Study Notes of Numerical Optimization

Pei Zhong

Update on May 25, 2024

Contents

1 Preface **2**

2 Knowledge review **3**

2.1 Taylor Expansion 3

2.2 Row space, column space and nullspace of matrix 4

2.3 Reference 4

I Convexity **5**

3 Convex Sets and Convex Functions **6**

3.1 Convex Sets and Convex Functions 6

3.2 Local Lipschitz Continuity of Convex Functions 7

3.3 Tests for Convexity 7

3.4 Reference 7

4 Affine and Convex Hulls **8**

4.1 Hyperplanes 8

4.2 Affine Set 8

4.3 Affine Hull 8

4.4 Convex Hull 8

4.5 Reference 8

5 The Directional Derivative and The Subdifferential **9**

5.1 Relative Interior 9

5.2 Convex Separation 9

5.3 Conjugate Functions 9

5.4 Subgradient of Convex Function 9

5.5 Directional Derivative 9

5.6 Reference 10

II Unconstrained Optimization 11

6 Basic Notions in Optimization 12
6.1 Types of optimization problems . . . 12
6.2 Constraints and feasible regions . . . 12
6.3 Types of optimal solutions . . . 12
6.4 Existence of solutions of optimization problems . . . 13
6.5 Reference . . . 13

7 Optimality conditions for unconstrained problems 14
7.1 Introduction . . . 14
7.2 Optimality conditions: the necessary and the sufficient . . . 14
7.3 Reference . . . 16

8 Line Search 17

9 Gradient Descent Method 18

10 Newton Method 19

11 Quasi Newton Method 20

12 Conjugate Gradient Method 21

13 Trust Region Method 22

14 Algorithms for nonlinear least squares problems 23

15 Sparse Optimization 24
15.1 Sparse optimization: motivation . . . 24
15.2 Sparse problem formulations . . . 24
15.3 Application: LASSO problem . . . 25
15.4 Application: matrix completion problem . . . 25
15.5 Reference . . . 26

16 Optimality Conditions for Unconstrained Nonsmooth Problems 27
16.1 First-order necessary and sufficient conditions for convex optimization problems . 27
16.2 First-order necessary conditions for composite optimization problems . . . 27

17 Proximal Algorithms 28

III Duality and Constrained Optimization 29

18 Optimality Conditions for Constrained Problems 30

18.1 Geometric constrained minimization 30

18.1.1 Examples 31

18.2 Reference 31

19 Normal and Tangent Cones 32

19.1 Reference 32

Chapter 1

Preface

- Convex Optimization: Fall 2015
- Convex Optimization: Theory, Algorithms, and Applications
- duality and admm
- Optimization for Modern Data Analysis

Chapter 2

Knowledge review

2.1 Taylor Expansion

Theorem 2.1: taylor formula for Peano-type remainder of a binary function

$f(x,y)$ is a function which is continuously derivable up to order n in some neighborhood of point (x_0,y_0). Then as $\rho=\sqrt{h^2+k^2}\to 0$,

$$\begin{aligned} f(x_0+h,y_0+k) &= f(x_0,y_0)+(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y})f(x_0,y_0)+\frac{1}{2!}(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y})^2 f(x_0,y_0) \\ &\quad +\cdots+\frac{1}{n!}(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y})^n f(x_0,y_0)+o(\rho^n) \end{aligned} \tag{2.1}$$

holds. (2.1) is called the taylor formula of order n with Peano-type remainder of f at (x_0,y_0).

Proposition 2.1

Let $f:X\to\mathbb{R}^n$, where $X\subset\mathbb{R}^n$ is open. f is differentiable at $\overline{x}$, then there exists a vector $\nabla f(\overline{x})$ and a function $\alpha(\overline{x},y):\mathbb{R}^n\to\mathbb{R}$ satisfying $\lim_{y\to 0}\alpha(\overline{x},y)=0$, such that for each $x\in X$

$$f(x)=f(\overline{x})+\nabla f(\overline{x})^T(x-\overline{x})+||x-\overline{x}||\alpha(\overline{x},x-\overline{x}). \tag{2.2}$$

Proof. Extending Theorem 2.1 to n dimensions, we can get

$$f(x)=f(\overline{x})+(\frac{\partial f(\overline{x})}{\partial x_1},\frac{\partial f(\overline{x})}{\partial x_2},\ldots,\frac{\partial f(\overline{x})}{\partial x_n})(x-\overline{x})+o(||x-\overline{x}||).$$

Since $o(||x-\overline{x}||)=||x-\overline{x}||o(1)(x\to\overline{x})$, let $\alpha(\overline{x},y)=o(1)(y=x-\overline{x}\to 0)$, $\lim_{y\to 0}\alpha(\overline{x},y)=0$. And let $\nabla f(x)^T=(\frac{\partial f(\overline{x})}{\partial x_1},\frac{\partial f(\overline{x})}{\partial x_2},\ldots,\frac{\partial f(\overline{x})}{\partial x_n})$, then (2.3) holds. □

Proposition 2.2

Let $f : X \to \mathbb{R}^n$, where $X \subset \mathbb{R}^n$ is open. f is twice differentiable at $\overline{x}$, then there exists a vector $\nabla f(\overline{x})$, an $n \times n$ symmetric matrix $H(\overline{x})$ and a function $\alpha(\overline{x}, y) : \mathbb{R}^n \to \mathbb{R}$ satisfying $\lim_{y \to 0} \alpha(\overline{x}, y) = 0$, such that for each $x \in X$

$$f(x) = f(\overline{x}) + \nabla f(\overline{x})^T(x - \overline{x}) + \frac{1}{2}(x - \overline{x})^T H(\overline{x})(x - \overline{x}) + ||x - \overline{x}||^2 \alpha(\overline{x}, x - \overline{x}). \quad (2.3)$$

2.2 Row space, column space and nullspace of matrix

2.3 Reference

- mathematical analysis SCNU version
- IOE 511/Math 652: Continuous Optimization Methods ch2

Part I

Convexity

Chapter 3

Convex Sets and Convex Functions

This chapter mainly introduces
(1) The definition of convex set and convex function
(2) The continuous property of convex function
(3) Methods for testing convex functions

3.1 Convex Sets and Convex Functions

Definition 3.1: Convex Sets and Functions: The Epi-graphical Perspective

(1) A set $C \subset \mathbb{R}^n$ is said to be a convex set if for every $x, y \in C$ and $\lambda \in [0,1]$, one has

$$(1-\lambda)x + \lambda y \in C. \tag{3.1}$$

(2) Let C be a convex subset of $\mathbb{R}^n$. (2) Given an extended real-valued function $f : \mathbb{R}^n \to \mathbb{R}_e := \mathbb{R} \cup \pm\infty$, the epi-graph and domain of f are given by

$$\text{epi}(f) := \{(x,r) \in \mathbb{R}^n \times \mathbb{R} | f(x) \leqslant r\} \tag{3.2}$$
$$\text{dom}(f) := \{x \in \mathbb{R}^n | f(x) \leqslant +\infty\}, \tag{3.3}$$

respectively.
(3) The function $f : \mathbb{R}^n \to \mathbb{R}_e$ is said to be convex if epi(f) is convex.

Lemma 3.1: Convexity and Secant Lines

The function $f : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ is convex if and only if, for every $x, y \in \text{dom} f$ and $\lambda \in [0,1]$, we have

$$f((1-\lambda)x + \lambda y) \leqslant (1-\lambda)f(x) + \lambda f(y). \tag{3.4}$$

That is, the secant line connecting $(x, f(x))$ and $(y, f(y))$ lies above the graph of f.

Proof. ($\Rightarrow$): If f is convex, then epi(f) is convex. That is, for $(x, f(x)), (y, f(y)) \in$ epi(f) and

$\lambda \in [0, 1]$,

$$(1-\lambda)(x, f(x)) + \lambda(y, f(y)) \in \text{epi}(f).$$

That is,

$$f((1-\lambda)x + \lambda y) \leqslant (1-\lambda)f(x) + \lambda f(y).$$

($\Leftarrow$): Let $(x, r_1), (y, r_2) \in \text{epi}(f)$ and $\lambda \in [0, 1]$. If (3.4) holds, then

$$f((1-\lambda)x + \lambda y) \leqslant (1-\lambda)f(x) + \lambda f(x) \leqslant (1-\lambda)r_1 + \lambda r_2.$$

Therefore, $((1-\lambda)x + \lambda y, (1-\lambda)r_1 + \lambda r_2) \in \text{epi}(f)$. Then $(1-\lambda)(x, r_1) + \lambda(y, r_2) \in \text{epi}(f)$. So epi$(f)$ is convex and f is convex. □

3.2 Local Lipschitz Continuity of Convex Functions

Recall that a function $F : \mathbb{R}^n \to \mathbb{R}^m$ is said to be Lipschitz continuous on a set $S \subset \mathbb{R}^n$ if there is a constant $L > 0$ such that

$$||F(x) - F(y)|| \leqslant L||x - y|| \quad \forall x, y \in S.$$

The function F is said to be locally Lipschitz on an open set $V \subset \mathbb{R}^n$ if for every $\overline{x} \in V$ there is an $\epsilon > 0$ and $L > 0$ such that

$$||F(x) - F(y)|| \leqslant L||x - y|| \quad \forall x, y \in \overline{x} + \epsilon\mathbb{B} \subset V,$$

where $\mathbb{B} := \{x : ||x|| \leqslant 1\}$. In this section we establish the remarkable fact that a convex function is locally Lipschitz continuous on the interior of its domain. It is possible to generalize this results to convex functions whose domains have no interior. But this requires an understanding of the relative topology of convex sets. We begin by establishing the local boundedness of a convex function on the interior of its domain.

3.3 Tests for Convexity

3.4 Reference

- convex analysis cuhk
- Convex Analysis and Optimization, Lecture 2
- Convex sets and convex functions: the fundamentals
- nonlinear optimization 5.1
- nonlinear optimization 5.3
- nonlinear optimization 5.4

Chapter 4

Affine and Convex Hulls

This chapter mainly introduces the definition of hyperplanes, affine set, affine and convex hull, which set the stage for the separation theorem and the existence theorem of subdifferentials in the next chapter.

4.1 Hyperplanes

4.2 Affine Set

4.3 Affine Hull

4.4 Convex Hull

4.5 Reference

- Hyperplanes
- Affine set
- Convex set

Chapter 5

The Directional Derivative and The Subdifferential

This chapter mainly introduces the separation theorem, the existence theorem of subdifferentials and the relation between directional derivative and the subdifferential.

5.1 Relative Interior

5.2 Convex Separation

5.3 Conjugate Functions

5.4 Subgradient of Convex Function

5.5 Directional Derivative

Lemma 5.1: Sublinearity of $f'(x;d)$

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex.
(1) Given $x \in \mathrm{dom} f$ and $d \in \mathbb{R}^n$ the difference quotient

$$\frac{f(x+td)-f(x)}{t}$$

is a non-decreasing function of t on $(0, +\infty)$.
(2) For every $x \in \mathrm{dom} f$ and $d \in \mathbb{R}^n$ the directional derivative $f'(x;d)$ always exists and is given by

$$f'(x;d) := \inf_{t>0} \frac{f(x+td)-f(x)}{t}.$$

(3) For every $x \in \text{dom} f$, the function $f'(x;\cdot)$ is sublinear, i.e. $f'(x;\cdot)$ is positively homogeneous,

$$f'(x;\alpha d) = \alpha f'(x;d) \quad \forall d \in \mathbb{R}^n, 0 \leqslant \alpha,$$

and subadditive,

$$f'(x;u+v) \leqslant f'(x;u) + f'(x;v).$$

In particular, for all $x \in \text{dom} f$, $f'(x;d)$ is a convex function of d.

5.6 Reference

- convex analysis cuhk
- Separating and Supporting Hyperplanes
- subgradient
- The derivatives and the subdifferential
- nonlinear optimization 5.2

Part II

Unconstrained Optimization

Chapter 6

Basic Notions in Optimization

6.1 Types of optimization problems

Unconstrained Optimization problem:

$$(\text{P}) \qquad \min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x})$$

where $f(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$.

6.2 Constraints and feasible regions

6.3 Types of optimal solutions

Consider a general optimization problem

$$(\text{P}) \qquad \min_{x \in S} f(x). \tag{6.1}$$

Recall: an ϵ-neighborhood of $\overline{x}$, or a ball centered at $\overline{x}$ with radius ϵ is the set: $B(\overline{x}, \epsilon) := \{x : ||x - \overline{x}|| \leqslant \epsilon\}$. We have the following definitions of local/global, strict/non-strict minimizers.

Definition 6.1

In the optimization problem (P),

- $x \in S$ is a global minimizer of (P) if $f(x) \leqslant f(y)$ for all $y \in S$.
- $x \in S$ is a strict global minimizer of (P) if $f(x) < f(y)$ for all $y \in S$, $y \neq x$.
- $x \in S$ is a local minimizer of (P) if there exists $\epsilon > 0$ such that $f(x) \leqslant f(y)$ for all $y \in B(x, \epsilon) \cap S$.
- $x \in S$ is a strict local minimizer of (P) if there exists $\epsilon > 0$ such that $f(x) < f(y)$ for all $y \in B(x, \epsilon) \cap S$, $y \neq x$.

6.4 Existence of solutions of optimization problems

6.5 Reference

- IOE 511/Math 652: Continuous Optimization Methods ch3

Chapter 7

Optimality conditions for unconstrained problems

7.1 Introduction

The definitions of global and local solutions of optimization problems are intuitive, but usually impossible to check directly. Hence, we will derive easily verifiable conditions that are either necessary for a point to be a local minimizer (thus helping us to identify candidates for minimizers), or sufficient (thus allowing us to confirm that the point being considered is a local minimizer), or, sometimes, both.

7.2 Optimality conditions: the necessary and the sufficient

$$(\text{P}) \qquad \min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x})$$

where $f(x) : \mathbb{R}^n \to \mathbb{R}$.

Necessary condition for local optimality: "if $\overline{x}$ is a local minimizer of (P), then $\overline{x}$ must satisfy…" Such conditions help us identify all candidates for local optimizers.

Definition 7.1: Directional Derivative

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $x, d \in \mathbb{R}$. The directional derivative of f at x in the direction d defined as

$$f'(x; d) = \lim_{t \downarrow 0} \frac{f(x + td) - f(x)}{t}. \tag{7.1}$$

It is important to observe that this is a one sided derivative since $t \downarrow 0$.

Lemma 7.1

Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $\overline{x} \in \mathbb{R}^n$ be a local solution to the problem $\min_{x\in\mathbb{R}^n} f(x)$. Then

$$f'(x; d) \geqslant 0 \tag{7.2}$$

for every direction $d \in \mathbb{R}^n$ for which $f'(x; d)$ exists.

Proof. Since $\overline{x}$ is a local solution, there exists $\epsilon > 0$ such that for all $x \in B(\overline{x}, \epsilon)$, we have

$$f(x) \geqslant f(\overline{x})$$

This means that for any sufficiently small $t > 0$, we have

$$\frac{f(\overline{x} + td) - f(\overline{x})}{t} \geqslant 0$$

Since $f'(x; d)$ exists, as $t \to 0^+$, we have

$$f'(\overline{x}; d) = \lim_{t\to 0^+} \frac{f(\overline{x} + td) - f(\overline{x})}{t} \geqslant 0.$$

□

Remark. If f is differentiable at x, we have $f'(x; d) = \nabla f(x)^T d$ for all $d \in R^n$. let's prove the statement: Since f is differentiable at x, we can use the gradient $\nabla f(x)$ to represent the linear approximation of f around x

$$f(x + h) = f(x) + \nabla f(x)^T h + o(\|h\|)$$

Then

$$\begin{aligned} f'(x; d) &= \lim_{t\to 0^+} \frac{f(x + td) - f(x)}{t} \\ &= \lim_{t\to 0^+} \frac{\nabla f(x)^T (td) + o(t\|d\|)}{t} \\ &= \nabla f(x)^T d + \lim_{t\to 0^+} \frac{\|d\| o(t)}{t} = \nabla f(x)^T d. \end{aligned}$$

Theorem 7.1: First-Order necessary condition for local optimality of Differentiable Functions

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at a point $\overline{x} \in \mathbb{R}^n$. If $\overline{x}$ is a local minimum of f, then $\nabla f(\overline{x}) = 0$.

Proof. Since f is differentiable at $\overline{x}$, we have $f'(\overline{x}; d) = \nabla f(x)^T d$ for all $d \in \mathbb{R}^n$. Then

$$0 \leqslant f'(\overline{x}; d) = \nabla f(\overline{x})^T d \text{for all} d \in \mathbb{R}^n.$$

Taking $f = -\nabla f(\overline{x})$ we get

$$0 \leqslant -\nabla f(\overline{x})^T \nabla f(\overline{x}) = -\|\nabla f(\overline{x})\|^2 \leqslant 0.$$

Therefore, $\nabla f(\overline{x}) = 0$. □

Theorem 7.2: Second-Order necessary condition for local optimality of Differentiable Functions

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable at a point $\overline{x} \in \mathbb{R}^n$. If $\overline{x}$ is a local minimum of f, then $\nabla f(\overline{x}) = 0$ and $\nabla^2 f(x)$ is positive semidefinite.

Proof. 1 □

Necessary conditions only allow us to come up with a list of candidate points for minimizers. Sufficient condition for local optimality: "if $\overline{x}$ satisfies …, then $\overline{x}$ is a local minimizer of (P).

7.3 Reference

- IOE 511/Math 652: Continuous Optimization Methods ch4
- nonlinear optimization 4.2,4.3

Chapter 8

Line Search

Chapter 9

Gradient Descent Method

Chapter 10

Newton Method

Chapter 11

Quasi Newton Method

Chapter 12

Conjugate Gradient Method

Chapter 13

Trust Region Method

Chapter 14

Algorithms for nonlinear least squares problems

Chapter 15

Sparse Optimization

15.1 Sparse optimization: motivation

Many applications need structured, approximate solutions of optimization formulations, rather than exact solutions.

- More Useful, More Credible
  - Structured solutions are easier to understand.
  - They correspond better to prior knowledge about the solution.
  - They may be easier to use and actuate.
  - Extract just the essential meaning from the data set, not the less important effects.
- Less Data Needed
  - Structured solution lies in lower-dimensional spaces $\Rightarrow$ need to gather / sample less data to capture it.
  - Choose good structure instead of "overfitting" to a particular sample.

15.2 Sparse problem formulations

The structural requirements have deep implications for how we formulate and solve these problems. A common type of desired structure is sparsity: We would like the approx solution $x \in \mathbb{R}^n$ to have few nonzero components. A sparse formulation of "$\min_x f(x)$" could be "Find an approximate minimizer $\overline{x} \in \mathbb{R}^n$ of f such that $||x||_0 \leqslant k$, where $||x||_0$ denotes cardinality: the number of nonzeros in x. But the zero-norm is a nonconvex discontinuous function, problems are hard to be solved. Use of $||x||_1$ has long been known to promote sparsity in x. Also, it can solve without discrete variables and maintains convexity. You can search the Internet to find out the rigorous justification of why does ℓ_1 Work.

ℓ_1-Constrained Formulation:

$$\begin{aligned} &\min_{x \in X} f(x) \\ &\text{s.t.} ||x||_1 \leqslant T, \end{aligned} \tag{15.1}$$

for some $T > 0$. Generally, smaller $T \Rightarrow$ sparser x.
Function-Constrained Formulation:

$$\begin{aligned} &\min_{x \in X} ||x||_1 \\ &\text{s.t.} f(x) \leqslant f \end{aligned} \tag{15.2}$$

for some $f \geqslant \min f(x)$.
Penalty Formulation:

$$\min_{x \in X} f(x) + \lambda ||x||_1 \tag{15.3}$$

for some parameter $\lambda \geqslant 0$. Generally, larger $\lambda \Rightarrow$ sparser x.

15.3 Application: LASSO problem

15.4 Application: matrix completion problem

The Matrix Completion Problem is a fundamental optimization and machine learning problem with a wide range of applications. The goal is to recover a complete low-rank matrix from a partially observed or sampled matrix.

Formally, let $M \in \mathbb{R}^{m \times n}$ be the partially observed matrix, where only a subset of the entries are known. The index set of the known entries is denoted as $\Omega \subset [m] \times [n]$. The objective is to find a low-rank matrix $X \in \mathbb{R}^{m \times n}$ that best approximates the known entries of M, while having low-rank structure.

The standard formulation of the Matrix Completion Problem is as follows:

$$\min_{X \in \mathbb{R}^{m \times n}} \|P_\Omega(X) - P_\Omega(M)\|_F^2 + \lambda \|X\|_*$$

Here, $P_\Omega(X)$ is the projection operator that keeps the entries of X in the set Ω and sets the rest to zero. The first term, $\|P_\Omega(X) - P_\Omega(M)\|_F^2$, is the data fidelity term that encourages the recovered matrix X to match the known entries of M. The second term, $\|X\|_*$, is the nuclear norm of X, which serves as a convex surrogate for the rank of X, encouraging a low-rank solution. The parameter $\lambda > 0$ controls the trade-off between these two terms.

The Matrix Completion Problem arises in a variety of applications, such as:

1. Recommender systems: Predicting user preferences for unrated items based on the partially observed user-item rating matrix. 2. Image processing: Recovering missing pixels in an image based on the known pixel values. 3. Signal processing: Reconstructing a signal from a limited number of measurements. 4. Bioinformatics: Inferring missing entries in biological data matrices, such as gene expression profiles.

Numerous algorithms have been developed to solve the Matrix Completion Problem, including convex optimization methods, low-rank matrix factorization, and iterative thresholding techniques. These approaches have demonstrated strong theoretical guarantees and practical performance in various applications.

15.5 Reference

- Methods for Sparse Optimization
- Sparse Optimization
- Sparse Optimization

Chapter 16

Optimality Conditions for Unconstrained Nonsmooth Problems

16.1 First-order necessary and sufficient conditions for convex optimization problems

16.2 First-order necessary conditions for composite optimization problems

Chapter 17

Proximal Algorithms

Part III

Duality and Constrained Optimization

Chapter 18

Optimality Conditions for Constrained Problems

18.1 Geometric constrained minimization

When we are solving an unconstrained optimization problem, the goal is clear: we want to find a point where the gradient vanishes. All of the algorithms we looked at over the last few lectures were in service of this condition. Once we add constraints, the optimality conditions are more complicated, and involve relationships between the gradient of the functional we are minimizing along with the gradi- ents of the constraints — these are the so-called Karush-Kuhn-Tucker (KKT) conditions.

We will build up to the KKT conditions slowly. We will first derive a general (and very easy to prove) geometric necessary and sufficient condition for x^* to be a minimizer of a constrained optimization program. We will then show how this simple result immediately yields the KKT conditions for certain kinds of constraints. In the next set of notes, we will derive the KKT conditions, show that they are always sufficient, and discuss conditions under which they are also necessary.

We start by considering the general constrained problem

$$\text{minimize}_{\boldsymbol{x}\in\mathcal{C}} f(\boldsymbol{x})$$

where $\mathcal{C}$ is a closed, convex set, and f is again a convex function. We have the following fundamental result:

Theorem 18.1

Let f be a differentiable convex function, and $\mathcal{C}$ be a close convex set. Then $\boldsymbol{x}^*$ is a minimizer of

$$\text{minimize}_{\boldsymbol{x}\in\mathcal{C}} f(\boldsymbol{x})$$

if and only if $\boldsymbol{x}^* \in \mathcal{C}$ and

$$\langle \boldsymbol{y} - \boldsymbol{x}^*, \nabla f(\boldsymbol{x}^*) \rangle \geqslant 0$$

for all $\boldsymbol{y} \in \mathcal{C}$.

Remark. This result is geometrically intuitive; it is saying that every vector from x^* to another point y in $\mathcal{C}$ must make an obtuse angle with $-\nabla f(x^*)$. That is, there cannot be any descent directions from x^* that lead to another point in $\mathcal{C}$. Here is a picture:

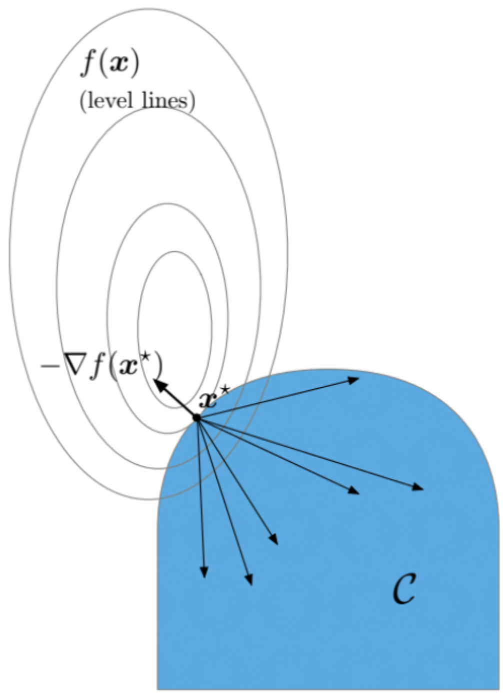

Figure 18.1

Proof. 1 □

18.1.1 Examples

The abstract geometrical result in the previous section will eventually lead us to the Karush-Kuhn-Tucker (KKT) conditions. But we will build up to this by looking at what it tells us in several important (and prevalent) cases. We assume throughout this section that f is convex, differentiable, and defined on all of $\mathbb{R}^n$.

Linear constraints

Non-negativity constraints

A single convex inequality constraint

18.2 Reference

- Constrained Minimization

Chapter 19

Normal and Tangent Cones

19.1 Reference

- **CS/ISyE/Math 730: Nonlinear Optimization 2 Lecture Notes**