

Nonlinear Optimization: Self-Study

Book: Nonlinear Programming 3rd Ed. - Bertsekas

Patrick Emami

Contents

1	Appendix A - Mathematical Background	2
2	Appendix B - Convex Analysis	2
2.1	Appendix B.1 - Convex Sets and Functions, 1/16/17	2
2.2	Appendix B.1 - Convex Sets and Functions, 1/18/17	4
2.3	Appendix B.2 - Hyperplanes, 1/20/17	5
2.4	Appendix B.3 - Cones and Polyhedral Convexity, 1/23/17 . .	7
2.5	Appendix B.4 - Extreme Points and LP, 1/25/17	8
2.6	Appendix B.5 - Differentiability Issues, 1/27/17	10
3	Chapter 1 - Unconstrained Optimization: Basic Methods	10
3.1	Chapter 1.1 - Optimality Conditions, 1/30/17	10

1 Appendix A - Mathematical Background

Definitions, propositions, and theorems useful for understanding the material presented in this book.

Proposition A.23. *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be twice continuously differentiable over an open sphere S centered at a vector x .*

(a) *For all y such that $x + y \in S$,*

$$f(x + y) = f(x) + y' \nabla f(x) + \frac{1}{2} y' \left(\int_0^1 \left(\int_0^t \nabla^2 f(x + \tau y) d\tau \right) dt \right) y.$$

(b) *For all y such that $x + y \in S$, there exists an $\alpha \in [0, 1]$ such that*

$$f(x + y) = f(x) + y' \nabla f(x) + \frac{1}{2} y' \nabla^2 f(x + \alpha y) y.$$

(c) *For all y such that $x + y \in S$ there holds,*

$$f(x + y) = f(x) + y' \nabla f(x) + \frac{1}{2} y' \nabla^2 f(x) y + o(\|y\|^2).$$

2 Appendix B - Convex Analysis

These results are presented without proof; most are actually omitted from the book anyways since the author refers the reader to his book on Convex Optimization.

2.1 Appendix B.1 - Convex Sets and Functions, 1/16/17

Some definitions and properties of convex sets and functions

A subset C of \mathbb{R}^n is called *convex* if

$$\alpha x + (1 - \alpha)y \in C, \quad \forall x, y \in C, \quad \forall \alpha \in [0, 1].$$

Some important properties of convex sets, presented without proof:

- (a) For any collection $\{C_i \mid i \in I\}$ of convex sets, the set intersection $\cap_{i \in I} C_i$ is convex.
- (b) The vector sum of two convex sets is convex.

- (c) The image of a convex set under a linear transformation is convex.
- (d) If C is a convex set and $f : C \rightarrow \mathbb{R}$ is a convex function, the level sets $\{x \in C \mid f(x) \leq \alpha\}$ and $\{x \in C \mid f(x) < \alpha\}$ are convex for all scalars α .

A function $f : C \rightarrow \mathbb{R}$ is called *convex* if

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y), \quad \forall x, y \in C, \quad \forall \alpha \in [0, 1].$$

The function f is *concave* if $-f$ is convex. The function f is called *strictly convex* if the above inequality is strict for all $x, y \in C$ with $x \neq y$, and all $\alpha \in (0, 1)$

A special case of Jensen's Inequality gives us the following

$$f\left(\sum_{i=1}^m \alpha_i x_i\right) \leq \sum_{i=1}^m \alpha_i f(x_i)$$

for $x_1, \dots, x_m \in C$, $\alpha_1, \dots, \alpha_m \geq 0$, and $\sum_{i=1}^m \alpha_i = 1$.

The following provides means for recognizing convex functions

- (a) A linear function is convex.
- (b) Any vector norm is convex.
- (c) The weighted sum of convex functions, with positive weights, is convex.

Characterizations of Differentiable Convex Functions

- (a) f is convex over C if and only if

$$f(z) \geq f(x) + (z - x)' \nabla f(x) \quad \forall x, z \in C$$

Note that one can easily picture this for the simple case of the quadratic function.

- (b) f is strictly convex over C if and only if the above inequality is strict whenever $x \neq z$

- (c) if $\nabla^2 f(x)$ is positive semidefinite for all $x \in C$, then f is convex over C .
- (d) if $\nabla^2 f(x)$ is positive definite for all $x \in C$, then f is strictly convex over C .
- (e) f is **strongly convex** if for some $\sigma > 0$, we have

$$f(y) \geq f(x) + \nabla f(x)'(y - x) + \frac{\sigma}{2} \|x - y\|^2$$

- (f) If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is continuously differentiable and strongly convex in that there is some σ satisfying the inequality from above, then f is strictly convex. If in addition, ∇f satisfies the Lipschitz condition

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathbb{R}^n,$$

for some $L > 0$, then we have for all $x, y \in \mathbb{R}^n$

$$(\nabla f(x) - \nabla f(y))'(x - y) \geq \frac{\sigma L}{\sigma + L} \|x - y\|^2 + \frac{1}{\sigma + L} \|\nabla f(x) - \nabla f(y)\|^2.$$

- (g) If f is twice continuously differentiable over \mathbb{R}^n , then f satisfies (e) if and only if the matrix $\nabla^2 f(x) - \sigma I$, where I is the identity, is positive semidefinite for every $x \in \mathbb{R}^n$

2.2 Appendix B.1 - Convex Sets and Functions, 1/18/17

Convex and Affine Hulls

Let X be a subset of \mathbb{R}^n . A *convex combination* of elements of X is a vector of the form $\sum_{i=1}^m \alpha_i x_i$, where x_1, \dots, x_m belong to X and $\alpha_1, \dots, \alpha_m$ are scalars such that

$$\alpha_i \geq 0, \quad i = 1, \dots, m, \quad \sum_{i=1}^m \alpha_i = 1.$$

The *convex hull* of X is the set of all convex combinations of elements of X . In particular, if X consists of a finite number of vectors x_1, \dots, x_m , its convex hull is

$$\text{conv}(X) = \left\{ \sum_{i=1}^m \alpha_i x_i \mid \alpha_i \geq 0, \ i = 1, \dots, m, \ \sum_{i=1}^m \alpha_i = 1 \right\}$$

The *affine hull* of a subspace S is the set of all affine combinations of elements of S ,

$$\text{aff}(S) = \left\{ \sum_{i=1}^m \alpha_i x_i \mid x_i \in S, \ \sum_{i=1}^m \alpha_i = 1 \right\}.$$

A set in a vector space is affine if it contains all of the lines generated by its points. The affine hull is also the intersection of all linear manifolds containing S , where linear manifolds are translations of a vector subspace. Note that $\text{aff}(S)$ is itself a linear manifold and it contains $\text{conv}(S)$.

Topological Properties of Convex Sets

Let C be a convex subset of \mathbb{R}^n . We say that x is a *relative interior point* of C if $x \in C$ and there exists a neighborhood N of x such that $N \cap \text{aff}(C) \subset C$, i.e., if x is an interior point of C relative to $\text{aff}(C)$. The *relative interior* of C is the set of all relative interior points of C .

If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex, then it is continuous. More generally, if $C \subset \mathbb{R}^n$ is convex and $f : C \rightarrow \mathbb{R}$ is convex, then f is continuous in the relative interior of C . Note that every function that is finite and convex on an open interval is continuous on that interval. The proof for this uses the fact that the left-hand and right-hand derivatives can be shown to exist at every point in the open interval. Alternatively, one can use the fact that ∇f satisfies the Lipschitz condition.

The set of minimizing points of a convex function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ over a closed convex set X is nonempty and compact if and only if all its level sets,

$$L_a = \{x \in X \mid f(x) \leq a\}, \quad a \in \mathbb{R},$$

are compact.

2.3 Appendix B.2 - Hyperplanes, 1/20/17

A *hyperplane* in \mathbb{R}^n is a set $H = \{x \mid a'x = b\}$, where a is a nonzero vector in \mathbb{R}^n and b is a scalar. Note that hyperplanes are convex sets. The hyperplane can also be described as an affine set that is parallel to the subspace $\{x \mid a'x = 0\}$. This is because one can describe a hyperplane also as

$$H = \bar{x} + \{x \mid a'x = 0\}$$

for $\bar{x} \in H$, or

$$H = \{x \mid a'x = a'\bar{x}\}.$$

Theorem 1 (Supporting Hyperplane). *If $C \subset \mathbb{R}^n$ is a convex set and \bar{x} is a point that does not belong to the interior of C , there exists a vector $a \neq 0$ such that*

$$a'x \geq a'\bar{x}, \quad \forall x \in C.$$

In fact, one can think of \bar{x} as being on the boundary of C . In general, a supporting hyperplane of a convex set C is one that entirely contains C in one of the two closed half-spaces bounded by the hyperplane. Also, C has at least one boundary-point on the hyperplane, and perhaps multiple supporting hyperplanes at a single boundary point.

Theorem 2 (Separating Hyperplane). *If C_1 and C_2 are two nonempty and disjoint convex subsets of \mathbb{R}^n , there exists a hyperplane that separates them, i.e., a vector $a \neq 0$ such that*

$$a'x_1 \leq a'x_2, \quad x_1 \in C_1, x_2 \in C_2.$$

Theorem 3 (Strict Separation Theorem). *If C_1 and C_2 are two nonempty and disjoint convex sets such that C_1 is closed and C_2 is compact, there exists a hyperplane that strictly separates them, i.e., a vector $a \neq 0$ and a scalar b such that*

$$a'x_1 < b < a'x_2, \quad x_1 \in C_1, x_2 \in C_2.$$

We can thus characterize a convex set as the intersection of the halfspaces that contain it.

Theorem 4 (Proper Separation). *(a) Let C_1 and C_2 be two nonempty convex subsets of \mathbb{R}^n . There exists a hyperplane that separates C_1 and C_2 , and does not contain both C_1 and C_2 if and only if*

$$ri(C_1) \cap ri(C_2) = \emptyset.$$

- (b) Let C and P be two nonempty convex subsets of \mathbb{R}^n such that P is the intersection of a finite number of closed halfspaces. There exists a hyperplane that separates C and P , and does not contain C if and only if

$$ri(C) \cap P = \emptyset.$$

See <http://www.unc.edu/~normanp/890part4.pdf> and <http://people.hss.caltech.edu/~kcb/Notes/SeparatingHyperplane.pdf> for proofs of various theorems and properties mentioned above.

2.4 Appendix B.3 - Cones and Polyhedral Convexity, 1/23/17

A subset C of a vector space V is a cone (sometimes called linear cone) if for each $x \in C$ and **positive** scalars α , the product αx is in C . A cone C is a **convex cone** if $\alpha x + \beta y$ belongs to C , for positive scalars α , β , and $x, y \in C$. The **polar cone** of C is given by

$$C^\perp = \{y \mid y'x \leq 0, \forall x \in C\}.$$

The polar cone of a subspace is the orthogonal complement, i.e., $C^\perp = -C^*$. A *finitely generated* cone has the form

$$C = \left\{ x \mid x = \sum_{j=1}^r \mu_j a_j, \mu_j \geq 0, j = 1, \dots, r \right\},$$

where a_1, \dots, a_r are some vectors. A cone C is *polyhedral* if it has the form

$$C = \{x \mid a'_j x \leq 0, j = 1, \dots, r\},$$

where a_1, \dots, a_r are some vectors. Note that all of these cones are convex.

Theorem 5 (Polar Cone Theorem). *For any nonempty closed convex cone C , we have $(C^\perp)^\perp = C$.*

(Farkas' Lemma) Let x, e_1, \dots, e_m , and a_1, \dots, a_r be vectors of \mathbb{R}^n . We have $x'y \leq 0$ for all vectors $y \in \mathbb{R}^n$ (i.e., x is in a polar cone), such that

$$y'e_i = 0, \forall i = 1, \dots, m \quad y'a_j \leq 0, \forall j = 1, \dots, r,$$

if and only if x can be expressed as

$$x = \sum_{i=1}^m \lambda_i e_i + \sum_{j=1}^r \mu_j a_j,$$

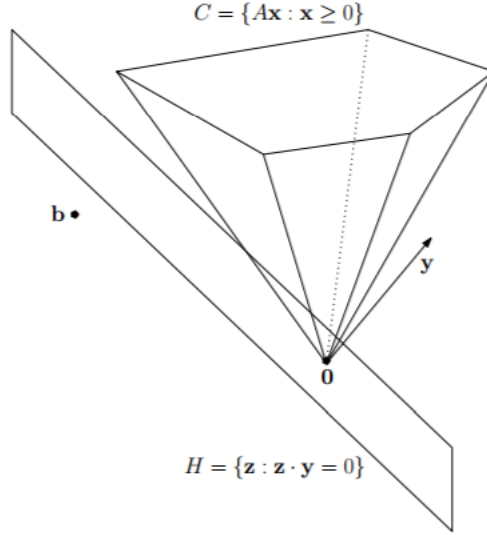


Figure 1: Farkas' Lemma (source: <http://www.sfu.ca/~mdevos/notes/misc/LP.pdf>)

where λ_i and μ_j are some scalars with $\mu_j \geq 0$ for all j . This is a result stating that a vector is either in a given convex cone or that there exists a hyperplane separating the vector from the cone- there are no other possibilities.

A subset of \mathbb{R}^n is a *polyhedral set* if it is nonempty and it is the intersection of a finite number of closed halfspaces, i.e., if it is of the form

$$P = \{x \mid a'_j x \leq b_j, \quad j = 1, \dots, r\},$$

where a_j are some vectors and b_j are some scalars. A set P is polyhedral if and only if it is the sum of a finitely generated cone and the convex hull of a finite set of points.

2.5 Appendix B.4 - Extreme Points and LP, 1/25/17

A vector x is said to be an extreme point of a convex set C if x belongs to C and there do not exist vectors $y, z \in C$, and a scalar $\alpha \in (0, 1)$ such that

$$y \neq x, \quad z \neq x, \quad x = \alpha y + (1 - \alpha)z,$$

Thinking about this, every point on a circle in \mathbb{R}^2 is an extreme point of the convex set consisting of these points.

Some important facts about extreme points:

1. if H is a hyperplane that passes through a boundary point of C and contains C in one of its halfspaces, then every extreme point of $C \cap H$ is also an extreme point of C .
2. C has at least one extreme point if and only if it does not contain a line, i.e., a set L of the form $L = \{x + \alpha d \mid \alpha \in \mathbb{R}\}$ with $d \neq 0$.
3. Let C be a closed convex subset of \mathbb{R}^n , and let C^* be the set of minima of a concave function $f : C \rightarrow \mathbb{R}$ over C . Then if C is closed and contains at least one extreme point, and C^* is nonempty, then C^* contains some extreme point of C .

Proposition B.19. *Let C be a closed convex set and let $f : C \rightarrow \mathbb{R}$ be a concave function. Assume that for some invertible $n \times n$ matrix A and some $b \in \mathbb{R}^n$ we have*

$$Ax \geq b, \quad \forall x \in C.$$

Then if f attains a minimum over C , it attains a minimum at some extreme point of C .

Now, important facts concerning polyhedral sets.

Let P be a polyhedral set in \mathbb{R}^n .

1. If P has the form

$$P = \{x \mid a'_j x \leq b_j, \quad j = 1, \dots, r\},$$

then a vector $v \in P$ is an extreme point of P if and only if the set

$$A_v = \{a_j \mid a'_j v = b_j, \quad j \in \{1, \dots, r\}\},$$

contains n linearly independent vectors.

2. If P has the form

$$P = \{x \mid Ax = b, \quad x \geq 0\},$$

where A is a given $m \times n$ matrix and b is a given vector, then a vector $v \in P$ is an extreme point of P if and only if the columns of A corresponding to the nonzero coordinates of v are linearly independent.

3. (*Fundamental Theorem of Linear Programming*) Assume that P has at least one extreme point. Then if a linear function attains a minimum over P , it attains a minimum at some extreme point of P .

Proof. For (3): Since P is polyhedral, it has a representation

$$P = \{x \mid Ax \geq b\},$$

for some $m \times n$ matrix A and some $b \in \mathbb{R}^m$. If A had rank less than n , then its nullspace would contain some nonzero vector \bar{x} , so P would contain a line parallel to \bar{x} , contradicting the existence of an extreme point. Thus A has rank n and hence it must contain n linearly independent rows that constitute an $n \times n$ invertible submatrix \hat{A} . If \hat{b} is the corresponding subvector of b , we see that every $x \in P$ satisfies $\hat{A}x \geq \hat{b}$. The result then follows by Prop. B.19. \square

2.6 Appendix B.5 - Differentiability Issues, 1/27/17

Given a convex function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, we say that a vector $d \in \mathbb{R}^n$ is a *subgradient* of f at a point $x \in \mathbb{R}^n$ if

$$f(z) \geq f(x) + (z - x)'d,$$

The set of all subgradients of a convex function is called the *subdifferential* of f at x , and is denoted by $\partial f(x)$.

A vector $x^* \in X$ minimizes f over a convex set $X \subset \mathbb{R}^n$ if and only if there exists a subgradient $d \in \partial f(x^*)$ such that

$$d'(z - x^*) \geq 0, \quad \forall z \in X.$$

For the special case where $X = \mathbb{R}^n$, we obtain a basic necessary and sufficient condition for unconstrained optimality of x^* , namely $0 \in \partial f(x^*)$.

3 Chapter 1 - Unconstrained Optimization: Basic Methods

3.1 Chapter 1.1 - Optimality Conditions, 1/30/17

Proposition 1.1.1 (Necessary Optimality Conditions). *Let x^* be an unconstrained local minimum of $f : \mathbb{R}^n \rightarrow \mathbb{R}$, and assume that f is continuously differentiable in an open set S containing x^* . Then*

$$\nabla f(x^*) = 0.$$

If in addition f is twice continuously differentiable within S , then

$$\nabla^2 f(x^*) : \text{positive semidefinite.}$$

Proof. Fix some arbitrary $d \in \mathbb{R}^n$. Then, using the chain rule to differentiate the function $g(\alpha) = f(x^* + \alpha d)$ of the scalar α , we have

$$0 \leq \lim_{\alpha \downarrow 0} \frac{f(x^* + \alpha d) - f(x^*)}{\alpha} = \frac{dg(0)}{d\alpha} = d'f(x^*),$$

where the inequality follows because we assume that x^* is a local minimum. Since d is arbitrary, we can replace it with $-d$ and the inequality still holds. Therefore, $d'\nabla f(x^*) = 0$ for all $d \in \mathbb{R}^n$, which shows that $\nabla f(x^*) = 0$.

Assume that f is twice continuously differentiable, and let d be any vector in \mathbb{R}^n . For all $\alpha \in \mathbb{R}$, the second order expansion yields

$$f(x^* + \alpha d) - f(x^*) = \alpha \nabla f(x^*)'d + \frac{\alpha^2}{2} d' \nabla^2 f(x^*) d + o(\alpha^2).$$

□