lambda \in (0, \delta_3)$. Thus, $x' + \lambda d \in S$ when $\lambda \in (0, \min(\delta_1, \delta_2, \delta_3))$. □

Note that $G_0 \subseteq D \subseteq G_0'$, where $G_0' = \{d \neq \bar{0}, \nabla g_i(x')^T d \leq 0, i \in I\}$. Also, $D = G_0$ if g_i, $i \in I$, are strictly ps.convex. $D = G_0'$ if they are strictly ps.concave.

Theorem (4.2.5, road to FJ)**.** *Let $x^* \in S$, g_i, $i \notin I$, continuous in x^*, g_i, $i \in I$, diff. at x^*. If x^* is local minimum then $F_0 \cap G_0 = \emptyset$.*

Proof. By Theorem 4.2.2 $F_0 \cap D = \emptyset$ and by Theorem 4.2.4 we have $F_0 \cap G_0 \subseteq F_0 \cap D$. □

Note that the condition is sufficient if f ps.convex at x^*, g_i, $i \in I$, strictly ps.convex at $N_\epsilon(x^*)$ for some $\epsilon > 0$.

Example (4.2.6)**.**

$$\begin{aligned} \min \quad & (x_1-3)^2+(x_2-2)^2 \\ s.t. \quad & x_1^2+x_2^2 \le 5, \\ & x_1+x_2 \le 3,\ x_1 \ge 0,\ x_2 \ge 0. \end{aligned}$$

$x^* = (2,1)$, $I = \{1,2\}$, $\nabla f(x^*) = -(2,2)$, $\nabla g_1(x^*) = (4,2)$, $\nabla g_2(x^*) = (1,1)$. *As should be* $F_0 \cap G_0 = \emptyset$*, which in general does not imply that* $F_0 \cap D = \emptyset$*. The problem does not satisfy the sufficient conditions since* g_2 *is not strictly ps. convex, and thus it cannot be said that* x^* *is a local optimum only by having* $F_0 \cap G_0 = \emptyset$*. However,* $F_0 \cap G_0' = \emptyset \Rightarrow F_0 \cap D = \emptyset$ *and with this we can say that* x^* *is a local minimum. The feasible set is convex and the objective is strictly convex, and thus* x^* *is actually a unique global minimum.*

The idea is to use the separation theorems (Gordan and Motzkin) with the geometric optimality to prove the algebraic conditions: the Fritz-John (FJ) and finally the Karush-Kuhn-Tucker (KKT) conditions. FJ conditions are more general but there are typically too many (nonoptimal) points that satisfy them. By making more assumptions to the problem and its constraints with so called constraint qualification (CQ) conditions, we can get rid of these nonoptimal points, and we get the KKT from the FJ conditions.

Note that we cannot use the same technique to the equality constraints with the following simple trick. We could define $h(x) = 0$ by $h(x) \le 0$ and $-h(x) \le 0$ but then the geometric optimality would not work since $G_0 = \emptyset$ for all points.

Theorem (4.2.8, FJ necessary)**.** *If* x^* *is a local minimum then* $\exists\ u_0, u_i$*,* $i \in I$*, s.t.*

$$\textit{(FJ1)} \quad \begin{cases} u_0 \nabla f(x^*) + \sum_{i \in I} u_i \nabla g_i(x^*) = \bar{0}, \\ u_0, u_i \ge 0,\ i \in I,\ u_j \ne 0 \text{ for some } j = 0 \text{ or } j = i \in I, \end{cases}$$

where the last one could be written as $(u_0, u_I) \ne (0, \bar{0})$*. If also* g_i*,* $i \notin I$ *differentiable at* x^* *then*

$$\textit{(FJ2)} \quad \begin{cases} u_0 \nabla f(x^*) + \sum_{i=1}^m u_i \nabla g_i(x^*) = \bar{0}, \\ u_i g_i(x^*) = 0,\ \forall i = 1, \dots, m, \\ u_0, u \ge 0,\ (u_0, u) \ne (0, \bar{0}). \end{cases}$$

Proof. Since x^* is a local minimum, Theorem 4.2.5 implies $F_0 \cap G_0 = \emptyset$. Let $m' \le m$ be the number of indexes in I, $A \in \mathbf{R}^{(m'+1)n}$ with rows of $\nabla f(x^*)^T$ and $\nabla g_i(x^*)^T$, $i \in I$. Geometric optimality now means that $\nexists d \in \mathbf{R}^n$ s.t. $Ad < \bar{0}$. Theorem 2.4.9 (Gordan) implies that $\exists p \ge \bar{0}$, $p \ne \bar{0}$, s.t. $A^T p = \bar{0}$, $p \in \mathbf{R}^{m'+1}$. Let us denote $p = (u_0, u_1, \dots, u_{m'})$. Thus, we have $(FJ1)$. The second equation in $(FJ2)$, the complementary slackness condition, means that $u_i = 0$, $i \notin I$, and it gives $(FJ2)$. □

Note that if $u_0 = 0$ then the conditions have no information about the objective.

Example. min $f(x)$ *s.t.* $g_1(x) \leq 0$ *and* $g_2(x) \leq 0$. *Now, any feasible* x' *with* $\nabla g_1(x') = -\nabla g_2(x') \Rightarrow G_0 = \emptyset$ *and* x' *is an FJ point.*

There are too many FJ points and more assumptions are needed.

Theorem (4.2.13, KKT necessary)**.** *Assume* $\nabla g_i(x^*)$ *are linearly independent. If* x^* *is a local minimum then* $\exists u_i \in \mathbf{R}$, $i \in I$ *s.t.*

$$\textit{(KKT1)} \begin{cases} \nabla f(x^*) + \sum_{i \in I} u_i \nabla g_i(x^*) = \bar{0}, \\ u_i \geq 0, \ i \in I. \end{cases}$$

If also g_i, $i \notin I$ *differentiable at* x^* *then*

$$\textit{(KKT2)} \begin{cases} \nabla f(x^*) + \nabla g(x^*)^T u = \bar{0}, \ \textit{(Lagrange optimality)} \\ u_i g_i(x^*) = 0, \ \forall i = 1, \ldots, m, \ \textit{(complementary slackness)} \\ u \geq 0. \ \textit{(dual feasibility)} \end{cases}$$

The scalars u_i are called the Lagrange multipliers or dual variables.

Example.

$$\begin{aligned} \min \quad & (x_1 - 3)^2 + (x_2 - 2)^2 \\ s.t. \quad & x_1^2 + x_2^2 \leq 5, \\ & x_1 + 2x_2 \leq 4, \ x_1 \geq 0, \ x_2 \geq 0. \end{aligned}$$

$x^* = (2, 1)$, $I = \{1, 2\}$. $\nabla g_1(x^*) = (4, 2)$, $\nabla g_2(x^*) = (1, 2)$. *We can choose the multipliers, e.g.,* $u_0 = 3 > 0$, $u_1 = 1 > 0$, $u_2 = 2 > 0$ *and* $u_3 = u_4 = 0$. *These satisfy both FJ and KKT conditions* (*Lagrange multipliers* $(1/3, 2/3)$).

Example.

$$\begin{aligned} \min \quad & -x_1 \\ s.t. \quad & x_2 - (1 - x_1)^3 \leq 0, \\ & -x_2 \leq 0. \end{aligned}$$

$x^* = (1, 0)$, $I = \{1, 2\}$. $\nabla g_1(x^*) = (0, 1)$, $\nabla g_2(x^*) = (0, -1)$. *The constraints gradients are linearly dependent. We can choose* $u_0 = 0$ *and* $u_1 = u_2$ *arbitrarily so that FJ conditions hold. Note that the optimum does not satisfy KKT conditions and there are no Lagrange multipliers.*

Sufficient conditions

Theorem (4.2.16, KKT sufficient)**.** *Assume* f *and* g_I *are convex. If* x^* *is a KKT point then* x^* *is a global minimum. If the convexities hold in* $N_\epsilon(x^*)$ *for some* $\epsilon > 0$ *then* x^* *is a local minimum.*

Extension: Production planning in continuous time*

This example is dynamic optimization and it is from Luenberger: optimization by vector space methods p.234. Let us examine a production planning problem where the decision variable is the production rate $r(t) = \dot{z}(t)$, $t \in (0,1)$ and $z(t)$ is the amount of products manufactured. It is assumed that there are no inventory costs and the demand rate $d(t) = \dot{s}(t)$ is known, where $s(t)$ is the amount of sold units. It is assumed that the demand must be met

$$z(0) + \int_0^t r(y)dy \geq \int_0^t d(y)dy \Leftrightarrow z(t) \geq s(t).$$

This means that the products available at time 0 plus production should be greater than demand at all time instances.

$$\begin{aligned} \min \quad & 1/2 \int_0^1 r^2(t)dt \\ s.t. \quad & \dot{z}(t) = r(t),\ z(t) \geq s(t),\ z(0) > 0, \end{aligned}$$

For example, $z(0) = 1/2$, $s(t) = \begin{cases} 2t,\ 0 \leq t \leq 1/2, \\ 1,\ 1/2 \leq t \leq 1, \end{cases}$. The sales rate is constant up to $t = 1/2$ and after that there is no sales. The space where the problem is solved is chosen as $X = Z = C[0,1]$, the space of continuous functions between 0 and 1, i.e., it is assumed that $z(t) = z(0) + \int_0^t r(k)dk$ is continuous. Note that the minimum may not be in this space if there could be jumps in the function. The dual space of continuous functions is $NBV[0,1]$, normalized bounded variation functions, which may have finite number of finite jumps. The Lagrange multiplier will belong to this space.

The Lagrange function is defined

$$\phi(r,u) = 1/2 \int_0^1 r^2(t)dt + \int_0^1 (s(t) - z(t))du(t),$$

where $u \in NBV[0,1]$ and u is nondecreasing. We can simplify the equation by Leibniz integration rule and integration by parts

$$\int_0^1 \int_0^t r(y)dydu(t) = /_0^1 \int_0^1 r(y)dyu(t) - \int_0^1 r(t)u(t)d(t).$$

Now, we get

$$\begin{aligned} \phi(r,u) &= 1/2 \int_0^1 r^2(t)dt + \int_0^1 (s(t) - z(0))du(t) - \int_0^1 \int_0^t r(y)du(du(t), \\ &= 1/2 \int_0^1 r^2(t)dt + \int_0^1 (s(t) - z(0))du(t) + \int_0^1 r(t)u(t)dt - u(1) \int_0^1 r(t)dt, \end{aligned}$$

since $u(0) = 0$ from normalization. The optimality conditions give

$$\begin{aligned}\frac{\partial \phi}{\partial r} &= r^*(t) + u^*(t) - u^*(1) \geq 0, \ \forall t, \\ & r^*(t)(r^*(t) + u^*(t) - u^*(1)) = 0, \ \forall t, \\ & u^*(t) \text{ varies only when } z(t) = s(t), \\ & u^*(t) \text{ is nondecreasing.}\end{aligned}$$

The economic interpretation of Lagrange multiplier is the same. Let J be the total cost then

$$\Delta J = \int_0^1 \Delta s(t) du(t) = -\int_0^1 \Delta \dot{s}(t) u(t) dt + \Delta s(t) u(1) - \Delta s(0) u(0).$$

Since $\Delta s(0) = 0$, $u(1) = 0$, we have

$$\Delta J = -\int_0^1 \Delta d(t) u(t) dt,$$

i.e., $-u(t)$ is the unit cost or the shadow price of extra demand. Now, this price is zero when $t > 1/2$.

6 Equality and inequality constrained problem

For geometric optimality and feasible directions, we need more restrictive assumptions on the equality constraints and more mathematical machinery. The next theorem gives the conditions that guarantee regularity in the constraints.

$$\begin{aligned}\min \quad & f(x) \\ s.t. \quad & g(x) \leq \bar{0} \in \mathbf{R}^m \\ & h(x) = \bar{0} \in \mathbf{R}^l.\end{aligned}$$

Definition 6.1. $H_0 = \{d, \ \nabla h_i(x)^T d = 0, \ i = 1, \ldots, l\}$.

Theorem (implicit function)**.** *If i)* $f(x_1, x_2) = \bar{0}$, $x_1 \in \mathbf{R}^n$, $x_2 \in \mathbf{R}^l$,
ii) f *continuous,*
iii) $\nabla_{x_2} f$ *continuous,*
iv) $\nabla_{x_2} f(x_1, x_2)$ *nonsingular, i.e.*

$$\begin{vmatrix} \frac{\partial f_1(x_1,x_2)}{\partial x_1} & \cdots & \frac{\partial f_1(x_1,x_2)}{\partial x_l} \\ & \ddots & \\ \frac{\partial f_l(x_1,x_2)}{\partial x_1} & \cdots & \frac{\partial f_l(x_1,x_2)}{\partial x_l} \end{vmatrix} \neq 0,$$

then
$\exists g : N_\epsilon(x_1) \mapsto \mathbf{R}^l$, $g(x_1) = x_2$ *and* $f(x_1, g(x_1)) = 0$.

If $\exists \nabla_{x_1} f$ then g is differentiable. If $p > 0$, f p-times continuously differentiable then g is also p-times continuously differentiable and

$$\nabla g(x_1) = -\nabla_x f(x_1, g(x_1))(\nabla_{x_2} f(x_1, g(x_1)))^{-1}, \ \forall x_1 \in N_\epsilon(x_1).$$

Theorem (geometric optimality)**.** *Let* $X \in \mathbf{R}^n$ *open,* $f, g_i, h_j : \mathbf{R}^n \mapsto \mathbf{R}$, $1 \leq i \leq m$, $1 \leq j \leq l$, f, g_i, $i \in I$ *diff. at* x^*, g_i, $i \notin I$, *continuous at* x^*, h_j *continuously differentiable at* $N_\epsilon(x^*)$ *for some* $\epsilon > 0$. *Assume that* $\nabla h_j(x^*)$ *are linearly independent. If* x^* *is a local minimum then* $F_0 \cap G_0 \cap H_0 = \emptyset$.

Proof. Assume $\exists y \in F_0 \cap G_0 \cap H_0$. Let us denote the Jacobian by $\nabla h(x) = \begin{bmatrix} \nabla h_1(x)^T \\ \vdots \\ \nabla h_l(x)^T \end{bmatrix}$. Since $y \in H_0$, $\nabla h(x^*)y = \bar{0}$. Let us check the conditions of implicit function theorem: i) $h(x^*) = 0$, ii) $h(x)$ is continuous, iii) $\nabla h(x)$ is continuous and iv) $\nabla h(x)$ is nonsingular since $\nabla h_i(x)$ are linearly independent. Thus, we get $\exists x : [-a, a] \mapsto \mathbf{R}^n$ which is continuously differentiable s.t. $x(0) = x^*$, $\dot{x}(0) = x'(0) = y$ and $h_i(x(t)) = 0$, $\forall t \in [-a, a]$. This means that we can move along $h(x) = \bar{0}$ at least for small distance. The feasibility and descent in objective goes as earlier.
Feasibility: $i \in I$: $\frac{d}{dt} g_i(x(t)) = \nabla g_i(x(t))^T \dot{x}(t)$. at $t = 0$ $\nabla g_i(x^*)^T y < 0$ ($y \in G_0$)
$i \notin I$: from continuity $g_i(x(t)) < 0$, $t \in (0, t_1)$
X open: $x(t) \in X$, $t \in (0, t_2)$
$x(t)$ feasible when $t \in (0, t')$ where $t' = \min(t_1, t_2, a)$.
Decrease: $\nabla f(x^*)^T y < 0$ ($y \in F_0$) $\Rightarrow f(x(t)) < f(x^*)$, $\forall t \in (0, t_3)$.
This contradicts the local optimality and we get the result. □

Theorem (4.3.2, FJ necessary)**.** g_i, $i \in I$ *continuous at* x^*, f, g_i, $i \in I$ *differentiable at* x^*, h_j *continuously differentiable at* $N_\epsilon(x^*)$ *for some* $\epsilon > 0$. *If* x^* *is a local minimum then* $\exists u_0, u_i$, $i \in I$ *and* v_j, $1 \leq j \leq l$ *s.t.*

$$\textit{(FJ1)} \ \begin{cases} u_0 \nabla f(x^*) + \sum_{i \in I} u_i \nabla g_i(x^*) + \sum_{j=1}^{l} v_j \nabla h_j(x^*) = \bar{0}, \\ u_0, u_i \geq 0, \ \forall i \in I, \ (u_0, u_I, v) \neq (0, \bar{0}, \bar{0}). \end{cases}$$

If also g_i, $i \notin I$ *differentiable at* x^* *then*

$$\textit{(FJ2)} \ \begin{cases} u_0 \nabla f(x^*) + u^T \nabla g(x^*) + v^T \nabla h(x^*) = \bar{0}, \\ u_i g_i(x^*) = 0, \ \forall i = 1, \ldots, m, \\ u_0, u \geq 0, \ (u_0, u, v) \neq (0, \bar{0}, \bar{0}). \end{cases}$$

Proof. Assume $\nabla h_i(x^*)$ are linearly dependent then $\exists v_i$ s.t. $\sum_{i=1}^{l} v_i \nabla h_i(x^*) = \bar{0}$ and some $v_i \neq 0$. Choose $u_0 = u_i = 0$, $i \in I$, and we get (FJ1).
Assume $\nabla h_i(x^*)$ are linearly independent then
denote $A_1 \in \mathbf{R}^{(m'+1)n} = \begin{bmatrix} \nabla f(x^*)^T \\ \nabla g_1(x^*)^T \\ \vdots \\ \nabla g_{m'}(x^*)^T \end{bmatrix}$ and $A_2 = \begin{bmatrix} \nabla h_1(x^*)^T \\ \vdots \\ \nabla h_l(x^*)^T \end{bmatrix}$. By Theorem 4.3.1, $\nexists d \in \mathbf{R}^n$ s.t. $A_1 d < \bar{0}$, $A_2 d = \bar{0}$. By Motzkin's theorem, $\exists p_1 \in \mathbf{R}^{m'+1}$, $p_2 \in \mathbf{R}^l$, $p_1 \geq \bar{0}$, $p_1 \neq \bar{0}$ s.t. $A_1^T p_1 + A_2^T p_2 = \bar{0}$, denote $p_1 = (u_0 \ u_1 \ \dots \ u_i)^T$ and $p_2 = v$ and we have (FJ1). □

Theorem (4.3.7, KKT necessary)**.** *Assume $\nabla g_i(x^*)$, $i \in I$ and $\nabla h_j(x^*)$, $1 \leq j \leq l$ are linearly independent. If x^* is a local minimum then $\exists u_i$, $i \in I$, v_j, $1 \leq j \leq l$ s.t.*

$$\textit{(KKT1)} \quad \begin{cases} \nabla f(x^*) + u_I^T \nabla g(x^*) + v^T \nabla h(x^*) = \bar{0}, \\ u_i \geq 0, \ \forall i \in I. \end{cases}$$

If also g_i, $i \notin I$ differentiable at x^* then

$$\textit{(KKT2)} \quad \begin{cases} \nabla f(x^*) + u^T \nabla g(x^*) + v^T \nabla h(x^*) = \bar{0}, \\ u_i g_i(x^*) = 0, \ \forall i = 1, \dots, m, \\ u \geq 0. \end{cases}$$

Proof. From FJ, $\exists u_0, u_i', v_i' \neq (0, \bar{0}, \bar{0})$. If $u_0 = 0$ then $(u_i', v_i') \neq (\bar{0}, \bar{0})$ and this contradicts the assumption of linear independence. Thus, $u_0 > 0$ and we can denote $u_i = u_i'/u_0$ and $v_i = v_i'/u_0$, and we have (KKT1). □

Note that there are other constraint qualification (CQ) or regularity conditions beside linear independence that guarantee that $u_0 > 0$.

Example.

$$\begin{aligned} \min \quad & x_1^2 + x_2^2 \\ s.t. \quad & x_1^2 + x_2^2 \leq 5, \\ & x_1 + 2x_2 = 4, \ x_1 \geq 0, \ x_2 \geq 0. \end{aligned}$$

$x^* = (4/5, 8/5)$, $I = \emptyset$. $\nabla f(x^*) = (8/5, 16/5)$, $\nabla h(x^*) = (1, 2)$. *The multiplier $v = -8/5$ solves the KKT conditions.*

Sufficient conditions

Theorem (4.3.8, KKT sufficient)**.** *Assume f, g_I convex, h_j, $j \in \{j, v_j > 0\}$ convex, h_j, $j \in \{j, v_j < 0\}$ concave. If x^* is a KKT point then x^* is a global minimum. If the convexities hold in $N_\epsilon(x^*)$ for some $\epsilon > 0$ then x^* is a local minimum.*

Proof. Shown in exercises by relating KKT to the variational inequality of the convex problem. KKT equals $\nabla f(x)^T(x - x^*) \geq 0$, for all feasible x. □

Note that the requirement for the convexity/concavity of h_j is not known before the KKT conditions are solved. One way to get around this is to assume that $h(x) = Ax + b$, i.e., the equality constraints are affine.

Definition 6.2. *The* **Lagrange function** *is* $\phi(x, u, v) = f(x) + u^T g(x) + v^T h(x)$. *The restricted Lagrangian is* $L(x) = \phi(x, u^*, v^*)$, *where* (u^*, v^*) *are the Lagrange multipliers that solve the KKT conditions (with* x^**).*

Theorem (4.4.1, second order sufficient)**.** *i) If* $\nabla^2 L(x)$ *is p.s.d.* $\forall x \in S$ *then KKT* x^* *is a global minimum.*
ii) If $\nabla^2 L(x)$ *p.s.d.* $\forall x \in S \cap N_\epsilon(x^*)$ *for some* $\epsilon > 0$ *then KKT* x^* *is a local minimum.*
iii) If $\nabla^2 L(x^*)$ *p.d. then KKT* x^* *is a unique local minimum.*

Proof. i) KKT $\Rightarrow \nabla L(x^*) = \bar{0}$. $\nabla^2 L(x)$ p.s.d. then $L(x)$ convex in $S \Rightarrow f(x^*) = L(x^*) \leq L(x) \leq f(x)$, $\forall x \in S$.
iii) $\nabla L(x^*) = \bar{0}$ and p.d. $\Rightarrow$ strict minimum for $L(x) \Rightarrow f(x^*) = L(x^*) < L(x) = f(x)$, $\forall x \neq x^* \in \{S \cap N_\epsilon(x^*)\}$. □

Definition 6.3. *Let* $I = \{i,\ g_i(x^*) = 0\}$, $I^+ = \{i \in I, u_i^* > 0\}$ *and* $I^0 = \{i \in I, u_i^* = 0\}$.

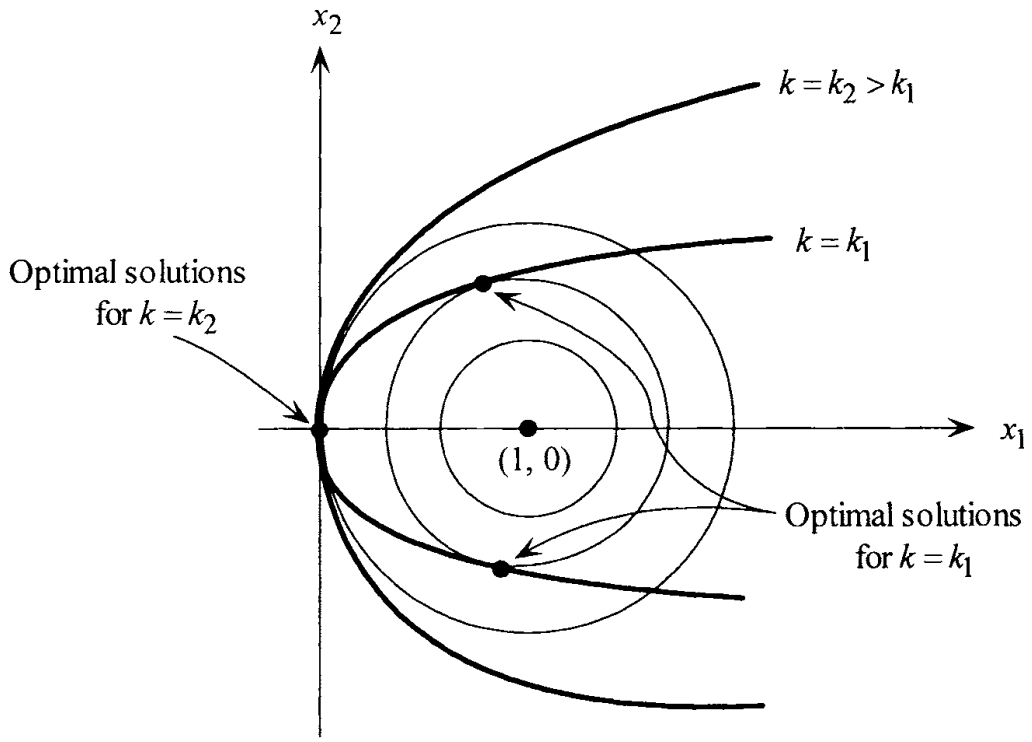
Theorem (4.4.2, second order sufficient)**.** *Let* $C = \{d \neq \bar{0} : \nabla g_i(x^*)^T d = 0$, *for* $i \in I^+$, $\nabla g_i(x^*)^T d \leq 0$, *for* $i \in I^0$, *and* $\nabla h_i(x^*)^T d = 0$, *for* $i = 1, \ldots, l\}$. *If* $d^T \nabla^2 L(x^*) d > 0$ *for all* $d \in C$, *then* x^* *is a strict local minimum.*

Theorem (4.4.3, second order necessary)**.** *Assume CQ. If* x^* *is a local minimum, then* x^* *is a KKT point and* $d^T \nabla^2 L(x^*) d \geq 0$ *for all* $d \in C$.

Example (4.4.4)**.**

$$\begin{aligned} \min \quad & (x_1 - 1)^2 + x_2^2 \\ s.t. \quad & 2kx_1 - x_2^2 \leq 0, \end{aligned}$$

where $k > 0$. *The objective is convex but the feasible set is not convex. The unconstrained minimum is not feasible, so the constraint must be binding. There are three KKT points depending on* k: $x^1 = (0, 0), u^1 = 1/k$ *for any* $k > 0$, *and for* $0 < k < 1$, $x^2 = (1 - k, \sqrt{2k(1-k)}), u^2 = 1$ *and* $x^3 = (1 - k, -\sqrt{2k(1-k)}), u^2 = 1$. *The constraint is not quasiconvex, so we cannot use the necessary conditions 4.2.16.*



$L(x) = (x_1 - 1)^2 + x_2^2 + u(2kx_1 - x_2^2)$ and $\nabla^2 L(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2(1-u) \end{bmatrix}$. $C = \{d \neq 0 : kd_1 = x_2 d_2\}$. Let us examine the necessary condition 4.4.3 first. For x^1, we have $d^T \nabla L(x) d = 2d_1^2 + 2(1 - 1/k)d_2^2$ and $d \in C$ means $d_1 = 0$. $d^T \nabla L(x) d \geq 0$ holds when $k > 1$ but is violated when $0 < k < 1$. We can conclude by 4.4.3 that x^1 is not a local minimum when $0 < k < 1$. $\nabla^2 L(x)$ are positive semidefinite at x^2 and x^3, and satisfy 2nd order necessary condition.

Now, we examine the sufficient condition 4.4.2. $\nabla^2 L(x^1)$ is p.d. when $k > 1$, so x^1 is then strict local minimum. For $k = 1$, we don't get this since $d^T \nabla L(x^1) d = 2d_1^2 = 0$. However, $\nabla^2 L(x^2)$ is not positive definite but $C = \{d \neq 0 : kd_1 = \sqrt{2k(1-k)} d_2\}$ and $d^T \nabla L(x^2) d = 2d_1^2 > 0$ for any $d \in C$. Thus, x^2 is strict local minimum for $0 < k < 1$ by 4.4.2. So, 4.4.1 didn't work and 4.4.2 was needed. Similarly, for x^3.

Definition 6.4. *The* **bordered Hessian** *is* $HL(x) = \begin{bmatrix} \bar{0} & B \\ B^T & \nabla^2 L(x) \end{bmatrix}$, *where* B *contains the constraints' gradients.*

Let us examine a two-dimensional problem with one equality constraint. We have $HL(x) = \begin{bmatrix} 0 & g_x & g_y \\ g_x & L_{xx} & L_{xy} \\ g_y & L_{yx} & L_{yy} \end{bmatrix}$.

Theorem (second order sufficient)**.** *If x^* is a KKT point and $det(HL(x^*)) < 0$, then x^* is a local minimum.*

Sensitivity analysis

Theorem (Bertsekas, Nonlinear Prog.)**.** *Consider the family of problems*

$$\min_{h(x)=t} f(x)$$

parameterized by $t \in \mathbf{R}^m$. Assume that for $t = \bar{0}$, this problem has a local minimum x^*, which is regular (satisfies some CQ) and together with its unique Lagrange multiplier v^* satisfies the sufficient second order KKT conditions for local minimum.

Then there exists an open sphere S centered at $t = \bar{0}$ such that for every $t \in S$, there is an $x(t)$ and a $v(t)$, which are a local minimum-Lagrange multiplier pair of the parameterized problem. Furthermore, $x(t)$ and $v(t)$ are continuously differentiable within S and we have $x(\bar{0}) = x^*$, $v(\bar{0}) = v^*$. In addition,

$$\nabla p(t) = -v(t), \quad \forall t \in S,$$

where $p(t)$ is the primal function $p(t) = f(x(t))$.

Proof. Apply the implicit function theorem to the system

$$\nabla f(x) + \nabla h(x)v = \bar{0}, \quad h(x) = t.$$

Let us check the conditions: i) for $t = \bar{0}$ the system has the solution (x^*, v^*), ii-iii) first and second derivatives need to be continuous for $f(x)$ and $h(x)$, iv) the Jacobian

$$J = \begin{pmatrix} \nabla^2 f(x^*) + v^T \nabla^2 h(x^*) & \nabla h(x^*) \\ \nabla h(x^*)^T & \bar{0} \end{pmatrix}$$

is nonsingular ($\nabla h(x) \nabla h(x)^T$ nonsingular since the constraints are linearly independent). Thus, for all $t \in S$ for some open sphere S centered at $t = \bar{0}$, there exist $x(t)$ and $v(t)$ such that $x(\bar{0}) = x^*$, $v(\bar{0}) = v^*$, the functions $x(t)$ and $v(t)$ are continuously differentiable, and

$$\nabla f(x(t)) + \nabla h(x(t))v(t) = \bar{0}, \quad h(x(t)) = v.$$

For t close to $t = \bar{0}$, using sufficiency conditions, $x(t)$ and $v(t)$ are a local minimum-Lagrange multiplier pair for the parameterized problem.

To derive $\nabla p(t)$, we i) differentiate $h(x(t)) = t \Rightarrow \nabla x(t) \nabla h(x(t)) = I$, and ii) differentiate the system $\Rightarrow \nabla x(t) \nabla f(x(t)) + \nabla x(t) \nabla h(x(t)) v(t) = \bar{0}$. Now, we have

$$\begin{aligned} \nabla p(t) &= \nabla_t f(x(t)) = \nabla x(t) \nabla f(x(t)), \\ &= -\nabla x(t) \nabla h(x(t)) v(t) = -v(t). \end{aligned}$$

□

Summary

The optimality conditions were derived using the geometric optimality and suitable separation theorems. In inequality constrained problem, the geometric optimality was $F_0 \cap G_0 = \emptyset$. The Gordan theorem was applied to this condition, and it gave the more general FJ conditions. There are some special cases when the optimum satisfies FJ but not KKT conditions. By assuming the linear independece constraint qualification condition, in FJ conditions we can guarantee that $u_0 > 0$ and we get the KKT conditions.

In equality constrained problem, we need to assume linear independence for $h_j(x)$ even in the geometric optimality $F_0 \cap G_0 \cap H_0 = \emptyset$. The suitable separation theorem is Motzkin and the theory goes like in the inequality constrained problem.

Note that there are no convexity assumptions in the necessary conditions. They appear only in the sufficient conditions. When the problem is convex and the constraint qualification holds, the KKT conditions turn out to be the same as the optimality conditions for the convex problem (variational inequality). Note also that the complementary slackness condition $u_i g_i(x^*) = 0$ does not mean that when $g_i(x^*) = 0 \Rightarrow u_i > 0$. When both $u_i = g_i(x^*) = 0$ it is said the the constraint is weakly active, and it means that removing the constraint does not alter the minimum. The constraint just happens to be active without restricting the optimal value.

Non-differentiable convex problem*

$$\begin{aligned} \inf \quad & f(x) \\ s.t. \quad & x \in S, \\ & g(x) \leq \bar{0}, \\ & Ax - b = \bar{0}, \end{aligned}$$

where $f, g_i,\ 1 \leq i \leq m : \mathbf{R}^n \mapsto (-\infty, \infty]$ convex, $S \subset \mathbf{R}^n$ convex, $b \in \mathbf{R}^p$. Also, define $L = \{x, Ax = b\}$.

Theorem (convex KKT, Eric Balder)**.** *Let x^* be a feasible point of the problem. i) x^* is a global minimum if $\exists u \in \mathbf{R}^m_+$, $v \in \mathbf{R}^p$ and $\eta \in \mathbf{R}^n$ s.t.*

$$u_i g_i(x^*) = 0, \quad i = 1, \ldots, m, \ \textit{(complementary slackness)}$$
$$\bar{0} \in \partial f(x^*) + \sum_{i \in I(x^*)} u_i \partial g_i(x^*) + A^T v + \eta, \ \textit{(normal Lagrange inclusion)}$$
$$\eta^*(x - x^*) \leq 0, \quad \forall x \in S. \ \textit{(obtuse angle property)}$$

ii) If x^* is a global minimum and if $x^* \in$ int dom $f \cap \cap_{i \in I(x^*)}$int dom g_i and int $S \cap L \neq \emptyset$ (regularity condition), then $\exists u_0 \in \{0,1\}$, $u \in \mathbf{R}^m_+$, $(u_0, u) \neq (0, \bar{0})$, $v \in \mathbf{R}^p$, $\eta \in \mathbf{R}^n$ s.t. CS, obtuse angle and

$$\bar{0} \in u_0 \partial f(x^*) + \sum_{i \in I(x^*)} u_i \partial g_i(x^*) + A^T v + \eta. \text{ (Lagrange inclusion)}$$

When $u_0 = 1$ it is said that the normal Lagrange inclusion occurs and the abnormal when $u_0 = 0$. The abnormal case is impossible with the regularity or constraint qualification conditions, like when A is of rank p and the Slater's condition holds: $\exists x' \in S \cap L$ s.t. $g_i(x') < 0$, for $i = 1, \ldots, m$.

7 Duality

There are many kinds of duality in mathematics; see polyhedral duality, where the role of vertices and faces is interchanged. Even in optimization, some classes of problems have much stronger duality theorems than others. A dual problem is another problem formulated with the data of the original problem that tells something about the original problem. In nonlinear optimization the dual gives lower (or upper) bounds for the original problem. This can be used in evaluation of how far the current solution is from the optimum. This will be especially useful in integer optimization, and this is demonstrated in the exercises.

For convex and linear problems, the results are much stronger. The dual may be faster to solve (or not, see Boyd: convex optimization), it may give some properties of optimal solution, or the dual can be used in proving the existence of a solution. For example, duality is used in solving large LP problems.

The primal problem P is

$$\begin{array}{lll} \min & f(x) & \\ s.t. & g(x) \leq \bar{0}, & \Leftrightarrow \\ & h(x) = \bar{0}, & \\ & x \in X, & \end{array} \qquad \min_{x \in X} \sup_{u \geq 0, v} \phi(x, u, v) \doteq L_p(x),$$

where $\phi(x, u, v) = f(x) + u^T g(x) + v^T h(x)$ is the Lagrange function,

$$L_p(x) = \begin{cases} f(x), & \text{x feasible,} \\ \infty, & \text{otherwise.} \end{cases}$$

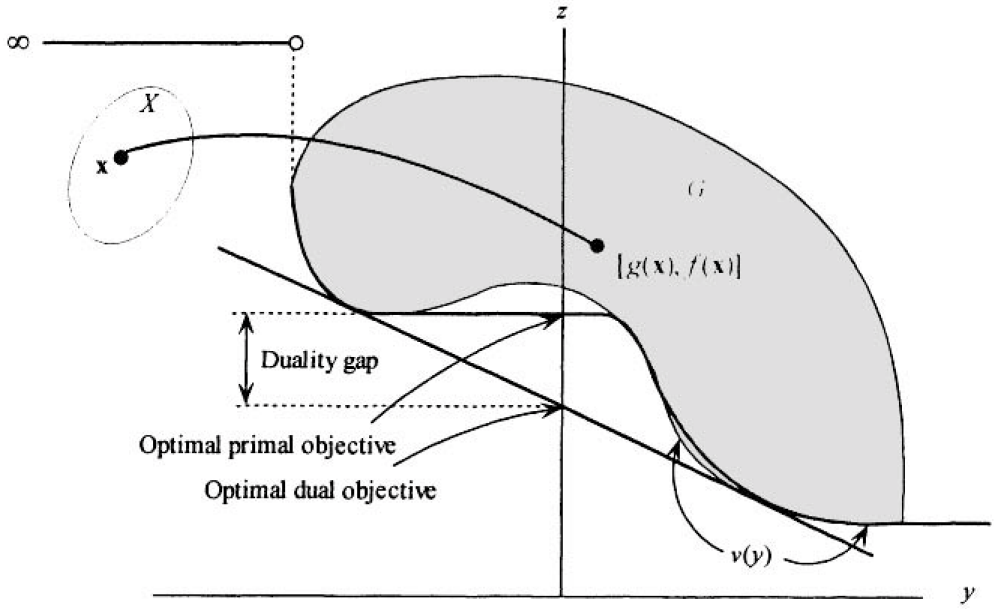
The Lagrange dual problem D is

$$\max_{u \geq \bar{0}, v} \inf_{x \in X} \phi(x, u, v) = \theta(u, v),$$

where $\theta(u, v)$ is the dual function. We can see that the primal and dual are taking the minimization over x and maximization over (u, v) in different order over the Lagrange function $\phi(x, u, v)$.

Geometric interpretation

Let us study a problem with one inequality constraint $g(x) \leq 0$. We can examine the points $x \in X$ in two-dimensional set $G = \{(y, z) = (g(x), f(x)), x \in X\}$. When we have $u \geq 0$ then $\theta(u) = \inf_{x \in X} f(x) + ug(x) = \min z + uy$, $(y, z) \in G$, and it is a line. The minimization moves this line as much down until it supports G from below. $\theta(u)$ is the intersection point with z axis. Now, we have the interpretation of the dual problem: find a slope $u \geq 0$ s.t. the supporting hyperplane of G intersects z axis as high as possible. (draw a figure)



Example. *Linear programming (LP) problem*

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

and we choose $X = \{x, x \geq 0\}$. *The dual function is*

$$\theta(v) = \inf_{x \in X} c^T x + v^T(b - Ax) = \inf_{x \in X} \{(c - A^T v)^T x\} + v^T b = \begin{cases} v^T b, & c - A^T v \geq 0, \\ -\infty, & \textit{otherwise.} \end{cases}$$

The dual is another LP problem:

$$\begin{aligned} \max \quad & b^T v \\ s.t. \quad & A^T v \leq c. \end{aligned}$$

Also, the dual of QP problem is another QP (exercises) and there are other duals beside Lagrange dual problem. For example, Fenchel (defined soon) and Wolfe dual max $\phi(x,u,v)$ s.t. $\nabla_x \phi(x,u,v) = 0$, $u \geq 0$.

Duality theorems

Theorem (6.2.1, weak duality)**.** *If x is feasible for P and (u,v) feasible for D then $f(x) \geq \theta(u,v)$.*

Proof.

$$\theta(u,v) = \inf_{y \in X} f(y) + u^T g(y) + v^T h(y) \leq f(x) + u^T g(x) + v^T h(x) \leq f(x).$$

□

The dual function gives lower bound estimates for the primal problem. This also raises a question whether $f(x) = \theta(u,v)$ for some (x,u,v). We also have the following corollaries:
i) If $\inf f(x)$ s.t. x feasible is strictly larger than $\sup \theta(u,v)$ s.t. $u \geq \bar{0}$, then it is said that the problem has a duality gap, which is the difference of these two values.
ii) If we find feasible (x',u',v') s.t. $f(x') = \theta(u',v')$ then x' solves p and (u',v') solves D.
iii) If $\sup_{u \geq \bar{0}, v} \theta(u,v) = \infty$ then P does not have a feasible point.
When is the duality gap zero?

Theorem (6.2.4, strong duality)**.** *If X open, convex, f, g_i, $i \in I$ convex, $h(x) = Ax - b$ (affine), Slater's CQ holds: $\exists x' \in X$ s.t. $g(x') < \bar{0}$, $h(x') = 0$ and x' regular, i.e., $\bar{0} \in int\ h(X) = int\{h(x), x \in X\}$, then*

$$\inf\{f(x), x \in X, g(x) \leq \bar{0}, h(x) = \bar{0}\} = \sup\{\theta(u,v), u \geq \bar{0}\}.$$

If inf *is finite then $\exists u \geq \bar{0}, v$ s.t.* sup *is achieved.*
If inf *is achieved at x_0 with Lagrange multipliers (u_0, v_0) then $u_0^T g(x_0) = 0$.*

Proof. Proofs by separation theorems. □

There is no duality gap for convex problems.

Definition 7.1. *(x_0, u_0, v_0) is a* **saddle point** *of ϕ if $x_0 \in X$, $u_0 \geq \bar{0}$ and*

$$\phi(x_0, u, v) \leq \phi(x_0, u_0, v_0) \leq \phi(x, u_0, v_0), \ \forall x \in X, \forall (u,v), u \geq \bar{0}.$$

See zero-sum games for an application of saddle point results.

Theorem (6.2.5). *(x_0, u_0, v_0) is a saddle point for $\phi \Leftrightarrow$ x_0 solves P, (u_0, v_0) solves D and there is no duality gap $\Leftrightarrow$*
i) $\phi(x_0, u_0, v_0) = \min_{x \in X} \phi(x, u_0, v_0)$,
ii) $g(x_0) \leq \bar{0}$, $h(x_0) = \bar{0}$,
iii) $u_0^T g(x_0) = 0$.

Corollary. With the assumptions of strong duality, there is no duality gap $\Rightarrow$ x_0 solves P$\Rightarrow$ $\exists u_0 \geq \bar{0}, v$ s.t. (x_0, u_0, v_0) is a saddle point for ϕ.

Theorem (6.2.6, KKT and saddle). *If x_0 is a KKT point with Lagrange multipliers (u_0, v_0), f, g_i, $i \in I$ convex, h_j affine for $v_j \neq 0$, then (x_0, u_0, v_0) is a saddle point for ϕ. Conversely, if (x_0, u_0, v_0) is a saddle point of ϕ with $x_0 \in int\ X$, $u_0 \geq \bar{0}$, then x_0 is a KKT point with Lagrange multipliers (u_0, v_0).*

Properties of dual function

Theorem (6.3.1). *Define $\beta = (g, h)$ and $w = (u, v)$. If $X \neq \emptyset$ compact, f, β continuous then*

$$\theta(w) = \inf_{x \in X} f(x) + w^T \beta(x), \text{ is concave in } w.$$

Since θ is concave, from Theorem 3.4.2 we have that all local optimum are also global optimum.

Theorem (6.3.4). *If also*

$$x_0 \in C(w_0) = \{y \in \arg\min_{x \in X} f(x) + w_0^T \beta(x)\},$$

then $\beta(x_0) \in \partial\theta(w_0)$. If $C(w_0) = \{x_0\}$ is a singleton then $\nabla\theta(w_0) = \beta(x_0)$.

Thus, the primal constraints $\beta(x_0)$ give a subgradient to the dual function, which could be used in generating ascent directions in numerical methods. In general, (Ruszczynski: Nonlinear optimization, p. 165)

$$\partial\theta(w_0) = conv(\cup_{x_0 \in C'(w_0)} \beta(x_0)),$$

where $C'(w_0) = \{x \in X, \phi(x_0, w_0) = \theta(w_0)\}$.

Theorem (6.3.11). *The steepest ascent direction of θ is ξ with the smallest Euclidian norm:*

$$d = \begin{cases} \bar{0}, & \xi = \bar{0}, \\ \xi / \|\xi\|, & \xi \neq \bar{0}. \end{cases}$$

Interpretations of Lagrange multipliers

The Lagrange multipliers have different interpretations in applications. In electric circuits, the decision variables can be currents in primal problem and the dual variables can then be voltages (exercises). In economics, if the primal variables are levels of consumption, then the dual variables can be prices of different products or services. In mechanics of materials, the primal variables can be stress levels (strain) of some elements (in bridges or buildings) and the dual variables are displacement of the element. The following gives an interpretation in mechanical spring system.

Example. *Let us examine three spring system with two blocks between two walls. The spring constants are $k_1, k_2, k_3 > 0$ and the distance between the walls is l. The blocks have width w and they are centered at locations x_1 and x_2. The system will be in equilibrium at point where the potential energy is at minimum:*

$$\min \; J = 1/2k_1x_1^2 + 1/2k_2(x_2 - x_1)^2 + 1/2k_3(l - x_2)^2,$$

s.t. the blocks and walls are rigid: $w/2 - x_1 \leq 0$, $w + x_1 - x_2 \leq 0$, $w/2 - l + x_2 \leq 0$.

So we have a QP problem with convex objective (check!) and linear constraints. A suitable CQ condition in this case is the Slater's CQ which says that $2w \leq l$, which means that the blocks must fit between the walls. Now, the sufficient KKT conditions are

$$\begin{bmatrix} k_1x_1 - k_2(x_2 - x_1) \\ k_2(x_2 - x_1) - k_3(l - x_2) \end{bmatrix} + u_1 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + u_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + u_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \bar{0},$$

$u_1(w/2 - x_1) = 0$, $u_2(w - x_2 + x_1) = 0$, $u_3(w/2 - l + x_2) = 0$, *and* $u_1, u_2, u_3 \geq 0$.

The interpretation of Lagrange optimality is that the forces are in equilibrium:

$$k_1x_1 - k_2(x_2 - x_1) - u_1 + u_2 = 0.$$

The complementary slackness conditions mean that the contact forces are active only when the blocks touch each other or the walls. The dual feasibility means that the contact forces are away from the contact surface.

We can see that the minimum potential solution equals the force balance equations (KKT), and this result has been used in the basic physics courses. We can now see the meaning of convexity for getting this result.

Fenchel conjugate duality

Conjugate function is the basic tool in convex analysis.

Definition 7.2. *A* **conjugate function** *is*

$$f^*(u) = \sup_{x \in \mathbf{R}^n} x^T u - f(x),\ u \in \mathbf{R}^n.$$

Example. *if* $f(x) = ax - b$ *then* $f^*(u) = \begin{cases} b, & u = a, \\ \infty, & u \neq a. \end{cases}$
if $f(x) = |x|$ *then* $f^*(u) = \begin{cases} 0, & |u| \leq 1, \\ \infty, & |u| > 1. \end{cases}$
if $f(x) = (c/2)x^2$ *then* $f^*(u) = u^2/(2c)$.

Let us derive the dual for the following problem

$$\min_{x \in X_1 \cap X_2} f_1(x) - f_2(x),$$

$$\max_{u \in \Omega_1 \cap \Omega_2} f_2^*(u) - f_1^*(u),$$

where $\Omega_1 = \{u, f_1^*(u) < \infty\}$ and $\Omega_2 = \{u, f_2^*(u) > -\infty\}$. This can be shown for example using Lagrange duality:

$$\min_{z=y} f_1(y) - f_2(z),$$

and the Lagrange dual function of this problem is

$$\begin{aligned} \theta(u) &= \inf_{y_i n X_1, z \in X_2} f_1(y) - f_2(z) + (z-y)^T u, \\ &= \inf_{z \in X_2} z^T u - f_2(z) + \inf_{y \in X_1} f_1(y) - y^T u, \\ &= f_2^*(u) - f_1^*(u). \end{aligned}$$

This is the classical interpretation of duality. (draw a figure)

8 Numerical methods for unconstrained problems

Optimization is one of the important fields in numerical computation, beside solving differential equations and linear systems. We can see that these fields are not independent and they share the algorithms and the ideas: solving (large) linear optimization problems equals to solving general linear equations ($Ax = b$), solving nonlinear unconstrained problems equals to solving a set of nonlinear equations ($f(x) = \bar{0}$), and solving dynamic optimization problems equals solving (partial) differential equations.

Next, we examine how to solve different types of optimization problems. What methods work in certain class of problems and why? Different approaches are presented for unconstrained and constrained, one-dimensional and multidimensional problems. The focus is on methods for finding local minimum, and the global methods or heuristics (simulated annealing, genetic algorithms etc.) are not presented on this course. Numerical methods are iterative algorithms that try to solve the problem using finite number of operations. The algorithms produce a sequence $\{x_t\}$, where the next solution is given by some rule and the information up to that point:

$$x_{t+1} = X_{t+1}(I_0, I_1, \ldots, I_t),$$

where I_t is the information from iteration t.

Definition 8.1. Algorithmic map *$A : X \to 2^X$ maps each point to a set of possible next iterates: $x_{k+1} \in A(x_k)$.*

A final iterate x^* is called **solution** and Ω is the solution set. Solution is **acceptable** if

- x^* is local optimum or FJ/KKT point
- $f(x^*) < b$ acceptable value
- $f(x^*) < LB + \epsilon$, LB some lower bound
- $f(x^*) < OPT + \epsilon$

Closed maps

Definition 8.2. *A map A is* **closed** *at $x \in X$ if $x_k \in X$, $\{x_k\} \to x$ and $y_k \in A(x_k)$, $\{y_k\} \to y$ implies that $y \in A(x)$. The map A is closed on $Z \subseteq X$ if it is closed at each point in Z.*

Definition 8.3. *A function $\alpha : X \to \mathbf{R}$ is a* **descent function** *if $\alpha(y) < \alpha(x)$ when $x \notin \Omega$ is not a solution and $y \in A(x)$.*

Theorem (7.2.3)**.** *If map A is closed over the complement of Ω and α is continuous descent function, then either the algorithm stops in a finite number of steps or it generates an infinite sequence $\{x_k\}$ such that*

- *every convergent subsequence of $\{x_k\}$ has a limit in Ω*
- *$\alpha(x_k) \to \alpha(x)$ for some $x \in \Omega$*

Note that the sequence must converge to the single value if Ω is a singleton.
Stopping condition

Typical stopping conditions are

- $\|x_{k+N} - x_k\| < \epsilon$
- $\|x_{k+1} - x_k\|/\|x_k\| < \epsilon$
- $\alpha(x_k) - \alpha(x_{k+N}) < \epsilon$
- $(\alpha(x_k) - \alpha(x_{k+N}))/|\alpha(x_k)| < \epsilon$

Criteria to compare the methods

We can classify and compare the methods using the following criteria:

1. The required information:
   - The zero-order methods use only the values of objective and constraint functions.
   - The first-order use also gradients of objective and constraints.
   - The second-order use also the Hessians.
2. The convergence properties:
   Let $\{s_k\}$ be a sequence and $s_k \to s'$, when $k \to \infty$.

   Definition 8.4. *The* **order of convergence** *is*

$$p = \sup\{q \in \mathbf{R}^+, \lim_{k\to\infty} \sup \frac{|s_{k+1} - s'|}{|s_k - s'|^q} < \infty\},$$

   where $\lim_{k\to\infty} \sup s_k = \lim_{k\to\infty} \sup\{s_m, m \geq k\}$.

   Definition 8.5. *The* **convergence ratio** *is*

$$\beta = \limsup \frac{|s_{k+1} - s'|}{|s_k - s'|^p}.$$

   - sublinear convergence: $p = 1, \beta = 1$,
   - linear convergence: $p = 1, \beta < 1$,
   - superlinear convergence: $p \geq 1, \beta = 0$, ($p = 1$ and $\beta = 0$, or $p > 1$)
   - quadratic convergence: $p = 2, \beta < \infty$.

Example. *Series* $s_k = 1/k$ *converges sublinearly as* $\beta = \lim k/k + 1 = 1$. *Series* $s_k = 1/k^k$ *converges at least superlinearly since*

$$\frac{s_{k+1}}{s_k} = \frac{k^k}{(k+1)^{k+1}} \leq \frac{k^k}{k^{k+1}} = \frac{1}{k} \to 0.$$

The higher convergence order and the smaller ratio is faster. Convergence is rather theoretical notion and it may be difficult to determine exactly for an algorithm.

3. The computational complexity:
   The required computational effort can be measured by the number of basic operations, like additions and multiplications.

   Definition 8.6. *A function* $f(x)$ *is* $O(g(x))$ *iff* $\exists c, n_0$ *s.t.* $|f(x)| < c|g(x)|$, *when* $x > n_0$.

4. The need for memory: does it need vectors or matrices to be stored?

5. The generality: does it solve all problems in certain class or just some specific cases?

6. Stability: how do the rounding errors during computation and inaccuracies in the original data affect the algorithm?

Line search methods

Difficult optimization problems are typically reduced to a set of easier problems. Constrained problems can be converted to a series of unconstrained problems with penalty and barrier functions. Multidimensional unconstrained problems can be solved by line search methods that generate a series of one-dimensional problems. Thus, solving line search problems efficiently is important for large class of optimization problems.

We examine a problem min $l(s) = f(x_k + sd_k)$, where s is the parameter to be optimized, which can be from some multidimensional minimization problem with objective $f(x)$, where x_k is the current iteration, d_k the descent (search) direction and s the step length. Typically, the step length is restricted to $s \in S = \{s, s \geq 0\}$ or $s \in [a, b]$ that is called the interval of uncertainty where the optimum lies.

Zero-order methods

Assume that $l(s)$ is strictly quasiconvex in s. The minimum can then be found with the following result. Let $\lambda < \mu$ then

$$\begin{aligned} i)\ l(\lambda) > l(\mu) &\Rightarrow l(z) \geq l(\mu),\ \forall z \leq \lambda, \\ ii)\ l(\lambda) \leq l(\mu) &\Rightarrow l(z) \geq l(\lambda),\ \forall z \geq \lambda, \end{aligned}$$

This means that in case i) the minimum cannot be between $a \leq z \leq \lambda$ and in case ii) between $\mu < z \leq b$, and the interval of uncertainty can be updated. This gives the following methods:

- **Uniform search**: choose points uniformly between $[a, b]$.
- **Dichotomous search**: choose $\delta > 0$, pick $\lambda = (a+b)/2-\delta$, $\mu)(a+b)/2+\delta$, evaluate $l(\lambda)$, $l(\mu)$ and update.
- **Golden section**: choose $\lambda = a + (1-\alpha)(b-a)$, $\mu = a + \alpha(b-a)$, $\alpha = (\sqrt{5}-1)/2 \approx 0.618$.
- **Fibonacci**: $F_0 = F_1 = 1$, $F_{i+1} = F_i + F_{i-1}$, choose $\lambda = a + F_{n-k-1}/F_{n-k+1}(b-a)$, $\mu = a + F_{n-k}/F_{n-k+1}(b-a)$.
- **Quadratic fit**: Using three points $s_1 < s_2 < s_3$, $l(s_1) \geq l(s_2)$, $l(s_3) \geq l(s_2)$, fit a second order polynomial (parable) $p(s)$ s.t. $p(s_i) = l(s_i)$ and find the minimum s^* for the parable. Evaluate $l(s^*)$, update and repeat.

Comparison: dichotomous has linear convergence with $\beta \approx \sqrt{1/2} \approx 0.71$, golden section and Fibonacci linear with $\beta \approx 0.618$ and quadratic fit superlinear with $p \approx 1.3$ (under certain assumptions).

First-order methods

Assume that $l(s)$ is differentiable, ps.convex, i.e. $l'(s_0) = 0 \Rightarrow s_0$ minimum.

- **Bisection** (Bolzano's method): Choose $s_k = (a+b)/2$. If $l'(s_k) < 0$ then $s^* > s_k$, or if $l'(s_k) > 0$ then $s^* < s_k$, otherwise $s_k = s^*$.
- **Cubic fit**: Calculate and fit according to $p(a) = l(a)$, $p'(a) = l'(a)$, $p(b) = l(b)$, $p'(b) = l'(b)$. Find the minimum for the third-degree polynomial $p(s)$, find the minimum s^* for $p(s)$, calculate $l'(s^*)$ and update.

Second-order methods

The Newton's method solves the quadratic approximation

$$\min \; l(s_k) + l'(s_k)(s-s_k) + 1/2 l''(s_k)(s-s_k)^2,$$

which gives an update

$$s_{k+1} = s_k - l'(s_k)/l''(s_k).$$

This can also be seen as solving the necessary condition $g(s) = l'(s) = 0$ by using the linear approximation $g(s) \approx g(s_k) + g'(s_k)(s-s_k) = 0$.

Comparison: Bisection method converges linearly with $\beta = 0.5$, cubic polynomial with quadratic convergence $p = 2$ (under certain assumptions) and Newton by quadratic convergence $p = 2$ (sufficiently close to the optimum).

Inexact line search

When the line search is solved as a subproblem of some larger problem, it is not necessary to find the minimum exactly but rather get fast some good enough solution. In terms of total complexity, it is better to use less computation and only few steps in each line search. The inexact line search methods define the sufficient conditions that the good enough solutions satisfy.

Definition 8.7. Armijo's rule: *The step length s is accepted and it descends enough if (draw a figure)*

$$f(x_k + sd_k) \leq f(x_k) + \epsilon s \nabla f(x_k)^T d_k, \ \epsilon \in (0, 1),$$

i.e., $l(s) \leq l(0) + \epsilon s l'(0)$. Typically, ϵ is small ($(10^{-5}, 10^{-1}), 0.2, 10^{-4}$ depending on the source). Note that the course book adds an additional requirement to prevent small step sizes: accept s if

$$l(\alpha s) \geq l(0) + \alpha \epsilon s l'(0), \ \alpha > 1,$$

for example $\alpha = 2$.

Definition 8.8. Goldstein rule: *accept s if*

$$l(0) + (1 - c)sl'(0) \leq l(s) \leq l(0) + csl'(0), \ c \in (0, 1/2).$$

Note that this equals the Armijo's rule when $l(s)$ is convex.

Definition 8.9. Wolfe's rule: *accept s if*

$$\begin{aligned} l(s) &\leq l(0) + \epsilon s l'(0), \ \textit{(Armijo)} \\ l'(s) &\geq \sigma l'(0), \ 0 < \epsilon < \sigma < 1 \\ |l'(s)| &\geq \sigma |l'(0)|. \ \textit{(strong Wolfe)} \end{aligned}$$

Multidimensional search

Two approaches are examined in solving multidimensional problems: line search methods (gradient, Newton and their modification) and trust-region methods. Typically, the line search methods generate a direction and do a search in this direction. The methods differ in how the search direction is chosen. Trust-region methods are also called as restricted step methods, where the objective is approximated often with a quadratic function that is minimized and the new point should

be inside the current trust region. The region is expanded depending on how well the quadratic function approximates the objective.

Zero-order methods

- **Cyclic coordinate method**: use coordinate axes as search directions and search them in order. The method does not work well if the function is sideways to the coordinate axes.
- **Hooke-Jeeves**: add an acceleration step to the previous method
- **Nelder-Meade Simplex**: update a simplex based on the function values at the corners (amoeba search)
- **Finite difference methods**: use higher order methods by using difference approximations

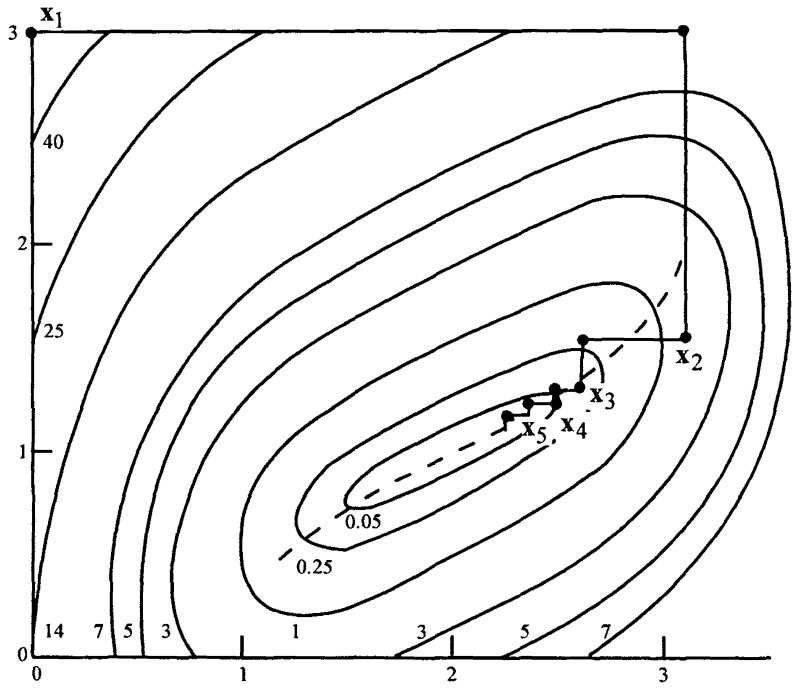


Figure 8.7 Cyclic coordinate method.

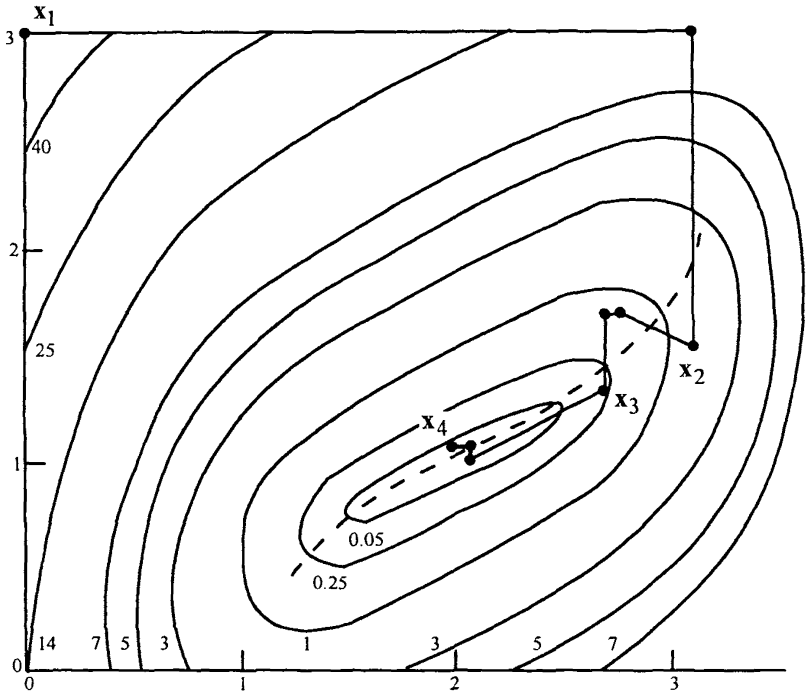


Figure 8.10 Method of Hooke and Jeeves using line searches. Method of Hooke and Jeeves with Discrete Steps

Gradient method

The gradient method is a first-order method that was originally proposed by Cauchy in 1847. When a function is differentiable then a direction d is a descent direction when

$$f'(x;d) = \lim_{s \to 0} \frac{f(x+sd) - f(x)}{s} = \nabla f(x)^T d < 0.$$

The gradient method uses the negative gradient as an update direction

$$x - x_k = -\nabla f(x_k).$$

Theorem (8.6.1)**.** *If $\nabla f(x) \neq \bar{0}$ then the steepest descent direction is*

$$\min_{\|d\| \leq 1} f'(x;d) \Rightarrow \bar{d} = \frac{-\nabla f(x)}{\|\nabla f(x)\|}.$$

Proof.

$$f'(x;d) = \nabla f(x)^T d \geq -\|\nabla f(x)\|\|d\| \geq -\|\nabla f(x)\|,$$

where the first is by Cauchy-Bunyakovsky-Schwarz inequality and the second hold as equality only if $d = \bar{d} = \frac{-\nabla f(x)}{\|\nabla f(x)\|}$. □

The steepest descent method does a line search

$$x_{k+1} = x_k - s_k \nabla f(x_k),$$

where $s_k \in \arg\min_{s \geq 0} f(x_k + s\nabla f(x_k))$ or some inexact line search, or simply $s_k = 1$ like in the gradient method. The stopping condition is for example when $\|\nabla f(x_k)\| < \epsilon$, for some $\epsilon > 0$.

Properties:

- with exact line search, $\nabla f(x_k + 1)^T \nabla f(x_k) = 0$, and it means zigzagging
- easy to program and reliable
- affected by change of variables $x' = Mx$
- example of convex problem where the method does not converge
- linear convergence that depends on condition number $\kappa = \lambda_n/\lambda_1$, where λ_n is the largest and λ_1 the smallest eigenvalue. $(\frac{\kappa-1}{\kappa+1})^2$ <converg. ratio< 1
- the eigenvectors, eigenvalues, and the condition number tells how the objective function is tilted and scaled in different directions

Theorem. *When steepest descent method (exact line search) is applied to*

$$f(x) = 1/2 x^T Q x - b^T x,$$

Q symmetric positive definite, then the error norm

$$1/2\|x - x^*\|_Q^2 = f(x) - f(x^*),$$

satisfies

$$\|x_{k+1} - x^*\|_Q^2 \leq \left[\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right]^2 \|x_k - x^*\|_Q^2 = \left[\frac{\kappa - 1}{\kappa + 1}\right]^2 \|x_k - x^*\|_Q^2,$$

where $0 < \lambda_1 \leq \ldots \leq \lambda_n$ are the eigenvalues of Q.

Definition 8.10. *The* **weighted norm** $\|x\|_P = (x^T P x)^{1/2} = \|P^{1/2} x\|_2$, *where P symm. pos.def.*

For quadratic function, the ratio is $r = \frac{k-1}{k+1}$ but in general larger than r^2.

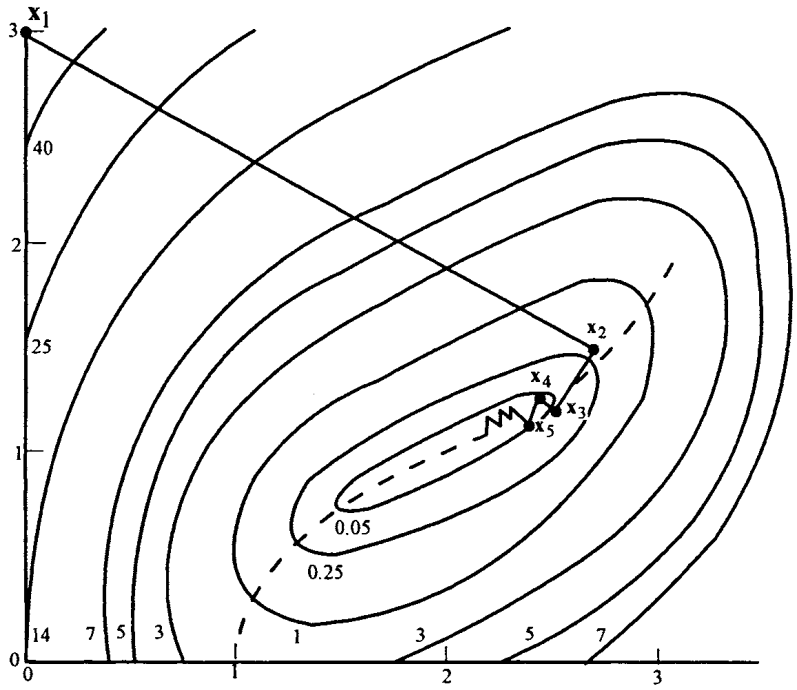


Figure 8.16 Method of steepest descent.

4.1. Twisted Function. We now describe the twisted function, whose level curves are shown in Fig. 2. The figure also shows the sequence generated by the Cauchy algorithm, to be described below. In the concluding remarks, we shall explain how the function was designed.

The twisted function is defined in $\mathbb{R}^2$, denoted $(x, y) \mapsto f(x, y)$. Let $\alpha \in (0, 0.25)$ be a constant, say $\alpha = 1/8$, and let

$$\Omega = \{z \in \mathbb{R}^2 \mid \|z\|_\infty < 1/\alpha\}.$$

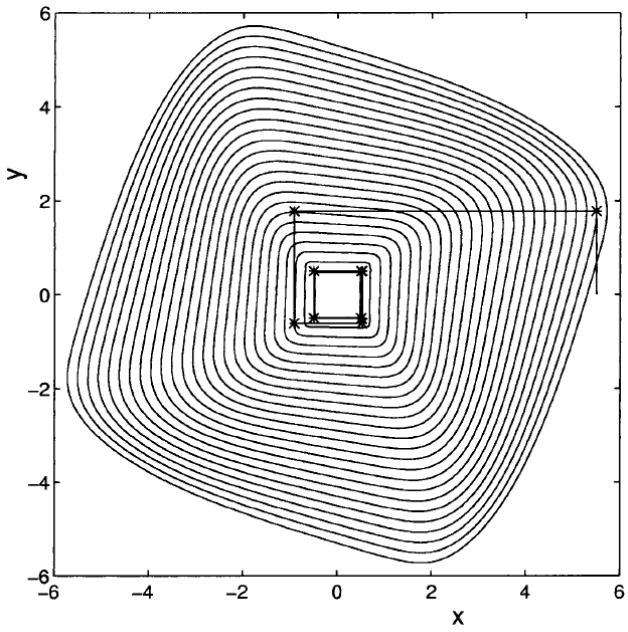


Fig. 2. Behavior of the Cauchy algorithm on the twisted function.

Given a constant $b > 0$, we define, for $\lambda \in \mathbb{R}$,

$$r(\lambda) = \lambda + b, \quad \text{if } \lambda < -b,$$
$$r(\lambda) = 0, \quad \text{if } \lambda \in [-b, b],$$
$$r(\lambda) = \lambda - b, \quad \text{if } \lambda > b.$$

Note that, if $b = 0$,

$$r(\lambda) \equiv \lambda.$$

Let us fix $b = 0.5$, and define

$$f(x, y) = f_1(x, y) + f_2(x, y),$$

where

$$f_1(x, y) = \{r(x)^2/[1 + \alpha y \operatorname{sign}(r(x))]\}^2,$$
$$f_2(x, y) = \{r(y)^2/[1 - \alpha x \operatorname{sign}(r(y))]\}^2.$$

This function is clearly defined in Ω, because by definition of Ω the denominators above are always positive. Its value is null in the set

$$\Omega^* = \{z = (x, y) \in \mathbb{R}^2 \mid r(x) = 0, r(y) = 0\}$$
$$= \{z \in \mathbb{R}^2 \mid \|z\|_\infty \leq 0.5\},$$

and is positive out of it. It follows that Ω^* is the optimal set.

The gradients of f_1 and f_2 at $z = (x, y)$ are

$$\nabla f_1(z) = \begin{bmatrix} 4r(x)^3/[1 + \alpha y \operatorname{sign}(r(x))]^2 \\ -2\alpha r(x)^4 \operatorname{sign}(r(x))/[1 + \alpha y \operatorname{sign}(r(x))]^3 \end{bmatrix},$$

$$\nabla f_2(z) = \begin{bmatrix} 2\alpha r(y)^4 \operatorname{sign}(r(y))/[1 - \alpha x \operatorname{sign}(r(y))]^3 \\ 4r(y)^3/[1 - \alpha x \operatorname{sign}(r(y))]^2 \end{bmatrix}.$$

Note that f_1 and f_2 are very similar. These gradients are obviously null in Ω^*.

The Hessian matrix for f_1 is given by

$$H_1(z) = \begin{bmatrix} 12r(x)^2/(1 \pm \alpha y)^2 & \mp 8\alpha r(x)^3/(1 \pm \alpha y)^3 \\ \mp 8\alpha r(x)^3/(1 \pm \alpha y)^3 & 6\alpha^2 r(x)^4/(1 \pm \alpha y)^4 \end{bmatrix},$$

where the signs depend on $\operatorname{sign}(r(x))$. The Hessian of f_2 is similar, and the Hessian determinants are

$$\det(H_1(z)) = 8\alpha^2 r(x)^6/(1 \pm \alpha y)^6,$$
$$\det(H_2(z)) = 8\alpha^2 r(y)^6/(1 \mp \alpha x)^6.$$

Convergence of steepest descent

Definition 8.11. *Function f is* **Lipschitz continuous** *with constant G if $\|f(x) - f(y)\| \leq G\|x - y\|$.*

As a line search algorithm, it will converge as long as f is continuous and differentiable and line search is exact.

A version of Armijo's rule is also guaranteed to converge as long as $\nabla f(x)$ is Lipschitz continuous with constant $G > 0$.

Newton and modified methods

Newton's method can be interpreted in the following ways:

1. Linear approximation to equations:
   Let us examine solving a nonlinear system of equations, $g : \mathbf{R}^n \mapsto \mathbf{R}^m$,

$g(x) = \bar{0}$. The linear approximation gives

$$g(x) \approx g(x_k) + H(x_k)(x - x_k) = \bar{0},$$

$$x_{k+1} = x_k - H_k^{-1} g(x_k).$$

We apply this to function $g(x) = \nabla f(x)$ and H_k is symmetric.

2. Minimize quadratic approximation:
   The above are the same as the necessary conditions for

$$\min q(x) = f(x_k) + \nabla f(x_k)^T (x - x_k) + 1/2 (x - x_k)^T H_k (x - x_k).$$

   The same equations can be interpreted as minimizing the quadratic (Taylor) approximation or solving the linear approximation of the necessary conditions.

   The idea is to take a suitable step s_k in the direction of $d_k = -H_k^{-1} f(x_k)$, i.e., $x_{k+1} = x_k + s_k d_k$. In minimization the direction is

$$d_k = -\nabla^2 f(x_k)^{-1} \nabla f(x_k),$$

   and the update can be written

$$\nabla^2 f(x_k)(x - x_k) = -\nabla f(x_k).$$

3. Steepest descent method in a local Hessian norm:
   Using the negative gradient is the steepest descent method in Euclidean norm. The Newton's method can be seen as steepest descent method in a norm induced by the local Hessian. In general, if we make a coordinate change by matrix P, the corresponding norm is $\|z\|_P = (z^T P z)^{1/2} = \|P^{1/2} z\|_2$, when P is symmetric and positive definite. The steepest descent method in this norm is

$$\Delta x_{sd} = -P^{-1} \nabla f(x),$$

   and $\|z\|_{\nabla^2 f(x)} = (z^T \nabla^2 f(x) z)^{1/2}$ which gives $\Delta x_{sd} = -\nabla^2 f(x)^{-1} \nabla f(x)$. This is very good search direction when $x \approx x^*$, it changes the condition by decreasing the eccentricity and converges in one step for quadratic function (like gradient method for function with condition $\kappa = 1$).

If the search direction is a descent direction, it is natural to use a line search. When $\nabla^2 f(x_k)$ is positive definite then d_k is a descent direction. Note that if $s_k = 1$, $\forall k$ then the method in general converges only locally. The problem is when $\nabla^2 f(x_k)$ is not invertible. Then modified Newton methods can be used,

where we replace $H_k = \nabla^2 f(x_k) + \epsilon_k I$, where ϵ_k is large enough so that H_k is positive definite. The update can be written:

$$(\nabla^2 f(x_k) + \epsilon_k I)(x - x_k) = -\nabla f(x_k).$$

See the connection to Levenberg-Marguardt method.

It can be seen that when ϵ_k is large the method is close to the steepest descent method, whereas when ϵ_k is small the method is close to the Newton's method.

Properties of Newton's method:

- quadratic convergence when started close enough to the optimum
- matrix inversion $O(n^3)$
- needs Hessian and memory for the matrices
- affine invariant $y = Px$
- many convergence results, e.g., if $\nabla^2 f$ positive definite and the lower level sets are bounded then exact, Armijo/Goldstein inexact methods converge to the unique global minimum.

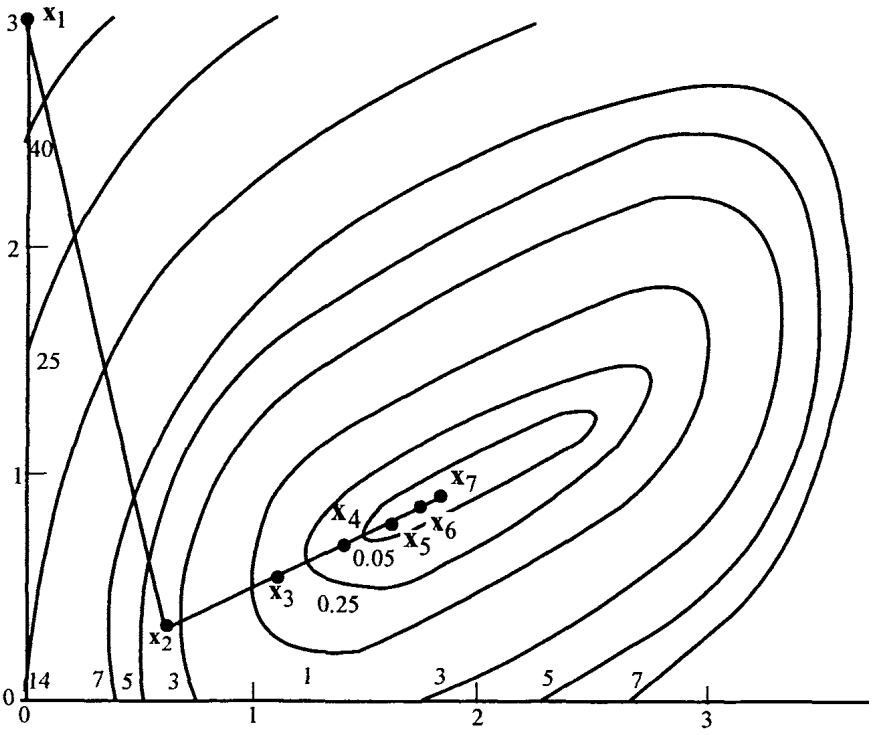


Figure 8.18 Method of Newton.

9 Conjugate gradient methods

The gradient method has the problem of zigzagging and slow convergence. The conjugate gradient methods try to solve this problem by using conjugate rather than orthogonal directions. These methods are especially useful for solving large problems. It is also an alternative to Gaussian elimination in solving linear systems.

Definition 9.1. *Let H be symmetric $n \times n$ matrix. Directions $d_1, \ldots, d_k$ are* **H-conjugate** *if $d_i^T H d_j = 0$, $\forall i \neq j$ and $d_1, \ldots, d_k$ linearly independent.*

Note that if H is positive definite and $d_i^T H d_j = 0$ then $d_1, \ldots, d_k$ are linearly independent. This means that it is advantageous to maintain positive definity in order to produce conjugate directions, which we see later on in quasi-Newton methods.

Theorem (8.8.3)**.** *Let $f(x) = 1/2 x^T H x + b^t x + c$, H symmetric, positive definite. If f is minimized consecutively in n H-conjugate directions then the minimum is found at most in n-th step.*

Proof. The sufficient condition is $\nabla f(x^*) = Hx^* + b = \bar{0}$ (1). Since H is pos.def, $d_1, \ldots, d_n$ are linearly independent. Thus, $\exists \beta_i$, $1 \leq i \leq n$ s.t. $x^* = x_0 + \sum_{i=1}^n \beta_i d_i$.

From (1), $Hx_0 + \sum \beta_i H d_i + b = \bar{0}$. Let us multiply this equation by d_j^T and we get $d_j^T H x_0 + \sum \beta_i d_j^T H d_i + d_j^T b = \bar{0}$ and $\beta_j = -\frac{(Hx_0+b)^T d_j}{d_j^T H d_j}$.

What do the line searches produce? s_j' s.t. $f'(x_j+s_j d_j) = 0 \Rightarrow \nabla f(x_{j+1})^T d_j = 0$ and $x_{j+1} = x_j + s_j + d_j$, $\nabla f(x_{j+1}) = Hx_{j+1} + b \Rightarrow (Hx_j + s_j H d_j + b)^T d_j = 0 \Leftrightarrow s_j = -\frac{(Hx_j+b)^T d_j}{d_j^T H d_j}$. Since $x_j = x_0 + \sum_{i=1}^{j-1} s_i d_i$ then

$$s_j = -\frac{(Hx_0 + \sum_{i=1}^{j} s_i H d_i + b)^T d_j}{d_j^T H d_j} = -\frac{(Hx_0 + b)^T d_j}{d_j^T H d_j} = \beta_j.$$

□

Note the connection to Krylov subspaces $\{d_0, Ad_0, A^2 d_0, \ldots, A^{i-1} d_0\}$. How do we produce the conjugate directions?

The algorithm for conjugate gradient (CG) methods:

$$x_{k+1} = x_k + s_k d_k,$$

where s_k is from exact or inexact line search. The search direction is

$$d_{k+1} = -\nabla f(x_{k+1}) + a_k d_k, \quad (1)$$

where a_k is given by some specific equation depending on which CG method is used. There can also be a restart in every n rounds when $d_k = -\nabla f(x_k)$ is set. There are three main CG methods: Hestens-Stiefel (HS), Polak-Ribiere (PR) and Fletcher-Reeves (FR), which can be derived by making certain assumptions on the objective function.

Multiplying (1) by Hd_k, we get $d_{k+1}^T H d_k = -\nabla f(x_{k+1})^T H d_k + a_k d_k^T H d_k$, from which

$$a_k = \frac{\nabla f(x_{k+1})^T H d_k}{d_k^T H d_k},$$

since d_k are H-conjugate directions. It is not efficient to determine H explicitly and Hd_k is often replaced with $\frac{\nabla f(x_{k+1}) - \nabla f(x_k)}{s_k}$, which are equal when the function is quadratic. With the substitution,

$$a_k = \frac{\nabla f(x_{k+1})(\nabla f(x_{k+1}) - \nabla f(x_k))}{d_k^T(\nabla f(x_{k+1}) - \nabla f(x_k))}. \text{ (HS)}$$

This is the Hestens-Stiefel (1952) update. This was used to solve linear equations $Ax = b$ when A is pos.def. If the **linesearch is exact**, then $d_k^T \nabla f(x_{k+1}) = 0$ and from (1): $-d_k^T \nabla f(x_k) = \nabla f(x_k)^T \nabla f(x_k) + a_{k-1} d_{k-1}^T \nabla f(x_k)$, where the last term is then zero. Now,

$$a_k = \frac{\nabla f(x_{k+1})(\nabla f(x_{k+1}) - \nabla f(x_k))}{\nabla f(x_k)^T \nabla f(x_k)}. \text{ (PR)}$$

Polak-Ribiere (1969) method is said to be the correct formula when the objective function is not quadratic. If f is **quadratic** then $\nabla f(x_{k+1})^T d_i = 0$, $\forall k$, $0 \leq k \leq n-1$, $0 \leq i \leq k$. Thus, $\nabla f(x_{k+1})^T d_k = -\nabla f(x_{k+1})^T \nabla f(x_k) + a_{k-1} \nabla f(x_{k+1})^T d_{k-1}$, where the first and last terms are zero. We get

$$a_k = \frac{\nabla f(x_{k+1})^T \nabla f(x_{k+1})}{\nabla f(x_k)^T \nabla f(x_k)} = \frac{\|\nabla f(x_{k+1})\|^2}{\|\nabla f(x_k)\|^2}. \text{ (FR)}$$

Fletcher-Reeves (1964) was used in solving nonlinear equations.

Theorem. *If H has only r distinct eigenvalues, then CG will terminate at the solution x^* in at most r iterations.*

Theorem. *If H has eigenvalues $\lambda_1 \leq \ldots \leq \lambda_n$,*

$$\|x_{k+1} - x^*\|_H^2 \leq \left[\frac{\lambda_{n-k} - \lambda_1}{\lambda_{n-k} + \lambda_1}\right]^2 \|x_0 - x^*\|_H^2.$$

The eigenvalues and their clustering determine the speed of convergence.

Example. *If the eigenvalues of H consist of m large values and the remaining $n - m$ smaller ones around 1. Then after $m + 1$ steps CG will produce a good estimate of the solution after only $m + 1$ steps.*

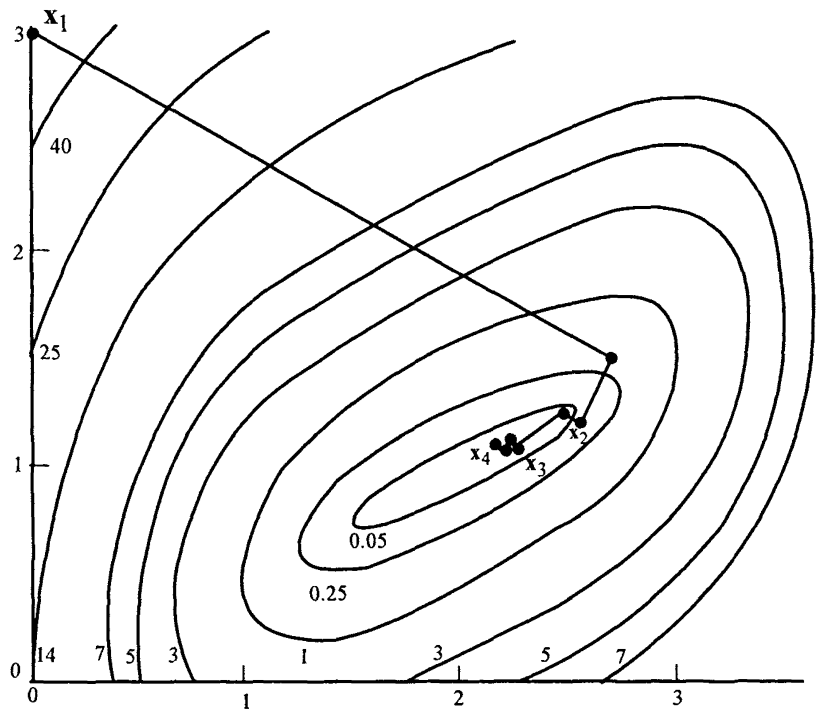


Figure 8.23 Method of Fletcher and Reeves.

Theorem (8.8.8). *If f is quadratic and (FR) is used then $d_1, \ldots, d_n$ are H-conjugate and descent directions.*

Properties:

- no need to store matrices, good for large problems
- exact line search critical for some methods
- if $\nabla f(x^*)$ pos.def. then superlinear convergence
- $(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1})^2$ <convergence ratio< 1 (compare to the gradient method)
- eigenvalues can be changed by preconditioning $x' = Cx$, C nonsingular
- if quadratic then quasi-Newton methods produce conjugate directions when using exact line search, and then the Hessian is approximated precisely after n steps.

Application: solve $Ax = b$, $A \in \mathbf{R}^{n\times n}$ invertible. solve $\min 1/2x^TA^TAx - b^TAx$ and its necessary condition $A^TAx - (b^TA)^T = \bar{0}$. Solution in at most n steps with CG method.

Quasi-Newton methods

The Newton method inverts a matrix and it requires a lot of computation. This can be improved by approximating the Hessian with the gradient information. These methods are also called as variable metric methods. The idea is to build a quadratic model which is sufficiently good to get superlinear convergence.

Definition 9.2. *H_k satisfies the* **quasi-Newton condition** *if*

$$H_k(x_{k+1} - x_k) = \nabla f(x_{k+1}) - \nabla f(x_k). \quad (H_1 = I)$$

Example.

$$H_{k+1} = \frac{f'(x_{k+1}) - f'(x_k)}{x_{k+1} - x_k},$$

$f : \mathbf{R} \mapsto \mathbf{R}$, $H_1 = 1$. *This is the secant method, which has a superlinear convergence ($p \approx 1.618$ under certain assumptions).*

Note that H_{k+1} is $n \times n$ matrix and update $H_{k+1} = H_k + M_k$, so quasi-Newton condition does not determine H_{k+1} uniquely. There are many quasi-Newton methods, and Broyden-Fletcher-Goldfarb-Shanno (BFGS) method can be derived by making the following assumptions. See course website for the history of discovering the method.

Denote $s_k = x_{k+1} - x_k$, $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$. Assume H_{k+1} is symmetric and positive definite $\Rightarrow H_{k+1} = J_{k+1}J_{k+1}^T$, where J is non-singular and $H_k = L_k L_k^T$ (Cholesky decomposition with lower triangular L). BFGS update solves

$$\begin{aligned} \min \quad & \|J_{k+1} - L_k\|_F \\ s.t. \quad & J_{k+1}J_{k+1}^T y_k = s_k, \end{aligned}$$

where $\|A\|_F = \sqrt{\sum_{i,j} a_{ij}^2}$ is the Fröbenius norm, and the constraint is the quasi-Newton condition. This has a unique solution as the objective is strictly convex and the constraint is affine.

$$H_{k+1} = H_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{H_k s_k s_k^T H_k}{s_k^T H_k s_k},$$

and similar update equation for the inverse of $H_k^{-1} = B_k$.

Davidon-Fletcher-Powell (DFP) method can be seen as a "dual" of BFGS where s_k and y_k are interchanged and a similar equation to B_k:

$$B_{k+1} = B_k + \frac{s_k s_k^T}{y_k^T s_k} - \frac{B_k y_k y_k^T B_k}{y_k^T B_k y_k}.$$

These two give a family of Broyden methods: $B_{k+1} = \alpha B_{k+1}^{BFGS} + (1-\alpha) B_{k+1}^{DFP}$.

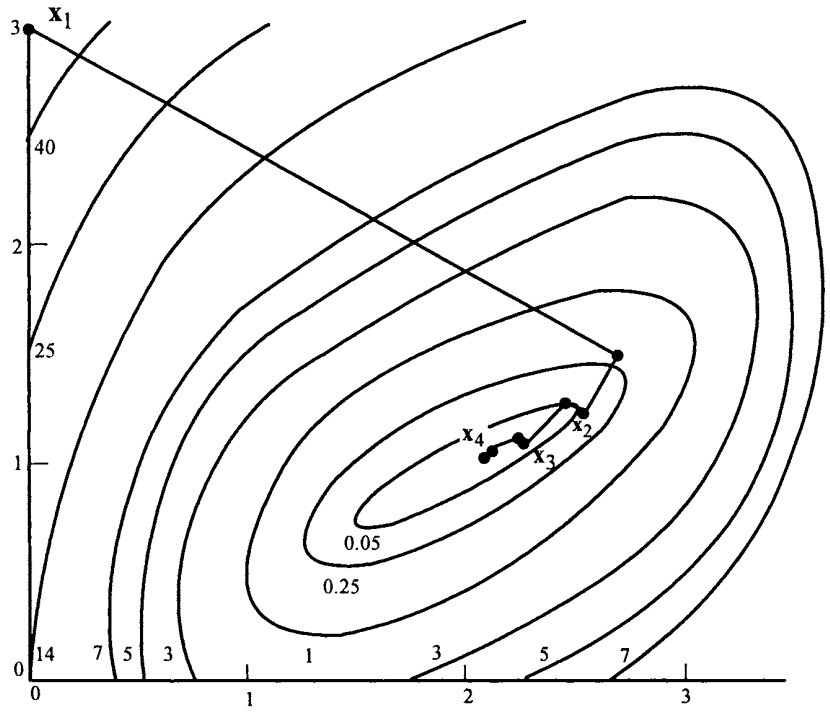


Figure 8.22 Davidon–Fletcher–Powell method.

Properties:

- local superlinear convergence
- update in $O(n^2)$
- if B_1 symmetric, pos.def. and exact line search then B_{k+1} symm. and pos.def. and d_k are descent directions
- it may happen that $B_k \not\to \nabla^2 f(x^*)$
- BFGS adapts better than DFP
- DFP is critical to exact line search
- same behavior for Broyden family for convex QP

Note the connection to Sherman-Morrison-Woodbury formula. If A^{-1} is known, it is easy to compute rank-1 update:

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^TA^{-1}}{1 + v^TA^{-1}u}.$$

Quasi-Newton update consists of two rank-one matrices, and thus H_k undergoes a rank-2 modification in each iteration.

Trust-region methods

Trust-region methods are good alternatives for line search methods. In these methods f is approximated with a quaratic function

$$q(x_k + s) = f(x_k) + \nabla f(x_k)^T s + 1/2 s^T H_k s,$$

where H_k is the Hessian or some (quasi-Newton) approximation. This approximation is relied only inside some trust region

$$\Omega_k = \{x, \|x - x_k\| \leq \Delta_k\},$$

where Δ_k is the radius. Depending on how well the quadratic function fits the objective, the size of trust region is updated:

$$\Delta_{k+1} = \begin{cases} 1/2\|x_{k+1} - x_k\|, & 0 \leq R_k < 0.25 \\ 2\Delta_k, & R_k > 0.75, \|s\| = \Delta_k, \\ \Delta_k, & \text{otherwise} \end{cases}$$

where

$$R_k = \frac{f(x_{k+1}) - f(x_k)}{q(x_{k+1}) - q(x_k)}.$$

The method solves

$$\begin{aligned} \min \quad & q(x_k + s) \\ s.t. \quad & \|s\| \leq \|\Delta_k\|^2, \end{aligned}$$

and its KKT conditions $\nabla f(x_k) + H_k s + 2vs = \bar{0}$. When $v = \bar{0}$, this gives (quasi-)Newton step, and otherwise $s = -(H + 2vI)^{-1}\nabla f(x_k)$. The advantage to the earlier methods it that H_k need not be positive definite. There are many variations, like dog-leg method.

Least squares application

One of the most important applications of unconstrained optimization are the least squares problems:

$$\min \ 1/2\|f(x)\|_2^2, \quad f : \mathbf{R}^n \mapsto \mathbf{R}^m.$$

Example. *Fit a model to some data. The data consists of measurements y_i, z_i, $i = 1, \ldots, m$, and the model $z = g(y, x)$ has parameters $x \in \mathbf{R}^n$. This gives*

$$f(x) = \begin{bmatrix} w_1(z_1 - g(y_1, x)) \\ \vdots \\ w_m(z_m - g(y_m, x)) \end{bmatrix},$$

where w_i are weights.

Let us calculate the gradients and Hessians:

$$\begin{aligned}\nabla(\|f(x)\|^2) &= 2\nabla f(x)^T f(x),\\ \nabla^2(\|f(x)\|^2) &= 2\nabla f(x)^T \nabla f(x) + 2S(x),\\ S(x) &= \sum_{i=1}^{m} f_i(x)\nabla^2 f_i(x).\end{aligned}$$

Newton: $\nabla^2(\|f(x_k)\|^2)s_k = -\nabla f(x_k)^T f(x_k)$
Gauß-Newton: $\nabla f(x_k)^T \nabla f(x_k)s_k = -\nabla f(x_k)^T f(x_k)$
Levenberg-Marquardt: $\nabla(f(x_k)^T \nabla f(x_k) + \mu_k I)s_k = -\nabla f(x_k)^T f(x_k)$, where μ_k s.t. the matrix is positive definite

The methods have both line search and trust-region variants and with quasi-Newton approximations.

Computation and accuracy*

Stability is related to an algorithm and a stable algorithm produces exact solutions for well-conditioned problem even though there are some rounding and floating point errors. Condition is related to the problem (or function if $f(x) = 0$ is to be solved): well-conditioned problem is such that when there are small deviations in x then there are small deviations in $f(x)$.

Definition 9.3. *The* **absolute condition number** *is*

$$\kappa' = \lim_{d\to 0} \sup_{\|\delta x\|\le d} \frac{\|\delta f\|}{\|\delta x\|},$$

where $\delta f = f(x + \delta x) - f(x)$.

Definition 9.4. *The* **relative condition number** *is*

$$\kappa = \lim_{d\to 0} \sup_{\|\delta x\|\le d} \frac{\|\delta f\|}{\|f(x)\|} / \frac{\|\delta x\|}{\|x\|} = \frac{\|J(x)\|\|x\|}{\|f(x)\|},$$

where the last holds if f *is differentiable and* δx *infinite decimal small.*

It is said that the problem/function is **well-conditioned** if κ small like $\kappa = 1, 10, 10^2$ and **ill-conditioned** if κ is large like $\kappa = 10^6, 10^{16}$.

Example. $f(x) = x/2$, $\kappa = \frac{\|J\|\|x\|}{\|f(x)\|} = \frac{1/2\cdot x}{x/2} = 1$,
$f(x) = \sqrt{x}$, $\kappa = \frac{1/2\cdot x\sqrt{x}}{\sqrt{(x)}} = 1/2$,
$f(x) = x_1 - x_2$, *in* $\|\cdot\|_\infty$ *norm,* $\kappa = \frac{2\max\{x_1,x_2\}}{|x_1-x_2|}$, *which is large if* $x_1 \approx x_2$ *and* x_1, x_2 *large.*

Example. *Computing eigenvalues of non-symmetric matrices: if* $A = \begin{pmatrix} 1 & 1000 \\ 0 & 1 \end{pmatrix}$ *and* $B = \begin{pmatrix} 1 & 1000 \\ 0.001 & 1 \end{pmatrix}$*, then* $\lambda_A = \{1, 1\}$ *and* $\lambda_B = \{0, 2\}$*. If the matrices are symmetric, then the problem is well-conditioned (*$\kappa' = 1$*,* $\kappa = \frac{\|A\|_2}{|\lambda|}$*)*

Example. *Solving linear equations:* $f(x) = Ax$*,* $\kappa = \|A\| \frac{\|x\|}{\|Ax\|}$*. If the matrix is non-singular and square, then* $\kappa \leq \|A\|\|A^{-1}\|$*. Solving* $Ax = b$*,* $\kappa = \|A^{-1}\| \frac{\|b\|}{\|x\|} \leq \|A\|\|A^{-1}\| = \kappa(A)$*, where* $\kappa(A)$ *is the condition number of A. In this case the condition expresses the eccentricity of hyperellipse (image of unit ball under mapping A), which is the ratio of* λ_n/λ_1*, since* $\|A\| = \lambda_n$ *the largest eigenvalue and* $\|A^{-1}\| = 1/\lambda_1$*, where* λ_1 *is the smallest eigenvalue.*

Solving linear equations*

Let $A \in \mathbf{R}^{m \times n}$. If $m < n$ the problem is underdetermined and the solution is a surface or larger dimensional set. If $m > n$ the problem is overdetermined and it is not necessarily possible to satisfy all equations and the problem is rather of least squares form $\|Ax - b\|$. If $m = n$ and A is non-singular then the solution is unique $x = A^{-1}b$.

The Gaussian elimination solves the problem in approx. $2/3n^3$ operations in two steps: forward elimination $Lx = b$ and back substitution $Ux = b$, where L is lower and U upper triangular matrix.

If A is symmetric positive definite, then **Cholesky decomposition** can be used:

0. Precondition the problem by switching rows $P^T AP$, where $P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$, to improve sparsity and stability.
1. Form Cholesky decomposition $A = LL^T$. (n^3)
2. Forward elimination $Lz = b$. (n^2)
3. Backward substitution $L^T x = z$. (n^2)

The total operations needed is $1/3n^3$ ($mn^2 + n^3/3$ if $A^T Ax = A^T b$). If multiple equations with the same A need to be solved, then the same decomposition can be used and only n^2 operations are needed.

If A is not positive definite then **QR decomposition** is more stable, where $Q^T Q = I$ is orthogonal as $Q^{-1} = Q^T$. There are different ways of computing the decomposition, like Gram-Schmidt or Householder's method. The steps are

0. $AP = QR$,
1. $z = Q^T b$,
2. $Rx = z$, which reduces to triangular matrix.

This requires $4/3n^3$ operations. ($2mn^2 + 2/3n^3$)

Even more stable method is the **singular value decomposition** (SVD): $A = U\Sigma V^T$, $U \in \mathbf{R}^{m\times m}$, $V \in \mathbf{R}^{n\times n}$ orthogonal, $\Sigma \in \mathbf{R}^{m\times n}$ non-negative diagonal. Then compute

1. $z = U^T b$,
2. $\Sigma w = z$ (diagonal),
3. $x = Vw$.

It needs $11n^3$ operations. ($2mn^2 + 11n^3$ if $m >> n$)

Exploiting structure in optimization*

In the next section some constrained optimization methods convert the constrained (difficult) problem into a series of easier problems. These can be QP problems (in SQP method), or unconstrained problems (in penalty and barrier function methods). Thus, it is important to have efficient methods to solve these easier problems as the more difficult problems rely on solving multiple instance of them. With some methods the structure of the problem is inherited to these easier problems and this may help dramatically. See for example Gondzio and Grothey for the largest optimization problems solved. Their method relies on solving efficiently linear equations that have a special structure. In solving linear equations the order of columns and sparsity play a significant role. There are many special structures that can be efficiently solved: arrowhead, bands, (tri)diagonals in $O(n)$, Toeplitz $O(n^2)$. Schur complement can be used when there is a subblock in the matrix that is easy to invert. Woodbury inversion formula can be used when the matrix is close to a matrix that is easy to invert $(A + pq^T)^{-1}$, where pq^T is rank-1 term. The matrix inversion is not the only operation that can be improved but also all matrix products may require a lot of computation.

Minimal volume ellipsoid covering a set*

Definition 9.5. *An* **ellipsoid** *has many representations, like*

$$e = \{x, x^T Ax + b^T x + c \leq 0, A \text{ symmetric pos.def.}\},$$

where the eigenvectors of A give the axis, axis lengths are given by the eigenvalues $1/\sqrt{\lambda_i}$*. The ellipsoid can also be seen as a unit ball mapped with an affine function*

$$e = \{x, \|Ax + b\|_2 \leq 1\} = \{x, x^T A^T Ax + 2(A^T b)^T x + b^T b - 1 \leq 0\}.$$

The volume of an ellipsoid is proportional to the product of its axis and thus $V \sim \prod_i \frac{1}{\sqrt{\lambda_i}}$. The determinant of a matrix A is also $\det A = \prod_i \lambda_i$ and $\det A^{-1} = \prod_i \lambda_i^{-1}$. Now, we can formulate a problem where a finite number of points need to be convered with an ellipsoid such that the ellipsoid has minimal volume

$$\begin{aligned} \min \quad & V \\ s.t. \quad & \|Ax + b\| \leq 1. \end{aligned}$$

The objective can be simplified

$$V = \sqrt{\det A^{-1}} \sim \det A^{-1} \sim \log \det A^{-1},$$

since both $\sqrt{x}$ and $\log(x)$ are monotone functions, and $\log \det A$ is convex function.

10 Numerical methods for constrained problems

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & h(x) = \bar{0}\ (l),\ g(x) \leq \bar{0}\ (m),\ x \in X \end{aligned}$$

The algorithms can be roughly divided the following way:

- **primal methods**: find descent direction keeping inside the feasible set (reduced gradient, method of feasible directions, active set)
- **barrier and penalty function methods**: solve sequence of unconstrained problems (augmented Lagrangian, primal-dual interior point method)
- **Lagrange multiplier methods** (augmented Lagrangian, dual methods)
- **SQP**: solve series of QP or solve KKT conditions with Newton's method

Primal methods

The method of feasible directions:

0. Find a feasible initial point.

1. Find a feasible descent direction, if not stop.
2. Determine the step length taking care of feasibility.

Zoutendijk method:

$$\begin{aligned} \min \quad & z \\ s.t. \quad & \nabla f(x_k)^T d - z \leq 0 \\ & \nabla g_i(x_k)^T d - z \leq 0,\ i \in I \\ & -1 \leq d_j \leq 1,\ j = 1, \ldots, n \end{aligned}$$

Let (z_k, d_k) be an optimal solution. If $z_k = \bar{0}$ stop and x_k is FJ point. If $z_k < 0$, do a line search:

$$\begin{aligned} \min \quad & f(x_k + s d_k) \\ s.t. \quad & 0 \leq s \leq s', \end{aligned}$$

where $s' = \sup\{s, g_i(x_k + s d_k) \leq 0,\ i = 1, \ldots, m\}$.

It may be difficult to find feasible direction, and the method may do zigzagging when new constraints become active. See also the gradient projection method of Rosen.

Active set method

The method is suitable for solving convex QP problems (used in SQP): $f(x) = 1/2 x^T Q x + c^T x$, where Q is symmetric and pos.def., $S = \{x, a_t^T x \leq b_i, 1 \leq i \leq m\}$. The method lists the active constraints $W_i = I(x_i)$ and minimizes

$$\begin{aligned} \min \quad & 1/2 d^T Q d + g_k^T d \\ s.t. \quad & a_i^T d = 0,\ i \in W_k. \end{aligned}$$

This gives the search direction d_k and the corresponding Lagrange multipliers v_i, $i \in W_k$.

- if $d_k = \bar{0}$ and $v_q = \min v_i \geq 0$ then x_k optimum, otherwise $W_{k+1} = W_k \setminus \{q\}$.
- if $d_k \neq \bar{0}$ and $x_k + d_k$ is feasible then take the step.
- if not feasible then find maximum step that is feasible and update W_{k+1}.

There may be a problem when the active set changes slowly and there are many constraints.

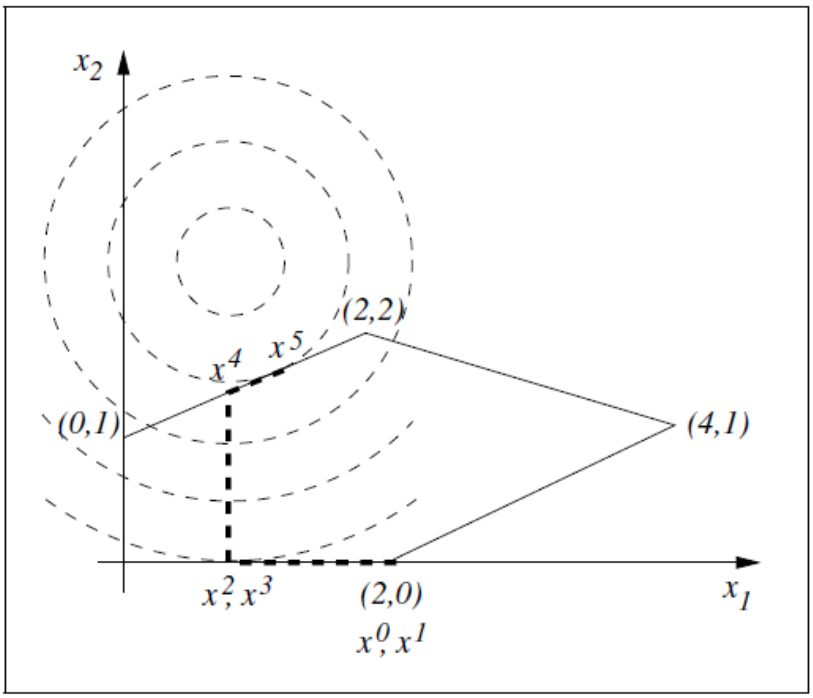


Figure 16.3 Iterates of the active-set method.

Reduced gradient methods (Frank-Wolfe)

Let $S = \{x, Ax = b, x \geq \bar{0}\}$, $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$, $m < n$. Assume that A is of rank m, i.e., any m columns are linearly independent. Then according to Theorem 2.6.4, every extreme point of S there are at least m strictly positive components (at most $n-m$ zeros). Let us reorder the variables and denote $x = (x_B, x_N)$, where $x_B > \bar{0}$ are the basic (dependent) variables and $x_N \geq \bar{0}$ the nonbasic variables. Similarly, we have $A = (B, N)$, where $B \in \mathbf{R}^{m\times m}$, $N \in \mathbf{R}^{m\times(n-m)}$, and $\nabla f(x) = (\nabla_B f(x), \nabla_N f(x))$.

The feasibility can be maintained by restricting the search direction to be feasible, i.e., requiring that $Ad = Bd_B + Nd_N = \bar{0}$, where $d = (d_B, d_N)$, and we can solve $d_B = -B^{-1}Nd_N$ since B is invertible by the rank assumption. We can compute the reduced gradient

$$\nabla f(x)^T d = \nabla_B f(x)^T d_B + \nabla_N f(x)^T d_N = (\nabla_N f(x)^T - \nabla_B f(x)^T B^{-1} N) d_N \doteq r_N^T,$$

where $r = (\bar{0}, r_N)$ is the reduced gradient.

If d_N is chosen s.t. $r_N^T d_N < 0$, $d_i \geq 0$, $i \in I$, then d is a feasible descent direction. d_N can, e.g., be chosen

$$d_j = \begin{cases} -r_j, & r_j \leq 0, \\ -x_j r_j, & r_j > 0, \end{cases}$$

and then $d_j \geq 0$ if $x_j = 0$ and it avoids small steps when $x_j > 0$. By Theorem 10.6.1. this choice means that $d_N = \bar{0} \Leftrightarrow$ x KKT point.

Algorithm: Initialization: Find x_1 satisfying $Ax_1 = b$, $x_1 \geq 0$.

1. Find the basic variables: I_k is the index set of the m largest components of x_k. Then from the columns of A: $B = \{a_j : j \in I_k\}$ and $N = \{a_j : j \notin I_k\}$. Compute $r^T = \nabla f(x_k)^T - \nabla_B f(x_k)^T B^{-1} A$. Form

$$d_j = \begin{cases} -r_j, & r_j \leq 0, \\ -x_j r_j, & r_j > 0, \end{cases}$$

   and compute $d_B = -B^{-1} N d_N$. Let $d_k^T = (d_B^T, d_N^T)$. If $d_k = 0$, stop and x_k is a KKT point.

2. Solve the line search $f(x_k + \lambda d_k)$ s.t. $0 \leq \lambda \leq \lambda_{max}$, where $\lambda_{max} = \min_{1 \leq j \leq n} -x_{jk}/d_{jk}$ for $d_{jk} < 0$, if $d_k \not\geq 0$, and $\lambda_{max} = \infty$ if $d_k \geq 0$. Update $x_{k+1} = x_k + \lambda_k d_k$. Goto step 1.

Summary of the Reduced Gradient Algorithm

We summarize below Wolfe's reduced gradient algorithm for solving a problem of the form to minimize $f(\mathbf{x})$ subject to $\mathbf{Ax} = \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$. It is assumed that all m columns of $\mathbf{A}$ are linearly independent and that every extreme point of the feasible region has m strictly positive components. As we show shortly, the algorithm converges to a KKT point, provided that the basic variables are chosen to be the m most positive variables, where a tie is broken arbitrarily.

Initialization Step Choose a point $\mathbf{x}_1$ satisfying $\mathbf{Ax}_1 = \mathbf{b}$, $\mathbf{x}_1 \geq \mathbf{0}$. Let $k = 1$ and go to the Main Step.

Main Step

1. Let $\mathbf{d}_k^t = (\mathbf{d}_B^t, \mathbf{d}_N^t)$ where $\mathbf{d}_N$ and $\mathbf{d}_B$ are obtained as below from (10.43) and (10.44), respectively. If $\mathbf{d}_k = \mathbf{0}$, stop; $\mathbf{x}_k$ is a KKT point. [The Lagrange multipliers associated with $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$ are, respectively, $\nabla_B f(\mathbf{x}_k)^t \mathbf{B}^{-1}$ and $\mathbf{r}$.] Otherwise, go to Step 2.

$$I_k = \text{index set of the } m \text{ largest components of } \mathbf{x}_k \tag{10.40}$$

$$\mathbf{B} = \{\mathbf{a}_j : j \in I_k\}, \quad \mathbf{N} = \{\mathbf{a}_j : j \notin I_k\} \tag{10.41}$$

$$\mathbf{r}^t = \nabla f(\mathbf{x}_k)^t - \nabla_B f(\mathbf{x}_k)^t \mathbf{B}^{-1}\mathbf{A} \tag{10.42}$$

$$\mathbf{d}_j = \begin{cases} -r_j & \text{if } j \notin I_k \text{ and } r_j \leq 0 \\ -x_j r_j & \text{if } j \notin I_k \text{ and } r_j > 0 \end{cases} \tag{10.43}$$

$$\mathbf{d}_B = -\mathbf{B}^{-1}\mathbf{N}\mathbf{d}_N. \tag{10.44}$$

2. Solve the following line search problem:

$$\begin{aligned} &\text{Minimize} \quad f(\mathbf{x}_k + \lambda \mathbf{d}_k) \\ &\text{subject to} \quad 0 \leq \lambda \leq \lambda_{\max}, \end{aligned}$$

where

$$\lambda_{\max} = \begin{cases} \min\limits_{1 \leq j \leq n} \left\{ \dfrac{-x_{jk}}{d_{jk}} : d_{jk} < 0 \right\} & \text{if } \mathbf{d}_k \neq \mathbf{0} \\ \infty & \text{if } \mathbf{d}_k \geq \mathbf{0} \end{cases} \tag{10.45}$$

10.6.2 Example

Consider the following problem:

$$\begin{aligned} \text{Minimize } & 2x_1^2 + 2x_2^2 - 2x_1x_2 - 4x_1 - 6x_2 \\ \text{subject to } & x_1 + x_2 + x_3 = 2 \\ & x_1 + 5x_2 + x_4 = 5 \\ & x_1, x_2, x_3, x_4 \geq 0. \end{aligned}$$

We solve this problem using Wolfe's reduced gradient method starting from the point $\mathbf{x}_1 = (0, 0, 2, 5)^t$. Note that

$$\nabla f(\mathbf{x}) = (4x_1 - 2x_2 - 4, 4x_2 - 2x_1 - 6, 0, 0)^t.$$

We shall exhibit the information needed at each iteration in tableau form similar to the simplex tableau of Section 2.7. However, since the gradient vector changes at each iteration, and since the nonbasic variables could be positive, we explicitly give the gradient vector and the complete solution at the top of each tableau. The reduced gradient vector $\mathbf{r}_k$ is shown as the last row of each tableau.

***Iteration* 1:**

Search Direction At the point $\mathbf{x}_1 = (0, 0, 2, 5)^t$, we have $\nabla f(\mathbf{x}_1) = (-4, -6, 0, 0)$. By (10.40), we have $I_1 = \{3, 4\}$, so that $\mathbf{B} = [\mathbf{a}_3, \mathbf{a}_4]$ and $\mathbf{N} = [\mathbf{a}_1, \mathbf{a}_2]$. From (10.42), the reduced gradient is given by

$$\mathbf{r}^t = (-4, -6, 0, 0) - (0, 0)\begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 5 & 0 & 1 \end{bmatrix} = (-4, -6, 0, 0).$$

By (10.16) we have $\mathbf{d}_N = (d_1, d_2)^t = (4, 6)^t$. We now compute $\mathbf{d}_B$ using (10.44) to get

$$\mathbf{d}_B = (d_3, d_4)^t = -\mathbf{B}^{-1}\mathbf{N}\mathbf{d}_N = -\begin{bmatrix} 1 & 1 \\ 1 & 5 \end{bmatrix}\begin{pmatrix} 4 \\ 6 \end{pmatrix} = (-10, -34)^t.$$

Note that $\mathbf{B}^{-1}\mathbf{N}$ is recorded under the variables corresponding to $\mathbf{N}$: namely, x_1 and x_2. The direction vector is, then, $\mathbf{d}_1 = (4, 6, -10, -34)^t$.

Line Search Starting from $\mathbf{x}_1 = (0, 0, 2, 5)^t$, we now wish to minimize the objective function along the direction $\mathbf{d}_1 = (4, 6, -10, -34)^t$. The maximum value of λ such that $\mathbf{x}_1 + \lambda\mathbf{d}_1$ is feasible is computed using (10.45), and we get

$$\lambda_{\max} = \min\left\{\frac{2}{10}, \frac{5}{34}\right\} = \frac{5}{34}.$$

The reader can verify that $f(\mathbf{x}_1 + \lambda\mathbf{d}_1) = 56\lambda^2 - 52\lambda$, so that λ_1 is the solution to the following problem:

$$\begin{aligned} &\text{Minimize} \quad 56\lambda^2 - 52\lambda \\ &\text{subject to} \quad 0 \le \lambda \le \frac{5}{34}. \end{aligned}$$

This yields $\lambda_1 = 5/34$, so that $\mathbf{x}_2 = \mathbf{x}_1 + \lambda_1\mathbf{d}_1 = (10/17, 15/17, 9/17, 0)^t$.

Iteration 2:

Search Direction At $\mathbf{x}_2 = (10/17, 15/17, 9/17, 0)^t$, from (10.40) we have $I_2 = \{1, 2\}$, $\mathbf{B} = [\mathbf{a}_1, \mathbf{a}_2]$, and $\mathbf{N} = [\mathbf{a}_3, \mathbf{a}_4]$. We also have $\nabla f(\mathbf{x}_2) = (-58/17, -62/17, 0, 0)^t$. The current information is recorded in the following tableau, where the

$$\mathbf{r}^t = \left(-\frac{58}{17}, -\frac{62}{17}, 0, 0\right) - \left(-\frac{58}{17}, -\frac{62}{17}\right)\begin{bmatrix} 1 & 0 & \frac{5}{4} & -\frac{1}{4} \\ 0 & 1 & -\frac{1}{4} & \frac{1}{4} \end{bmatrix} = \left(0, 0, \frac{57}{17}, \frac{1}{17}\right).$$

From (10.43), then, $d_3 = -(9/17)(57/17) = -513/289$ and $d_4 = 0$, so that $\mathbf{d}_N = (-513/289, 0)^t$. From (10.44), we get

$$\mathbf{d}_B = (d_1, d_2)^t = -\begin{bmatrix} \frac{5}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} \end{bmatrix}\begin{pmatrix} -\frac{513}{289} \\ 0 \end{pmatrix} = \begin{bmatrix} \frac{2565}{1156} \\ -\frac{513}{1156} \end{bmatrix}.$$

The new search direction is therefore given by $\mathbf{d}_2 = (2565/1156, -513/1156, -513/289, 0)^t$.

Line Search Starting from $\mathbf{x}_2 = (10/17, 15/17, 9/17, 0)^t$, we wish to minimize the objective function along the direction $\mathbf{d}_2 = (2565/1156, -513/1156, -513/289, 0)^t$. The maximum value of λ such that $\mathbf{x}_2 + \lambda\mathbf{d}_2$ is feasible is computed using (10.45), and we get

$$\lambda_{\max} = \min\left\{\frac{-15/17}{-513/1156}, \frac{-9/17}{-513/289}\right\} = \frac{17}{57}.$$

The reader can verify that $f(\mathbf{x}_2 + \lambda\mathbf{d}_2) = 12.21\lambda^2 - 5.95\lambda - 6.436$, so that λ_2 is obtained by solving the following problem:

$$\begin{aligned} &\text{Minimize} \quad 12.21\lambda^2 - 5.9\lambda - 6.436 \\ &\text{subject to} \quad 0 \le \lambda \le \frac{17}{57}. \end{aligned}$$

This can be verified to yield $\lambda_2 = 68/279$, so that $\mathbf{x}_3 = \mathbf{x}_2 + \lambda_2\mathbf{d}_2 = (35/31, 24/31, 3/31, 0)^t$.

***Iteration* 3:**

Search Direction Now $I_3 = \{1, 2\}$, so that $\mathbf{B} = [\mathbf{a}_1, \mathbf{a}_2]$ and $\mathbf{N} = [\mathbf{a}_3, \mathbf{a}_4]$. Since $I_3 = I_2$, the tableau at Iteration 2 can be retained. However, we now have $\nabla f(\mathbf{x}_3) = (-32/31, -160/31, 0, 0)^t$.

		x_1	x_2	x_3	x_4
Solution $\mathbf{x}_3$		$\frac{35}{31}$	$\frac{24}{31}$	$\frac{3}{31}$	0
$\nabla f(\mathbf{x}_3)$		$-\frac{32}{31}$	$-\frac{160}{31}$	0	0
$\nabla_B f(\mathbf{x}_3) = \begin{bmatrix} -\frac{32}{31} \\ -\frac{160}{31} \end{bmatrix}$	x_1	1	0	$\frac{5}{4}$	$-\frac{1}{4}$
	x_2	0	1	$-\frac{1}{4}$	$\frac{1}{4}$
$\mathbf{r}$		0	0	0	$\frac{32}{31}$

From (10.42) we get

$$\mathbf{r}^t = \left(-\frac{32}{31}, -\frac{160}{31}, 0, 0\right) - \left(-\frac{32}{31}, -\frac{160}{31}\right) \begin{bmatrix} 1 & 0 & \frac{5}{4} & -\frac{1}{4} \\ 0 & 1 & -\frac{1}{4} & \frac{1}{4} \end{bmatrix} = \left(0, 0, 0, \frac{32}{31}\right).$$

From (10.43), $\mathbf{d}_N = (d_3, d_4)^t = (0,0)^t$; and from (10.44) we also get $\mathbf{d}_B = (d_1, d_2)^t = (0,0)^t$. Hence, $\mathbf{d} = \mathbf{0}$, and the solution $\mathbf{x}_3$ is a KKT solution and therefore optimal for this problem. The optimal Lagrange multipliers associated with the equality constraints are $\nabla_B f(\mathbf{x}_3)^t \mathbf{B}^{-1} = (0, -32/31)^t$, and those associated with the nonnegativity constraints are $(0, 0, 0, 1)^t$. Table 10.5 gives a summary of the computations, and the progress of the algorithm is shown in Figure 10.19.

Penalty function methods

The idea of penalty functions is to move the constraints into the objective function and make it unconstrained problem.

Example. $\min \ f(x)$ *s.t.* $h(x) = 0 \Rightarrow \min f(x) + \mu h(x)^2$. *when* μ *is big, then* $h(x) \approx 0$. $(\max(0, g(x))^2$ *for inequality constraints)*

Definition 10.1. *A* **penalty function** *is* $\alpha(x) = \sum_{i=1}^{m} \phi(g_i(x)) + \sum_{i=1}^{l} \psi(h_i(x))$,

$$\phi(y) = \begin{cases} 0, & y \leq 0 \\ > 0, & y > 0 \end{cases}, \quad \psi(y) = \begin{cases} 0, & y = 0 \\ > 0, & y \neq 0 \end{cases}.$$

Definition 10.2. *A* **penalty function problem** $\min_{x \in X} f(x) + \mu\alpha(x)$, $\mu > 0$.

Algorithm:

1. Solve $x_{k+1} \in \arg\min_{x \in X} f(x) + \mu_k \alpha(x)$.
2. If $\mu_k \alpha(x_{k+1}) < \epsilon$ stop, otherwise $\mu_{k+1} = \beta \mu_k$, $\beta > 1$.

Problems: may stop prematurely or converge slowly. $\nabla^2(f(x) + \mu\alpha(x))$ is almost singular when μ is large, and thus the convergence properties are poor.

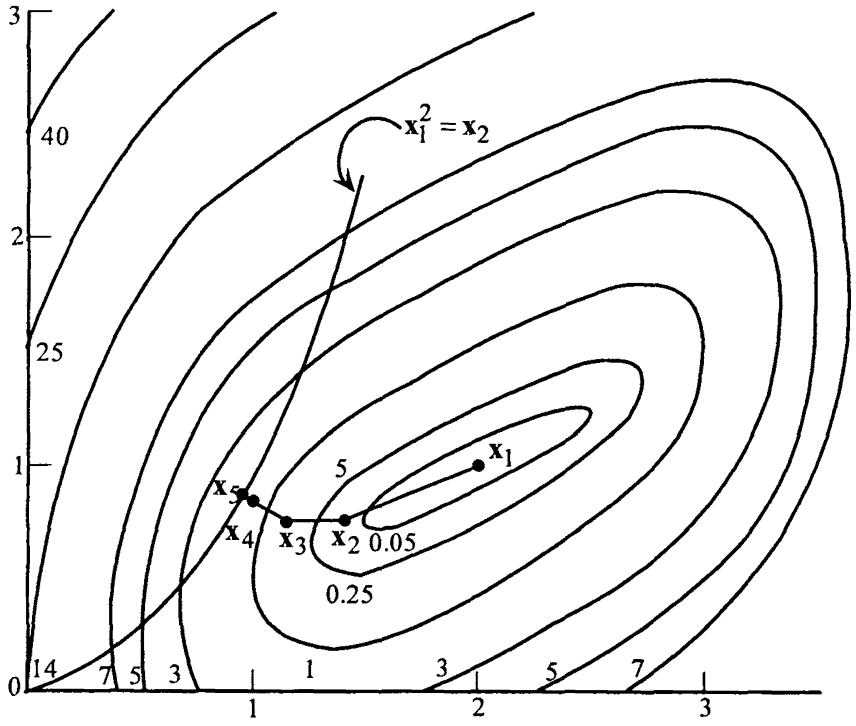


Figure 9.5 Penalty function method.

Theorem (9.2.2)**.** *Assume f,g,h continuous.* $\theta(\mu) = \min_{x \in X} f(x) + \mu\alpha(x) = f(x_\mu) + \mu\alpha(x_\mu)$. *Assume* x_μ *belongs to a compact subset for all* $\mu > 0$. *Then*

$$\inf\{f(x), g(x) \leq \bar{0}, h(x) = \bar{0}, x \in X\} = \lim_{\mu \to \infty} \theta(\mu),$$

and $x_\mu \to x^*$, $\mu\alpha(x_\mu) \to 0$, *when* $\mu \to \infty$.

How large μ is then needed?

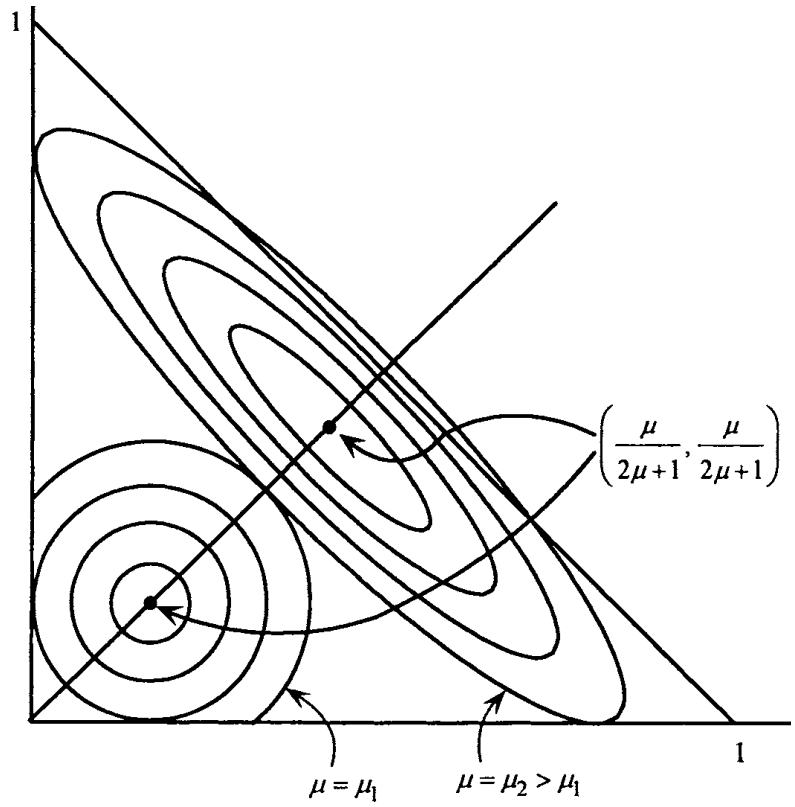
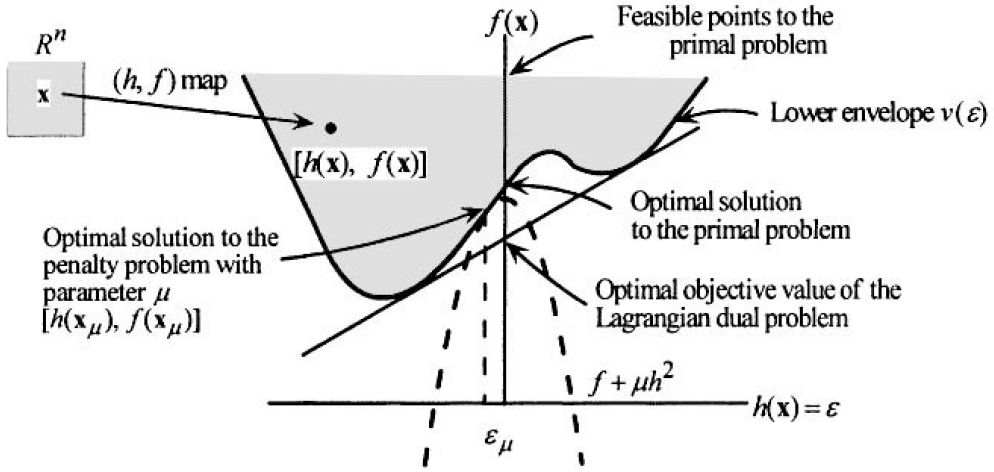


Figure 9.4 Ill-conditioning effect of a large μ value.



Penalty function with nonconvex problem

Definition 10.3. *A penalty function is* **exact** *when* $\exists \bar{\mu}$ *s.t.* x^* *is achieved with all* $\mu \geq \bar{\mu}$.

Example. *Absolute value* (l_1) *penalty function* $\alpha(x) = \mu(\sum \max\{0, g_i(x)\} + \sum |h_i(x)|)$ *is exact.*

Theorem (9.3.1)**.** *If* f, g_i, $i \in I$, *convex,* h *affine, then* x^* *minimizes* $\theta(\mu)$ *with absolute value penalty function when* $\mu \geq \max(u_i, |v|_j)$

Note that this penalty function is not smooth.

Augmented Lagrangian method

Another exact penalty function is the **augmented Lagrangian penalty function** (ALAG):

$$f_{ALAG}(x, v) = f(x) + v^T h(x) + \mu \sum_{i=1}^{l} h_i^2(x).$$

The inequality constraints are not a problem and they can be handled with slack variables. The penalty function convexifies the problem locally. If (x^*, v^*) is a KKT point, then $\nabla_x f_{ALAG}(x^*, v^*) = \nabla f(x^*) + v^{*T} \nabla h(x^*) + 2\mu \sum h_i(x^*) \nabla h_i(x^*) = \bar{0}$ and if $\mu > \bar{\mu}$ then x^* minimizes the penalty function problem.

Example. min x^3 *s.t.* $x+1 = 0$. *The solution is* $x^* = -1$, $v^* = 3$. $f_{ALAG}(x, v^*) = x^3 - 3(x+1) + \mu(x+1)^2$, $f'_{ALAG}(x, v^*) = 3x^2 - 3 + 2\mu x + 1$, $f''_{ALAG}(x, v^*) = 6x - 2\mu$. *When* $\mu \geq v^*$ *then the penalty function is convex at* x^*.

Algorithm:

- $VIOL(x) = \max(|h_i(x)|, i = 1, \ldots, l)$.

- **Inner loop**: Solve min $f_{ALAG}(x, v')$. If $VIOL(x_k) = 0$ then stop and x_k is KKT point. If $VIOL(x_k) \leq VIOL(x_{k-1})/4$ then go to outer loop. Otherwise, $\mu_i = \beta\mu_i$, $\beta > 1$, for all i that violate the above condition and repeat the inner loop.
- **Outer loop**: Update $v_i' = v_i' + 2\mu_i h_i(x_k)$. Return to the inner loop.

Note:

- $x_k \to x^*$ only if $v_k \to v^*$
- how do you know v^*? Guess?
- update of v affects the convergence
- problems when μ is too large or small
- what method is used in the inner loop?

Barrier function methods

Barrier function methods approach the optimum from inside of the feasible set.

Example. min $f(x)$ *s.t.* $g(x) \leq 0 \Rightarrow \min f(x) - \mu \log(-g(x))$. *when* μ *is small, then* $g(x)$ *can get close to zero.*

Definition 10.4. *A* **barrier function** $B(x)$ *is a continuous function s.t.* $B(x) \geq 0$ *when* $g(x) < \bar{0}$*, and* $B(x) \to \infty$ *when* $g_i(x) \to 0^+$.

Example. *These condition are satisfied by* $B(x) = -\sum \ln(\min(1, -g_i(x)))$*, and Frisch barrier* $B(x) = -\sum \ln(-g_i(x))$.

Definition 10.5. *A* **penalty function problem** min $\theta(\mu) = f(x) + \mu B(x)$, $\mu > 0$.

Algorithm: starting point x_0 s.t. $g(x_0) < \bar{0}$.

1. Solve min $f(x) + \mu_k B(x)$.
2. If $\mu_k B(x_{k+1}) < \epsilon$ stop and otherwise $\mu_{k+1} = \beta\mu_k$, $\beta \in (0, 1)$.

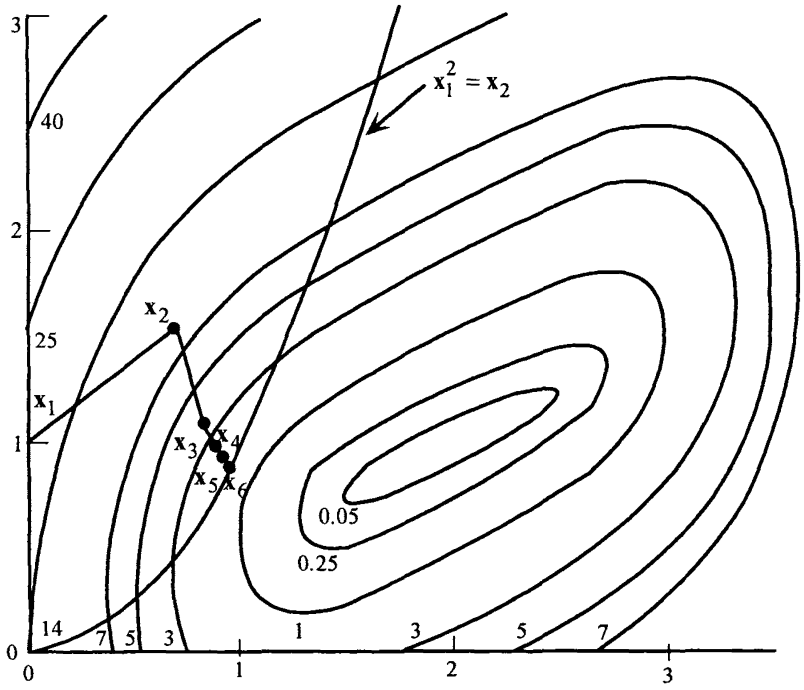


Figure 9.9 Barrier function method.

Note:

- the method needs strictly feasible starting point
- maintaining feasibility may be difficult (slow convergence)
- numeric problems at the boundary
- convergence as with penalty function method

Theorem (9.4.3). *Let f, g be continuous, $\{x \in X, g(x) \leq 0\}$ non-empty. Assume that for any neighborhood N around x^*, there is $x \in X \cap N$ s.t. $g(x) < \bar{0}$, then*

$$\min f(x),\ s.t.\ x \in X, g(x) \leq \bar{0} = \lim_{\mu \to 0^+} \theta(\mu) = \inf_{\mu > 0} \theta(\mu),$$

and $\mu B(x_\mu) \to 0$ when $\mu \to 0^+$.

11 Primal-dual interior point method

Primal-dual method is a barrier function method that is a linear-time algorithm for LP problem. Let us develop the method for convex problem:

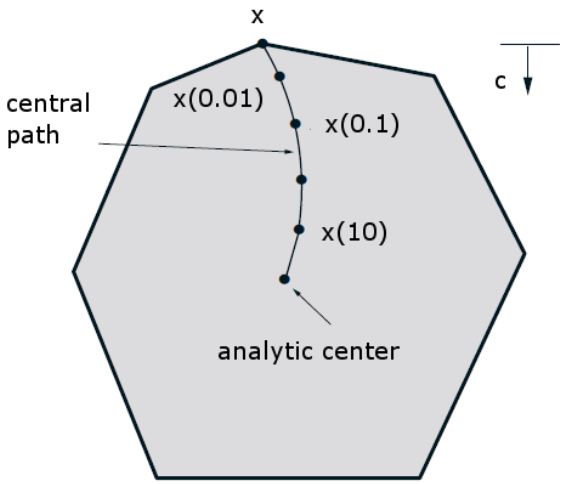
$$\begin{array}{ll} \min & f(x) \\ s.t. & g(x) \leq \bar{0}, \quad (m) \\ & Ax + b = \bar{0}, \quad (l) \end{array}$$

where f, g_i are convex and h is affine. The barrier problem is

$$\begin{aligned} \min \quad \beta(x;\mu) &= f(x) - \mu \sum_{i=1}^{m} \ln(-g_i(x)) \\ s.t.\ Ax + b &= \bar{0} \end{aligned}$$

The idea is to solve x_μ and have $\mu \to 0^+$. Does this approach x^*?

Definition 11.1. *A sequence* $\{x_\mu\}$, $\mu > 0$ *is the* **central path** *and as* $\mu \to \infty$, $x_\mu \to x^A$ *is the* **analytic center**.



The barrier function satisfies the KKT conditions:

$$\begin{aligned} \nabla_x \beta(x_\mu;\mu) + A^T v_\mu &= \bar{0}, \\ Ax_\mu + b &= \bar{0}, \end{aligned}$$

where $\nabla_x \beta(x_\mu, \mu) = \nabla f(x) + \mu \sum -\frac{\nabla g_i(x)}{g_i(x)} = f(x) + \nabla g(x)^T D^{-1} e$, where $D = diag(-g_j(x))$, $D^{-1} = diag(-1/g_j(x))$ and $e = (1, \ldots, 1)^T$. Let us denote $u_\mu = -\mu D^{-1} e$, i.e., $u_{\mu_i} = -\mu / g_i(x)$. This vector approximates the Lagrange multipliers of the original problem.

Theorem. *Duality gap:* $f(x_\mu) - \theta(u_\mu, v_\mu) = m\mu \to 0$, *when* $\mu \to 0$.

Proof. From KKT conditions, $\nabla_x \phi(x_\mu, u_\mu, v_\mu) = \bar{0}$, where $\phi(x, u, v) = f(x) + (Ax + b)^T v + g(x)^T u$ is convex, i.e., x_μ minimizes $\phi(x, u_\mu, v_\mu)$. Thus,

$$\theta(u_\mu, v_\mu) = \min \ \phi(x, u_\mu, v_\mu) = \phi(x_\mu, u_\mu, v_\mu) =$$
$$= f(x_\mu) + (Ax_\mu + b)^T v_\mu + g(x_\mu)^T u_\mu) = f(x_\mu) - m\mu,$$

since $Ax_\mu + b = \bar{0}$ and $g(x_\mu)^T u_\mu = \sum \frac{-g_i(x)\mu}{g_i(x)}$. □

The algorithm for LP problem, where $f(x) = c^T x$, $g(x) = -x$:

0. Choose $x_0, u_0, \mu_0 > \bar{0}$, v, $t \in (0, 1)$, $\epsilon > 0$.
1. Solve $x_{k+1}, v_{k+1}, u_{k+1}$ from the KKT conditions of the barrier function problem with the Newton's method. (these are derived in below)
2. If $c^T x_{k+1} - b^T v_{k+1} = n\mu_k < \epsilon$ (duality gap for LP) then stop, otherwise $\mu_{k+1} = t\mu_k$ and repeat.
3. Possible rounding to a feasible point.

The primal and dual prolems are

$$\begin{array}{ll} \min & c^T \ \ s.t.\ Ax = b,\ x \geq \bar{0} \\ \max & b^T \ \ s.t.\ A^T v + u = c,\ u \geq \bar{0} \end{array}$$

The KKT conditions:

$$\begin{aligned} A^T v + u &= c, \\ Ax &= b, \\ x &\geq \bar{0}, \\ u &\geq \bar{0}, \\ XUe &= \bar{0}, \end{aligned}$$

where $X = diag(x)$, $U = diag(u)$, $e = (1, \ldots, 1)^T$. The first equation is the Lagrange optimality, next two primal feasibility, then dual feasibility and finally the complementary slacness condition. Let us denote the equality constraints

$$F(x, v, u) = \begin{bmatrix} A^T v + u - c \\ Ax - b \\ XUe \end{bmatrix}.$$

The Newton update $\nabla F(x)\Delta x = -F(x)$ for this system is

$$\begin{bmatrix} \bar{0} & A^T & I \\ A & \bar{0} & \bar{0} \\ U & \bar{0} & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta v \\ \Delta u \end{bmatrix} = \begin{bmatrix} \bar{0} \\ \bar{0} \\ -XUe \end{bmatrix},$$

if (x_k, v_k, u_k) is primal-dual feasible, i.e., $A^T v_k + u_k = c$ and $Ax_k = b$. If not, then the two terms on the right-hand side were not zero. This direction is called as the affine scaling direction.

The logarithmic barrier function problem is

$$\min \quad c^T x - \mu \sum_{i=1}^{m} \log(x_i) \quad s.t.\ Ax = b,$$

and its KKT conditions

$$\begin{aligned} A^T v + \mu X^{-1} e &= c, \\ Ax &= b, \end{aligned}$$

and if we denote $u = \mu X^{-1} e$ then

$$\begin{aligned} A^T v + u &= c, \\ Ax &= b, \\ XUe &= \mu e. \end{aligned}$$

Now, we can see that the logarithmic barrier function relaxes the only nonlinear equation in the system, the complementary slackness condition, from zero to μ. The Newton update is

$$\begin{bmatrix} \bar{0} & A^T & I \\ A & \bar{0} & \bar{0} \\ U & \bar{0} & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta v \\ \Delta u \end{bmatrix} = \begin{bmatrix} -r_c \\ -r_b \\ -XUe + \mu e \end{bmatrix},$$

where $r_c = A^T v + u - c$ and $r_b = Ax - b$. Sometimes, a centering parameter $\sigma \in (0, 1)$ is added: $\sigma\mu e$. When $\sigma = 1$, the step is called **centering** step and when $\sigma = 0$ the step is called **Newton** or **affine scaling** direction.

Note that the primal-dual infeasibility is not a problem for the initial solution x_0, v_0, u_0. If a full Newton step is taken then $r_c = r_b = 0$ after that step, since these equations are linear.

There are many variants of the interior point method. Karmarkar's algorithm in 1984 was first practical polynomial time algorithm, following Khaciyan's ellipsoid algorithm which worked only in theory. Mehrotra presented the primal-dual

predictor-corrector in 1989 that used the same Cholesky decomposition to find two different directions.

Sequential quadratic programming (SQP) method

The sequential quadratic programming method can be seen as doing Newton step to the KKT conditions, i.e., making a quadratic approximation with linearized constraints.

Let us examine an equality constrained problem

$$\min \quad f(x) \quad s.t.\ h(x) = \bar{0} \quad (l)$$

The Lagrange function is $\phi(x, v) = f(x) + h(x)^T v$, $L(x) = \phi(x, v_k)$ and the KKT conditions:

$$\begin{aligned} \nabla_x \phi(x, v) &= \nabla f(x) + h(x)^T v = \bar{0}, \quad (n) \\ h(x) &= \bar{0}, \quad (l) \end{aligned}$$

and this system is denoted by $W(x, v) = \bar{0}$. Appyling the Newton update

$$W(x_k, v_k) + \nabla_{x,v} W(x_k, v_k)(x - x_k, v - v_k)^T = \bar{0},$$

where the Jacobian is

$$\nabla_{x,v} W(x, v) = \begin{bmatrix} \nabla^2_{xx} \phi(x, v) & \nabla h(x)^T \\ \nabla h(x) & \bar{0} \end{bmatrix}$$

This gives so-called Newton-Lagrange equations

$$\begin{aligned} \nabla f(x) + \nabla h(x)^T v_k + \nabla^2_{xx} \phi(x_k, v_k)(x - x_k) + \nabla h(x_k)^T (v - v_k) &= \bar{0}, \\ h(x_k) + \nabla h(x_k)^T (x - x_k) &= \bar{0}. \end{aligned}$$

Note that these are the same as the KKT conditions for the following problem

$$\begin{aligned} \min \quad & 1/2 d_k^T \nabla^2_{xx} \phi(x_k, v_k) d_k + \nabla f(x_k)^T (x - x_k) \\ s.t. \quad & h(x_k) + \nabla h(x_k)(x - x_k) = \bar{0}, \end{aligned}$$

where $d_k = x - x_k$. So, the search direction for (x, v) is solved from a QP problem and if $d_k = \bar{0}$ then x_k is a KKT point. Otherwise, the point is updated or a line search is performed in the search direction. If the problem has inequality constraints then they appear in the Lagrange function and the corresponding linearized constraints are

$$g(x_k) + \nabla g(x_k)^T d_k \leq \bar{0}.$$

There are many variants of the SQP method. Quasi-Newton approximation B_k can be used to replace $\nabla^2_{xx}\phi(x_k, v_k)$, which is updated. Then $s_k = x_{k+1} - x_k$, $y_k = \nabla L(x_{k+1}) - \nabla L(x_k)$. Note that the QP problem is then strictly convex because B_k is positive definite.

There is however problem that these methods only converge locally. Global convergence can be achieved by using **merit function** $m(x)$, $\hat{f}(x) = f(x) + m(x)$, in the line search step. Examples of merit functions are absolute value merit function $\mu(\sum \max(0, g_i(x)) + \sum |h_i(x)|)$ or augmented Lagrangian merit function.

The problem with a merit function is so called **Maratos effect**, where the merit function may decline a direction that takes towards the optimum, and this may result in slow convergence. It can happen that even if $\|x_k + d_K - x^*\| < \|x_k - x^*\|$ then it may be that $\hat{f}(x_k + d_k) > \hat{f}(x_k)$. This can be solved by adding second order correction terms or by choosing a suitable merit function.

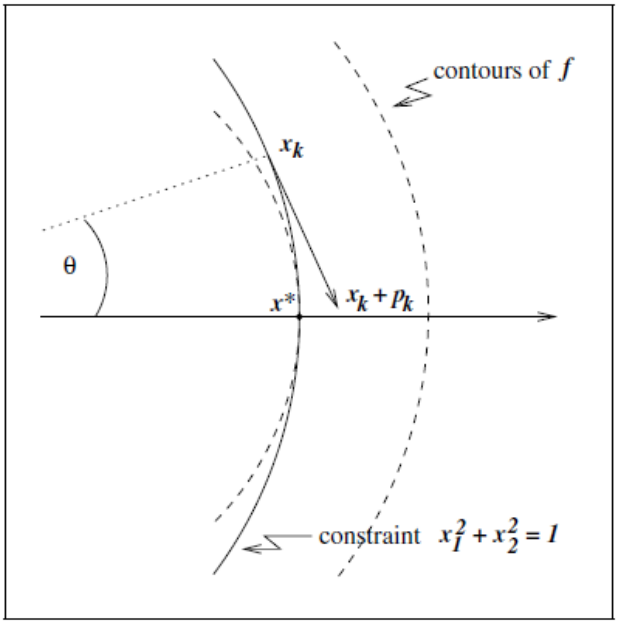


Figure 15.8
Maratos Effect: Example 15.4. Note that the constraint is no longer satisfied after the step from x_k to $x_k + p_k$, and the objective value has increased.

Observations:

- QP may be infeasible
- solution to QP may go to infinity
- Lagrange multipliers need to be updated, so the method needs to give the multipliers too. For example, active set method can be used

One variant is SL_1QP trust-region variant

$$\begin{aligned} \min \quad & \nabla f(x_k)^T d + 1/2 d^T B_k d + \mu(\sum \max(0, g_i(x) + \nabla g_i(x_k)^T d) + \\ & + \sum |h_i(x_k) + \nabla h_i(x_k)^T d|) \\ s.t. \quad & -\Delta_k \leq d \leq \Delta_k, \end{aligned}$$

which can be converted into QP problem. The method has the benefit that it is feasible and bounded, so it at least has a solution. However, the Maratos effect is still possible for this variant.