Convex Optimization

Contents

1 Introduction **1**

1.1 Mathematical optimization 1
1.2 Least-squares and linear programming 1
1.3 Convex optimization 1
1.4 Nonlinear optimization 1
1.5 Outline 1
1.6 Notation 1

I Theory **3**

2 Convex sets **5**

2.1 Affine and convex sets 5
2.2 Some important examples 6
2.3 Operations that preserve convexity 7
2.4 Generalized inequalities 7
2.5 Separating and supporting hyperplanes 8
2.6 Dual cones and generalized inequalities 8

3 Convex functions **11**

3.1 Basic properties and examples 11
3.2 Operations that preserve convexity 13
3.3 The conjugate function 14
3.4 Quasiconvex functions 16
3.5 Log-concave and log-convex functions 17
3.6 Convexity with respect to generalized inequalities 18

4 Convex optimization problems **19**

4.1 Optimization problems 19
4.2 Convex optimization 23
4.3 Linear optimization problems 25
4.4 Quadratic optimization problems 26
4.5 Geometric programming 27
4.6 Generalized inequality constraints 27
4.7 Vector optimization 27

5 Duality **29**
5.1 The Lagrange dual function . . . 29
5.2 The Lagrange dual problem . . . 32
5.3 Geometric interpretation . . . 36
5.4 Saddle-point interpretation . . . 36
5.5 Optimality conditions . . . 36
5.6 Perturbation and sensitivity analysis . . . 41
5.7 Examples . . . 41
5.8 Theorems of alternatives . . . 41
5.9 Generalized inequalities . . . 41

II Algorithms **43**

6 Unconstrained minimization **45**
6.1 Unconstrained minimization problems . . . 45
6.2 Descent methods . . . 47
6.3 Gradient descent method . . . 49
6.4 Steepest descent method . . . 50
6.5 Newton's method . . . 50
6.6 Self-concordance . . . 53
6.7 Implementation . . . 53

7 Equality constrained minimization **55**
7.1 Equality constrained minimization problems . . . 55
7.2 Newton's method with equality constraints . . . 56
7.3 Infeasible start Newton method . . . 58
7.4 Implementation . . . 61

8 Interior-point methods **63**
8.1 Inequality constrained minimization problems . . . 63
8.2 Logarithmic barrier function and central path . . . 64
8.3 The barrier method . . . 67
8.4 Feasibility and phase I methods . . . 70
8.5 Complexity analysis via self-concordance . . . 70
8.6 Problems with generalized inequalities . . . 70
8.7 Primal-dual interior-point methods . . . 70
8.8 Implementation . . . 70

A Mathematical background **73**
A.1 Norms . . . 73
A.2 Analysis . . . 75
A.3 Functions . . . 75
A.4 Derivatives . . . 75
A.5 Linear algebra . . . 75

Chapter 1

Introduction

1.1 Mathematical optimization

1.2 Least-squares and linear programming

1.3 Convex optimization

1.4 Nonlinear optimization

1.5 Outline

1.6 Notation

Part I

Theory

Chapter 2

Convex sets

2.1 Affine and convex sets

2.1.1 Lines and line segments

A *line* passing through x_1 and x_2 on $\mathbb{R}^n$ ($x_1 \neq x_2$) has the form

$$\begin{aligned} y &= \theta x_1 + (1-\theta)x_2 \\ &= x_2 + \theta(x_1 - x_2) \end{aligned}$$

It becomes *line segment* when $0 \leq \theta \leq 1$

2.1.2 Affine sets

Set $C \subseteq \mathbb{R}^n$ is *affine* if the line through any two distinct points in C lies in C, *i.e.*,
for any $x_1, x_2 \in C$ and $\theta \in \mathbb{R}$, we have $\theta x_1 + (1-\theta)x_2 \in C$
The *affine combination* of $x_1, \ldots, x_k$ is

$$\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k \quad \text{where} \quad \theta_1 + \theta_2 + \cdots + \theta_k = 1$$

If $C \subseteq \mathbb{R}^n$ is affine and $x_1, \ldots, x_k \in C$, then the affine combination of $x_1, \ldots, x_k$ is in C. This property can be proved by induction.

If C is affine and $x_0 \in C$, then $V = C - x_0 = \{x - x_0 \mid x \in C\}$ is a subspace.

Proof. let $v_1 = x_1 - x_0 \in V$ and $v_2 = x_2 - x_0 \in V$

$$\begin{aligned} \alpha v_1 + \beta v_2 &= \alpha(x_1 - x_0) + \beta(x_2 - x_0) \\ &= \alpha x_1 + \beta x_2 + (1 - \alpha - \beta)x_0 - x_0 \in V \end{aligned}$$

since $\alpha x_1 + \beta x_2 + (1-\alpha-\beta)x_0 \in C$. Therefore, V is a subspace. □

- subspace V associated with an affine set C does not depend on choice of x_0
- definition of the *dimension* of an affine set C is the dimension of subspace $V = C - x_0$

Example 2.1 (*Solution set of linear equations*)**.** The solution set of $C = \{x \mid Ax = b\}$ is an affine set. Let $x_1, x_2 \in C$, then $Ax_1 = A_2 = b$

$$A(\theta x_1 + (1-\theta)x_2) = \theta A x_1 + (1-\theta)Ax_2 = b$$

- the subspace associated with the affine set C is the nullspace of A
- converse: every affine set can be expressed as the solution set of a system of linear equations

The *affine hull* of C is the set of all affine combinations

$$\textbf{aff}\, C = \{\theta_1 x_1 + \cdots + \theta_k x_k \mid x_1, \ldots, x_k \in C,\ \sum_{i=1}^{k} \theta_i = 1\}$$

- affine hull is the smallest affine set that contains C
- if S is affine set with $C \subseteq S$, then $\textbf{aff}\, C \subseteq S$

2.1.3 Affine dimension and relative interior

Example 2.2.

2.1.4 Convex sets

2.1.5 Cones

2.2 Some important examples

2.2.1 Hyperplanes and halfspaces

2.2.2 Euclidean balls and ellipsoids

2.2.3 Norm balls and norm cones

Example 2.3 (*Second-order cone*)**.**

2.2.4 Polyhedra

Example 2.4 (*Nonnegative orthant*)**.**

Example 2.5 (*Some common simplexes*)**.**

2.2.5 The positive semidefinite cone

Example 2.6 (*Positive semidefinite cone in* $\mathbb{S}^2$)**.**

2.3 Operations that preserve convexity

2.3.1 Intersection

Example 2.7.

Example 2.8.

2.3.2 Affine functions

Example 2.9 (*Polyhedron*)**.**

Example 2.10 (*Solution set of linear matrix inequality*)**.**

Example 2.11 (*Hyperbolic cone*)**.**

Example 2.12 (*Ellipsoid*)**.**

2.3.3 Linear-fractional and perspective functions

Remark 2.1.

Remark 2.2.

Example 2.13 (*Conditional probabilities*)**.**

2.4 Generalized inequalities

2.4.1 Proper cones and generalized inequalities

Example 2.14 (*Nonnegative orthant and componentwise inequality*).

Example 2.15 (*Positive semidefinite cone and matrix inequality*).

Example 2.16 (*Cone of polynomials nonegative on* $[0, 1]$).

2.4.2 Minimum and minimal elements

Example 2.17.

Example 2.18 (*Minimum and minimal elements of a set of symmetric matrices*).

2.5 Separating and supporting hyperplanes

2.5.1 Separating hyperplane theorem

Example 2.19 (*Separation of an affine and a convex set*).

Example 2.20 (*Strict separation of a point and a closed convex set*).

Example 2.21 (*Theorem of alternatives for strict linear inequalities*).

2.5.2 Supporting hyperplanes

2.6 Dual cones and generalized inequalities

2.6.1 Dual cones

The *dual cone* of a cone K is defined by

$$K^* = \{y \mid x^T y \geq 0, \ \forall x \in K\} \tag{2.1}$$

- dual cone is a cone
- dual cone is always convex (even if origin cone is not)
- $y \in K^*$ if and only if $-y$ is the normal of a hyperplane supports K at the origin

Example 2.22 (*Subspace*)**.** The dual cone of a subspace $V \subseteq \mathbb{R}^n$ is its orthogonal complement $V^\perp = \{y \mid y^T v = 0, \ \forall v \in V\}$.

Proof.

(i) trivially, $V^\perp \subseteq V^*$

(ii) start here

□

Example 2.23 (*Nonnegative orthant*)**.**

Example 2.24 (*Positive semidefinite cone*)**.**

Example 2.25 (*Dual of a norm cone*)**.**

2.6.2 Dual generalized inequalities

Example 2.26 (*Theorem of alternatives for linear strict generalized inequalities*)**.**

2.6.3 Minimum and minimal elements via dual inequalities

Example 2.27 (*Pareto optimal production frontier*)**.**

Chapter 3

Convex functions

3.1 Basic properties and examples

3.1.1 Definition

A fuction $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* if $\mathbf{dom}\, f$ is convex and

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y) \quad \forall x, y \in \mathbf{dom}\, f,\ 0 \le \theta \le 1 \tag{3.1}$$

- *strictly convex*: strictly inequality holds in (3.1) when $x \ne y$ and $0 < \theta < 1$
- *concave*: $-f$ is convex
- *strictly concave*: $-f$ is strictly convex
- affine function is both convex and concave

A function is convex if and only if it is convex when restricted to any line that intersects its domain. *i.e.*,

$$f \text{ is convex} \Leftrightarrow \forall x \in \mathbf{dom}\, f,\ \forall v,\ g(t) = f(x+tv) \text{ is convex}$$

on its domain $\{t \mid x + tv \in \mathbf{dom}\, f\}$.

3.1.2 Extended-value extensions

If f is convex, define *extended-value extension* $\tilde{f} : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ by

$$\tilde{f}(x) = \begin{cases} f(x) & x \in \mathbf{dom}\, f \\ \infty & x \notin \mathbf{dom}\, f \end{cases}$$

- recover original domain as $\mathbf{dom}\, f = \{x \mid \tilde{f}(x) < \infty\}$
- express (3.1) with "*any* x and y"
- for $f = f_1 + f_2$, we need write $\mathbf{dom}\, f$ but need not write $\mathbf{dom}\, \tilde{f}$ explicitly

- the rest of this book implicitly use extended-value extensions
- for extended-value extension of a concave function, define $-\infty$

Example 3.1 (*Indicator function of a convex set*). Let $C \subseteq \mathbb{R}^n$ be a convex set and consider the convex function $I_C(x) = 0 \quad \forall x \in C$, we have

$$\tilde{I}_C(x) = \begin{cases} 0 & x \in C \\ \infty & x \notin C \end{cases}$$

The function $\tilde{I}_C$ is the *indicator function* of the set C.
Minimize f ($\mathbf{dom}\, f = \mathbb{R}^n$) on C is the same as minimize $f + \tilde{I}_C$ on $\mathbb{R}^n$

- indeed, $f + \tilde{I}_C$ is f restricted to the set C

3.1.3 First-order conditions

Suppose f is differentiable (*i.e.*, ∇f exists on $\mathbf{dom}\, f$, $\mathbf{dom}\, f$ is open). Then f is convex if and only if $\mathbf{dom}\, f$ is convex and

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \forall x, y \in \mathbf{dom}\, f \tag{3.2}$$

- convex function $\Leftrightarrow$ first-order Taylor approximation is a *global underestimator*
- if $\nabla f(x) = 0$, then $f(y) \geq f(x) \quad \forall y \in \mathbf{dom}\, f$, *i.e.*, x is global minimizer of f
- strictly convex: $f(y) > f(x) + \nabla f(x)^T (y - x) \quad \forall x, y \in \mathbf{dom}\, f$
- for concave and strictly concave, change the inequality direction

Proof of first-order convexity condition

Proof. **Add in future** □

3.1.4 Second-order conditions

Example 3.2 (*Quadratic functions*).

Remark 3.1.

3.1.5 Examples

3.1.6 Sublevel sets

Example 3.3.

3.1.7 Epigraph

Example 3.4 (*Matrix fractional function*).

3.1.8 Jensen's inequality and extensions

Remark 3.2.

3.1.9 Inequalities

3.2 Operations that preserve convexity

3.2.1 Nonnegative weighted sums

3.2.2 Composition with an affine mapping

3.2.3 Pointwise maximum and supremum

Example 3.5 (*Piecewise-linear functions*).

Example 3.6 (*Sum of r largest components*).

Example 3.7 (*Support function of a set*).

Example 3.8 (*Distance to farthest point of a set*).

Example 3.9 (*Least-squares cost as a function of weights*).

Example 3.10 (*Maximum eigenvalue of a symmetric matrix*).

Example 3.11 (*Norm of a matrix*).

3.2.4 Composition

Example 3.12.

Example 3.13 (*Simple composition results*).

Remark 3.3.

Example 3.14 (*Vector composition examples*).

3.2.5 Minimization

Example 3.15 (*Schur complement*).

Example 3.16 (*Distance to a set*).

Example 3.17.

3.2.6 Perspective of a function

Example 3.18 (*Euclidean norm squared*).

Example 3.19 (*Negative logarithm*).

Example 3.20.

3.3 The conjugate function

3.3.1 Definition and examples

Let $f : \mathbb{R}^n \to \mathbb{R}$, then the *conjugate* of f, denoted by f^* is defined as

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} \{y^T x - f(x)\} \tag{3.3}$$

- $\mathbf{dom}\, f^*$ consists of $y \in \mathbb{R}^n$ for which supremum is finite, *i.e.*, $y^T x - f(x)$ is bounded above on $\mathbf{dom}\, f$
- f^* is a convex function, since it is the pointwise supremum of affine functions of y
- when f is convex, the subscript $x \in \mathbf{dom}\, f$ is not necessary since $y^T x - f(x) = -\infty$ for $x \notin \mathbf{dom}\, f$

Example 3.21. A few of conjugate examples.

- *Affine function.* $f(x) = ax + b$. The function $xy - (ax + b)$ is bounded at $\{a\}$. Therefore, $f^*(a) = -b$ with $\mathbf{dom}\, f^* = \{a\}$.
- *Negative logarithm.* $f(x) = -\log(x)$, $\mathbf{dom}\, f = \mathbb{R}_{++}$. The function $xy + log(x)$ is unbounded above if $y \geq 0$, and maximum if $y < 0, ; x = -1/y$. Therefore, $f^*(y) = -\log(-y) - 1$ with $\mathbf{dom}\, f^* = -\mathbb{R}_{++}$.
- *Exponential.* $f(x) = e^x$. The function $xy - e^x$ is unbounded if $y \leq 0$, maximum if $y > 0$, $x = \log(y)$, and maximum 0 if $y = 0$. Therefore, $f^*(y) = y\log(y) - y$ with $\mathbf{dom}\, f^* = \mathbb{R}_+$.
- *Negative entropy.* $f(x) = x\log(x)$, $\mathbf{dom}\, f = \mathbb{R}_+$, $f(0) = 0$. The function $xy - x\log(x)$ is bounded above for all y, and attains maximum at $x = e^{y-1}$. Therefore, $f^*(y) = e^{y-1}$ with $\mathbf{dom}\, f^* = \mathbb{R}$.
- *Inverse.* $f(x) = 1/x$, $\mathbf{dom}\, f = \mathbb{R}_{++}$. The function $xy - 1/x$ is unbounded above if $y > 0$, maximum if $y < 0$, $x = (-y)^{-1/2}$, and maximum 0 if $y = 0$. Therefore, $f^*(y) = -2(-y)^{1/2}$ with $\mathbf{dom}\, f^* = -R_+$.

Example 3.22 (*Strictly convex quadratic function*)**.**

Example 3.23 (*Log-determinant*)**.**

Example 3.24 (*Indicator function*)**.**

Example 3.25 (*Log-sum-exp function*)**.**

Example 3.26 (*Norm*)**.**

Example 3.27 (*Norm squared*)**.**

Example 3.28 (*Revenue and profit functions*)**.**

3.3.2 Basic properties

3.4 Quasiconvex functions

3.4.1 Definition and examples

Example 3.29.

Example 3.30 (*Length of a vector*).

Example 3.31.

Example 3.32 (*Linear-fractional function*).

Example 3.33 (*Distance ratio function*).

Example 3.34 (*Internal rate of return*).

3.4.2 Basic properties

Example 3.35 (*Cardinality of a nonegative vector*).

Example 3.36 (*Rank of positive semidefinite matrix*).

3.4.3 Differentiable quasiconvex functions

3.4.4 Operations that preserve quasiconvexity

Example 3.37 (*Generalized eigenvalue*).

3.4.5 Representation via family of convex functions

Example 3.38 (*Convex over concave function*).

3.5 Log-concave and log-convex functions

3.5.1 Definition

A function $f : \mathbb{R}^n \to \mathbb{R}$ is *logarithmically concave* or *log-concave* if $f(x) > 0$ for all $x \in \mathbf{dom}\, f$ and $\log f$ is concave. Similarly, f is *logarithmically convex* or *log-convex* if $\log f$ is convex.

- f is log-convex if and only if $1/f$ is log-concave
- it is convenient to allow f to take on the value zero with $\log f(x) = -\infty$

A function $f : \mathbb{R}^n \to \mathbb{R}$ with convex domain and $f(x) > 0$ for all $x \in \mathbf{dom}\, f$, is log-concave if and only if

$$f(\theta x + (1-\theta)y) \geq f(x)^\theta f(y)^{1-\theta} \quad \forall x, y \in \mathbf{dom}\, f$$

where $0 \leq \theta \leq 1$.

Example 3.39 (*Some simple examples of log-concave and log-convex functions*).

Example 3.40 (*Log-concave density functions*).

3.5.2 Properties

Twice differentiable log-convex/concave functions

Suppose f is twice differentiable, with $\mathbf{dom}\, f$ convex, so

$$\nabla^2 \log f(x) = \frac{1}{f(x)} \nabla^2 f(x) - \frac{1}{f(x)^2} \nabla f(x) \nabla f(x)^T$$

Hence f is log-convex if and only if for all $x \in \mathbf{dom}\, f$

$$f(x) \nabla^2 f(x) \succeq \nabla f(x) \nabla f(x)^T$$

and log-concave if and only if for all $x \in \mathbf{dom}\, f$

$$f(x) \nabla^2 f(x) \preceq \nabla f(x) \nabla f(x)^T$$

Example 3.41 (*Laplace transform of a nonnegative function and the moment and cumulant generating functions*).

Example 3.42.

Example 3.43 (*Yield function*).

Example 3.44 (*Volume of polyhedron*).

3.6 Convexity with respect to generalized inequalities

3.6.1 Monotonicity with respect to a generalized inequality

Example 3.45 (*Monotone vector functions*).

Example 3.46 (*Matrix monotone functions*).

3.6.2 Convexity with respect to a generalized inequality

Example 3.47 (*Convexity with respect to componentwise inequality*).

Example 3.48 (*Matrix convexity*).

Example 3.49.

Chapter 4

Convex optimization problems

4.1 Optimization problems

4.1.1 Basic terminology

Problem notation

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0,\ i = 1, \ldots, m \\ & h_i(x) = 0,\ i = 1, \ldots, p \end{array} \tag{4.1}$$

- *optimization variable*: $x \in \mathbb{R}^n$
- *objective/cost function*: $f_0 : \mathbb{R}^n \to \mathbb{R}$
- *inequality constraints function*: $f_i(x) : \mathbb{R}^n \to \mathbb{R}$
- *equality constraints function*: $h_i(x) : \mathbb{R}^n \to \mathbb{R}$
- if there are no constraints, we say problem (4.1) is *unconstrained*
- domain of problem (4.1): $\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \ \cap \ \bigcap_{i=0}^{p} \mathbf{dom}\, h_i$
- a point $x \in \mathcal{D}$ is *feasible* if it satisfies all constraints
- *feasible/constraint set*: the set of all feasible points

The *optimal value* p^* is defined by

$$p^* = \inf\{f_0(x) \mid f_i(x) \leq 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, p\}$$

- allow p^* take on extend values $\pm\infty$
- if the problem is *infeasible*, then $p^* = \inf\{\varnothing\} = \infty$
- if the problem is *unbounded below*, then there are feasible points x_k with $f_0(x_k) \to -\infty$ as $k \to \infty$, $p^* = -\infty$

Optimal and locally optimal points

The *optimal point* x^*: x^* is feasible and $f_0(x) = p^*$. Hence, *optimal set* is denoted by

$$X_{\text{opt}} = \{x \mid f_i(x) \leq 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, p,\ f_0(x) = p^*\}$$

- if there exists optimal point in (4.1), then the optimal value is *attained* or *achieved*
- the problem is then *solvable*

A feasible point x with $f_0(x) \leq p^* + \epsilon$, $\epsilon > 0$ is called ϵ-*suboptimal*, and the set of all ϵ-*suboptimal* points is called the ϵ-*suboptimal set*.

A point x is *locally optimal* if there is an $R > 0$ such that

$$f_0(x) = \inf\{f_0(z) \mid f_i(z) \leq 0,\ i = 1, \ldots, m,\ h_i(z) = 0,\ i = 1, \ldots, p,\ \|z - x\|_2^2 \leq R\}$$

or, x solves the problem

$$\begin{array}{ll} \text{minimize} & f_0(z) \\ \text{subject to} & f_i(z) \leq 0,\ i = 1, \ldots, m \\ & h_i(z) = 0,\ i = 1, \ldots, p \\ & \|z - x\|_2^2 \leq R \end{array}$$

- if x is feasible and $f_i(x) = 0$, then the constraint is *active* at x
- if x is feasible and $f_i(x) < 0$, then the constraint is *inactive* at x
- if x is feasible, then all the equality constraints are *active*
- a constraint is *redundant* if deleting it does not change the feasible set

Example 4.1. A few simple unconstrained problems with $x \in \mathbb{R}$ and $\mathbf{dom}\, f_0 = \mathbb{R}_{++}$

- $f_0(x) = 1/x$: $p^* = 0$, but the optimal value is not achieved (x^* does not exist)
- $f_0(x) = -\log x$: $p^* = -\infty$, the problem is unbounded below
- $f_0(x) = x \log x$: $p^* = -1/e$ with unique optimal point $x^* = 1/e$

Feasibility problems

If the objective function is identically zero, then p^* is either 0 or ∞ (if feasible set is empty). The *feasibility problem* is

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

The problem is to determine whether the constraints are consistent.

4.1.2 Expressing problems in standard form

We refer to (4.1) as an optimization problem in standard form.

Example 4.2 (*Box constraints*)**.** The problem

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

where $x \in \mathbb{R}^n$. Its feasible set is a box, and the standard form is

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & l_i - x_i \leq 0,\ i = 1, \ldots, n \\ & x_i - u_i \leq 0,\ i = 1, \ldots, n \end{array}$$

Maximization problems

The maximization problems can be solved by minimizing $-f_0$ and have the form

$$\begin{array}{ll} \text{maximize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0,\ i = 1, \ldots, m \\ & h_i(x) = 0,\ i = 1, \ldots, p \end{array} \tag{4.2}$$

- optimal value: $p^* = \sup\{f_0(x) \mid f_i(x) \leq 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, p\}$
- a feasible point x is ϵ-suboptimal if $f_0(x) \geq p^* - \epsilon$

4.1.3 Equivalent problems

Informally, we call two problems *equivalent* if from a solution of one, a solution of the other is readily found, and vice versa. A simple example,

$$\begin{array}{ll} \text{minimize} & \tilde{f}(x) = \alpha_0 f_0(x) \\ \text{subject to} & \tilde{f}_i(x) = \alpha_i f_i(x) \leq 0,\ i = 1, \ldots, m \\ & \tilde{h}_i(x) = \beta_i h_i(x) = 0,\ i = 1, \ldots, p \end{array} \tag{4.3}$$

where $\alpha_i > 0$ and $\beta_i \neq 0$.
(4.1) and (4.3) are equivalent but not the same.

Change of variables

Suppose $\phi : \mathbb{R}^n \to \mathbb{R}^n$ is one-to-one with $\phi(\mathbf{dom}\,\phi) \supseteq \mathcal{D}$ (domain of **??**). Define $\tilde{f}_i$ and $\tilde{h}_i$ as

$$\tilde{f}_i(z) = f_i(\phi(z)),\ i = 0, \ldots, m \qquad \tilde{h}_i(z) = h_i(\phi(z)),\ i = 1, \ldots, p$$

Consider the problem

$$\begin{array}{ll} \text{minimize} & \tilde{f}_0(x) \\ \text{subject to} & \tilde{f}_i(x) \leq 0,\ i = 1, \dots, m \\ & \tilde{h}_i(x) = 0,\ i = 1, \dots, p \end{array} \tag{4.4}$$

with variable z. Then (4.1) and (4.4) are equivalent.

Transformation of objective and constraint functions

Suppose that

1. $\psi_0 : \mathbb{R} \to \mathbb{R}$ is monotone increasing.

2. $\psi_1, \dots, \psi_m : \mathbb{R} \to \mathbb{R}$ satisfy $\psi_i(u) \leq 0 \Leftrightarrow u \leq 0$.

3. $\psi_{m+1}, \dots, \psi_{m+p} : \mathbb{R} \to \mathbb{R}$ satisfy $\psi_i(u) = 0 \Leftrightarrow u = 0$.

Define $\tilde{f}_i$ and $\tilde{h}_i$ as

$$\tilde{f}_i(x) = \psi_i(f_i(x)),\ i = 0, \dots, m \qquad \tilde{h}_i(x) = \psi_{m+i}(h_i(x)),\ i = 1, \dots, p$$

The problem

$$\begin{array}{ll} \text{minimize} & \tilde{f}_0(x) \\ \text{subject to} & \tilde{f}_i(x) \leq 0,\ i = 1, \dots, m \\ & \tilde{h}_i(x) = 0,\ i = 1, \dots, p \end{array}$$

and (4.1) are equivalent. This is the general rule of (4.3).

Example 4.3 (*Least-norm and least-norm-squared prolems*)**.** Consider an unconstrained Euclidean norm minimization problem

$$\text{minimize} \quad \|Ax - b\|_2 \tag{4.5}$$

with variable $x \in \mathbb{R}^n$. Consider another problem

$$\text{minimize} \quad \|Ax - b\|_2^2 = (Ax - b)^T(Ax - b) \tag{4.6}$$

The problems (4.5) and (4.6) are equivalent but not the same.

For example, $\|Ax - b\|_2$ is not differentiable at any x with $Ax - b = 0$, whereas $\|Ax - b\|_2^2$ is differentiable for all x.

Slack variables

There exists $s_i \geq 0$ that satisfies $f_i(x) + s_i = 0$, and (4.1) is equivalent to the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & s_i \geq 0,\ i = 1, \ldots, m \\ & f_i(x) + s_i = 0,\ i = 1, \ldots, m \\ & h_i(x) = 0,\ i = 1, \ldots, p \end{array} \tag{4.7}$$

- variables: $n \to n + m$
- inequality constraints: $m \to m$
- equality constraints: $p \to p + m$

Eliminating equality constraints

Eliminating linear equality constraints

Introducing equality constraints

Optimizing over some variables

Example 4.4 (*Minimizing a quadratic function with constraints on some variables*).

Epigraph problem form

Implicit and explicit constraints

4.1.4 Parameter and oracle problem descriptions

4.2 Convex optimization

A *convex optimization problem* has the form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0,\ i = 1, \ldots, m \\ & a_i^T x = b_i,\ i = 1, \ldots, p \end{array} \tag{4.8}$$

Requirements:

- the objective function is convex
- the inequality constraint functions are convex
- the equality constraint function $h_i(x) = a_i^T x - b$ is affine

Properties:

- its feasible set is convex (assume constraints are consistent)
- if f_0 is quasiconvex, then (4.8) is a standard form *quasiconvex optimization problem*
- for both convex and quasiconvex problems, the optimal set is convex
- if f_0 is strictly convex, then the optimal set contains at most one point

Concave maximization problems

Problem (4.8) with maximizing concave function $f_0(x)$ is also a convex optimization problem.

$$\begin{array}{ll} \text{maximize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \; i = 1, \dots, m \\ & a_i^T x = b_i, \; i = 1, \dots, p \end{array} \tag{4.9}$$

Abstract form convex optimization problem

4.2.1 Convex optimization problems in standard form

4.2.2 Local and global optima

For a convex optimization problem, **any locally optimal point is also globally optimal**.

Proof. proof ....:

$$f_0(x) = \inf\{f_0(z) \mid z \text{ feasible}, \; \|z - x\|_2 \le R\} \tag{4.10}$$

for some $R > 0$ □

4.2.3 An optimality criterion for differentiable f_0

Suppose f_0 is differentiable, we have (3.2)

$$f_0(y) \ge f_0(x) + \nabla f_0(x)^T (y - x) \quad x, y \in \mathbf{dom}\, f_0 \tag{4.11}$$

Let X denote the feasible set, then x is optimal if and only if $x \in X$ and

$$\nabla f_0(x)^T (y - x) \ge 0 \quad \forall y \in X \tag{4.12}$$

Proof of optimality condition

Unconstrained problems

Problems with equality constraints only

Example 4.5 (*Unconstrained quadratic optimization*)**.** Consider $f_0 = (1/2)x^T P x + q^T x + r$ where $P \in \mathbb{S}_+^n$ (which makes f_0 convex). Then x is a minimizer of f_0 if and only if

$$\nabla f_0(x) = Px + q = 0$$

- if $q \notin \mathcal{R}(P)$, then there is no solution. (f_0 is unbounded below)
- if $P \succ 0$ (which make f_0 strictly convex), then there is a unique minimizer $x^* = -P^{-1}q$
- if P is singular but $q \in \mathcal{R}(P)$, then $X_{\text{opt}} = -P^{\dagger}q + \mathcal{N}(P)$

Example 4.6 (*Analytic centering*).

4.2.4 Equivalent convex problems

4.2.5 Quasiconvex optimization

4.3 Linear optimization problems

A *linear programming* (LP) problem has the form

$$\begin{array}{ll} \text{minimize} & c^T x + d \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{array} \tag{4.13}$$

where $G \in \mathbb{R}^{m \times n}$ and $A \in \mathbb{R}^{p \times n}$.

Standard and inequality form linear programs

A *standard form LP* is

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array} \tag{4.14}$$

A *inequality form LP* is

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \end{array} \tag{4.15}$$

Converting LPs to standard form

4.3.1 Examples

4.3.2 Linear-fractional programming

A *linear-fractional program* has the form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{array} \tag{4.16}$$

where $f_0(x) = \dfrac{c^T x + d}{e^T x + f}$ and $\mathbf{dom}\, f_0 = \{x \mid e^T x + f > 0\}$.

Example 4.7 (*Von Neumann growth problem*).

4.4 Quadratic optimization problems

A *quadratic program* (QP) has the form

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{array} \tag{4.17}$$

where $P \in \mathbb{S}_+^n$, $G \in \mathbb{R}^{m\times n}$, and $A \in \mathbb{R}^{p\times n}$. A *quadratically constrained quadratic program* (QCQP) has the form

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P_0 x + q_0^T x + r_0 \\ \text{subject to} & (1/2)x^T P_i x + q_i^T x + r_i \leq 0,\ i = 1, \ldots, m \\ & Ax = b \end{array} \tag{4.18}$$

where $P_i \in \mathbb{S}_+^n,\ i = 0, \ldots, m$.

- QP with $P = 0$ is LP
- QCQP with $P_i = 0,\ i = 1, \ldots, m$ is QP

4.4.1 Examples

4.4.2 Second-order cone programming

Example 4.8 (*Portfolio optimization with loss risk constraints*).

4.5 Geometric programming

4.5.1 Monomials and posynomials

4.5.2 Geometric programming

4.5.3 Geometric program in convex form

4.5.4 Examples

4.6 Generalized inequality constraints

4.6.1 Conic form problems

4.6.2 Semidefinite programming

4.6.3 Examples

4.7 Vector optimization

4.7.1 General and convex vector optimization problems

Example 4.9 (*Best linear unbiased estimator*).

4.7.2 Optimal points and values

4.7.3 Pareto optimal points and values

4.7.4 Scalarization

Example 4.10 (*Minimal upper bound on a set of matrices*).

4.7.5 Multicriterion optimization

4.7.6 Examples

Chapter 5

Duality

5.1 The Lagrange dual function

5.1.1 The Lagrangian

The optimization problem from (4.1)

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \; i = 1, \ldots, m \\ & h_i(x) = 0, \; i = 1, \ldots, p \end{array} \tag{5.1}$$

with variable $x \in \mathbb{R}^n$, $\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \; \cap \; \bigcap_{i=0}^{p} \mathbf{dom}\, h_i$ is not empty, and optimal value is p^*.

We define *Lagrangian* $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ associated with (5.1) as

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

with $\mathbf{dom}\, L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$.

- λ_i is *Lagrange multiplier* associated with the ith inequality constraints
- ν_i is *Lagrange multiplier* associated with the ith equality constraints
- λ and ν are *dual variables* or *Lagrange multiplier vectors*

5.1.2 The Lagrange dual function

We define *Lagrange dual function* $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ as

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

where $\lambda \in \mathbb{R}^m$ and $\nu \in \mathbb{R}^p$.

- when the Lagrangian is unbounded below in x, the dual function takes on $-\infty$
- the dual function is concave since it is the pointwise infimum of affine functions

5.1.3 Lower bounds on optimal value

For any $\lambda \succeq 0$, any ν and optimal value p^* in (5.1) we have

$$g(\lambda, \nu) \leq p^* \tag{5.2}$$

Proof. Suppose $\tilde{x}$ is a feasible point, *i.e.*, $f_i(\tilde{x}) \leq 0$ and $h_i(\tilde{x}) = 0$, then

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

where $\lambda \succeq 0$, therefore

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

Hence, we have $g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x})$ which leads to (5.2) $\square$

- (5.2) is vacuous when $g(\lambda, \nu) = -\infty$
- The dual function $g(\lambda, \nu)$ gives a nontrivial lower bound on p^* only when $\lambda \succeq 0$ and $(\lambda, \nu) \in \textbf{dom}\, g$, and for the sufficient pair (λ, ν) is called *dual feasible*.

5.1.4 Linear approximation interpretation

Add in the future.

5.1.5 Examples

Least-squares solution of linear equations

$$\begin{array}{ll} \text{minimize} & x^T x \\ \text{subject to} & Ax = b \end{array} \tag{5.3}$$

where $A \in \mathbb{R}^{p \times n}$

The Lagrangian is $L(x, \nu) = x^T x + \nu^T(Ax - b)$ with domain $\mathbb{R}^n \times \mathbb{R}^p$, and the dual function is $g(\nu) = \inf_x L(x, \nu)$. Since $L(x, \nu)$ is a convex quadratic function of x, the optimal condition is

$$\nabla_x L(x, \nu)|_{x = x^*} = 2x^* + A^T \nu = 0$$

which yields $x^* = -\frac{1}{2} A^T \nu$, therefore $g(\nu) = L(x^*, \nu) = -\frac{1}{4} \nu^T A A^T \nu - b^T \nu$ which is a concave quadratic function with domain $\mathbb{R}^p$.

The lower bound property (5.2) states that for any $\nu \in \mathbb{R}^p$, we have

$$-\frac{1}{4} \nu^T A A^T \nu - b^T \nu \leq \inf\{x^T x \mid Ax = b\}$$

Standard form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array} \tag{5.4}$$

The n inequalities are $f_i(x) = -x_i \leq 0$, and the Lagrangian becomes

$$L(x, \lambda, \nu) = c^T x - \sum_{i=1}^{n} \lambda_i x_i + \nu^T (Ax - b) = -b^T \nu + (c + A^T \nu - \lambda)^T x$$

The dual function is

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = -b^T \nu + \inf_x (c + A^T \nu - \lambda)^T x$$

Note that a linear function is bounded below only when it is identically zero, *i.e.*,

$$g(\lambda, \nu) = \begin{cases} -b^T \nu, & c + A^T \nu - \lambda = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

The lower bound property (5.2) is nontrivial only when $\lambda \succeq 0$ and $A^T \nu + c = \lambda \succeq 0$. When this occurs, $-b^T \nu$ is a lower bound on the optimal value of the LP (5.4).

Two-way partitioning problem

Consider the nonconvex problem

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \; i = 1, \ldots, n \end{array} \tag{5.5}$$

where $W \in \mathbb{S}^n$.

For large n, this problem is difficult to solve since the feasible set grows exponentially. We can interpret the problem in another way.

- partition the feasible set: $\{1, \ldots, n\} = \{i \mid x_i = -1\} \cup \{i \mid x_i = 1\}$
- view $W_i j$ as the cost of having elements i and j in the same partition
- view $-W_i j$ as the cost of having elements i and j in the different partition
- view $x^T W x$ as total cost over all pairs of elements

The Lagrangian is

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

The dual function is

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

We can get the bound on the optimal value p^*. For example, choose $\nu = -\lambda_{\min}(W)\mathbf{1}$ which is dual feasible since $W + \mathbf{diag}(\nu) = W - \lambda_{\min}(W)I \succeq 0$. Therefore,

$$p^* \geq -\mathbf{1}^T \nu = n\lambda_{\min}(W) \tag{5.6}$$

Remark 5.1. Obtain the lower bound without using Lagrange dual function.
Modify the problem

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & \displaystyle\sum_{i=1}^{n} x_i^2 = n \end{array} \tag{5.7}$$

The problem (5.5) imply the constraint in (5.7). Therefore, optimal value of (5.7) is a lower bound on optimal value of (5.5), *i.e.*, $n\lambda_{\min}(W) \leq p^*$.

5.1.6 The Lagrange dual function and conjugate functions

Add in the future.

Equality constrained norm minimization

Add in the future.

Entropy maximization

Add in the future.

Minimum volume covering ellipsoid

Add in the future.

5.2 The Lagrange dual problem

The Lagrange dual function gives us a lower bound on the optimal value p^*. The problem

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array} \tag{5.8}$$

is therefore called *Lagrange dual problem* associated with the problem (5.1).

- problem (5.1) is called the *primal problem*
- a pair (λ, ν) is *dual feasible* ($\lambda \succeq 0$ and $g(\lambda, \nu) > -\infty$) implies it is feasible for the dual problem
- (λ^*, ν^*) is *dual optimal* or *optimal Lagrange multipliers*
- (5.8) is a convex optimization problem since it maximize a concave function with convex constraint functions (even if (5.1) is not convex)

5.2.1 Making dual constraints explicit

It is common for the domain of the dual function

$$\mathbf{dom}\, g = \{(\lambda, \nu) \mid g(\lambda, \nu) > -\infty\}$$

to have dimension smaller than $m + p$. We identify the affine hull of $\mathbf{dom}\, g$ and describe it as a set of linear equality constraints.

Lagrange dual of standard form LP

The standard form LP is

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array} \tag{5.9}$$

And the Lagrange dual problem of the standard form LP (5.9) is

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) = \begin{cases} -b^T \nu, & A^T \nu + c - \lambda = 0 \\ -\infty, & \text{otherwise} \end{cases} \\ \text{subject to} & \lambda \succeq 0 \end{array} \tag{5.10}$$

Problem (5.10) is equivalent to

$$\begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & A^T \nu + c - \lambda = 0 \\ & \lambda \succeq 0 \end{array} \tag{5.11}$$

Thus equivalent to

$$\begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & A^T \nu + c \succeq 0 \end{array} \tag{5.12}$$

which is an LP in inequality form. With abuse of the terminology, we call (5.11) or (5.12) is the Lagrange dual of (5.9).

Lagrange dual of inequality form LP

The inequality form LP is

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \end{array} \tag{5.13}$$

The Lagrangian is

$$L(x, \lambda) = c^T x + \lambda^T (Ax - b) = -b^T \lambda + (A^T \lambda + c)^T x$$

The dual function is

$$g(\lambda) = \inf_x L(x, \lambda) = -b^T \lambda + \inf_x (A^T \lambda + c)^T x = \begin{cases} -b^T \lambda, & A^T \lambda + c = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

And the Lagrange dual of the inequality form LP (5.13) is

$$\begin{array}{ll} \text{maximize} & -b^T \lambda \\ \text{subject to} & A^T \lambda + c = 0 \\ & \lambda \succeq 0 \end{array} \tag{5.14}$$

which is an LP in standard form. Note that the dual of a standard form LP is an inequality form LP, and vice versa.

5.2.2 Weak duality

The *weak duality* property

$$d^* \leq p^* \tag{5.15}$$

where d^* is the optimal value of the Lagrange dual problem.

The inequality (5.15) holds when d^* and p^* are infinite.

- if the primal problem is unbounded below, then the Lagrange dual problem is infeasible

$$p^* = -\infty \Rightarrow d^* = -\infty$$

- if the Lagrange dual problem is unbounded above, then the primal problem is infeasible

$$d^* = \infty \Rightarrow p^* = \infty$$

- the difference $p^* - d^*$ is called *optimal duality gap* of the original problem
- the *optimal duality gap* is always nonnegative
- (5.15) can be used to find a lower bound on the optimal value of a difficult problem

Note that the dual problem of of (5.5) is an SDP,

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

5.2.3 Strong duality and Slater's constraint qualification

The *strong duality* property is the inequality

$$d^* = p^* \tag{5.16}$$

holds, *i.e.*, *optimal duality gap* is zero.

- strong duality does not, in general, hold
- if the primal problem (5.1) is convex, *i.e.*,

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0,\ i = 1, \ldots, m \\ & Ax = b \end{array} \tag{5.17}$$

  where $f_0, \ldots, f_m$ are convex, we usually (but not always) have strong duality

- *constraint qualifications*: strong duality holds under a condition, beyond convexity

Slater's condition: There exists an $x \in \textbf{relint}\, \mathcal{D}$ such that

$$f_i(x) < 0,\ i = 1, \ldots, m,\ Ax = b \tag{5.18}$$

then the point is called *strictly feasible*. Slater's theorem states that strong duality holds, if Slater's condition holds (and the problem is convex). And we prove it in §5.3.2.
Refined Slater's condition: $f_1, \ldots, f_k$ are affine, there exists an $x \in \textbf{relint}\, \mathcal{D}$ such that

$$f_i(x) \le 0,\ i = 1, \ldots, k,\ f_i(x) < 0,\ i = k+1, \ldots, m,\ Ax = b \tag{5.19}$$

It reduces to feasibility when the constraints are all linear and $\textbf{dom}\, f_0$ is open.

5.2.4 Examples

Least-squares solution of linear equations

Recall the problem (5.3)

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

And its dual problem (unconstrained concave quadratic maximization problem)

$$\text{maximize} \quad -\frac{1}{4}\nu^T A A^T \nu - b^T \nu$$

Slater's condition is that the primal problem is feasible, so $p^* = d^*$ provided $b \in \mathcal{R}(A)$. In fact, strong duality holds even if $p^* = \infty$ ($b \notin \mathcal{R}(A)$). This leads to that there is a z with $A^T z = 0,\ b^T z \neq 0$, and dual function is unbounded above along $\{tz \mid t \in \mathbb{R}\}$, so $d^* = \infty$.

Lagrange dual of LP

By the weaker Slater's condition, strong duality holds for any LP (both standard and inequality form) provided the primal problem is feasible. Apply the result to the duals, strong duality holds for any LP provided the dual problem is feasible. Therefore, strong duality only fails when both the primal and dual problems are infeasible.

Lagrange dual of QCQP

Entropy maximization

Minimum volume covering ellipsoid

A nonconvex quadratic problem with strong duality

5.2.5 Mixed strategies for matrix games

5.3 Geometric interpretation

5.3.1 Weak and strong duality via set of values

5.3.2 Proof of strong duality under constraint qualification

5.3.3 Multicriterion interpretation

5.4 Saddle-point interpretation

5.4.1 Max-min characterization of weak and strong duality

5.4.2 Saddle-point interpretation

5.4.3 Game interpretation

5.4.4 Price or tax interpretation

5.5 Optimality conditions

Remind that we do not assume the problem (5.1) is convex.

5.5.1 Certificate of suboptimality and stopping criteria

If we can find a dual feasible (λ, ν), then we establish a lower bound of the primal problem. We call (λ, ν) provides a *proof* or *certificate* that $p^* \geq g(\lambda, \nu)$.

- strong duality means there exist arbitrarily good certificates

- Dual feasible points allow us to bound how suboptimal (§4.1.1) a given feasible point is without knowing p^*. If x is primal feasible and (λ, ν) is dual feasible, then we have

$$\begin{cases} f_0(x) \leq p^* + f_0(x) - g(\lambda, \nu) \\ g(\lambda, \nu) \geq d^* - (f_0(x) - g(\lambda, \nu)) \end{cases}$$

  i.e., x and (λ, ν) are ϵ-suboptimal for its problem where $\epsilon = f_0(x) - g(\lambda, \nu)$.

- the difference $f_0(x) - g(\lambda, \nu)$ is called the *duality gap* with the two feasible points

- a primal dual feasible pair x, (λ, ν) localizes the optimal value of the problems

$$p^* \in [g(\lambda, \nu), f_0(x)],\ d^* \in [g(\lambda, \nu), f_0(x)]$$

  the width of the interval is the duality gap

- If the duality gap is zero, then both x and (λ, ν) are optimal. Therefore, we can think of (λ, ν) as a certificate that proves x is optimal, and vice versa.

The nonheuristic stopping criteria in the optimization algorithm

$$f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)}) \leq \epsilon_{abs}$$

where $\epsilon_{abs} > 0$ guarantees that when the algorithm terminates, $x^{(k)}$ is ϵ_{abs}-suboptimal with $(\lambda^{(k)}, \nu^{(k)})$ as a certificate. Strong duality must hold if this method is to work for arbitrarily small tolerances ϵ_{abs}.

A similar condition is to use relative accuracy $\epsilon_{rel} > 0$. If

$$g(\lambda^{(k)}, \nu^{(k)}) > 0,\ \frac{f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)})}{g(\lambda^{(k)}, \nu^{(k)})} \leq \epsilon_{rel}$$

holds, or

$$f_0(x^{(k)}) < 0,\ \frac{f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)})}{-f_0(x^{(k)})} \leq \epsilon_{rel}$$

holds, then $p^* \neq 0$ and the relative error

$$\frac{f_0(x^{(k)}) - p^*}{|p^*|}$$

is guaranteed to be less than or equal to ϵ_{rel}.

5.5.2 Complementary slackness

Suppose strong duality holds, then

$$\begin{aligned} f_0(x^*) &= g(\lambda^*, \nu^*) && \text{(strong duality)} \\ &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x) \right) && \text{(definition)} \\ &\le f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*) && (\text{when } x = x^*) \\ &\le f_0(x^*) \end{aligned}$$

the fourth line is derived by

- (λ^*, ν^*) is dual feasible, then $\lambda^* \succeq 0$
- x^* is optimal, then $f_i(x^*) \le 0,\ i = 1, \ldots, m,\ h_i(x^*) = 0,\ i = 1, \ldots, p$

Therefore, the last two inequalities becomes equalities, and we have some conclusions here

- x^* minimizes $L(x, \lambda^*, \nu^*)$ over x (L can have other minimizers)
- the *complementary slackness* condition

$$\lambda_i^* f_i(x^*) = 0,\ i = 1, \ldots, m \tag{5.20}$$

- when strong duality holds, the *complementary slackness* condition holds for any primal optimal x^* and dual optimal (λ^*, ν^*)

We can express the complementary slcakness condition as

$$\lambda_i^* > 0 \Rightarrow f_i(x^*) = 0$$

or, equivalently

$$f_i(x^*) < 0 \Rightarrow \lambda_i^* = 0$$

This means the ith optimal Lagrange multiplier is zero unless the ith constraint is active at the optimum.

5.5.3 KKT optimality conditions

We now assume that $f_0, \ldots, f_m, h_1, \ldots, h_p$ are differentiable (and therefore have open domains).

KKT conditions for nonconvex problems

Let x^* and (λ^*, ν^*) be any primal and dual optimal points. Since x^* minimizes $L(x, \lambda^*, \nu^*$ as discussed in §5.5.2, its gradient must vanish at x^*.

Therefore, we have *Karush-Kuhn-Tucher* (KKT) conditions.

$$\begin{aligned} f_i(x^*) &\leq 0, \ i = 1, \ldots, m \\ h_i(x^*) &= 0, \ i = 1, \ldots, p \\ \lambda_i^* &\geq 0, \ i = 1, \ldots, m \\ \lambda_i^* f_i(x^*) &= 0, \ i = 1, \ldots, m \\ \nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p \nu_i^* \nabla h_i(x^*) &= 0 \end{aligned} \tag{5.21}$$

To summarize, for *any* optimization problems with

1. differentiable objective and constraint functions
2. strong duality holds

then any pair of primal and dual optimal points must satisfy the KKT conditions (5.21).

KKT conditions for convex problems

KKT conditions become sufficient for optimal points when the primal problem is convex, *i.e.*, if f_i are convex, h_i are affine, and any points $\tilde{x}$, $\tilde{\lambda}$, $\tilde{\nu}$ satisfy

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

then $\tilde{x}$ and $(\tilde{\lambda}, \tilde{\nu})$ are primal and dual optimal with zero duality gap.

Proof. First two conditions states that $\tilde{x}$ is primal feasible. Since $\tilde{\lambda} \succeq 0$, $L(x, \tilde{\lambda}, \tilde{\nu})$ is convex in x. Thus the last condition states that $\tilde{x}$ minimizes $L(x, \tilde{\lambda}, \tilde{\nu})$ over x. We have

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

where we use the third and the fourth KKT conditions.

We show that $\tilde{x}$ and $(\tilde{\lambda}, \tilde{\nu})$ have zero duality gap, and are primal and dual optimal. □

To summarize, for any *convex* optimization problems with

1. differentiable objective and constraint functions
2. any points that satisfy the KKT conditions

the points are primal and dual optimal, and have zero duality gap.

If a convex optimization problem with differentiable objective and constraint functions satisfies Slater's condition (optimal duality gap is zero), then KKT conditions provide necessary and sufficient conditions for optimality, *i.e.*, x is optimal if and only if there exists (λ, ν) that satisfies KKT conditions with x.

Example 5.1 (*Equality constrained convex quadratic minimization*)**.** Consider

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Ax = b \end{array} \tag{5.22}$$

where $P \in \mathbb{S}^n_+$. The KKT conditions are

$$Ax^* = b, \quad Px^* + q + A^T \nu^* = 0$$

i.e.,

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$$

Solving x^* and ν^* gives the optimal primal and dual variables for (5.22).

Example 5.2 (*Water-filling*)**.** Consider

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^{n} \log(\alpha_i + x_i) \\ \text{subject to} & x \succeq 0, \ \mathbf{1}^T x = 1 \end{array}$$

where $\alpha_i > 0$. This problem arises in information theory

- x_i represents the transmitter power allocated to ith channel
- $\log(\alpha_i + x_i)$ gives the capacity or communication rate of the channel
- then the problem is to allocate total power of one to the channels to maximize the communication rate

Introducing $\lambda^* \in \mathbb{R}^n$ and $\nu \in \mathbb{R}$, then the KKT conditions are

$$\begin{gathered} x^* \succeq 0, \quad \mathbf{1}^T x = 1, \quad \lambda^* \succeq 0, \quad \lambda_i^* x_i^* = 0, \ i = 1, \dots, n \\ -1/(\alpha_i + x_i^*) - \lambda_i^* + \nu^* = 0, \ i = 1, \dots, n \end{gathered}$$

Eliminating λ^*,

$$x^* \succeq 0, \quad \mathbf{1}^T x = 1, \quad x_i^*(\nu^* - 1/(\alpha_i + x_i^*)) = 0,\ i = 1, \ldots, n$$
$$\nu^* \geq 1/(\alpha_i + x_i^*),\ i = 1, \ldots, n$$

5.5.4 Mechanics interpretation of KKT conditions

5.5.5 Solving the primal problem via the dual

5.6 Perturbation and sensitivity analysis

5.6.1 The perturbed problem

5.6.2 A global inequality

5.6.3 Local sensitivity analysis

5.7 Examples

5.7.1 Introducing new variables and equality constraints

5.7.2 Transforming the objective

5.7.3 Implicit constraints

5.8 Theorems of alternatives

5.8.1 Weak alternatives via the dual function

5.8.2 Strong alternatives

5.8.3 Examples

5.9 Generalized inequalities

5.9.1 The Lagrange dual

5.9.2 Optimality conditions

5.9.3 Perturbation and sensitivity analysis

5.9.4 Theorems of alternatives

Part II

Algorithms

Chapter 6

Unconstrained minimization

6.1 Unconstrained minimization problems

We discuss methods for solving the problem

$$\text{minimize} \quad f(x) \tag{6.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is convex and twice continuously differentiable ($\mathbf{dom}\, f$ is open). The necessary and sufficient condition for a point x^* to be optimal is

$$\nabla f(x^*) = 0 \tag{6.2}$$

We usually solve the problem by iterative algorithm except for some special cases. A *minimizing sequence* is $x^{(0)}, x^{(1)}, \cdots \in \mathbf{dom}\, f$ with $f(x^{(k)}) \to p^*$ as $k \to \infty$. The algorithm is terminated when $f(x^{(k)}) - p^* \leq \epsilon$ where $\epsilon > 0$ is tolerance.

Initial point an sublevel set

The methods described in this chapter require that the start point $x^{(0)}$ lies in $\mathbf{dom}\, f$, and the sublevel set

$$S = \{x \in \mathbf{dom}\, f \mid f(x) \leq f(x^{(0)})\} \tag{6.3}$$

must be closed.

- this condition is satisfied for all $x^{(0)} \in \mathbf{dom}\, f$ if f is closed (definition of closed function in §A.3.3)
- continuous functions with $\mathbf{dom}\, f = \mathbb{R}^n$ are closed
- continuous functions with open domains, for which $f(x)$ tends to infinity as x approaches $\mathbf{bd}\,\mathbf{dom}\, f$, are closed

6.1.1 Examples

Quadratic minimization and least-squares

The general quadratic minimization form

$$\text{minimize} \quad (1/2)x^TPx + q^Tx + r \tag{6.4}$$

where $P \in \mathbb{S}^n_+, q \in \mathbb{R}^n, r \in \mathbb{R}$.

- if $P \succ 0$ then there is a unique solution $x^* = -P^{-1}q$
- if $P \not\succ 0$ then any solution of $Px^* = -q$ is optimal
- if $Px^* = -q$ does not have a solution then the problem is unbounded below

The special cases of quadratic minimization problem, known as least-squares problem

$$\text{minimize} \quad \|Ax - b\|_2^2 = x^T(A^TA)x - 2(A^Tb)^Tx + b^Tb$$

The optimal condition $A^TAx^* = A^Tb$ are called the *normal equations* of the problem.

Unconstrained geometric programming

The unconstrained geometric program in convex form

$$\text{minimize} \quad f(x) = \log\left(\sum_{i=1}^m \exp(a_i^Tx + b_i)\right)$$

The optimal condition is

$$\nabla f(x^*) = \frac{1}{\sum_{j=1}^m \exp(a_j^Tx^* + b_j)} \sum_{i=1}^m \exp(a_i^Tx^* + b_i)a_i = 0$$

which in general has no analytical solution. For this problem $\mathbf{dom}\, f = \mathbb{R}^n$, so any point can be chosen as $x^{(0)}$.

Analytic center of linear inequalities

Consider the problem

$$\text{minimize} \quad f(x) = -\sum_{i=1}^m \log(b_i - a_i^Tx) \tag{6.5}$$

where the domain of f is the open set

$$\mathbf{dom}\, f = \{x \mid a_i^Tx < b_i,\ i = 1, \dots, m\}$$

- the objective function f is called the *logarithmic barrier* for inequalities $a_i^Tx \le b_i$
- the solution of (6.5) is called the *analytic center* of the inequalities
- the initial point must satisfy the strict inequalities $a_i^Tx^{(0)} < b_i,\ i = 1, \dots, m$
- since f is closed, the sublevel set S for any such point is closed

Analytic center of a linear matrix inequalities

Consider the problem

$$\text{minimize} \quad f(x) = \log\det F(x)^{-1} \tag{6.6}$$

where $F : \mathbb{R}^n \to \mathbb{S}^p$ is affine, *i.e.*,

$$F(x) = F_0 + x_1 F_1 + \cdots + x_n F_n$$

with $F_i \in \mathbb{S}^p$, and $\mathbf{dom}\, f = \{x \mid F(x) \succ 0\}$.

- the objective function f is called *logarithmic barrier* for the linear matrix inequality $F(x) \succeq 0$
- the solution of (6.6) is called the *analytic center* of the linear matrix inequality
- the initial point must satisfy the strict linear matrix inequality $F(x^{(0)}) \succ 0$
- since f is closed, the sublevel set S of any such point is closed

6.1.2 Strong convexity and implications

Add in the future!

Upper bound on $\nabla^2 f(x)$

Condition number of sublevel sets

Example 6.1 (*Condition number of an ellipsoid*).

The strong convexity constants

6.2 Descent methods

Minimizing sequence where

$$x^{(k+1)} = x^{(k)} + t^{(k)}\Delta x^{(k)},\ t^{(k)} > 0 \text{ when } x^{(k)} \text{ is not optimal}$$

- $\Delta x^{(k)}$ is called the *step* or *search direction*
- $t^{(k)} > 0$ is called the *step size* or *step length*
- we use lighter notation $x^+ = x + t\Delta x$ or $x := x + t\Delta x$
- *descent methods* means that $f(x^{(k+1)}) < f(x^{(k)})$, *i.e.*, all $x^{(k)} \in S$ (initial sublevel set)

We know that $\nabla f(x^{(k)})^T(y - x^{(k)}) \geq 0$ implies $f(y) \geq f(x^{(k)})$ from convexity property in §3.1.3, so the search direction must satisfy

$$\nabla f(x^{(k)})^T \Delta x^{(k)} < 0$$

The gradient descent methods is as follows.

Algorithm 6.1 (*General descent method*).
given a starting point $x \in \textbf{dom}\, f$.
repeat

1. *Determine a descent direction* Δx.
2. *Line search.* Choose a step size $t > 0$.
3. *Update.* $x := x + t\Delta x_{\text{nt}}$

until stopping criterion is satisfied

The stopping criterion is often of the form $\|\nabla f(x)\|_2 \leq \eta$ where η is small and positive, as suggested by the suboptimality condition (9.9).

Exact line search

Step size t is chosen to minimize f along the ray $\{x + \Delta x \mid t \geq 0\}$

$$t = \operatorname*{argmin}_{s \geq 0} f(x + s\Delta x) \tag{6.7}$$

This method can be used in some special cases, discussed in §9.7.1.

Backtracking line search

The *backtracking* line search needs two constants: $0 < \alpha < 0.5$ and $0 < \beta < 1$.

Algorithm 6.2 (*Backtracking line search*).
given a descent direction Δx for f at $x \in \textbf{dom}\, f$, $\alpha \in (0, 0.5)$, $\beta \in (0, 1)$.
$t := 1$
while $f(x + t\Delta x) > f(x) + \alpha t \Delta f(x)^T \Delta x$, $t := \beta t$.

It starts with unit step size and then reduces it by the factor β until

$$f(x + t\Delta x) \leq f(x) + \alpha t \nabla f(x)^T \Delta x$$

holds. Since Δx is a descent direction, we have small enough t such that

$$f(x + t\Delta x) \approx f(x) + t\nabla f(x)^T \Delta x < f(x) + \alpha t \nabla f(x)^T \Delta x$$

which shows that the backtracking line search eventually terminates. The α can be interpreted as the fraction of the decrease in f predicted by linear extrapolation that we will accept. The figure 6.1 suggests that the backtracking line search stops with t that satisfies

$$t = 1 \quad \text{or} \quad t \in (\beta t_0, t_0]$$

Thus the step length obtained by the backtracking line search satisfies $t \geq \min\{1, \beta t_0\}$.

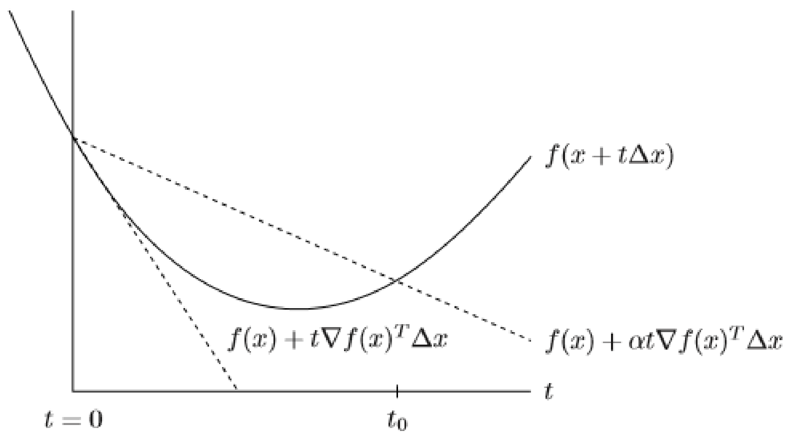

Figure 6.1 *Backtracking line search.* The backtracking condition is that f lies below the upper dashed line, *i.e.*, $0 \le t \le t_0$.

Note that we need to check $x + t\Delta x \in \mathbf{dom}\, f$ when using backtracking line search in practical implementation since f is infinite outside its domain by our convention. Then we start to check whether the inequality $f(x + t\Delta x) \le f(x) + \alpha t \nabla f(x)^T \Delta x$ holds.

The parameter α is typically chosen between 0.01 and 0.3, meaning that we accept a decrease in f between 1% and 30% of the prediction based on the linear extrapolation. The parameter β is often chosen to be between 0.1 (very crude search) and 0.8 (less crude search).

6.3 Gradient descent method

Choose $\Delta x = -\nabla f(x)$, then *gradient algorithm* or *gradient descent method* is

---

Algorithm 6.3 (*Gradient descent method*)**.**
given a starting point $x \in \mathbf{dom}\, f$.
repeat

1. $\Delta x := -\nabla f(x)$.
2. *Line search.* Choose step size t via exact or backtracking line search.
3. *Update.* $x := x + t\Delta x$

until stopping criterion is satisfied.

The criterion is usually $\|\nabla f(x)\|_2 \le \eta$ where $\eta > 0$ is small, and the condition is checked after step 1 in most implementations.

6.3.1 Convergence analysis

Add in the future!

6.3.2 Examples

A quadratic problem in $\mathbb{R}^2$

Consider minimizing $f(x) = \dfrac{1}{2}(x_1^2 + \gamma x_2^2)$ where $\gamma > 0$.

A nonquadratic problem in $\mathbb{R}^2$

A problem in $\mathbb{R}^{100}$

Gradient method and condition number

Conclusions

6.4 Steepest descent method

Add in the future!

6.4.1 Steepest descent for Euclidean and quadratic norms

Steepest descent for Euclidean norm

Steepest descent for quadratic norm

Interpretation via change of coordinates

6.4.2 Steepest descent for ℓ_1-norm

6.4.3 Convergence analysis

6.4.4 Discussion and examples

Choice of norm for steepest descent

Examples

6.5 Newton's method

6.5.1 The Newton step

The *Newton step* is

$$\Delta x_{nt} = -\nabla^2 f(x)^{-1} \nabla f(x)$$

Positive definiteness of $\nabla^2 f(x)$ implies that

$$\nabla f(x)^T \Delta x_{nt} = -\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) < 0$$

unless $\nabla f(x) = 0$, so the Newton step is a descent direction.
The Newton step can be interpreted and motivated in several ways.

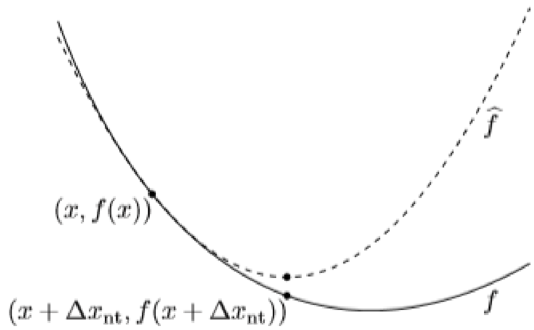

Figure 6.2 The function f and its second-order approximation $\hat{f}$.

Minimizer of second-order approximation

The second-order Taylor approximation $\hat{f}$ of f at x is

$$\hat{f}(x + v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v \tag{6.8}$$

which is a convex quadratic function of v, and is minimized when $v = \Delta x_{nt}$. Thus, Δx_{nt} must be added to the point x to minimize the second-order approximation of f at x as illustrated in figure 6.2. The insights of Newton step are

- if f is quadratic, then $x + \Delta x_{nt}$ is the exact minimizer of f
- if f is nearly quadratic, then $x + \Delta x_{nt}$ is a good estimate of the minimizer of f
- if x is near x^*, then $x + \Delta x_{nt}$ is a good estimate of x^*

Steepest descent direction in Hessian norm

Add in the future!

Solution of linearized optimality condition

Add in the future!

Affine invariance of the Newton step

The Newton step is independent of linear (or affine) changes of coordinates. Let $T \in \mathbb{R}^{n \times n}$ is nonsingular and $\bar{f}(y) = f(Ty)$, we have

$$\nabla \bar{f}(y) = T^T \nabla f(x), \quad \nabla \bar{f}(y) = T^T \nabla^2 f(x) T$$

where $x = Ty$. The Newton step for $\bar{f}$ at y is

$$\begin{aligned} \Delta y_{nt} &= -(T^T \nabla^2 f(x) T)^{-1} (T^T \nabla f(x)) \\ &= -T^{-1} \nabla^2 f(x)^{-1} \nabla f(x) \\ &= T^{-1} \Delta x_{nt} \end{aligned}$$

where Δx_{nt} is the Newton step for f at x, therefore $x + \Delta x_{nt} = T(y + \Delta y_{nt})$.

The Newton decrement

The *Newton decrement* is

$$\lambda(x) = (\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x))^{1/2}$$

The relation between f and the Newton decrement is

$$f(x) - \inf_y \hat{f}(y) = f(x) - \hat{f}(x + \Delta x_{nt}) = \frac{1}{2}\lambda(x)^2$$

where $\hat{f}$ is the second-order approximation of f at x.

- $\lambda^2/2$ is an estimate of $f(x) - p^*$, based on the quadratic approximation of f at x
- The Newton decrements can also expressed as

  $$\lambda(x) = (\Delta x_{nt}^T \nabla^2 f(x) \Delta x_{nt})^{1/2} = \|\Delta x_{nt}\|_{\nabla^2 f(x)} \tag{6.9}$$

  this shows that λ is the quadratic norm of the Newton step defined by Hessian $\nabla^2 f(x)$.
- Using in backtracking line search, we have

  $$-\lambda(x)^2 = \nabla f(x)^T \Delta x_{nt} = \frac{d}{dt} f(x + \Delta x_{nt} t)|_{t=0} \tag{6.10}$$

  this constant can be interpreted as the directional derivative of f at x in the direction of the Newton step.
- Newton decrement is, like the Newton step, affine invariant, *i.e.*, the Newton decrement of $\bar{f}(y) = f(Ty)$ at y is the same as the Newton decrement of f at $x = Ty$ provided T is nonsinngular.

6.5.2 Newton's method

The *damped/guarded* Newton method is outlined below, and the *pure* Newton method is to fix step size $t = 1$.

Algorithm 6.4 (*Newton's method*).
given a starting point $x \in \mathbf{dom}\, f$, tolerance $\epsilon > 0$.
repeat

1. *Compute the Newton step and decrement.*

   $$\Delta x_{\mathrm{nt}} := -\nabla^2 f(x)^{-1} \nabla f(x), \quad \lambda^2 := \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)$$

2. *Stopping criterion.* **quit** if $\lambda^2/2 \le \epsilon$
3. *Line search.* Use backtracking line search.
4. *Update.* $x := x + t\Delta x_{\mathrm{nt}}$

The minor difference from descent method is the position of the stopping criterion.

6.5.3 Convergence analysis

6.5.4 Examples

6.6 Self-concordance

6.7 Implementation

Chapter 7

Equality constrained minimization

7.1 Equality constrained minimization problems

Consider a convex optimization problem with equality constraints

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array} \tag{7.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is convex and twice continuously differentiable, and $A \in \mathbb{R}^{p \times n}$ with $\mathbf{rank}\, A = p < n$. This means that there are fewer equality constraints than variables, and the equality constraints are independent.

Recall (from §4.2.3 or §5.5.3) that a point $x^* \in \mathbf{dom}\, f$ is optimal for (7.1) if and only if there is a $\nu^* \in \mathbb{R}^p$ such that

$$Ax^* = b, \quad \nabla f(x^*) + A^T \nu^* = 0 \tag{7.2}$$

- (7.1) is equivalent to (7.2) which is a set of $n + p$ equations in the $n + p$ variables
- the first set are called the *primal feasibility equations* which is linear
- the second set are called the *dual feasibility equations* which is in general nonlinear

A special case to solve (7.2) is described in §7.1.1, and the elimination and dual methods are described in §7.1.2 and §7.1.3.

7.1.1 Equality constrained convex quadratic minimization

Consider

$$\begin{array}{ll} \text{minimize} & f(x) = (1/2)x^T P x + q^T x + r \\ \text{subject to} & Ax = b \end{array} \tag{7.3}$$

where $P \in \mathbb{S}_+^n$ and $A \in \mathbb{R}^{p \times n}$. The KKT conditions are

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix} \tag{7.4}$$

- this set is called the *KKT system* for the problem (7.3)
- the coefficient matrix is called the *KKT matrix*
- if KKT matrix is nonsingular, then there is a unique optimal (x^*, ν^*)
- if KKT matrix is singular and KKT system is solvable, then any solution yields an optimal (x^*, ν^*)
- if KKT matrix is singular and KKT system is unsolvable, then the problem is unbounded below or infeasible

In the last case there exist $v \in \mathbb{R}^n$ and $w \in \mathbb{R}^p$ such that

$$Pv + A^T w = 0, \quad Av = 0, \quad -q^T v + b^T w > 0$$

Let $\hat{x}$ be any feasible point. The point $x = \hat{x} + tv$ is feasible for all t and

$$\begin{aligned} f(\hat{x} + tv) &= f(\hat{x}) + t(v^T P\hat{x} + q^T v) + (1/2)t^2 v^T P v \\ &= f(\hat{x}) + t(-\hat{x}^T A^T w + q^T v) - (1/2)t^2 w^T A v \\ &= f(\hat{x}) + t(-b^T w + q^T v) \end{aligned}$$

which decreases without bound as $t \to \infty$.

Nonsingularity of the KKT matrix

There are several conditions equivalent to nonsingularity of the KKT matrix

- $\mathcal{N}(P) \cap \mathcal{N}(A) = \{0\}$, *i.e.*, P and A have no nontrivial common nullspace
- $Ax = 0,\ x \neq 0 \Rightarrow x^T P x > 0$, *i.e.*, P is positive definite on the nullspace of A
- $F^T P F \succ 0$ where $F \in \mathbb{R}^{n \times (n-p)}$ is a matrix for which $\mathcal{R}(F) = \mathcal{N}(A)$

An important special case: if $P \succ 0$, then the KKT matrix must be nonsingular.

7.1.2 Eliminating equality constraints

7.1.3 Solving equality constrained problems via the dual

7.2 Newton's method with equality constraints

The initial point must be feasible (*i.e.*, satisfy $x \in \mathbf{dom}\, f$ and $Ax = b$), and the definition of Newton step is modified to take the equality constraints into account. In particular, the Newton step Δx_{nt} is a feasible direction, *i.e.*, $A\Delta x_{nt} = 0$.

7.2.1 The Newton step

Definition via second-order approximation

We replace the objective with its second-order Taylor approximation near x

$$\begin{array}{ll} \text{minimize} & \hat{f}(x+v) = f(x) + \nabla f(x)^T v + (1/2) v^T \nabla^2 f(x) v \\ \text{subject to} & A(x+v) = b \end{array} \tag{7.5}$$

with variable v. This is a convex quadratic minimization problem with equality constraints. We define Δx_{nt} is the solution, and assume the KKT matrix is nonsingular. From §7.1.1 we know the Newton step Δx_{nt} is characterized by

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{nt} \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix} \tag{7.6}$$

where w is the associated optimal dual variable.

- when f is exactly quadratic, the Newton update $x + \Delta x_{nt}$ exactly solves the equality constrained minimization problem
- in the unconstrained case, when f is nearly quadratic, $x + \Delta x_{nt}$ should be a good estimate of x^*

Solution of linearized optimality conditions

The Newton step Δx_{nt} and the vector w can be interpreted as the solutions of a linearized approximation of the optimality conditions

$$Ax^* = b, \quad \nabla f(x^*) + A^T \nu^* = 0$$

Let $x^* = x + \Delta x_{nt}$, $\nu^* = w$, replace ∇f by its linearized approximation near x, and using $Ax = b$. We can get the equations (7.6).

The Newton decrement

The Newton decrement is

$$\lambda(x) = (\Delta x_{nt}^T \nabla^2 f(x) \Delta x_{nt})^{1/2} = \|\Delta x_{nt}\|_{\nabla^2 f(x)} \tag{7.7}$$

Let $\hat{f}(x+v)$ be the second-order Taylor approximation of f at x as shown in (7.5), we have

$$f(x) - \inf\{\hat{f}(x+v) \mid A(x+v) = b\} = \lambda(x)^2/2 \tag{7.8}$$

This means $\lambda(x)$ serves as the basis of a good stopping criterion.

The Newton decrement comes up in the line search as well, since the directional derivative of f in the direction Δx_{nt} is

$$\left.\frac{d}{dt} f(x + t\Delta x_{nt})\right|_{t=0} = \nabla f(x)^T \Delta x_{nt} = -\lambda(x)^2 \tag{7.9}$$

as in the unconstrained one.

Feasible descent direction

Affine invariance

7.2.2 Newton's method with equality constraints

The algorithm is the same as unconstrained problem

Algorithm 7.1 (*Newton's method for equality constrained minimization*).
given a starting point $x \in \mathbf{dom}\, f$ with $Ax = b$, tolerance $\epsilon > 0$.
repeat

1. *Compute the Newton step and decrement.*

$$\Delta x_{\text{nt}} := -\nabla^2 f(x)^{-1} \nabla f(x), \quad \lambda^2 := \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)$$

2. *Stopping criterion.* **quit** if $\lambda^2/2 \leq \epsilon$

3. *Line search.* Use backtracking line search.

4. *Update.* $x := x + t\Delta x_{\text{nt}}$

The method is called a *feasible descent method*, it requires that the KKT matrix be invertible at each x.

7.2.3 Newtons method and elimination

7.2.4 Convergence analysis

7.3 Infeasible start Newton method

7.3.1 Newton step at infeasible points

Again start with the optimality conditions for the equality constrained problem

$$Ax^* = b, \quad \nabla f(x^*) + A^T \nu^* = 0$$

Let $x \in \mathbf{dom}\, f$ denote the current point, which we do not assume to befeasible. The goal is to find Δx such that $x + \Delta x \approx x^*$. We substitute $x + \Delta x$ for x^* and w for ν^*, and use the first-order approximation.

$$\nabla f(x + \Delta x) \approx \nabla f(x) + \nabla^2 f(x) \Delta x$$

to obtain

$$A(x + \Delta x) = b, \quad \nabla f(x) + \nabla^2 f(x) \Delta x + A^T w = 0$$

Therefore, the set becomes

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ w \end{bmatrix} = -\begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix} \tag{7.10}$$

- $Ax - b$ is the residual vector for the linear equality constraints which vanishes when x is feasible, *i.e.*, (7.6)
- if x is feasible, then $\Delta x = \Delta x_{nt}$ with no confusion

Interpretation as primal-dual Newton step

We update both primal variable and dual variable in order to (approximately) satisfy the optimality conditions. We express optimality as $r(x^*, \nu^*) = 0$

$$
\begin{aligned}
r(x, \nu) &= (r_{dual}(x, \nu), r_{pri}(x, \nu)) \\
r_{dual}(x, \nu) &= \nabla f(x) + A^T \nu \\
r_{pri}(x, \nu) &= Ax - b
\end{aligned}
$$

where $r : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n \times \mathbb{R}^p$, r_{dual} is called *dual residual*, and r_{pri} is called *primal residual*. The first-order Taylor approximation of r near current estimate y is

$$r(y + z) \approx \hat{r}(y + z) = r(y) + Dr(y)z$$

where $Dr(y) \in \mathbb{R}^{(n+p)\times(n+p)}$ is the derivative of r. We define the primal-dual Newton step Δy_{pd} as the step z for which $\hat{r}(y + z)$ vanishes, *i.e.*,

$$Dr(y)\Delta y_{pd} = -r(y) \tag{7.11}$$

where $\Delta y_{pd} = (\Delta x_{pd}, \Delta \nu_{pd})$. Then express (7.11) as

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{pd} \\ \Delta \nu_{pd} \end{bmatrix} = - \begin{bmatrix} r_{dual} \\ r_{pri} \end{bmatrix} = - \begin{bmatrix} \nabla f(x) + A^T \nu \\ Ax - b \end{bmatrix} \tag{7.12}$$

i.e.,

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{pd} \\ \nu^+ \end{bmatrix} = - \begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix} \tag{7.13}$$

where $\nu^+ = \nu + \Delta \nu_{pd}$.

- (7.12) and (7.13) are the same as (7.10) with the relation

  $$\Delta x_{nt} = \Delta x_{pd}, \quad w = \nu^+ = \nu + \Delta \nu_{pd}$$

  The (infeasible) Newton step is the same as the primal part of the primal-dual step, and the associated dual vector w is the updated primal-dual variable ν^+.

- (7.12) shows that the Newton step and the associated dual step are obtained by solving a set of equations with residuals as the righthand side.
- (7.13) shows that how we originally defined the Newton step, gives the Newton step and the updated dual variable. The current value of the dual variable is not needed to compute the primal step, or the updated value of the dual variable.

Residual norm reduction property

The Newton direction at an infeasible point is not necessarily a descent direction

$$\begin{aligned}\frac{d}{dt}f(x+t\Delta x)\bigg|_{t=0} &= \nabla f(x)^T\Delta x \\ &= -\Delta x^T\nabla^2 f(x)\Delta x + (Ax-b)^T w\end{aligned}$$

which is not necessarily negative (unless x is feasible, *i.e.*, $Ax = b$). The primal-dual interpretation shows that the norm of the residual decreases in the Newton direction, *i.e.*,

$$\frac{d}{dt}\|r(y+t\Delta y_{pd})\|_2^2\bigg|_{t=0} = 2r(y)^T Dr(y)\Delta y_{pd} = -2r(y)^T r(y) < 0$$

Take the derivative of the square, we have

$$\frac{d}{dt}\|r(y+t\Delta y_{pd})\|_2\bigg|_{t=0} = -\|r(y)\|_2 \tag{7.14}$$

Therefore $\|r\|_2$ can be used to measure the progress of the infeasible start Newton method.

Full step feasibility property

The Newton step in (7.10) has the property

$$A(x+\Delta x_{nt}) = b \tag{7.15}$$

It follows that, if a step length of one is taken using the Newton step, the following iterates will be feasible. More generally, with a step length $t \in [0,1]$, the next iterate of equality residual r_{pri} is

$$r_{pri}^{+} = A(x+\Delta x_{nt}t) - b = (1-t)(Ax-b) = (1-t)r_{pri}$$

using (7.15). Therefore, we have

$$r^{(k)} = \left(\prod_{i=0}^{k-1}(1-t^{(i)})\right) r^{(0)}$$

where $r^{(i)} = Ax^{(i)} - b$ is the residual of $x^{(i)} \in \mathbf{dom}\, f$, and $t^{(i)} \in [0,1]$.

- the primal residual at each step is in the direction of the initial primal residual, and is scaled down at each step.
- once a full step ($t^{(i)} = 1$) is taken, all future iterates are primal feasible

7.3.2 Infeasible start Newton method

The extension of Newton's method, with $x^{(0)} \in \mathbf{dom}\, f$, $\Delta\nu_{nt} = \Delta\nu_{pd} = w - \nu$

Algorithm 7.2 (*Infeasible start Newton method*).
given a starting point $x \in \mathbf{dom}\, f$, ν, tolerance $\epsilon > 0$, $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$.
repeat

1. *Compute primal and dual Newton steps* Δx_{nt} *and* $\Delta \nu_{nt}$.
2. *Backtracking line search on* $\|r\|_2$.
   (a) $t := 1$
   (b) **while** $\|r(x + t\Delta x_{nt}, \nu + t\Delta \nu_{nt})\|_2 > (1 - \alpha t)\|r(x, \nu)\|_2,\ t := \beta t$
3. *Update.* $x := x + t\Delta x_{\rm nt},\ \nu := \nu + t\Delta \nu_{\rm nt}$

until $Ax = b$ and $\|r(x, \nu)\|_2 \le \epsilon$

The difference between this algorithm and the standard Newton method are

- the search direction include the extra correction terms (primal residual)
- the line search is carried out using the norm of the residual
- the algorithm terminates when primal feasibility has been achieved, and the norm of the (dual) residual is small

There are some properties here

- the line search must terminate in a finite number of steps since (7.14) shows that the exit condition is satisfied for small t
- once the step length is chosen to be one, the following search direction coincides with the one for the (feasible) Newton method
- another variation is to switch to the (feasible) Newton method once feasibility is achieved, *i.e.*, change line search and exit condition

7.3.3 Convergence analysis

7.3.4 Convex-concave games

7.3.5 Examples

7.4 Implementation

7.4.1 Elimination

7.4.2 Solving KKT systems

7.4.3 Examples

Chapter 8

Interior-point methods

8.1 Inequality constrained minimization problems

Consider a convex optimization problem with inequality constraints

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0,\ i = 1, \ldots, m \\ & Ax = b \end{array} \tag{8.1}$$

where $f_i : \mathbb{R}^n \to \mathbb{R},\ i = 0, \ldots, m$ are convex and twice continuously differentiable, and $A \in \mathbb{R}^{p \times n}$ with $\textbf{rank}\, A = p < n$. We assume that

- the problem is solvable, *i.e.*, an optimal x^* exists
- the problem is strictly feasible, *i.e.*, there exists $x \in \mathcal{D}$ that satisfies

$$Ax = b, \quad f_i(x) < 0,\ i = 1, \ldots, m$$

Then the Slater's constraint qualification holds, and there exists dual optimal $\lambda^* \in \mathbb{R}^m$ and $\nu^* \in \mathbb{R}^p$, together with x^*, satisfy the KKT conditions (§5.5.3)

$$\begin{aligned} f_i(x^*) &\leq 0,\ i = 1, \ldots, m \\ Ax^* &= b \\ \lambda^* &\succeq 0 \\ \lambda_i^* f_i(x^*) &= 0,\ i = 1, \ldots, m \\ \nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + A^T \nu^* &= 0 \end{aligned} \tag{8.2}$$

Interior-point methods solve the problem (8.1) (or the KKT conditions (8.2)) by applying Newton's method to a sequence of equality constrained problems, or to a sequence of modified versions of the KKT conditions.

Examples

Add in the future.

8.2 Logarithmic barrier function and central path

Rewrite (8.1) as an equality constrained problem

$$\begin{aligned} &\text{minimize} && f_0(x) + \sum_{i=1}^{m} I_-(f_i(x)) \\ &\text{subject to} && Ax = b \end{aligned} \tag{8.3}$$

where $I_- : \mathbb{R} \to \mathbb{R}$ is the indicator function for the nonpositive reals

$$I_-(u) = \begin{cases} 0, & u \le 0 \\ \infty, & u > 0 \end{cases}$$

The problem (8.3) has no inequality constraints, but the objective function is not differentiable (Newton's method cannot be applied).

8.2.1 Logarithmic barrier

Approximate the indicator function I_- by the function

$$\hat{I}_-(u) = -(1/t)\log(-u), \ \mathbf{dom}\, \hat{I}_- = -\mathbb{R}_{++}$$

where $t > 0$. The function $\hat{I}_-$ takes on ∞ for $u > 0$ by our convention, and then it is

- convex and nondecreasing
- differentiable and closed
- the approximation becomes more accurate as t increases
- figure 8.1 shows the function

The new approximation becomes

$$\begin{aligned} &\text{minimize} && f_0(x) + \sum_{i=1}^{m} -(1/t)\log(-f_i(x)) \\ &\text{subject to} && Ax = b \end{aligned} \tag{8.4}$$

where the objective function is convex. The *logarithmic barrier* function for problem (8.1) is

$$\phi(x) = -\sum_{i=1}^{m} \log(-f_i(x)) \tag{8.5}$$

where $\mathbf{dom}\, \phi = \{x \in \mathbb{R}^n \mid f_i(x) < 0, \ i = 1, \dots, m\}$.

- the quality of the approximation improves as t grows

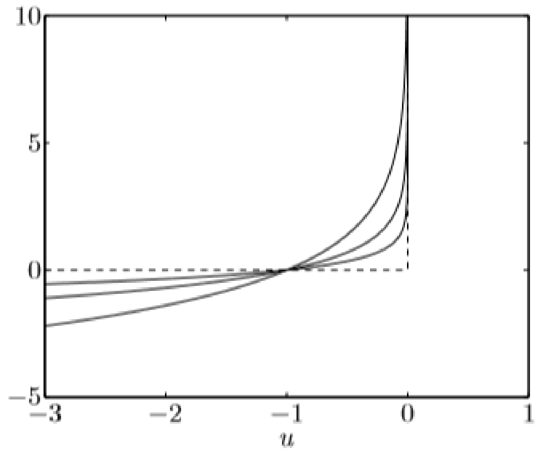

Figure 8.1 The dashed lines show the function $I\_(u)$, and the solid curves show the function $\hat{I}\_(u)$ for $t = 0.5, 1, 2$. The best approximation is $t = 2$.

- if t is large, the function $f_0 + (1/t)\phi$ is difficult to minimize by Newton's method
- the above problem can be circumvented by the algorithm discussed later

The gradient and Hessian of the logarithmic barrier function ϕ are (§A.4.2 and §A.4.4)

$$\nabla\phi(x) = \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla f_i(x)$$
$$\nabla^2\phi(x) = \sum_{i=1}^{m} \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T + \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla^2 f_i(x)$$

8.2.2 Central path

Consider the equivalent problem of (8.4)

$$\begin{array}{ll} \text{minimize} & t f_0(x) + \phi(x) \\ \text{subject to} & Ax = b \end{array} \tag{8.6}$$

We assume (8.6) can be solved via Newton's method, and it has a unique solution for each $t > 0$. Define $x^*(t)$ as the solution of (8.6), and the *central path* associated with (8.1) is defined as the set of points $x^*(t)$, $t > 0$, which we call the *central points*.

We have the property that $x^*(t)$ is strictly feasible, *i.e.*,

$$Ax^*(t) = b, \quad f_i(x^*(t)) < 0, \ i = 1, \dots, m$$

if and only if there exists a $\nu \in \mathbb{R}^p$ such that

$$\begin{aligned} 0 &= t\nabla f_0(x^*(t)) + \nabla\phi(x^*(t)) + A^T\hat{\nu} \\ &= t\nabla f_0(x^*(t)) + \sum_{i=1}^{m} \frac{1}{-f_i(x^*(t))} \nabla f_i(x^*(t)) + A^T\hat{\nu} \end{aligned} \tag{8.7}$$

which is called *centrality condition*.

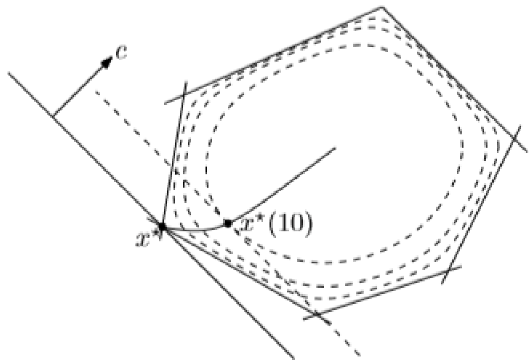

Figure 8.2 The dashed curves show three contour lines of ϕ. The line $c^T x = c^T x^*(10)$ is tangent to the contour line of ϕ through $x^*(10)$.

Example 8.1 (*Inequality form linear programming*)**.** Consider the inequality form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \end{array} \tag{8.8}$$

The logarithmic barrier function is

$$\phi(x) = \sum_{i=1}^{m} \log(b_i - a_i^T x), \ \mathbf{dom}\, \phi = \{x \mid Ax \prec b\}$$

where $a_1^T, \ldots, a_m^T$ are the rows of A. The gradien and Hessian are

$$\nabla\phi(x) = \sum_{i=1}^{m} \frac{1}{b_i - a_i^T x} a_i, \quad \nabla^2\phi(x) = \sum_{i=1}^{m} \frac{1}{(b_i - a_i^T x)^2} a_i a_i^T$$

or, more compactly,

$$\nabla\phi(x) = A^T d, \quad \nabla^2\phi(x) = A^T \,\mathbf{diag}(d)^2 A$$

where $d \in \mathbb{R}^m$, $d_i = 1/(b_i - a_i^T x)$. Since x is strictly feasible, we have $d \succ 0$, so Hessian is nonsingular if and only if A has rank n.
The centrality condition (8.7) is

$$tc + \sum_{i=1}^{m} \frac{1}{b_i - a_i^T x} a_i = tc + A^T d = 0 \tag{8.9}$$

We have $\nabla\phi(x^*(t)) = -tc$, *i.e.*, at a point $x^*(t)$ on the central path the gradient must be parallel to $-c$. In other words, the hyperplane $c^T x = c^T x^*(t)$ is tangent to the level set of ϕ through $x^*(t)$. Figure 8.2 shows an example with $m = 6$ and $n = 2$.

Dual points from central path

From (8.7) we know that every central point yields a dual feasible point, and hence a lower bound of p^*. Define

$$\lambda_i^*(t) = -\frac{1}{t f_i(x^*(t))}, \; i = 1, \dots, m, \quad \nu^*(t) = \frac{\hat{\nu}}{t} \tag{8.10}$$

we claim that the pair $\lambda^*(t)$, $\nu^*(t)$ is dual feasible.

Proof. We have $\lambda^*(t) \succ 0$ since $f_i(x^*(t)) < 0$. The optimality conditions (8.7) becomes

$$\nabla f_0(x^*(t)) + \sum_{i=1}^m \lambda_i^*(t) \nabla f_i(x^*(t)) + A^T \nu^*(t) = 0$$

we see that $x^*(t)$ minimizes the Lagrangian

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \nu^T (Ax - b)$$

for $\lambda = \lambda^*(t)$ and $\nu = \nu^*(t)$, which means that $\lambda^*(t)$, $\nu^*(t)$ is a dual feasible pair. □

The Lagrangian dual function

$$\begin{aligned} g(\lambda^*(t), \nu^*(t)) &= f_0(x^*(t)) + \sum_{i=1}^m \lambda_i^*(t) f_i(x^*(t)) + \nu^*(t)^T (Ax^*(t) - b) \\ &= f_0(x^*(t)) - m/t \end{aligned}$$

The duality gap associated with $x^*(t)$, $\lambda^*(t)$ and $\nu^*(t)$ is m/t. An important result states

$$f_0(x^*(t)) - p^* \leq m/t$$

i.e., $x^*(t)$ is no more than m/t-suboptimal. This confirms the intuitive idea that $x^*(t)$ converges to an optimal point as $t \to \infty$.

Interpretation via KKT conditions

Force field interpretation

8.3 The barrier method

We know that $x^*(t)$ is m/t-suboptimal, and that a certificate of this accuracy is provided by $\lambda^*(t)$ and $\nu^*(t)$. We take $t = m/\epsilon$ and solve the problem using Newton's method.

$$\begin{array}{ll} \text{minimize} & (m/\epsilon) f_0(x) + \phi(x) \\ \text{subject to} & Ax = b \end{array}$$

This method could be called the *unconstrained minimization method*.

8.3.1 The barrier method

A simple extension of the unconstrained minimization method does work well. It is based on solving a sequence of unconstrained (or linearly constrained) minimization problems. This method was first proposed by Fiacco and McCormick in the 1960s, and was called *sequential unconstrained minimization technique* (SUMT). Today the method is called the *barrier method* or *path-following method*.

Algorithm 8.1 (*Barrier method*).
given strictly feasible x, $t := t^{(0)}$, $\mu > 1$, tolerance $\epsilon > 0$
repeat

1. *Centering step.* Get $x^*(t)$ by minimizing $tf_0 + \phi$, subject to $Ax = b$, starting at x.
2. *Update.* $x := x^*(t)$.
3. *Stopping criterion.* **quit** if $m/t < \epsilon$.
4. *Increase t.* $t := \mu t$.

8.3.2 Examples

8.3.3 Convergence analysis

8.3.4 Newton step for modified KKT equations

8.4 Feasibility and phase I methods

8.4.1 Basic phase I method

8.4.2 Phase I via infeasible start Newton method

8.4.3 Examples

8.5 Complexity analysis via self-concordance

8.5.1 Self-concordance assumption

8.5.2 Newton iterations per centering step

8.5.3 Total number of Newton iterations

8.5.4 Feasibility problems

8.5.5 Combined phase I/phase II complexity

8.5.6 Summary

8.6 Problems with generalized inequalities

8.6.1 Logarithmic barrier and central path

8.6.2 Barrier method

8.6.3 Examples

8.6.4 Complexity analysis via self-concordance

8.7 Primal-dual interior-point methods

8.7.1 Primal-dual search direction

8.7.2 The surrogate duality gap

8.7.3 Primal-dual interior-point method

8.7.4 Examples

8.8 Implementation

8.8.1 Standard form linear programming

8.8.2 ℓ_1-norm approximation

Appendices

Appendix A

Mathematical background

A.1 Norms

A.1.1 Inner product, Euclidean norm, and angle

Define *standard inner product* on $\mathbb{R}^n$ by

$$\langle x, y\rangle = x^T y = \sum_{i=1}^{n} x_i y_i$$

Define *Euclidean norm*, ℓ_2-norm on $x \in \mathbb{R}^n$ by

$$\|x\|_2 = (x^T x)^{1/2} = (x_1^2 + \cdots + x_n^2)^{1/2} \tag{A.1}$$

The *Cauchy-Schwartz inequality*

$$|x^T y| \leq \|x\|_2 \|y\|_2$$

Define *standard inner product* on $\mathbb{R}^{m\times n}$

$$\langle X, Y\rangle = \mathbf{tr}(X^T Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij} Y_{ij}$$

A.1.2 Norms, distance, and unit ball

A function $f : \mathbb{R}^n \to \mathbb{R}$ with $\mathbf{dom}\, f = \mathbb{R}^n$ is a norm if

- f is nonnegative: $f(x) \geq 0 \quad \forall x \in R^n$
- f is definite: $x = 0 \Rightarrow f(x) = 0$
- f is homogeneous: $f(tx) = |t| f(x), \quad \forall x \in \mathbb{R}^n$ and $t \in \mathbb{R}$
- f satisfies the triangle inequality: $f(x + y) \leq f(x) + f(y), \quad \forall x, y \in \mathbb{R}^n$

A.1.3 Examples

Define *Chebyshev norm*, ℓ_∞-norm by

$$\|x\|_\infty = max\{|x_1|, \ldots .|x_n|\}$$

A.1.4 Equivalence of norms

A.1.5 Operator norms

Suppose $\|\cdot\|_a$ and $\|\cdot\|_b$ are norms on $\mathbb{R}^m$ and $\mathbb{R}^n$. Define *operator norm* of $X \in \mathbb{R}^{m\times n}$, induced by $\|\cdot\|_a$ and $\|\cdot\|_b$, as

$$\|X\|_{a,b} = \sup\{\|Xu\|_a \mid \|u\|_b \leq 1\}.$$

The *spectral norm*/ℓ_2*-norm* of X is its *maximum singular value*

$$\|X\|_2 = \sigma_{\max}(X) = (\lambda_{\max}(X^TX))^{1/2}.$$

The ℓ_∞*-norm* of X is the *max-row-sum norm*

$$\|X\|_\infty = \sup\{\|Xu\|_\infty \mid \|u\|_\infty \leq 1\} = \max_{i=1,\ldots,m} \sum_{j=1}^{n} |X_{ij}|.$$

The ℓ_1*-norm* of X is the *max-column-sum norm*

$$\|X\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^{m} |X_{ij}|.$$

A.1.6 Dual norm

Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$, then the associated *dual norm* is

$$\|z\|_* = \sup\{z^Tx \mid \|x\| \leq 1\}$$

A.2 Analysis

A.2.1 Open and closed sets

A.2.2 Supremum and infimum

A.3 Functions

A.3.1 Function notation

A.3.2 Continuity

A.3.3 Closed functions

A.4 Derivatives

A.4.1 Derivative and gradient

A.4.2 Chain rule

A.4.3 Second derivative

A.4.4 Chain rule for second derivative

A.5 Linear algebra

A.5.1 Range and nullspace

A.5.2 Symmetric eigenvalue decomposition

A.5.3 Generalized eigenvalue decomposition

A.5.4 Singular value decomposition

A.5.5 Schur complement