Notes on Optimization:

Vishal Raman

We present expository notes and detailed solutions to problems from several texts and papers on optimization. This is meant to be an expansive set of notes on optimization topics that I have studied or am currently studying. Any typos or mistakes are my own - please redirect them to my email.

Contents

Part I.
Convex Optimization

We present expository notes on *Convex Optimization* by Boyd and Vandenberge. Solutions to some exercises are presented, but many topics are currently left out. I might fill in some of the applications in the future, but the main goal of the first reading is to have a fundamental understanding of the theory, algorithms, and their analysis.

Some useful additional references are notes from EECS 227A/227B from UC Berkeley.

1. Convex Sets

1.1. Affine and Convex Sets

Definition 1.1 (Affine)**.** A set $C \subset \mathbb{R}^n$ is affine if the line through any two distinct points in C lies in C; that is, for any $x_1, x_2 \in C$ and $\theta \in \mathbb{R}$, we have $\theta x_1 + (1-\theta)x_2 \in C$.

More generally, if we have $\theta_1 + \cdots + \theta_k = 1$, $\theta_1 x_1 + \cdots + \theta_k x_k$ is an affine combination of the points $x_1, \ldots, x_k$, and an affine set contains every affine combination of its points.

Given an affine set C and a point $x_0 \in C$, note that

$$V = C - x_0 = \{x - x_0 : x \in C\}$$

is a linear subspace: closed under sums and scalar multiplication. It follows that the affine set C can be expressed as $C = V + x_0$, and this subspace does not depend on the choice of x_0. We define $\dim(C) = \dim(V)$, the dimension of the corresponding subspace.

Definition 1.2 (Affine Hull)**.** the set of all affine combinations of points in a set $C \subset \mathbb{R}^n$ is called the affine hull of C, denoted $\operatorname{aff} C$:

$$\operatorname{aff} C = \{\theta_i x^i : x_1, \ldots, x_k \in C, \sum_{i=1}^{k} \theta_i = 1.\}$$

> **Remark 1.3.** Note the usage of Einstein summation convention in the definition - I will probably use this without a remark later in the notes whenever it is clear.

We define the affine dimension of a set C as the dimension of its affine hull. This is a useful definition in the context of convex analysis, but it is important to note that it is not always consistent with other definitions of dimension.

Example 1.4 (Affine dimension of $\mathbb{S}^1$)
Consider $\mathbb{S}^1 = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$. Note that $\dim(\operatorname{aff}(\mathbb{S}^1)) = \dim(\mathbb{R}^2) = 2$, while $\dim(\mathbb{S}^1) = 1$ under most definitions.

Definition 1.5 (Relative Interior)**.** We define the relative interior of a set C, denoted $\operatorname{relint} C$ as the interior relative to $\operatorname{aff} C$:

$$\operatorname{relint} C = \{x \in C : B(x, r) \cap \operatorname{aff} C \subseteq C \text{ for some } r > 0\}.$$

The important thing to note with the relative interior is that the we now consider the norm assosiated with $\operatorname{aff} C$ for $C \subset \mathbb{R}^n$, which have affine dimension less than n.

Definition 1.6 (Convex Set)**.** A set C is convex if the line segment between any two points in C lies in C: for $x_1, x_2 \in C$, and $\theta \in [0, 1]$, we have

$$\theta x_1 + (1-\theta)x_2 \in C.$$

As before this generalizes to multiple points. We denote $\operatorname{conv} C$ as the convex hull of a set C.

Definition 1.7 (Cone)**.** A set C is called a cone, or nonnegative homogeneous, if for every $x \in C$ and $\theta \geq 0$, we have $\theta x \in C$. A set C is a convex cone if it is convex and a cone, which means that for any $x_1, x_2 \in C$ and $\theta_1, \theta_2 \geq 0$, we have

$$\theta_1 x_1 + \theta_2 x_2 \in C.$$

Points of this form have an apex at 0 with edges passing through x_1 and x_2. We can similarly define conic combinations and a conic hull.

1.2. Important Examples

We start with simple examples:

- The empty set, a singleton, and the whole space are affine.
- Any line if affine. If it passes through zero, it is a subspace, and hence a convex cone.
- A line segment is convex but not affine (unless it reduces to a point).
- A ray, which has the form $\{x_0 + \theta v | \theta \geq 0\}$ with $v \neq 0$ is convex but not affine. It is a convex cone if $x_0 = 0$.
- Any subspace is affine and a convex cone, hence convex.

Definition 1.8 (Hyperplane)**.** A hyperplane is a set of the form $\{x | a^\intercal x = b\}$, where $a \in \mathbb{R}^n \setminus 0$, $b \in \mathbb{R}$.

Geometrically, we can interpret this as the set of points witha constant inner product to a given vector a, or as a hyplerplane with normal vector a and offset from the origin b.

Definition 1.9 (Positive Semidefinite Cone)**.** We use the notation $\mathcal{S}^n$ to denote the set of symmetric $n \times n$ matrices, a vectorspace with dimension $n(n+1)/2$. Then, we have

$$\mathcal{S}^n_+ = \{X \in \mathcal{S}^n : X \succcurlyeq 0\},$$

$$\mathcal{S}^n_{++} = \{X \in \mathcal{S}^n : X \succ 0\}$$

,the set of of symmetric positive semidefinite and symmetric positive definite matrices respectively. The set $\mathcal{S}^n_+$ is a convex cone.

1.3. Generalized Inequalities

Definition 1.10 (Proper Cone)**.** A cone $K \subset \mathbb{R}^n$ is called proper if it satisfies the following properties:

- K is convex.
- K is closed.
- K is solid, which means that it has nonempty interior.
- K is pointed, which means that it contains no line (or equivalently, $x \in K, -x \in K \Rightarrow x = 0$).

Definition 1.11 (Generalized Inequality)**.** We associate a partial ordering on $\mathbb{R}^n$ on a proper cone K, which is called a generalized inequality. This is defined by

$$x \preceq_K y \iff y - x \in K.$$

This is associated with a strict partial ordering given by

$$x \prec_K y \iff y - x \in K^\circ.$$

The generalized inequality satisfies the following properties:

- Preserved under addition: if $x \preceq_K y, u \preceq_K v$, then $x + u \preceq_K y + v$.
- Transitive.
- Preserved under non-negative scaling.
- Reflexive.
- Antisymmetric.
- Preserved under limits: if $x_i \preceq_K y_i$ and $x_i \to x, y_i \to y$, then $x \preceq_K y$.

The strict version satisfies fewer properties:

- $x \prec_K y \Rightarrow x \preceq_K y$
- $x \prec_K y$ and $u \preceq_K v$, then $x + u \preceq_K y + v$
- $x \prec_K y$ and $\alpha > 0$, then $\alpha x \prec_K \alpha y$
- $x \not\prec_K x$
- if $x \prec_K y$, then for u and v small enough(wrt the norm), $x + u \prec_K y + v$.

The notion of minimum and maximum elements are more complicated in the context of generalized inequalities:

Definition 1.12 (Minimum Element)**.** $x \in S$ is the minimum element of S if for every $y \in S$, $x \preceq_K y$. Alternatively, $x \in S$ is the minimum element if and only if

$$S \subseteq x + K.$$

Definition 1.13 (Minimal Element)**.** $x \in S$ is a minimal element of S if $y \preceq x$ only if $y = x$. Alternatively, $x \in S$ is minimal if and only if

$$(x - K) \cap S = \{x\}.$$

Remark 1.14. We can define maximum and maximal analogously. A set can have at most one minimum/maximum element. However, it can have many minimal or maximal elements.

1.4. Supporting and Separating Hyperplanes

We begin with the statement of the separating hyperplane theorem:

Theorem 1.15 (Separating Hyperplane Theorem)
Suppose C and D are nonempty disjoint convex sets, $C \cap D = \emptyset$. Then there exists $a \neq 0$ and b such that $a^\intercal x \leq b$ for all $x \in C$ and $a^\intercal x \geq b$ for all $x \in D$. In other words, the affine function $a^\intercal x - b$ is nonpositive on C and nonnegative on D. The hyperplane $\{x : a^\intercal x = b\}$ is called a separating hyperplane for the sets C and D.

Proof. We begin by proving a special case and leave the general case for a future exercise.

Define $\text{dist}(C, D) = \inf\{\|u - v\|_2 | u \in C, v \in D\}$. We assume $\text{dist}(C, D) > 0$ and there exist points $c \in C, d \in D$ achieving the minimum distance (this is satisfied when C, D are closed as one set is bounded).

Define $a = d - c$, $b = \frac{\|d\|_2^2 - \|c\|_2^2}{2}$. We can show that the function $f(x) = a^\intercal x - c = (d-c)^\intercal(x - (1/2)(d+c))$ is nonpositive on C and nonnegative on D.

It suffices to show that f is nonnegative on D (we can swap C and D and consider $-f$ for the other case). Suppose there was a point $u \in D$ such that $f(u) < 0$. We can express $f(u)$ as

$$f(u) = (d-c)^\intercal(u-d) + (1/2)\|d-c\|_2^2,$$

which implies that $(d-c)^\intercal(u-d) < 0$. Now, note that

$$\frac{d}{dt}\|d + t(u-d) - c\|_2^2\Big|_{t=0} = 2(d-c)^\intercal(u-d) < 0.$$

It follows that there exists some $t > 0$ with $t \leq 1$ such that

$$\|d + t(u-d) - c\|_2 < \|d - c\|_2,$$

so the point $d_0 = d + t(u-d)$ is closer to c than d. However, $d + t(u-d) \in D$ as a point on the line segment between d and u, which contradicts the fact that $d \in D$ is the closest point to C. □

Remark 1.16. Note that we may not have strict separation of the sets C and D, even if the sets are closed.

If we impose a stronger condition, we also have a converse.

Theorem 1.17 (Converse Separating Hyperplane Theorem)
Any two convex sets C and D, at least one of which is open, are disjoint if and only if there exists a separating hyperplane.

Definition 1.18 (Supporting Hyperplane)**.** Suppose $C \subset \mathbb{R}^n$ and $x_0 \in \partial C$, the topological boundary. If $a \neq 0$ satisfies $a^\intercal x \leq a^\intercal x_0$ for all $x \in C$, then the hyperplane $\{x : a^\intercal x = a^\intercal x_0\}$ is called a supporting hyperplane to C at the point x_0. Equivalently, we have that the point x_0 and the set C are separated by the hyperplane $\{x : a^\intercal x = a^\intercal x_0\}$.

Geometrically, we have that the hyperplane $\{x : a^\intercal x, a^\intercal x_0\}$ is tangent to C at x_0 and the halfspace $\{x : a^\intercal x \leq a^\intercal x_0\}$ contains C.

Theorem 1.19 (Supporting Hyperplane Theorem)
For any nonempty convex set C and $x_0 \in \partial C$, there exists a supporting hyplerplane to C at x_0.

This follows directly from applying the separating hyperplane theorem. There also exists a partial converse that we will show in a future exercise.

Theorem 1.20 (Partial converse of Supporting Hyperplane Theorem)
If a set is closed, has nonempty interior, and has a supporting hyperplane at every point in its boundary, then it is convex.

1.5. Dual Cones

Definition 1.21 (Dual Cone)**.** Let K be a cone. The set

$$K^* = \{y : x^\intercal y \geq 0 \forall x \in K\}$$

is called the dual cone of K. It is easy to show that K^* is always a cone. Moreover, K^* is always convex, even if the original one is not.

Example 1.22 (Dual Cone of a Subspace)
The dual cone of a subspace $V \subseteq \mathbb{R}^n$ is the orthogonal complement $V^\perp = \{y : v^\intercal y = 0 \forall v \in V\}$.

Example 1.23 (Nonnegative Orthant)
The cone R^n_+ is its own dual:

$$x^\intercal y \geq 0 \forall x \succeq 0 \Leftrightarrow y \succeq 0.$$

We call such a cone self-dual.

Example 1.24 (Positive semidefinite cone)
Note that on $\mathcal{S}^n$, we use the standard inner product $\text{tr}(XY) = \sum_{i,j=1}^n X_{ij}Y_{ij}$. We claim that the positive semidefinite cone $\mathcal{S}^n_+$ is self-dual.

Proof. Suppose $Y \notin \mathcal{S}^n_+$. Then, there exists $q \in \mathbb{R}^n$ such that

$$q^\intercal Y q = \text{tr}(qq^\intercal Y) < 0.$$

It follows that $Y \notin (\mathcal{S}^n_+)^*$ since $\text{tr}(XY) < 0$ where $X = qq^\intercal$.

Conversely, suppose $X, Y \in \mathcal{S}^n_+$. Recall the eigenvalue decomposition: $X = \sum_{i=1}^n \lambda_i q_i q_i^\intercal$. Hence, we have

$$\text{tr}(XY) = \text{tr}\left(Y \sum_{i=1}^n \lambda_i q_i q_i^\intercal\right) = \sum_{i=1}^n \lambda_i q_i^\intercal Y q_i \geq 0.$$

□

Dual cones satisfy the following properties:

- K^* is closed and convex.
- $K_1 \subseteq K_2$ implies $K_2^* \subseteq K_1^*$.
- If K has nonempty interior, then K^* is pointed(contains no line)
- If the closure of K is pointed, then K^* has nonempty interior.
- K^{**} is the closure of the convex hull of K.

These properties show that if K is a proper cone(closed, convex, nonempty interior, pointed), then so is its dual K^* and $K^{**} = K$.

1.5.1. Dual generalized inequalities

Suppose K^* is proper so that it induces a generalized inequality $\preceq_K$. Then K^* is proper, and induces its own generalized inequality.

- $x \preceq y$ if and only if $\lambda^\intercal x \leq \lambda^\intercal y$ for all $\lambda \succeq_{K^*} 0$.
- $x \prec y$ if and only if $\lambda^\intercal x < \lambda^\intercal y$ for all $\lambda \succeq_{K^*} 0$, $\lambda \neq 0$.

We can use these to also establish dual characterizations of minimum/minimal elements.

Proposition 1.25

$x \in S$ is the minimum element of S with respect to $\preceq_K$ if and only if for all $\lambda \succ_{K^*} 0$, x is the unique minimizer of $\lambda^\intercal z$ over $z \in S$. Geometrically, this means that for all $\lambda \succ_{K^*} 0$, the hyperplane

$$\{z : \lambda^\intercal (z - x) = 0\}$$

is a strict supporting hyperplane to S at x.

Remark 1.26. Note that we did not make any assumptions about the convexity of S - this is not required.

Proposition 1.27

If $\lambda \succ_{K^*} 0$ and $x \in S$ minimizes $\lambda^\intercal z$ over $z \in S$, then x is minimal over S.

Proposition 1.28

If $x \in S$ is minimal and S is convex, then there exists a nonzero $\lambda \succeq_{K^*} 0$ such that x minimizes $\lambda^\intercal z$ over $z \in S$.

1.6. Solutions to selected exercises

Exercise 1.29 (2.3)**.** A set C is midpoint convex if $a, b \in C$ implies that $(a+b)/2 \in C$. Show that if C is closed and midpoint convex, then C is convex.

Proof. Take $t \in [0,1]$ and $x, y \in C$. It suffices to show that $z = tx + (1-t)y \in C$. This essentially follows from binary search on z - we can find a sequence of midpoints $\{z_n\} \to z$ and $z \in C$ by the definition of closedness. □

Exercise 1.30 (2.4)**.** Show that the $\operatorname{conv} C = \bigcap\{S \supset C : S \text{ is convex}\}$.

Proof. It is clear that $\operatorname{conv} C \supset \bigcap\{S \supset C : S \text{ is convex}\}$ since $\operatorname{conv} C$ is convex. Moreover, if $x \in \operatorname{conv} C$, then $x = \theta_i x^i$ a convex combination of elements in $C \subset S$, so $x \in S$ for all convex $S \supset C$. □

Exercise 1.31 (2.10)**.** Show that the solution set of a quadratic inequality:

$$C = \{x \in \mathbb{R}^n : x^\intercal A x + b^\intercal x + c \leq 0\}$$

with $A \in \mathcal{S}^n, b \in \mathbb{R}^n, c \in R$ is convex is $A \succeq 0$. Moreover, show that the intersection of C and the hyperplane defined by $g^\intercal x + h = 0$ (where $g \neq 0$) is convex if $A + \lambda g g^\intercal \succeq 0$ for some $\lambda \in \mathbb{R}$. Is the converse of any of these statements true?

Proof. Suppose $A \succeq 0$. If $x, y \in C$, and $t \in [0,1]$ and we define $z = tx + (1-t)y$, it follows that

$$\begin{aligned} z^\intercal A z + b^\intercal z + c &= (tx + (1-t)y)^\intercal A(tx + (1-t)y) + b^\intercal(tx + (1-t)y) + c \\ &= t(tx^\intercal A x + b^\intercal x + c) + (1-t)((1-t)y^\intercal A y + b^\intercal y + c) + 2t(1-t)x^\intercal A y \\ &\leq 2t(1-t)x^\intercal A y - t(1-t)x^\intercal A x - t(1-t)y^\intercal A y \\ &= -t(1-t)(x+y)^\intercal A(x+y) \leq 0. \end{aligned}$$

The converse is not true - consider $A = -1$, $b = 0$ and $c = 0$. Then, $C = \mathbb{R}$ which is convex while $A \not\succeq 0$.

Another proof is as follows: the intersection of C with the hyperplane is given by

$$\{x \in \mathbb{R}^n : x^\intercal A x + b^\intercal x + c \leq 0\}.$$

Recall that a set C is convex if the intersection with any arbitrary line is convex. Let $L = \{x + tv : t \in \mathbb{R}\}$ for some fixed $v \in \mathbb{R}^n$. Then, note that

$$(x + tv)^\intercal A(x + tv) + b^\intercal(x + tv) + c = \alpha t^2 + \beta t + \gamma,$$

where $\alpha = v^\intercal A v$, $\beta = b^\intercal v + 2x^\intercal A v$, $\gamma = c + b^\intercal x + x^\intercal A x$. It follows that

$$C \cap L = \{x + tv : \alpha t^2 + \beta t + \gamma \leq 0\}.$$

Note that a sufficient condition for convexity is $\alpha \geq 0$ which happens when $v^\intercal A v \geq 0$ or $A \succeq 0$.

For the second problem, we have the added condition that $g^\intercal x + h = 0$. We can define $\delta = g^\intercal v$ and $\epsilon = g^\intercal x + h$, taking the intersection of the hyperplane with the line. Without loss of generality, we may assume that $\epsilon = 0$. Then, we have the intersection

$$\{x + tv : \alpha t^2 + \beta t + \gamma \leq 0, \delta t = 0\}.$$

It $\delta \neq 0$, then the intersection is the singleton set $\{x\}$. Otherwise, $\delta = g^\intercal v = 0$, and the set reduces to

$$\{x + tv : \alpha t^2 + \beta t + \gamma \leq 0\}.$$

As before, a sufficient condition for the convexity is $\alpha > 0$. Therefore, the set is convex if $g^\intercal v$ implies that $v^\intercal A v \geq 0$. If there exists λ such that $A + \lambda g g^\intercal v \succeq 0$, then we have

$$v^\intercal A v = v^\intercal (A + \lambda g g^\intercal) v \geq 0,$$

which proves the result. □

Remark 1.32. Checking the intersection with all lines is a useful trick that is also helpful in the case of proving (or numerically checking) convexity of functions.

Exercise 1.33 (2.11)**.** Show that the hyperbolic set $\mathbb{H}^2 = \{x \in \mathbb{R}^2_+ : x_1 x_2 \geq 1\}$ is convex. Show the same result for $\mathbb{H}^n$.

Proof. Suppose $a, b \in \mathbb{H}^n$. We wish to show that for $t \in [0, 1]$ $c = ta + (1-t)b \in \mathbb{H}^n$. Note that

$$\begin{aligned}
\prod_{i=1}^n c_i &= \prod_{i=1}^n (t a_i + (1-t) b_i) \\
&\geq \prod_{i=1}^n a_i^t b_i^{1-t} \\
&= (\prod_{i=1}^n a_i)^t (\prod_{i=1}^n b_i)^{1-t} \\
&\geq 1.
\end{aligned}$$

□

Exercise 1.34 (2.20)**.** Finding a strictly positive solution of linear equations.

Proof. We first prove the hint. If there exists λ such that $c = A^\intercal \lambda$ and $d = b^\intercal \lambda$, then if $Ax = b$,

$$c^\intercal x = \lambda^\intercal A x = \lambda^\intercal b = d^\intercal = d.$$

Conversely, suppose that $c^\intercal x = d$ for all x satisfying $Ax = b$ - in other words $Ax = b$ implies that $c^\intercal x = d$.

If $\text{rank}(A) = r < n$, then we can find $F \in \mathbb{R}^{n \times (n-r)}$ with $R(F) = N(A)$ so that any x satisfying $Ax = b$ is of the form $x = Fy + x_0$ for all $y \in \mathbb{R}^{n-r}$. Note that we have

$$c^\intercal (Fy + x_0) = c^\intercal F y + c^\intercal x_0 = d.$$

This is only possible for all y if $c \in N(F^\intercal) = R(A^\intercal)$, which proves that $c = A^\intercal \lambda$ for some λ. It follows that $d = \lambda^\intercal A x = \lambda^\intercal b$, which proves the result.

Suppose there exists λ so that $A^\intercal \lambda \succeq 0$, $A^\intercal \lambda \neq 0$ and $b^\intercal \lambda \leq 0$. Then, by the hint, we have that $c^\intercal x = d$ for all x satisfying $Ax = b$ where $d = b^\intercal \lambda$ and $c = A^\intercal \lambda$. It follows that $x \not\succ 0$ since $c^\intercal x = d \leq 0$. Conversely, if there exists $x \succ 0$ with $Ax = b$, then if there existed λ with $c := A^\intercal \lambda \succeq 0$, $c \neq 0$ and $d := b^\intercal \lambda \leq 0$, then we have $c^\intercal x_0 = d$ for all x satisfying $Ax_0 = b$. This is a contradiction as before since if we have $x \succ 0$, then $c^\intercal x = d \succ 0$, a contradiction. □

Exercise 1.35 (2.21)**.** Suppose C and D are disjoint subsets of $\mathbb{R}^n$. Consider the set of $(a, b) \in \mathbb{R}^{n+1}$ for which $a^\intercal x \leq b$ for all $x \in C$ and $a^\intercal x \geq b$ for all $x \in D$. Show that this set is a convex cone.

Proof. Let S denote the set of separating hyperplanes. First, it is clear that if $(a, b) \in S$, then $k(a, b) \in S$ for all $k \geq 0$. If $(a, b), (c, d) \in S$, $t_1, t_2 \geq 0$, then

$$(t_1 a + t_2 c)^\intercal x \leq t_1 b + t_2 d, \quad x \in C$$

$$(t_1 a + t_2 c)^\intercal x \geq t_1 b + t_2 d, \quad x \in C$$

□

Exercise 1.36 (2.22)**.** Completing the proof of the separating hyperplane theorem.

Suppose C, D are disjoint convex sets and consider $S = C - D = \{x - y | x \in C, y \in D\}$, which is convex and does not contain the origin.

Suppose $0 \notin \overline{S}$. Then, we can apply the special case of the separation theorem to obtain a with $a^\intercal(x - y) > 0$ for $x \in C, y \in D$. In particular, $a^\intercal x > a^\intercal y$ for all $x \in C$, $y \in D$, so we can take $b = \sup_{y \in D} a^\intercal y$ so that $a^\intercal x \geq b$ for all $x \in C$ and $a^\intercal y \leq b$ for all $y \in D$.

Now, suppose $0 \in \overline{S}$. Suppose $(\overline{S})^\circ = \emptyset$. Then, $\overline{S}$ must be contained in a hyperplane $H = \{x : a^\intercal x = 0\}$ that contains 0. But this implies the result since for all $x \in C, y \in D$, we have $a^\intercal(x - y) = 0$, or $a^\intercal x = a^\intercal y$, a trivial separating hyperplane.

Otherwise, $(\overline{S})^\circ \neq \emptyset$. For all $\lambda > 0$, define the set $S_\lambda = \{x \in S : \text{dist}(x, \partial S) > \lambda\}$. Since $\overline{S}$ has nonempty interior, there exists some $\epsilon > 0$ so that for all $0 < \lambda < \epsilon$, $S_\lambda \neq \emptyset$. Moreover, note that $\overline{S_\lambda}$ is closed and disjoint from the origin. Let $\lambda_i \to 0$ be a decreasing sequence with $\lambda_1 < \epsilon$. By the special case of the separating hyperplane theorem, we can find $a_i \neq 0$ such that $a_i^\intercal x > 0$ for $x \in S_{\lambda_i}$ and without loss of generality, we can assume that $\|a_i\|_2 = 1$ through normalization. However, note that $\{a_i\} \subset \{x : \|x\|_2 \leq 1\}$, which is compact, so it follows that it has a convergent subsequence to a point a, which must satisfy

$$a^\intercal(x - y) \geq 0, \quad x - y \in S.$$

Exercise 1.37 (2.24)**.** Compute the supporting hyperplanes for $\mathbb{H}^2$.

Proof. For each $t > 0$, we can compute the supporting hyperplane for $\mathbb{H}^2$ at $(t, 1/t)$ - the tangent line has slope $-\frac{1}{t^2}$, which gives the corresponding line

$$\frac{y - \frac{1}{t}}{t - x} = \frac{1}{t^2},$$

which can be simplified to

$$x + t^2 y = 2t.$$

Hence, we can express $\mathbb{H}^2$ as

$$\mathbb{H}^2 = \bigcap_{t>0} \{(x, y) \in \mathbb{R}^2 : x + t^2 y \geq 2t\}.$$

□

Exercise 1.38 (2.26)**.** The support function of $C \subset \mathbb{R}^n$ is defined by

$$S_C(y) = \sup_{x \in C} y^\intercal x,$$

where S_C can take the value $+\infty$. Suppose C, D are closed convex sets. Show that $C = D$ if and only if their support functions are equal.

Proof. The forward direction is obvious. For the reverse, suppose $S_C(y) = S_D(y)$ for all $y \in \mathbb{R}^n$. Suppose there is a point $x \in C \setminus D$ (the case with $D \setminus C$ is the same). It follows that x_0 is strictly separated from D so there exists (a, b) with $a \neq 0$ so that $a^\intercal x < b$ for $x \in D$, $a^\intercal x_0 > b$. But this implies that $S_C(a) \geq b$, while $S_D(a) < b$, a contradiction. □

Exercise 1.39 (2.27)**.** Converse supporting hyperplane theorem: suppose C is closed, solid, and has a supporting hyperplane at every point in its boundary. Show that C is convex.

Proof. We prove this by showing that the intersection of all all spaces defined by supporting hyperplanes of C is exactly C. It is clear that the intersection contains C since each supporting hyperplane contains C.

To show the other direction, we show that if a point $p \notin C$, then it is not in the intersection of all the half spaces defined by the supporting hyperplanes. Suppose without loss of generality that $0 \in C^\circ$. There exists some $t \in (0, 1)$ so that $tp \in \partial C$. Let $a^\intercal(x - tp) = 0$ be the supporting hyperplane at tp. Note that since $0 \in C^\circ$, we have that $a^\intercal(-tp) < 0$ so $a^\intercal p > 0$. This implies the result since

$$a^\intercal(p - tp) = (1 - t)a^\intercal p > 0,$$

so p is not contained in the corresponding half space. □

2. Convex Functions

2.1. Definition and Basic Properties

Definition 2.1. A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if $\operatorname{dom} f$ is convex and if for all $x, y \in \operatorname{dom} f$ and $\theta \in [0, 1]$, we have

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y).$$

f is strictly convex if the inequality is strict for $x \neq y$.

Geometrically, this means that the line segment between $(x, f(x))$ and $(y, f(y))$ is above the graph of f.

One of the most useful ways to characterize convexity of a function is through line-restriction.

Theorem 2.2 (Line-Restriction)
A function f is convex if and only if for all $x \in \operatorname{dom} f$ and all v, the function $g(t) = f(x + tv)$ is convex (on its domain).

Proposition 2.3 (First-order condition)
Suppose f is differentiable. Then f is convex if and only if $\operatorname{dom} f$ is convex and

$$f(y) \geq f(x) + \nabla f(x)^\intercal (y - x)$$

for all $x, y \in \operatorname{dom} f$. Similarly, f is strictly convex if and only if the inequality is strict for $x \neq y$.

Remark 2.4. This corresponds to the first-order Taylor approximation of f near x and the result says that this is a global underestimator for a convex function.

Proof. First, consider $n = 1$. Assume f is convex and take $x, y \in \operatorname{dom} f$. Since $\operatorname{dom} f$ is convex, for $t \in [0, 1]$, $x + t(y - x) \in \operatorname{dom} f$, so by the convexity of f,

$$f(x + t(y - x)) \leq (1 - t)f(x) + tf(y).$$

Equivalently, we have

$$f(y) \geq f(x) + \frac{f(x + t(y - x)) - f(x)}{t} = f(x) + f'(x)(y - x) + o(t).$$

Taking the limit as $t \to 0$ gives the result. For the converse, choose $x \neq y$, $\theta \in [0, 1]$ and let $z = \theta x + (1 - \theta)y$. We have

$$f(x) \geq f(z) + f'(z)(x - z), \quad f(y) \geq f(z) + f'(z)(y - z).$$

Then, we have

$$\theta f(x) + (1 - \theta)f(y) \geq f(z),$$

which proves the result.

To prove the general case, consider $x, y \in \mathbb{R}^n$ and define $g(t) = f(ty + (1-t)x)$, so that $g'(t) = \nabla f(ty + (1-t)x)^\intercal (y - x)$. If f is convex, g is convex, so $g(1) \geq g(0) + g'(0)$, which proves the result. If the inequality holds for all x, y, then $ty + (1-t)x \in \operatorname{dom} f$, and $\widehat{t}y + (1-\widehat{t})x \in \operatorname{dom} f$, so

$$g(t) \geq g(\widehat{t}) + g'(\widehat{t})(t - \widehat{t}),$$

which implies that g is convex.

□

Theorem 2.5 (Second-order condition)

Suppose f is twice differentiable. Then f is convex if and only if $\operatorname{dom} f$ is convex and its Hessian is positive semidefinite: for all $x \in \operatorname{dom} f$,

$$\nabla^2 f(x) \succeq 0.$$

Some common examples:

- Exponential: e^{ax} is convex on $\mathbb{R}$ for all $a \in \mathbb{R}$.
- Powers: x^a is convex on $\mathbb{R}_{++}$ when $a \geq 1$ or $a \leq 0$, and concave for $a \in [0, 1]$.
- Powers of absolute value: $|x|^p$ for $p \geq 1$ is convex on $\mathbb{R}$.
- Logarithm: $\log x$ is concave on $\mathbb{R}_{++}$.
- Negative entropy: $x \log x$ is convex on $\mathbb{R}_{++}$.
- Norms: every norm on $\mathbb{R}^n$ is convex.
- Max function: $f(x) = \max\{x_1, \ldots, x_n\}$ is convex on $\mathbb{R}^n$.
- Quadratic-over-linear: $f(x, y) = x^2/y$ over $\mathbb{R} \times \mathbb{R}_{++}$.
- Log-sum-exp: $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex on $\mathbb{R}^n$.
- Geometric mean: $f(x) = (\prod_{i=1}^n x_i)^{1/n}$ is concave on $\mathbb{R}^n_{++}$.
- Log-determinant: $f(X) = \log\det X$ is concave on $\mathcal{S}^n_{++}$.

Definition 2.6 (α-sub/superlevel sets)**.** The α-sublevel set of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$C_\alpha = \{x \in \operatorname{dom} f : f(x) \leq \alpha\}.$$

The α-superlevel set given by

$$C^\alpha = \{x \in \operatorname{dom} f : f(x) \geq \alpha\}$$

Sublevel sets of a convex function are convex for any value of α. The converse is not true: consider $f(x) = -e^x$. Similarly, if f is concave, then C^α is convex for any value of α.

Definition 2.7 (Epigraph)**.** The epigraph of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\operatorname{epi} f = \{(x, t) : x \in \operatorname{dom} f, f(x) \leq t\}.$$

Definition 2.8 (Hypograph)**.** The epigraph of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\operatorname{hypo} f = \{(x, t) : x \in \operatorname{dom} f, f(x) \geq t\}.$$

The last very useful result is Jensen's inequality:

Theorem 2.9 (Jensen's Inequality)
If f is convex, $x_1, \ldots, x_k \in \operatorname{dom} f$, and $\theta_1, \ldots, \theta_k \geq 0$ with $\theta_1 + \cdots + \theta_k = 1$, then

$$f(\theta_i x^i) \leq \theta_i f(x^i).$$

2.2. Convexity-preserving Operations

We have the following examples:

- Conic combination: If f_i are convex functions, and $w_i \geq 0$, then $w_i f^i$ is convex.
- Nonnegative integrals: If $f(x, y)$ is convex in x for each $y \in \mathcal{A}$, and $w(y) \geq 0$ for each $y \in \mathcal{A}$, then
$$g(x) = \int_{\mathcal{A}} w(y) f(x, y)\, dy$$
is convex.
- Composition with affine mapping: If f is convex, then $g(x) = f(Ax + b)$ is convex.
- Composite theorem: If $f : D_1 \to \mathbb{R}$ is convex, and $g : D_2 \to \mathbb{R}$ is non-decreasing, with $\operatorname{range}(f) \subset D_2$, then $g \circ f$ is convex.
- Pointwise supremum: if for each $y \in \mathcal{A}$, $f(x, y)$ is convex in x, then the function g, defined as
$$g(x) = \sup_{y \in \mathcal{A}} f(x, y)$$
is convex in x. This follows immediately from the epigraph characterization of convexity.

There is also a converse to the last result: almost every convex function can be expressed as the pointwise supremum of a family of affine functions.

Theorem 2.10
If f is a lower semicontinuous convex function, then for all $x \in \mathbb{R}^n$,

$$f(x) = \sup\{g(x) | g \text{ affine}, g \leq f\}.$$

We also have the following result relating to minimization:

Theorem 2.11
If f is convex in (x, y) and C is a convex nonempty set, then the function

$$g(x) = \inf_{y \in C} f(x, y)$$

is convex.

2.3. The conjugate function

Definition 2.12 (Conjugate)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$. The conjugate function $f^* : \mathbb{R}^n \to \mathbb{R}$ is defined by

$$f^*(y) = \sup_{x \in \operatorname{dom} f} (y^\intercal x - f(x)).$$

The domain of the conjugate function is $y \in \mathbb{R}^n$ so that the supremum is finite.

As an immediate corollary of the definition, we obtain the following results.

Theorem 2.13 (Fenchel's Inequality)

For all x, y, we have

$$f(x) + f^*(y) \geq x^\intercal y.$$

This is sometimes called Young's inequality when f is differentiable.

Theorem 2.14

If f is convex and epi f is closed, then $f^{**} = f$. In general, $f^{**} \leq f$.

Example 2.15 (Legendre Transform)

The conjugate of a differentiable function f is also called the Legendre transform. Suppose f is convex and differentiable, with $\operatorname{dom} f = \mathbb{R}^n$. Any maximizer x^* of $y^\intercal x - f(x)$ satisfies $y = \nabla f(x^*)$. Conversely, if x^* satisfies $y = \nabla f(x^*)$, then x^* maximizes $y^\intercal x - f(x)$. Therefore, if $y = \nabla f(x^*)$, we have

$$f^*(y) = (x^*)^\intercal \nabla f(x^*) - f(x^*).$$

Thus, we can determine $f^*(y)$ for any y for which we can solve $y = \nabla f(z)$ for z.

Alternatively, let $z \in \mathbb{R}^n$ be arbitrary and define $y = \nabla f(z)$. Then,

$$f^*(y) = z^\intercal \nabla f(z) - f(z).$$

2.4. Quasiconvexity

Definition 2.16 (Quasiconvex Function)**.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is quasiconvex if its domain and all sublevel sets C_α for $\alpha \in \mathbb{R}$ are convex.

Theorem 2.17 (Jensen's Characterization)

A function f is quasiconvex if and only if $\operatorname{dom} f$ is convex and for any $x, y \in \operatorname{dom} f$ and $\theta \in [0, 1]$

$$f(\theta x + (1 - \theta)y) \leq \max\{f(x), f(y)\}.$$

In $\mathbb{R}$ we have a simple characterization of quasiconvex functions: Namely, a continuous function $f : \mathbb{R} \to \mathbb{R}$ is quasiconvex if and only if at least one of the conditions holds:

- f is nondecreasing,

- f is nonincreasing,
- there is a point $c \in \operatorname{dom} f$ such that for $t \leq c$, f is nonincreasing, and for $t \geq c$, f is nondecreasing.

The point c can be chosen as a global minimizer of f.

Theorem 2.18 (First-order conditions)
Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. Then f is quasiconvex if and only if $\operatorname{dom} f$ is convex and for all $x, y \in \operatorname{dom} f$,

$$f(y) \leq f(x) \implies \nabla f(x)^\intercal (y - x) \leq 0.$$

Geometrically, when $\nabla f(x) \neq 0$, this says that $\nabla f(x)$ defines a supporting hyperplane to the sublevel set $\{y | f(y) \leq f(x)\}$.

Remark 2.19. It is important to note the differences between the first-order conditions for convexity and quasiconvexity. Notably, if f is convex and $\nabla f(x) = 0$, then x is a global minimizer of f. This is false for quasiconvex functions.

Theorem 2.20 (Second-order conditions)
Suppose f is twice differentiable. If f is quasiconvex, then for all $x \in \operatorname{dom} f$, and all $y \in \mathbb{R}^n$, we have

$$y^\intercal \nabla f(x) = 0 \implies y^\intercal \nabla^2 f(x) y \geq 0.$$

As a partial converse, if f satisfies

$$y^\intercal \nabla f(x) = 0 \implies y^\intercal \nabla^2 f(x) y > 0$$

for all $x \in \operatorname{dom} f$ and all $y \in \mathbb{R}^n$, $y \neq 0$, then f is quasiconvex.

The condition is more complex to interpret. It says that when $\nabla f(x) = 0$, then $\nabla^2 f(x) \succeq 0$. When $\nabla f(x) \neq 0$, it means that $\nabla^2 f(x)$ is positive semidefinite on the space $\nabla f(x)^\perp$ - this means that $\nabla^2 f(x)$ can have at most one negative eigenvalue. The converse says that $\nabla^2 f(x)$ is positive definite whenever $\nabla f(x) = 0$ and for all other points, $\nabla^2 f(x)$ is positive definite on $\nabla f(x)^\perp$.

2.5. Solutions to selected problems

Exercise 2.21 (3.4)**.** Show that a continuous function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for every line segment, its average value on the segment is less than or equal to the average of its values at the endpoints of the segment: for every $x, y \in \mathbb{R}^n$,

$$\int_0^1 f(x + \lambda(y - x))\, dy \leq \frac{f(x) + f(y)}{2}.$$

Proof. If f is convex, then $f(x + t(y - x)) \leq (1 - t)f(x) + tf(y)$, so it follows that

$$\int_0^1 f(x + \lambda(y - x))\, d\lambda \leq \int_0^1 (1 - \lambda)f(x) + \lambda f(y)\, dy = \frac{f(x) + f(y)}{2}.$$

□

Exercise 2.22 (3.5)**.** Suppose $f : \mathbb{R} \to \mathbb{R}$ is convex with $\mathbb{R}_+ \subset \operatorname{dom} f$. Show that its running average F, defined as

$$F(x) = \frac{1}{x} \int_0^x f(x)\, dt, \quad \operatorname{dom} F = \mathbb{R}_{++}.$$

Proof. Note that we can rewrite

$$F(x) = \int_0^1 f(sx)\, ds$$

which proves the result. □

Exercise 2.23 (3.7)**.** Show that a bounded convex function on $\mathbb{R}^n$ is constant.

Proof. We first show the result for $n = 1$. Suppose $f < M$. If we have $f(x) \neq f(y)$, then if we take the line through $f(x)$ and $f(y)$, this must lie below f for all $z \notin [x, y]$. But this is impossible as the line intersects the vertical line $y = M$.

To generalize the result, note that if f is convex and bounded, then $g_v(t) = f(x + tv)$ is convex and bounded for any v, which implies that g_v is constant. But this follows for all $v \in \mathbb{R}^n$ which proves the result. □

Exercise 2.24 (3.8)**.** Prove the second-order convexity condition.

Proof. We first prove the case for $n = 1$. Note that by the second-order Taylor expansion

$$f''(x) = \frac{2}{h^2} \left(f(x + h) - f(x) - f'(x)h + o(h^2) \right)$$

By the first order characterization $f(x + h) - f(x) - f'(x)h \geq 0$, which implies that

$$f''(x) \geq o(1) \xrightarrow{h \to 0} 0.$$

Conversely, by the mean-value theorem version of Taylor's theorem, we have

$$f(y) = f(x) + f'(x)(y - x) + \frac{1}{2} f''(z)(y - x)^2, \quad z \in [x, y],$$

which implies the result.

To generalize, define $g(t) = f(x + tv)$ for some v. If f is convex, then g is. Conversely, it suffices to show that

$$g''(t) = v^\intercal \nabla^2 f(x + tv) v \geq 0.$$

But this follows directly from the definition of positive-semidefiniteness. □

Exercise 2.25 (3.9)**.** Let $F \in \mathbb{R}^{n\times m}$, $\widehat{x} \in \mathbb{R}^n$. The restriction of f to $\{Fz + \widehat{x} : z \in \mathbb{R}^m\}$ is defined as the function $\tilde{f} : \mathbb{R}^m \to \mathbb{R}$ with

$$\tilde{f}(z) = f(Fz + \widehat{x}), \quad \operatorname{dom} f = \{z : Fz + \widehat{x} \in \operatorname{dom} f\}.$$

Suppose f is twice differentiable and convex. Suppose $A \in \mathbb{R}^{p\times n}$ is a matrix whose nullspace is equal to the range of F and $\operatorname{rank} A = n - \operatorname{rank} F$. Show that $\tilde{f}$ is convex if for all $z \in \operatorname{dom} \tilde{f}$ there exists a $\lambda \in \mathbb{R}$ such that

$$\nabla^2 f(Fz + \widehat{x}) + \lambda A^\intercal A \succeq 0.$$

Proof. The convexity follows from the hint that $B \in \mathcal{S}^n$ and $A \in \mathbb{R}^{p\times n}$, then $x^\intercal Bx \geq 0$ for all $x \in N(A)$ if there exists λ such that $B + \lambda A^\intercal A \succeq 0$. So we prove the hint. But this is obvious because

$$0 \leq x^\intercal (B + \lambda A^\intercal A)x = x^\intercal Bx + \lambda x A^\intercal Ax = x^\intercal Bx,$$

since $x \in N(A)$. □

Exercise 2.26 (3.11)**.** A function $\psi : \mathbb{R}^n \to \mathbb{R}$ is called a monotone mapping if for all $x, y \in \operatorname{dom} \psi$,

$$(\psi(x) - \psi(y))^\intercal (x - y) \geq 0.$$

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable and convex. Show that ∇f is monotone. Is this converse true?

Proof. This follows from the first order characterization:

$$f(y) \geq f(x) + \nabla f(x)^\intercal (y - x),$$

$$f(x) \geq f(y) + \nabla f(y)^\intercal (x - y),$$

and adding the two gives the result.

The converse is false - essentially follows from the fact that not all vector fields conservative. It is easy to find monotone vector fields that are not conservative - take $\psi(x, y) = (x + y, y)$. □

Exercise 2.27 (3.12)**.** Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is convex, $g : \mathbb{R}^n \to \mathbb{R}$ is concave, $\operatorname{dom} f = \operatorname{dom} g = \mathbb{R}^n$ and $g \leq f$. Show that there exists an affine function h such that $g \leq h \leq f$.

Proof. A nice geometric argument: $(\operatorname{epi} f)^\circ$ and $(\operatorname{hypo} g)^\circ$ are both nonempty, disjoint convex sets. By the separating hyperplane theorem, there exists a separating hyperplane between the two sets, which defines the affine function between f and g. □

Exercise 2.28 (3.14)**.** *Convex-concave functions and saddle-points* A function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is convex-concave if $f(x, z)$ is a concave function of z for each fixed x and a convex function of x for each fixed z. We also require the domain to have the product form $\operatorname{dom} f = A \times B$, where A, B are convex.

- Give a second-order condition for $f \in C^2$ to be convex-concave in terms of its Hessian.

  Proof. We require $\nabla^2_{xx} f \succeq 0$ and $\nabla^2_{zz} f \preceq 0$. □

- Suppose f is convex-concave and differentiable with $\nabla f(x_0, z_0) = 0$. Show the saddle-point property holds: for all x, z,

$$f(x_0, z) \leq f(x_0, z_0) \leq f(x, z_0).$$

Show that this implies the strong max-min property:

$$\sup_z \inf_x f(x, z) = \inf_x \sup_z f(x, z).$$

Proof. By the first order characterization: $f(x, z_0) \geq f(x_0, z_0)$ and $f(x_0, z) \leq f(x_0, z_0)$ which proves the result. This implies the strong max-min property because we can independently maximize in z and minimize in x regardless of order. □

- Suppose f is differentiable and the saddle-point property holds at x_0, y_0. Show that $\nabla f(x_0, z_0) = 0$.

Proof. Note that if $f(x_0, z_0) \leq f(x, z_0)$ for all x, then $\nabla_x f(x_0, z_0) = 0$ since x_0 is a global minimizer of $f(\cdot, z_0)$. We can argue the same way for z to obtain the result. □

Exercise 2.29 (3.28)**.** Let $f : D \to \mathbb{R}$ be a convex function with $\operatorname{dom} f = D \neq \mathbb{R}^n$. Define $\tilde{f} : D \to \mathbb{R}$ to be the pointwise supremum of all affine functions that are global underestimators of f:

$$\tilde{f} = \sup\{g : g \text{ affine}, g \leq f\}.$$

Show that $f(x) = \tilde{f}(x)$ for all $x \in D^\circ$. Furthermore, show that if f is closed(epi f is closed), then $f = \tilde{f}$.

Proof. We first show that $f(x) = \tilde{f}(x)$ for all $x \in D^\circ$. Note that the point $(x, f(x)) \in \partial D$. There is a supporting hyperplane to epi f at $(x, f(x))$ - there exists $a \in \mathbb{R}^n, b \in \mathbb{R}$ so that $a \neq 0$, $b \geq 0$ and

$$a^\intercal y + bt \leq a^\intercal x + bf(x)$$

for all $(y, t) \in \operatorname{epi} f$. In particular, note that we cannot have $b = 0$ since this would gives $a^\intercal(y - x) \geq 0$ for all $y \in D$, which implies that $x \in \partial D$, a contradiction. Therefore, we have

$$t \geq f(x) + (a/b)^\intercal(x - y), \quad (y, t) \in \operatorname{epi} f$$

so if we define the affine underestimator $g(y) = f(x) + (a/b)^\intercal(x - y)$, then $g \leq \tilde{f} \leq f$, but $g(x) = f(x)$, which implies that $\tilde{f}(x) = f(x)$.

When f is closed, note that epi f is a closed convex set, so it is the intersection of all the half-spaces that contain it. Define

$$H = \{(a, b, c) \in \mathbb{R}^{n+2} : (a, b) \neq 0, \inf_{(x,t) \in \operatorname{epi} f} (a^\intercal x + bt) \geq c\}.$$

Note that each triple (a, b, c) corresponds to a half-space that contains epi f. It follows that

$$\operatorname{epi} f = \bigcap_{(a,b,c) \in H} \{(x, t) : a^\intercal x + bt \geq c\}.$$

It is clear that $b \geq 0$. It suffices to show that

$$\bigcap_{(a,b,c) \in H} \{(x, t) : a^\intercal x + bt \geq c\} = \bigcap_{(a,b,c) \in H, b > 0} \{(x, t) : a^\intercal x + bt \geq c\},$$

since each of the half-spaces on the right correspond to the epigraph of an affine underestimator of f.

It is obvious that

$$\bigcap_{(a,b,c)\in H} \{(x,t) : a^\intercal x + bt \geq c\} \subset \bigcap_{(a,b,c)\in H, b>0} \{(x,t) : a^\intercal x + bt \geq c\}.$$

Suppose (x,t) satisfies $a^\intercal x + bt \geq c$ for all nonvertical half-spaces ($b > 0$) that contain $\operatorname{epi} f$, but there exists a_0, c_0 with $(a_0, 0, c_0) \in H$ so that $a_0^\intercal x < c_0$.

Note that H contains at some element (a_1, b_1, c_1) with $b_1 \neq 0$, so consider the halfspace defined by $(a_0 + \epsilon a_1, \epsilon b_1, c_0 + \epsilon c_1)$ where $\epsilon < \frac{c_0 - a_0^\intercal x}{|a_t^\intercal x + b_1 t|}$ or ($\epsilon = 1$ if $|a_t^\intercal x + b_1 t| = 0$) which contains $\operatorname{epi} f$, since

$$(a_0 + \epsilon a_1)^\intercal y + \epsilon b_1 s \geq a_0^\intercal y + \epsilon(a_1^\intercal y + b_1 s) \geq c_0 + \epsilon c_1,$$

for all $(y,t) \in \operatorname{epi} f$. But note that

$$(a_0 + \epsilon a_1)^\intercal x + \epsilon b_1 t = a_0^\intercal x + \epsilon(a_1^\intercal x + b_1 t) < c_0$$

but this contradicts the fact that (x,t) is in the intersection of all nonvertical half-spaces containing $\operatorname{epi} f$. □

Exercise 2.30 (3.30)**.** The convex hull of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$g(x) = \inf\{t : (x,t) \in \operatorname{conv} \operatorname{epi} f\}.$$

Show that if h is convex and $h \leq f$ then $h \leq g$.

Proof. This follows immediately from considering the epigraph of h and noting that $a \leq b$ implies that $\operatorname{epi} a \supset \operatorname{epi} b$, where a, b are functions. □

Exercise 2.31 (3.31)**.** Let f be a convex function. Define the function g as

$$g(x) = \inf_{\alpha > 0} \frac{f(\alpha x)}{\alpha}.$$

- Show that g is homogeneous.

  Proof. First, note that $g(0) = 0$ since we take the limit as $\alpha \to \infty$. Hence, the result is clear for $t = 0$. For $t > 0$, note that

  $$g(tx) = \inf_{\alpha > 0} \frac{f(\alpha t x)}{\alpha} = \inf_{\beta > 0} t \frac{f(\beta x)}{\beta} = t g(x).$$

  □

- Show that g is the largest homogeneous underestimator of f.

  Proof. If h is homogeneous and $h \leq f$, then note that $h(\alpha x) = \alpha h(x) \leq f(\alpha x)$, which implies that $h(x) \leq \frac{f(\alpha x)}{\alpha}$ for $\alpha > 0$. This proves the result since g is the infimum of $\frac{f(\alpha x)}{\alpha}$ over $\alpha > 0$. □

- Note that

$$\begin{aligned} g(tx+(1-t)y) &= \inf_{\alpha>0} \frac{f(\alpha tx+\alpha(1-t)y)}{\alpha} \\ &\geq \inf_{\alpha>0} \frac{f(\alpha tx)+f((1-t)\alpha y)}{\alpha} \\ &\geq \inf_{\alpha>0} \frac{f(\alpha tx)}{\alpha} + \inf_{\beta>0} \frac{f((1-t)\beta y)}{\beta} \\ &= g(tx)+g((1-t)y) \\ &= tg(x)+(1-t)g(y). \end{aligned}$$

Exercise 2.32 (3.36)**.** Derive the conjugates of the following functions:

- $f(x) = \max_{i=1,\dots,n} x_i$ on $\mathbb{R}^n$.

  Proof. We have

$$f^*(y) = \begin{cases} 0, & y \succeq 0, \mathbf{1}^\intercal y = 1 \\ \infty, & \text{otherwise} \end{cases}$$

  If some entry $y_k < 0$, then note that we can take $x_j = -t\delta_{jk}$, so that

$$y^\intercal x - \max_i x_i = -ty_k \xrightarrow{t\to\infty} \infty.$$

  If $y \succeq 0$ but $\mathbf{1}^\intercal y > 1$, then taking $x = t\mathbf{1}$ gives

$$y^\intercal x - \max_i x_i = t(\mathbf{1}^\intercal y - 1) \xrightarrow{t\to\infty} \infty.$$

  Similarly, if $y \succeq 0$ and $\mathbf{1}^\intercal y < 1$, then taking $x = -t\mathbf{1}$ gives

$$y^\intercal x - \max_i x_i = t(-1^\intercal y + 1) \xrightarrow{t\to\infty} \infty.$$

  Finally, if $y \succeq 0$ and $\mathbf{1}^\intercal y = 1$, then $y^\intercal x \leq \max_i x_i$ so $y^\intercal x - \max_i x_i \leq 0$ with equality if $y^\intercal x = \max_i x_i$ giving a value $f^*(y) = 0$. □

- $f(x) = \sum_{i=1}^r x_{(i)}$ on $\mathbb{R}^n$.

  Proof. If $y_k < 0$ for some k, then taking $x_j = -t\delta_{jk}$ gives $y^\intercal x - 0 = -ty_k \xrightarrow{t\to\infty} \infty$. So we have $y \succeq 0$.

  If $y_k > 1$ for some k, then taking $x_j = t\delta_{jk}$ gives $y^\intercal x - t = t(y_k - 1) \xrightarrow{t\to\infty} \infty$. So we have $y \preceq 1$.

  If $\mathbf{1}^\intercal y > r$ then taking $x = t\mathbf{1}$ gives $y^\intercal x - tr = t(\mathbf{1}^\intercal y - r) \xrightarrow{t\to\infty} \infty$. We have a similar case for $\mathbf{1}^\intercal y < r$ giving $\mathbf{1}^\intercal y = r$.

  It follows that

$$f^*(y) = \begin{cases} 0, & 0 \preceq y \preceq 1, \mathbf{1}^\intercal y = r \\ \infty, & \text{otherwise} \end{cases}.$$

  □

- $f(x) = x^p$ on $\mathbb{R}_{++}$ for $p > 1$.

  Proof. Note that for fixed $y > 0$, $xy - x^p$ has a maximum at $x = (y/p)^{\frac{1}{p-1}}$ (find the critical point and note the concavity of the function). It follows that

  $$f^*(y) = x(y - x^{p-1}) = y(y/p)^{\frac{1}{p-1}}(1 - 1/p) = \frac{(y/p)^q}{pq} = (1-p)(y/p)^q, \qquad \frac{1}{p} + \frac{1}{q} = 1.$$

  □

Exercise 2.33 (3.37)**.** Show that the conjugate of $f(X) = \operatorname{tr}(X^{-1})$ with $\operatorname{dom} f = \mathcal{S}^n_{++}$ is given by

$$f^*(Y) = -2\operatorname{tr}(-Y)^{1/2}, \qquad \operatorname{dom} f^* = -S^n_+.$$

Proof. Note the definition of the conjugate in this case is given by

$$f^*(Y) = \sup_{X \in \operatorname{dom} f} (\langle Y, X \rangle - f(X)).$$

in other words, we take the induced inner product when defining the conjugate. d The rest of the problem follows from simply taking the derivative. □

Exercise 2.34 (3.38)**.** Let $f : \mathbb{R} \to \mathbb{R}$ be increasing with $f(0) = 0$ and let g be its inverse. Define

$$F(x) = \int_0^x f(a)\, da, \qquad G(y) = \int_0^y g(a)\, da.$$

Show that F and G are conjugates. Give a simple graphical interpretation of Young's inequality,

$$xy \leq F(x) + G(y).$$

Proof. Draw the graphs of F, G to see the inequality. The equality case exactly shows that F and G are conjugates. □

Exercise 2.35 (3.39)**.** In this problem, we derive some properties of conjugate functions.

- Define $g(x) = f(x) + c^\intercal x + d$, where f is convex. Express g^* in terms of f^*.

  Proof. Note that

  $$g^*(y) = \sup_{x \in \operatorname{dom} g} (y^\intercal x - c^\intercal x - d - f(x)) = \sup((y-c)^\intercal x - f(x)) - d = f^*(y-c) - d.$$

  □

- Let $f(x, z)$ be convex in (x, z) and define $g(x) = \inf_z f(x, z)$. Find g^*.

  Proof. We have

  $$g^*(y) = \sup_x (y^\intercal x - \inf_z f(x, z)) = \sup_{x,z} (y^\intercal x - f(x, z)) = f^*(y, 0).$$

  □

Exercise 2.36 (Conjugate of conjugate)**.** Show that a conjugate of the conjugate of a closed convex function is itself: $f = f^{**}$ if f is closed and convex.

Proof. Note that $f^{**} \leq f$ - this follows from Fenchel's Inequality:

$$f(x) \geq x^\intercal y - f^*(y)$$

and taking the supremum over y gives $f \geq f^{**}$.

Let f be closed and convex. Suppose $f^{**} < f$. Take $(x, f^{**}(x)) \notin \operatorname{epi} f$. Since $\operatorname{epi} f$ is a closed convex set, there exists a strict separating hyperplane (a, b) with $a \neq 0$ so that

$$a^\intercal y + bt < c, \quad (y, t) \in \operatorname{epi} f$$

while

$$a^\intercal x + bf^{**}(x) > c.$$

Note that in particular $b \leq 0$, otherwise $t \to \infty$ gives a contradiction. Subtracting the two inequalities gives

$$a^\intercal(y - x) + b(t - f^{**}(x)) < 0$$

or equivalently

$$z^\intercal y - z^\intercal x - t + f^{**}(x) < 0$$

where we define $z = (a/-b)$. Taking the supremum over $y, t \in \operatorname{epi} f$ gives

$$f^*(z) + f^{**}(x) < z^\intercal x,$$

but this contradicts Fenchel's inequality.

If $b = 0$, then if $a_0 \in \operatorname{dom} f^*$ we claim that for sufficiently small $\epsilon > 0$, the pair $(a + \epsilon a_0, -\epsilon)$ defines a strict separating hyperplane:

$$\begin{aligned}(a + \epsilon a_0)^\intercal(y - x) - \epsilon(t - f^{**}(x)) &= a^\intercal(y - x) + \epsilon(a_0^\intercal(y - x) - t + f^{**}(x)) \\ &\leq a^\intercal(y - x) + \epsilon(a_0^\intercal(y - x) - f(y) + f^{**}(x)) \\ &\leq a^\intercal(y - x) + \epsilon(f^*(a_0) + f^{**}(x) - a_0^\intercal x).\end{aligned}$$

Since $a^\intercal(y - x) < 0$, we can take $\epsilon > 0$ sufficiently small so that we have

$$a^\intercal(y - x) + \epsilon(f^*(a_0) + f^{**}(x) - a_0^\intercal x) < 0.$$

Hence, we can apply the same argument as with $b < 0$.

□

Exercise 2.37 (3.40)**.** Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is convex and twice continuous differentiable. Suppose x and y are related by $y = \nabla f(x)$ and that $\nabla^2 f(x) \succ 0$.

- Show that $\nabla f^*(y) = x$.
- Show that $\nabla^2 f^*(y) = \nabla^2 f(x)^{-1}$.

Proof. Recall the implicit function theorem: Suppose $F \in C^1(\mathbb{R}^n \times \mathbb{R}^m; \mathbb{R})$ satisfies $F(\overline{u}, \overline{v}) = 0$ and $D_v F(u, v)$ is nonsingular in a neighborhood of $(\overline{u}, \overline{v})$. Then, there exists $\varphi \in C^1(\mathbb{R}^n; \mathbb{R}^m)$ that satisfies $\overline{v} = \varphi(\overline{u})$ and

$$F(u, \varphi(u)) = 0$$

in a neighborhood of $\overline{u}$.

Define $F(x,y) = \nabla f(x) - y$. Note that the conditions of the implicit function theorem are satisfied. Hence, there exists a function $\varphi \in C^(\mathbb{R}^n)$ so that $\overline{x} = \varphi(\overline{y})$ and $F(\varphi(y), y) = 0$ in a neighborhood of $(\overline{x}, \overline{y})$.

Now, note that

$$\nabla(\nabla f(g(y))) = \nabla^2 f(g(y))Dg(y) = I,$$

which implies that $Dg(y) = (\nabla^2 f(g(y)))^{-1}$ in a neighborhood of $\overline{y}$. It follows that $\nabla^2 f^*(\overline{y}) = (\nabla^2 f(x))^{-1}$.

Now, since $\nabla^2 f(\overline{x}) \succ 0$, it follows that $x = g(y)$ is a unique maximizer of the conjugate function, so

$$f^*(y) = y^\intercal g(y) - f(g(y)).$$

Taking the derivative gives

$$\begin{aligned}\nabla f^*(y) &= \nabla(y^\intercal g(y)) - \nabla f(g(y)) \\ &= g(y)\nabla y^\intercal + y^\intercal \nabla g(y) - \nabla f(g(y)) \\ &= g(y) + y^\intercal \nabla g(y) - y^\intercal \nabla g(y) \\ &= g(y).\end{aligned}$$

It follows that $\nabla f^*(\overline{y}) = g(\overline{y}) = \overline{x}$. □

Remark 2.38. I also remember seeing this implicit theorem method in the context of several complex variables in a lecture on the Legendre transform by Prof. Maciej Zworski (and of course in several other areas).

Exercise 2.39 (3.42)**.** Let $f_0, \ldots, f_n : \mathbb{R} \to \mathbb{R}$ be continuous functions. For $x \in \mathbb{R}^n$, we say that $f = \sum_{i=1}^n x_i f_i$ approximates f_0 with tolerance $\epsilon > 0$ over the interval $[0, T]$ if $|f(t) - f_0(t)| \leq \epsilon$ for $0 \leq t \leq T$. Now we choose a fixed tolerance $\epsilon > 0$ and define the approximation width as the largest T such that f approximates f_0 over the interval $[0, T]$:

$$W(x) = \sup\{T : \left|\sum_{i=1}^n x_i f_i(t) - f_0(t)\right| \leq \epsilon \text{ for } 0 \leq t \leq T\}.$$

Show that W is quasiconcave.

Proof. Note that $W(x) \geq \alpha$ if and only if for all $0 \leq t \leq \alpha$,

$$\left|\sum_{i=1}^n x_i f_i(t) - f_0(t)\right| \leq \epsilon.$$

Notice that for each $t \in [0, \alpha)$ this corresponds to the intersection of two half-spaces, so we have the intersection of infinitely many half-spaces which is convex. □

Exercise 2.40 (3.43)**.** Prove the first-order condition for quasiconvexity: a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, with $\operatorname{dom} f$ convex, is quasiconvex if and only if for all $x, y \in \operatorname{dom} f$,

$$f(y) \leq f(x) \implies \nabla f(x)^\intercal (y - x) \leq 0.$$

Proof. We first show the result for $n = 1$. We can use the characterization of quasiconvex functions in $\mathbb{R}$:

- if f is nondecreasing, then $f'(x) \geq 0$ so $f(y) \leq f(x)$ implies that $y \leq x$ so it follows that $f'(x)(y - x) \leq 0$.

- if f is nonincreasing, then $f'(x) \leq 0$ so $f(y) \leq f(x)$ implies that $x \leq y$ so it follows that $f'(x)(y-x) \leq 0$.
- if x, y are both on the nondecreasing side, or both on the nonincreasing side, the result follows the above analysis. Otherwise, $f'(x)$ and $y - x$ have opposite signs which proves the result.

Alternate proof of sufficiency: Suppose $f(x) \geq f(y)$. Then, $f(x + t(y - x)) \leq f(x)$ for all $0 < t \leq 1$. It follows that

$$\frac{f(x + t(y - x)) - f(x)}{t} \to f'(x)(y - x) \leq 0.$$

Proof of necessity: Suppose $f(y) \leq f(x)$. By assumption, this implies that $\nabla f(x)^\intercal (y - x) \leq 0$. Without loss of generality suppose $x \leq y$. It suffices to show that for $z \in [x, y]$, we have

$$f(z) \leq f(x).$$

Suppose to the contrary that there exists $z_0 \in [x, y]$ with $f(z_0) > f(x)$. It follows that we can also find $z_1 \in [x, y]$ with $f(z_1) > f(x)$ with $f'(z_1) < 0$. But this contradicts the assumption that $f(z_1) > f(x)$ implies $f'(z_1)(x - z) \leq 0$.

Now, define $g(t) = f(x+tv)$ for $x \in \operatorname{dom} f$ and v arbitrary. If f is quasiconvex, then so is g. Conversely, if $g(s) \leq g(t)$ implies that $g'(t)(s-t) \leq 0$, we have $f(x+sv) \leq f(x+tv)$, and it suffices to show that $(t - s)\nabla f(x + tv)^\intercal v \leq 0$. But this follows immediately from the fact that $g'(t) = \nabla f(x + tv)^\intercal v$. □

3. Convex Optimization

3.1. Notation and Definitions

Definition 3.1 (Standard Form)**.** We will use the notation

$$\begin{aligned} &\operatorname{argmin} f_0(x) \\ &\quad \text{s.t. } f_i(x) \leq 0, \quad 1 \leq i \leq m \\ &\qquad h_i(x) = 0, \quad 1 \leq l \leq p \end{aligned}$$

to describe the problem of finding x that minimizes $f_0(x)$ among all x satisfying the conditions. We call $x \in \mathbb{R}^n$ the **optimization variable**, $f_0 : \mathbb{R}^n \to \mathbb{R}$ the **objective function**, $f_i(x) \leq 0$ the **inequality constraints**, and $h_i(x) = 0$, the **equality constraints**.

The domain of the optimization problem is defined as

$$\mathcal{D} = \bigcap_{i=1}^{m} \operatorname{dom} f_i \cap \bigcap_{i=1}^{m} \operatorname{dom} h_i.$$

Definition 3.2 (Optimal Value)**.** The optimal value p^* of the problem (given the above notation) is defined as

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

Definition 3.3 (Optimal Point)**.** We say that x^* is an optimal point if x^* is feasible and $f_0(x^*) = p^*$.

Definition 3.4 (ε-suboptimal)**.** A feasible point x with $f_0(x) \leq p^* + \varepsilon$ is called ε-suboptimal.

Definition 3.5 (Locally Optimal)**.** A feasible point x is locally optimal if there is an $R > 0$ such that

$$\begin{aligned} f_0(x) = \inf\{ & f_0(z) : f_i(z) \leq 1 \leq i \leq m \\ & h_i(z) = 0, 1 \leq i \leq p, \|z - x\|_2 \leq R\}. \end{aligned}$$

3.2. Equivalent problems

We call two problems equivalent if from a solution of one, a solution of the other is readily found, and vice versa.

> **Remark 3.6.** A more formal definition of equivalence is possible to give, but is slightly complicated and unnecessary for our purposes.

Ways we can form equivalent optimization problems are as follows:

- Change of variables: Suppose $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ is one-to-one, with image covering the problem domain $\mathcal{D}$, i.e., $\varphi(\operatorname{dom} \varphi) \supset \mathcal{D}$. We define functions $\tilde{f}_i$ and $\tilde{h}_i$ as

$$\tilde{f}_i(z) = f_i(\varphi(z)), \qquad \tilde{h}_i(z) = h_i(\varphi(z)).$$

- Transformation of objective and constraints: Suppose $\psi_0 : \mathbb{R} \to \mathbb{R}$ is monotone increasing, $\psi_1, \ldots, \psi_m : \mathbb{R} \to \mathbb{R}$ satisfy $\psi_i(u) \leq 0$ if and only if $u \leq 0$, and $\psi_{m+1}, \ldots, \psi_{m+p} : \mathbb{R} \to \mathbb{R}$ satisfy $\psi_i(u) = 0$ if and only if $u = 0$. We define $\tilde{f}_i$ and $\tilde{h}_i$ as the compositions:

$$\tilde{f}_i = \psi_i \circ f_i, \qquad \tilde{h}_i = \psi_{m+i} \circ h_i.$$

- Slack variables: Note that $f_i(x) \leq 0$ if and only if there is $s_i \geq 0$ that satisfies $f_i(x) + s_i = 0$. Hence, we can transform inequality constrainsts to a non-negativity and an equality constraint.

- Eliminating equality constrainsts: If we can explicitly parameterize the solutions of the equality constrainsts $h_i(x) = 0$ using a parameter z, then we can eliminate them as follows: Suppose $\varphi : \mathbb{R}^k \to \mathbb{R}^n$ is such that x satisfies $h_i(x) = 0$ if and only if there is some $z \in \mathbb{R}^k$ with $x = \varphi(z)$. Then, the optimization problem

$$\begin{aligned} \text{argmin} \quad & \tilde{f}_0(z) = f_0(\varphi(z)) \\ \text{s.t.} \quad & \tilde{f}_i(z) = f_i(\varphi(z)) \leq 0 \end{aligned}$$

  is equivalent to the standard form problem.

- Eliminating linear equality constraints: Suppose all the equality constraints are linear: $Ax = b$. If $Ax = b$ is inconsistent ($b \notin R(A)$), then the original problem is infeasible. Otherwise, let x_0 denote any solution of the equality constraints. Let $F \in \mathbb{R}^{n \times k}$ be any matrix with $\mathcal{R}(F) = \mathcal{N}(A)$, so that the general solution of $Ax = b$ is given by $Fz + x_0$ for $z \in \mathbb{R}^k$ (we can choose F to be full rank so that $k = n - \text{rank } A$). Then, we can set $\varphi(z) = Fz + x_0$, and repeat as above.

- Introducing equality constrainsts: instead of the general case, we present a useful example. Consider the problem:

$$\begin{aligned} &\text{argmin} f_0(A_0 x + b_0) \\ &\text{s.t. } f_i(A_i x + b_i) \leq 0, \quad 1 \leq i \leq m \\ &\qquad h_i(x) = 0, \quad 1 \leq l \leq p \end{aligned}$$

  where $x \in \mathbb{R}^n, A_i \in \mathbb{R}^{k_i \times n}$ and $f_i : \mathbb{R}^{k_i} \to \mathbb{R}$. It is convenient to intrudoce $y_i = A_i x + b_i$, to form the equivalent problem

$$\begin{aligned} &\text{argmin} f_0(y_0)) \\ &\text{s.t. } f_i(y_i) \leq 0, \quad 1 \leq i \leq m \\ &\qquad y_i = A_i x + b_i, i = 0, \ldots, m \\ &\qquad h_i(x) = 0, \quad 1 \leq l \leq p \end{aligned}$$

3.3. Convex Optimization

Definition 3.7 (Convex Optimization Problem)**.** A convex optimization problem is one of the form

$$\begin{aligned} &\text{argmin} f_0(x) \\ &\text{s.t. } f_i(x) \leq 0, \quad 1 \leq i \leq m \\ &\qquad a_i^\intercal x = b_i, \quad 1 \leq l \leq p \end{aligned}$$

where $f_0, \ldots, f_m$ are convex.

Note that the domain $\mathcal{D}$ is convex as the intersection of the convex level sets. If f_0 is quasiconvex, we say the problem is a quasiconvex optimization problem.

One of the most important properties of convex optimization solutions is that any locally optimzal point is also globally optimal.

Theorem 3.8 (First-order Optimality Criterion)
Suppose $f_0 \in C^1$, so that for all $x, y \in \operatorname{dom} f_0$,

$$f_0(y) \geq f_0(x) + \nabla f_0(x)^\intercal (y - x).$$

Let X denote the feasible set. Then x is optimal if and only if $x \in X$ and

$$\nabla f_0(x)^\intercal (y - x) \geq 0, \qquad \forall y \in X.$$

We will see how this theorem presents itself in the context of various convex optimization problems:

- Unconstrained problems: We claim the problem reduces to the well known necessary and sufficient condition $\nabla f_0(x) = 0$. Suppose x is optimzal. For all y sufficiently close to x, they are feasible since $\operatorname{dom} f_0$ is open. Take $y = x - t\nabla f_0(x)$. For t small and positive, y is feasible, so

  $$\nabla f_0(x)^\intercal (y - x) = -t\|\nabla f_0(x)\|_2^2 \geq 0,$$

  which implies that $\nabla f_0(x) = 0$.

- Problems with equality constraints only: Suppose we only have the constraint $Ax = b$, and the feasible set is nonempty. The optimality condition is $\nabla f_0(x)^\intercal (y - x) \geq 0$ for all y satisfying $Ay = b$. Since x is feasible, we have $y = x + v$ for some $v \in \mathcal{N}(A)$, so we have

  $$\nabla f_0(x)^\intercal v \geq 0, \qquad \forall v \in \mathcal{N}(A).$$

  But if a linear function is non-negative on a subspace, it must be zero on the subspace (since it is closed under scalar multiplication), so we have

  $$\nabla f_0(x) \perp \mathcal{N}(A)$$

  or equivalently $\nabla f_0(x) \in \mathcal{R}(A^\intercal)$: there exists $\eta \in \mathbb{R}^p$ such that

  $$\nabla f_0(x) + A^\intercal \eta = 0.$$

  The is the classical Lagrange multiplier optimality condition, which we will derive in the next chapter.

- Minimization over the nonnegative orthant: Suppose we have $\min f_0(x)$ over $x \succeq 0$. Note that the first order condition is of the form $\nabla f_0(x)^\intercal (y - x) \geq 0$ for $x, y \succeq 0$, so it is easy to see that $\nabla f_0(x)^\intercal x = 0$ exactly. But this is the sum of non-negative numbers, so they must each be 0, or in other words:

  $$x \succeq 0, \qquad \nabla f_0(x) \succeq 0, \qquad x_i(\nabla f_0(x))_i = 0.$$

  The last condition is called complementarity, since it means that the sparcity patterns of x and $\nabla f_0(x)$ complement each other. We will encounter this again in the next chapter.

Theorem 3.9 (First-order optimality for quasiconvex problems)
If $f_0 \in C^1$, then x is optimal if

$$x \in X, \qquad \nabla f_0(x)^\intercal (y - x) > 0 \qquad \forall y \in X \setminus \{x\}.$$

Some important distinctions between this condition and the one for convex optimization problems:

- This is only a sufficient condition for optimality - it is easy to show that this is not necessary for any optimal point.
- The condition requires ∇f_0 to be nonzero.

A general approach to solving quasiconvex optimization problems relies on the representation of the sublevel sets as a family of convex functions. Let $\varphi_t : \mathbb{R}^n \to \mathbb{R}$, $t \in \mathbb{R}$ a family of convex functions that satisfy

$$f_0(x) \leq t \Leftrightarrow \varphi_t(x) \leq 0,$$

and for each x, $\varphi_t(x)$ is nonincreasing: $\varphi_s(x) \leq \varphi_t(x)$ for $s \geq t$. Let p^* denote the optimal value of the quasiconvex optimization problem. If the problem

$$\begin{array}{ll} \text{find} & x \\ \text{s.t.} & \varphi_t(x) \leq 0 \\ & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

is feasible, then $p^* \leq t$. Conversely, if the problem is not feasible, then $p^* \geq t$. This allows us to essentially binary search on p^* via bisection.

Remark 3.10. This chapter mainly introduced notation without much interesting theory. Though, the first-order condition had interesting corollaries. Problems are mostly simple examples. Onward to chapter 5, which is arguably the most important one.

4. Duality

4.1. The Lagrangian

Consider the standard form optimization problem

$$\begin{aligned} &\operatorname{argmin} f_0(x) \\ &\text{s.t. } f_i(x) \leq 0, \quad 1 \leq i \leq m \\ &\quad\; h_i(x) = 0, \quad 1 \leq l \leq p \end{aligned}$$

We define the Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ associated with the problem as

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

where $\operatorname{dom} L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$. We refer to λ_i as the Lagrange multiplier associated with the ith inequality constraint, and analogously with ν_i. The vectors λ, ν are called the dual variables, or Lagrange multiplier vectors associated with the problem.

Definition 4.1 (Lagrange Dual)**.** We define the dual function $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ as follows: for $\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p$

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

Note that the dual function is concave as the pointwise infimum of a family of affine functions of λ, ν even when the original problem is not convex.

Theorem 4.2 (Lower bounds on optimal value)

Suppose p^* is the optimal value of the standard form problem. For any $\lambda \succeq 0$ and any ν we have

$$g(\lambda, \nu) \leq p^*.$$

Proof. Suppose $\tilde{x}$ is feasible ($f_i(\tilde{x}) \leq 0, h_i(\tilde{x}) = 0$) and $\lambda \succeq 0$. Then,

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

This implies that

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x}).$$

Since $\tilde{x}$ is an arbitrary feasible point, the original statement follows. □

Example 4.3 (Linear approximation interpretation)
Consider the equivalent unconstrained problem

$$\min \qquad f_0(x) + \sum_{i=1}^{m} I_-(f_i(x)) + \sum_{i=1}^{p} I_0(h_i(x)),$$

where I_0 is infinity for $u \neq 0$ and 0 otherwise and $I_-(u)$ is infinity for $u > 0$ and 0 otherwise.

Suppose we replace $I_-(u)$ with $\lambda_i u$ for some $\lambda_i \geq 0$ and $I_0(u)$ with $\nu_i u$. We can interpret these as a soft approximation of the indicator functions. Although this is a poor approximation, it is at least an underestimator of the original function, which immediately yields the lower bound property.

Now, we present some basic examples.

- Least-squares solution of linear equations: We consider the problem argmin $x^\intercal x$ subject to $Ax = b$. The lagrangian is $L(x, \nu) = x^\intercal x + \nu^\intercal(Ax - b)$. Since $L(x, \nu)$ is convex and quadratic, we can find the minimizing x via the optimilaity condition:

  $$\nabla_x L(x, \nu) = 2x + A^\intercal v = 0,$$

  which gives $x = -(1/2)A^\intercal \nu$. Therefore, the dual function is given by

  $$g(\nu) = L(-(1/2)A^\intercal \nu, \nu) = -(1/4)\nu^\intercal A A^\intercal \nu - b^\intercal \nu,$$

  which is a concave quadratic function.

- Standard form LP: Consider the standard form problem

  $$\begin{aligned} \text{argmin} \quad & c^\intercal x \\ \text{s.t.} \quad & Ax = b \\ & x \succeq 0 \end{aligned}$$

  The lagrangian is given by

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

  The dual function is given by

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

  which is easily determined analytically since a linear function is bounded below only when it is identically zero. So we obtain $g(\lambda, \nu) = -\infty$ except when $c + A^\intercal \nu - \lambda = 0$, in which case it is $-b^\intercal \nu$.

- Two-way partitioning problem: consider the nonconvex problem

  $$\begin{aligned} \text{argmin} \quad & x^\intercal W x \\ \text{s.t.} \quad & x_i^2 = 1 \end{aligned}$$

for $W \in \mathcal{S}^n$. Notice that we restrict x_i to $1, -1$. Since the feasible set is finite, we can simply check the objective value of each feasible point - but this grows exponentially with n.

Note that W_{ij} can be interpreted as the cost of x_i, x_j having the same sign and $-W_{ij}$ can be interpreted as the cost of x_i, x_j has opposite signs. Now, we derive the dual. The Lagrangian is given by

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

Now,

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

This yields lower bounds on the optimal value of the problem. For example, taking $\nu = -\lambda_{min}(W)\mathbf{1}$ which is dual-feasible, we obtain a bound on the optimal value p^*, given by

$$p^* \geq -1^\intercal \nu = n\lambda_{min}(W).$$

4.2. Lagrange dual and conjugate function

Consider the problem:

$$\begin{aligned} \text{argmin} \quad & f(x) \\ \text{s.t.} \quad & x = 0 \end{aligned}$$

The dual function is given by

$$g(\nu) = \inf_x (f(x) + \nu^\intercal x) = -\sup_x ((-\nu)^\intercal x - f(x)) = -f^*(-\nu).$$

More generally, consider an optimization problem of the form:

$$\begin{aligned} \text{argmin} \quad & f_0(x) \\ \text{s.t.} \quad & Ax \preceq b \\ & Cx = d. \end{aligned}$$

We have

$$\begin{aligned} g(\lambda, \nu) &= \inf(f_0(x) + \lambda^\intercal (Ax - b) + \nu^\intercal (Cx - d)) \\ &= -b^\intercal \lambda - d^\intercal \nu + \inf_x (f_0(x) + (A^\intercal \lambda + C^\intercal \nu)^\intercal x) \\ &= -b^\intercal \lambda - d^\intercal \nu - f_0^*(-A^\intercal \lambda - C^\intercal \nu). \end{aligned}$$

Note that

$$\text{dom}\, g = \{(\lambda, \nu) | -A^\intercal \lambda - C^\intercal \nu \in \text{dom}\, f_0^*\}.$$

Now, we present some examples:

- Equality constrained norm minimization: Consider the problem argmin $\|x\|$ subject to $Ax = b$. Note that the conjugate function of $f_0 = \|\cdot\|$ is given by

$$f_0^*(y) = \begin{cases} 0, & \|y\|_* \leq 1, \\ \infty, & \text{otherwise} \end{cases}$$

  Using the above result, we obtain

$$g(\nu) = -b^\intercal \nu - f_0^*(-A^\intercal \nu) = \begin{cases} -b^\intercal \nu, & \|A^\intercal \nu\|_* \leq 1 \\ -\infty, & \text{otherwise} \end{cases}$$

- Entropy maximization: Consider the problem argmin $f_0(x) = \sum_{i=1}^n x_i \log x_i$ subject to $Ax \preceq b, \mathbf{1}^\intercal x = 1$, where $\operatorname{dom} f_0 = \mathbb{R}^n_{++}$. The conjugate function is given by

$$f_0^*(y) = \sum_{i=1}^n e^{y_i - 1},$$

  with $\operatorname{dom} f_0^* = \mathbb{R}^n$. It follows that t he dual function is given by

$$g(\lambda, \nu) = -b^\intercal \lambda - \nu - e^{-\nu - 1} \sum_{i=1}^n e^{-a_i^\intercal \lambda},$$

  where a_i is the ith column of A.

4.3. Lagrange dual problem

We start with a natural question: what is the best lower bound that can be obtained from the Lagrange dual function? The leads us to the *Lagrange dual problem*:

$$\begin{aligned} \text{argmax} \quad & g(\lambda, \nu) \\ & \lambda \succeq 0. \end{aligned}$$

The original standard form problem is called *primal* in this context. Note that the Lagrange dual problem is a convex optimization problem, whether or not the primal problem is convex.

Example 4.4 (Making dual constraints explicit)

Note that it is not uncommon for $\dim \operatorname{dom} g < m + p$. In this case, we identify the affine hull of $\operatorname{dom} g$, defining it as a set of linear constraints. This means that we can identify the equality constraints that are hidden in the original Lagrange dual problem.

Recall the Lagrange dual function for the standard form LP is given by $g(\lambda, \nu) = -b^\intercal \nu$ for $A^\intercal \nu - \lambda + c = 0$, and $-\infty$ otherwise. Then, notice that we can form an equivalent dual problem as

$$\begin{aligned} \text{argmax} \quad & -b^\intercal \nu \\ & \lambda = A^\intercal \nu + c \succeq 0 \end{aligned}$$

Similarly, if we have argmin $c^\intercal x$ subject to $Ax \preceq b$, the dual function is given by $g(\lambda) = -b^\intercal \lambda$ for $A^\intercal \lambda + c = 0$ and $-\infty$ otherwise. The Lagrange dual of the LP is given by

$$\begin{aligned} \text{argmax} \quad & -b^\intercal \lambda \\ & A^\intercal \lambda + c = 0 \\ & \lambda \succeq 0 \end{aligned}$$

4.4. Weak and Strong Duality

Theorem 4.5 (Weak Duality)

Let d^* denote the optimal value of the Lagrange dual problem, which is the best lower bound on p^*, the optimal value of the primal. Then, we have

$$d^* \leq p^*.$$

In particular, note that this inequality holds even when d^*, p^* are infinite. The difference $p^* - d^*$ is referred to as the duality gap.

We say that *strong duality* holds if the duality gap is 0. It is important to note that strong duality does not hold in general. If the problem is convex, it usually holds, but not always.

4.4.1. Slater's condition

Consider the problem given by

$$\begin{aligned} \text{argmin} \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b, \end{aligned}$$

where $f_0, \ldots, f_m$ are convex. We have the following important condition for strong duality.

Theorem 4.6 (Slater's Condition)
If there exists $x \in \operatorname{relint} \mathcal{D}$ such that

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

then we have strong duality: $d^* = p^*$. Such a point satisfying the condition is called *strictly feasible*.

If we have affine constraints, then Slater's condition can be relaxed to the following: if the first k constraint functions $f_1, \dots, f_k$ are affine, then if there exists $x \in \operatorname{relint} \mathcal{D}$ such that

$$f_i(x) \leq 0, \quad i = 1, \dots, k, \qquad f_i(x) < 0, \quad i = k+1, \dots, m, \qquad Ax = b.$$

Remark 4.7. Slater's condition (and the refinement) also implies that the dual optimal value is attained when $d^* > -\infty$ - that is there exists (λ^*, ν^*) with $g(\lambda^*, \nu^*) = d^* = p^*$.

Before showing the proof, we present some examples.

- Least-squares solution of linear equations: Recall the problem

$$\begin{aligned} \text{argmin} \quad & x^\intercal x \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

with the associated dual problem

$$\text{argmax} \quad -(1/4)\nu^\intercal A A^\intercal \nu - b^\intercal \nu,$$

which is an unconstrained concave quadratic maximization problem.

Slater's condition reduces to feasibility, so $p^* = d^*$ provided that $b \in \mathcal{R}(A)$. In particular, even when $b \notin \mathcal{R}(A)$, there is z with $A^\intercal z = 0$, $b^\intercal z \neq 0$. It follows that the dual function is unbounded above alon $\{tz : t \in \mathbb{R}\}$ so $d^* = \infty$.

- Lagrange dual of LPs: Note that the weaker form of Slater's conditiond implies that strong duality holds for any LP provided that the primal problem is feasible. We can also apply the result to the dual to show that strong duality holds for LPs if the dual is feasible. Hence, the only case where strong duality fails is if both the primal and dual are infeasible. We will show an example of this in a future exercise.

- Entropy maximization: Consider the problem

$$\begin{aligned} \text{argmin} \quad & \sum_{i=1}^n x_i \log x_i \\ \text{s.t.} \quad & Ax \preceq b \\ & \mathbf{1}^\intercal x = 1 \end{aligned}$$

with $\mathcal{D} = \mathbb{R}^n_+$. We derived the dual problem before as

$$\text{argmax} \quad -b^\intercal - \nu - e^{-\nu - 1} \sum_{i=1}^n e^{-a_i^\intercal \lambda}, \qquad \lambda \succeq 0.$$

Slater's condition says that the duality gap is zero if there exists an $x \succ 0$ with $Ax \preceq b$ and $\mathbf{1}^\intercal x = 1$.

- Minimum volume covering ellipsoid: the problem is given by

$$\begin{array}{ll}\text{argmin} & \log\det X^{-1} \\ & a_i^\intercal X a_i \leq 1, \quad i = 1, \ldots, m,\end{array}$$

with domain $\mathcal{D} = \mathcal{S}^n_{++}$. The dual problem can be expressed as

$$\begin{array}{ll}\text{argmax} & \log\det\left(\sum_{i=1}^m \lambda_i a_i a_i^\intercal\right) - \mathbf{1}^\intercal\lambda + n \\ & \lambda \succeq 0\end{array}$$

where we take $\log\det X = -\infty$ if $X \not\succ 0$.

Slater's condition for the problem is that there exists $X \in \mathcal{S}^n_{++}$ with $a_i^\intercal X a_i \leq 1$ for $i = 1, \ldots, m$. This is always satisfied, so strong duality is always obtained.

Example 4.8 (Nonconvex problem with strong duality)

A nonconvex quadratic problem with strong duality: On rare occasions, strong duality is obtained for a nonconvex problem. Consider the problem of minimizing a nonconvex quadratic function over the unit ball:

$$\begin{array}{ll}\text{argmin} & x^\intercal A x + 2b^\intercal x \\ & x^\intercal x \leq 1,\end{array}$$

where $A \in \mathcal{S}^n$, $A \not\succeq 0$, $b \in \mathbb{R}^n$. This is somtimes called the *trust region problem*.

The Lagrangian is given by

$$L(x, \lambda) = x^\intercal A x + 2b^\intercal x + \lambda(x^\intercal x - 1) = x^\intercal(A + \lambda I)x + 2b^\intercal x - \lambda,$$

so the dual is given by

$$g(\lambda) = \begin{cases} -b^\intercal(A + \lambda I)^\dagger b - \lambda, & A + \lambda \succeq 0, \quad b \in \mathcal{R}(A + \lambda I) \\ -\infty, & \text{otherwise} \end{cases}$$

Note that the corresponding Lagrange dual problem can be expressed as

$$\begin{array}{ll}\text{argmax} & -\sum_{i=1}^n (q_i^\intercal b)^2/(\lambda_i + \lambda) - \lambda \\ & \lambda \geq -\lambda_{\min}(A),\end{array}$$

where λ_i and q_i are the eigenvalues and corresponding orthonormal eigenvectors of A, and we interpret $(q_i^\intercal b)^2/0$ as 0 is $q_i^\intercal b = 0$ and ∞ otherwise.

Although the original problem is not convex, we always have zero optimal duality gap for this problem. In fact, strong duality holds for any optimization problem with quadratic objective and one quadratic inequality constraint, provided Slater's condition holds. This is proved using the S-procedure, but this requires more technology than we currently have.

Remark 4.9. See the images on Pg. 233-234 for a geometric interpretation of weak duality. I particularly like figures 5.3 and 5.4.

Now, we prove Slater's condition.

Proof. Consider the original primal problem with $f_0, \dots, f_m$ convex, and assume Slater's condition holds: There exists $\tilde{x} \in \operatorname{relint} \mathcal{D}$ with $f_i(\tilde{x}) < 0$ and $A\tilde{x} = b$. To simplify the proof, we assume that $\operatorname{relint} \mathcal{D} = \mathcal{D}^\circ$, and $\operatorname{rank} A = p$. We also assume that p^* is finite.

Consider the set

$$\mathcal{A} = \mathcal{G} + (\mathbb{R}^m_+ \times \{0\} \times \mathbb{R}_+),$$

where

$$\mathcal{G} = \{(f_1(x), \dots, f_m(x), h_1(x), \dots, h_p(x), f_0(x)) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} : x \in \mathcal{D}\}.$$

It is clear that $\mathcal{A}$ is convex if the underlying problem is convex. We define a second convex set $\mathcal{B}$ as

$$\mathcal{B} = \{(0, 0, s) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} | s < p^*\}.$$

First, we claim that $\mathcal{A} \cap \mathcal{B} = \emptyset$. To prove this, suppose $(0, 0, t) \in \mathcal{A} \cap \mathcal{B}$. Since $(0, 0, t) \in \mathcal{B}$ we have $t < p^*$. Since $(0, 0, t) \in \mathcal{A}$, there exists x with $f_i(x) \leq 0, i = 1, \dots, am\, Ax - b = 0$, and $f_0(x) \leq t < p^*$, which is impossible since p^* is the optimal value of the primal.

By the separating hyperplane theorem, there exists $(\tilde{\lambda}, \tilde{\nu}, \mu) \neq 0$ and α such that we have

$$(u, v, t) \in \mathcal{A} \implies \tilde{\lambda}^\intercal u + \tilde{\nu}^\intercal v + \mu t \geq \alpha,$$
$$(u, v, t) \in \mathcal{B} \implies \tilde{\lambda}^\intercal u + \tilde{\nu}^\intercal v + \mu t \leq \alpha,$$

Note that we must have $\tilde{\lambda} \succeq 0$ and $\mu \geq 0$ in order the first inequality to hold. We must also have $\mu t \leq \alpha$ for all $t < p^*$ by the second inequality, so $\mu p^* \leq \alpha$. Therefore, for any $x \in \mathcal{D}$, we have

$$\sum_{i=1}^m \tilde{\lambda}_i f_i(x) + \tilde{\nu}^\intercal (Ax - b) + \mu f_0(x) \geq \alpha \geq \mu p^*.$$

If $\mu > 0$, then we can divide by μ to obtain that for all $x \in \mathcal{D}$,

$$L(x, \tilde{\lambda}/\mu, \tilde{\nu}/\mu) \geq p^*.$$

Now, we minimize over x to obtain $g(\lambda, \nu) \geq p^*$ (after rescaling). Since weak duality implies that $g(\lambda, \nu) \leq p^*$, we have $g(\lambda, \nu) = p^*$.

Now, suppose $\mu = 0$. For all $x \in \mathcal{D}$, we have

$$\sum_{i=1}^m \tilde{\lambda} f_i(x) + \tilde{\nu}^\intercal (Ax - b) \geq 0.$$

Applying this to $\tilde{x}$ which satisfies Slater's condition, we have

$$\sum_{i=1}^m \tilde{\lambda}_i f_i(\tilde{x}) \geq 0.$$

Since $f_i(\tilde{x}) < 0$ and $\tilde{\lambda}_i \geq 0$, we can conclude that $\tilde{\lambda} = 0$. Since $(\tilde{\lambda}, \tilde{\nu}, \mu) \neq 0$, and $\tilde{\lambda} = 0$, $\mu = 0$, we can conclude that $\tilde{\nu} \neq 0$. It follows that for all $x \in \mathcal{D}$, $\tilde{\nu}^\intercal (Ax - b) \geq 0$. But $\tilde{x}$ satisfies $\tilde{\nu}^\intercal (A\tilde{x} - b) = 0$, and since $\tilde{x} \in D^\circ$, there are points in $\tilde{\nu}^\intercal (Ax - b) < 0$ (take $x = \tilde{x} - \epsilon y$ for small $\epsilon > 0$ and $\tilde{\nu}^\intercal Ay \geq 0$) unless $A^\intercal \tilde{\nu} = 0$, but this contradicts that $\operatorname{rank} A = p$. □

4.4.2. Max-min duality

For simplicity, assume there are no equality constraints. We can easily extend the results to cover these cases.

First, note that

$$\sup_{\lambda \succeq 0} L(x, \lambda) = \sup_{\lambda \succeq 0} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right) = \begin{cases} f_0(x), & f_i(x) \leq 0 \quad i = 1, \ldots, m \\ \infty, & \text{otherwise} \end{cases}$$

Proof. Suppose x is not feasible and $f_i(x) > 0$ for some i. Then $\sup_{\lambda \succeq 0} L(x, \lambda) = \infty$ since we can choose $\lambda_j = 0$ for $j \neq i$ and $\lambda_i \to \infty$. On the otherhand, if $f_i(x) \leq 0$, $i = 1, \ldots, m$, then the optimal choice of λ is $\lambda = 0$ and $\sup_{\lambda \succeq 0} L(x, \lambda) = f_0(x)$. □

Therefore, the optimal value of the primal problem is

$$p^* = \inf_x \sup_{\lambda \succeq 0} L(x, \lambda).$$

Recall that by the definition of the dual function

$$d^* = \sup_{\lambda \succeq 0} \inf_x L(x, \lambda).$$

Therefore, weak duality can be expressed as the inequality

$$\sup_{\lambda \succeq 0} \inf_x L(x, \lambda) \leq \inf_x \sup_{\lambda \succeq 0} L(x, \lambda)$$

and strong duality is given by

$$\sup_{\lambda \succeq 0} \inf_x L(x, \lambda) = \inf_x \sup_{\lambda \succeq 0} L(x, \lambda)$$

Note that the first inequality does not depend on any properties of L by the max-min inequality: For any $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $W \subset \mathbb{R}^n, Z \subset \mathbb{R}^m$, we have

$$\sup_Z \inf_W f(w, z) \leq \inf_W \sup_Z f(w, z).$$

When equality holds, we say that f satisfies a strong max-min property or the saddle-point property. In the case of strong duality, this corresponds to the case where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is the Lagrangian of a problem with $W = \mathbb{R}^n$ and $Z = \mathbb{R}^m_+$.

Definition 4.10 (Saddle-point)**.** We refer to a pair $\tilde{w} \in W$, $\tilde{z} \in Z$ a saddle-point for f if

$$f(\tilde{w}, z) \leq f(\tilde{w}, \tilde{z}) \leq f(w, \tilde{z})$$

for all $w \in W$, $z \in Z$.

Note that if x^* and λ^* are primal and dual optimal points for a problem in which strong duality obtains, they form a saddle-point for the Lagrangian. The converse is also true: if (x, λ) is a saddle-point of the Lagrangian, then x is primal optimal, λ is dual optimal, and the duality gap is 0 (we proved this in a previous exercise). A useful

theorem to keep in mind, which is a generalization of the famous Von Neumann minimax theorem is as follows:

Theorem 4.11 (Sion's minimax theorem)

Let X be a compact convex subset of a topological vector space and Y a convex subset of a topological vector space. If f is a real-valued function on $X \times Y$, then f satisfies the strong min-max property if both of the following properties hold:

- $f(x, \cdot)$ is upper semicontinuous and quasi-concave on Y for all $x \in X$,
- $f(\cdot, y)$ is lower semicontinuous and quasi-convex on X for all $y \in R$.

4.5. Optimality conditions

4.5.1. Certificates of suboptimality

If we can find a dual feasible (λ, ν), then we establish a lower bound on the optimal value of the primal by definition. Therefore, a dual feasible point provides a *proof* or *certifiacte* that $p^* \geq g(\lambda, \nu)$. Strong duality implies the existence of arbitrarily good certificates.

This allows us to bound how suboptimal a give feasible point is: if x is primal feasible and (λ, ν) is dual feasible, then

$$f_0(x) - p^* \leq f_0(x) - g(\lambda, \nu),$$

which implies that x is ε-suboptimal where $\varepsilon := f_0(x) - g(\lambda, \nu)$. Similarly, (λ, ν) is ε-suboptimal for the dual problem.

In particular, this is useful in order to establish stopping criterion for algorithms based on the relative or absolute suboptimality gap.

4.5.2. Complementary slackness

Suppose the primal and dual optimal values are attaned and equal. Let x^* be primal optimal and (λ^*, ν^*) be dual optimal. Then,

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

But this implies all the inequalities are equalities. We can draw several interesting conclusions from this:

- x^* minimizes $L(x, \lambda^*, \nu^*)$ over x.
- $\sum_{i=1}^m \lambda_i^* f_i(x^*) = 0$ which implies that $\lambda_i^* f_i(x^*) = 0$, $i = 1, \ldots, m$. The second condition is known as *complementary slackness*. It can be restated as

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

or

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

Roughly speaking, it says that the ith Lagrange multiplier is zero unless the ith constraint is active at the optimum.

4.5.3. KKT conditions

Now, we make the assumption that $f_0, \ldots, f_m, h_1, \ldots, h_m$ are differentiable, but we still make no assumptions about convexity.

Now, let x^* and (λ^*, ν^*) be any primal and dual optimal points with zero duality gap. Since x^* minimizes $L(x, \lambda^*, \nu^*)$ over x, it follows that the gradient must vanish at x^*:

$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) = 0.$$

This implies that

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

These are called the Karush-Kuhn-Tucker (KKT) optimality conditions. Any optimization problem with differentiable objective and constraint functions for which strong duality obtains, any pair of primal and dual optimal points must satisfy the KKT conditions.

Proposition 4.12 (KKT for Convex Problems)

When the primal is convex, the KKT conditions are also sufficient for the points to be primal and dual optimal.

Proof. Let $\tilde{x}$, $(\tilde{\lambda}, \tilde{\nu})$ be the points in question. Note that the first two conditions correspond to the primal feasibility of $\tilde{x}$. Since $\tilde{\lambda}_i \geq 0$, and $L(x, \tilde{\lambda}, \tilde{\nu})$ is convex in x, the last condition says that the gradient of L vanishes at $x = \tilde{x}$, so $\tilde{x}$ minimizes $L(x, \tilde{\lambda}, \tilde{\nu})$ over x. Therefore, we can conclude that

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

This proves that we have zero duality gap, which implies the result. □

Remark 4.13. See examples 5.3, 5.4 in the text for some cases where we can use strong duality and existence of optimal dual solutions to construct optimal primal solutions if the minimizer of $L(x, \lambda^*, \nu^*)$ is unique. A typical case for this is for convex problems when Lagrangian is a strictly convex function of x.

4.6. Perturbation and sensitivity analysis

When strong duality is obtained, it is interesting to consider the sensitivity of the optimal value with respect to perturbations of the constraints. Consider the problem given by

$$\begin{aligned} \text{argmin} \quad & f_0(x) \\ & f_i(x) \leq u_i, \quad i = 1, \ldots, m \\ & h_i(x) = v_i, \quad i = 1, \ldots, p \end{aligned}$$

When $u_i > 0$, this corresponds to a relaxation of the ith constraint. Similarly, when $u_i < 0$, this corresponds to a tightening of the constraint.

We define

$$p^*(u, v) = \inf\{f_0(x) : \exists x \in \mathcal{D}, f_i(x) \leq u_i, i = 1, \ldots, m, h_i(x) = v_i, i = 1, \ldots, p\}.$$

Note that we can perturb the problem so that it becomes infeasible, *i.e.*, $p^*(u, v) = \infty$. Also note that if the original problem is convex, p^* is a convex function of u and v - this can be seen by noting that the epigraph is the closure of the set $\mathcal{A}$ that we defined previously (we will show this in a future exercise).

Theorem 4.14

Suppose that strong duality holds, and that the dual optimum is attained. Let (λ^*, ν^*) be optimal for the dual of the unperturbed problem. Then, for all u, v, we have

$$p^*(u, v) \geq p^*(0, 0) - (\lambda^*)^\intercal u - (\nu^*)^\intercal v.$$

Proof. Suppose x is a feasible point for the perturbed problem. Then, by strong duality, we have

$$\begin{aligned} p^*(0, 0) = g(\lambda^*, \nu^*) &\leq f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \\ &\leq f_0(x) + (\lambda^*)^\intercal u + (\nu^*)^\intercal v \end{aligned}$$

It follows that for any feasible x for the perturbed problem, we have

$$f_0(x) \geq p^*(0, 0) - (\lambda^*)^\intercal u - (\nu^*)^\intercal v.$$

□

Remark 4.15. Note that the inequality and the conclusions we can derive from it give a lower bound on the optimal perturbed optimal value, but no upper bound. It follows that the results we obtain are not symmetric with respect to loosening or tighening a constraint.

If $p^*(u, v)$ is differentiable at $(0, 0)$, then provided strong duality holds, we have

$$\lambda_i^* = -\frac{\partial p^*(0, 0)}{\partial u_i}, \qquad \nu_i^* = -\frac{\partial p^*(0, 0)}{\partial v_i}.$$

This provides us with a characterization of the local sensitivities of the optimal value.

Remark 4.16. Section 5.7 in the text presents several examples where equivalent reformulations of a problem can lead to very different (and potentially much more useful) dual problems. We encourage the reader of these notes to read over these examples.

4.7. Theorem of alternatives

In this section, we consider applying Lagrange duality theory to the problem of determining feasibility of a system of inequalities and equalities, *i.e.,* the standard problem with objective $f_0 = 0$.

The dual function is given by

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

Note that this is homogeneous in (λ, ν), which implies that the optimal value of the dual problem is given by

$$d^* = \begin{cases} \infty, & \lambda \succeq 0,\ g(\lambda, \nu) > 0 \text{ is feasible,} \\ 0, & \lambda \succeq 0,\ g(\lambda, \nu) > 0 \text{ is infeasible.} \end{cases}$$

By weak duality, $d^* \leq p^*$, so it follows that if the inequality system $\lambda \succeq 0$, and $g(\lambda, \nu) > 0$, then the original problem is infeasible. This is an example of weak alternatives.

Definition 4.17 (Weak alternatives)**.** Two systems of inequalities (and equalities) are called weak alternatives if at most one of the two is feasible.

Note that this holds whether or not the inequalities are convex. Moreover, the alternate inequality system is always convex.

4.7.1. Strict inequalities

We can also consdier the feasibility of the strict inequality system given by

$$f_i(x) < 0, \quad i = 1, \ldots, m \qquad h_i(x) = 0, \quad i = 1, \ldots, p.$$

The alternate inequality system is given by

$$\lambda \succeq 0, \quad \lambda \neq 0, \quad g(\lambda, \nu) \geq 0.$$

We first show directly that the two systems are weak alternatives. Suppose there exists $\tilde{x}$ with $f_i(\tilde{x}) \leq 0$, $h_i(\tilde{x}) = 0$. Then, for any $\lambda \succeq 0$, $\lambda \neq 0$, and ν,

$$\lambda_1 f_i(\tilde{x}) + \cdots + \lambda_m f_m(\tilde{x}) + \nu_1 h_1(\tilde{x}) + \cdots + \nu_p h_p(\tilde{x}) < 0.$$

It follows that

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \left(\sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right) \leq \sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) < 0.$$

Therefore, feasibility of the primal system implies there does not exist (λ, ν) satisfying the dual system.

Definition 4.18 (Strong alternatives)**.** Two systems of inequalities (and equalities) are called strong alternatives if exactly one of the two is feasible.

When the original inequality system is convex, and some type of constraint qualification holds, then the weak alternatives are actually strong alternatives. First, we prove this in the context of the strict inequality system presented earlier.

Proof. We need a technical condition: there exists $x \in \operatorname{relint} \mathcal{D}$ with $Ax = b$. This is automatically satisfied by the consistency of equality constraints when $\mathcal{D} = \mathbb{R}^n$.

Consider the related problem

$$\begin{array}{ll} \text{argmin} & s \\ & f_i(x) - s \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

The optimal value p^* is negative if and only if there exists a solution to the strict primal inequality system.

The Lagrange dual function fo the problem is

$$\inf_{x \in \mathcal{D}, s} \left(s + \sum_{i=1}^{m} \lambda_i (f_i(x) - s) + \nu^\intercal (Ax - b) \right) = \begin{cases} g(\lambda, \nu) & \mathbf{1}^\intercal \lambda = 1 \\ -\infty & \text{otherwise} \end{cases}$$

Therefore, we can express the dual problem as

$$\begin{array}{ll} \text{argmax} & g(\lambda, \nu) \\ & \lambda \succeq 0, \quad \mathbf{1}^\intercal \lambda = 1. \end{array}$$

Now, we observe that Slater's condition holds for the related problem: by hypothesis, there exists $\tilde{x} \in \operatorname{relint} \mathcal{D}$ with $A\tilde{x} = b$. Choosing any $\tilde{s} > \max_i f_i(\tilde{x})$ yields a point $(\tilde{x}, \tilde{s})$ which is strictly feasible. Therefore, $d^* = p^*$ and the optimal d^* is obtained.

Now, suppose that the strict inequality system is infeasible, which means that $p^* \geq 0$. Then, (λ^*, ν^*) which obtain the dual optimum satisfy the alternate inequality system. Similarly, if the alternate inequality system is feasible, then $d^* = p^* \geq 0$, which shows that the strict inequality system is infeasible. Therefore, the inequality systems are strong alternatives. □

The result easily generalizes for nonstrict inequality systems with the additional assumption that p^* is attained.

4.7.2. Example: Intersection of Ellipsoids

Consider m ellipsoids, described as

$$\mathcal{E}_i = \{x : f_i(x) \leq 0\}$$

with $f_i(x) = x^\intercal A_i x + 2b_i^\intercal x + c_i$, $i = 1, \ldots, m$, where $A_i \in \mathcal{S}^n_{++}$. We ask when the intersection of these ellipsoids has nonempty interior. This is equivalent to the feasibility of the set of strict quadratic inequalities:

$$f_i(x) = x^\intercal A_i x + 2b_i^\intercal x + c_i < 0, \quad i = 1, \ldots, m.$$

The dual function g is given by

$$\begin{aligned} g(\lambda) &= \inf_x \left(x^\intercal A(\lambda)x + 2b(\lambda)^\intercal x + c(\lambda)\right) \\ &= \begin{cases} -b(\lambda)^\intercal A(\lambda)^\dagger b(\lambda) + c(\lambda), & A(\lambda) \succeq 0, b(\lambda) \in \mathcal{R}(A(\lambda)) \\ -\infty, & \text{otherwise} \end{cases} \end{aligned}$$

where

$$A(\lambda) = \sum_{i=1}^m \lambda_i A_i, \quad b(\lambda) = \sum_{i=1}^m \lambda_i b_i, \quad c(\lambda) = \sum_{i=1}^m \lambda_i c_i.$$

Note that for $\lambda \succeq 0$, $\lambda \neq 0$, we have $A(\lambda) \succ 0$, so it follows that

$$g(\lambda) = -b(\lambda)^\intercal A(\lambda)^{-1} b(\lambda) + c(\lambda).$$

Therefore, the strong alternative of the system is

$$\lambda \succeq 0, \quad \lambda \neq 0, \quad -b(\lambda)^\intercal A(\lambda)^{-1} b(\lambda) + c(\lambda) \geq 0.$$

We can interpret the strong alternatives geometrically as follows: for any nonzero $\lambda \succeq 0$, the ellipsoid

$$\mathcal{E}_\lambda = \{x : x^\intercal A(\lambda)x + 2b(\lambda)^\intercal x + c(\lambda) \leq 0\}$$

contains $\mathcal{E}_1 \cap \ldots, \mathcal{E}_m$, since $f_i(x) \leq 0$ implies that $\sum_{i=1}^m \lambda_i f_i(x) \leq 0$.

Now, $\mathcal{E}_\lambda$ has empty interior if and only if

$$\inf_x (x^\intercal A(\lambda)x + 2b(\lambda)^\intercal x + c(\lambda)) = -b(\lambda)^\intercal A(\lambda)^{-1} b(\lambda) + c(\lambda) \geq 0.$$

Therefore, the alternate system implies that $\mathcal{E}_\lambda$ has empty interior.

Weak duality is obvious: if the strong alternative holds, then $\mathcal{E}_\lambda$ contains $\mathcal{E}_1 \cap \cdots \cap \mathcal{E}_m$ and has empty interior, so the intersection has empty interior. The fact that these are strong alternatives states the fact that if the intersection $\mathcal{E}_1 \cap \cdots \cap \mathcal{E}_m$ has nonempty interior, then we can construct an ellipsoid $\mathcal{E}_\lambda$ that contains the intersection and has empty interior.

Theorem 4.19 (Farkas' Lemma)

The system of inequalities

$$Ax \preceq 0, \quad c^\intercal x < 0$$

where $A \in \mathbb{R}^{m \times n}$ and $c \in \mathbb{R}^n$, and the systems of equalities and inequalities

$$A^\intercal y = c = 0, \quad y \succeq 0$$

are strong alternatives.

Remark 4.20. I might refer back to the exercises if I find some future chapters challenging or don't follow some arguments. Most of the end of chapter problems feel like basic exercises and translating definitions.

5. Approximation and fitting

5.1. Norm Approximation

The simpliest version of a norm approximation problem is of the form

$$\text{argmin} \quad \|Ax - b\|,$$

where $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$. A solution oof the problem is sometimes called an approximate solution of $Ax \approx b$ in the norm $\|\cdot\|$.

Note that norm approximation is always convex and solvable. The optimal value is zero if and only if $b \in \mathcal{R}(A)$, but the problem is more interesting otherwise. We assume without loss of generality that $m \geq n$. For $m = n$, the optimal point is $A^{-1}b$, so we can assume $m > n$.

There are many ways we can interpret the problem:

- Approximate interpretation: By expressing

  $$Ax = x_1a_1 + \ldots x_na_n,$$

  where $a_1, \ldots, a_m \in \mathbb{R}^m$ are columns of A, we see that the goal of norm approximation is to appximate b by a linear combination of columns of A, with deviation measured in the norm $\|\cdot\|$.

- Estimation interpretation: Consider the linear measurement model

  $$y = Ax + v,$$

  where $y \in \mathbb{R}^m$, $x \in \mathbb{R}^n$ is a vector that is to be estimated, and v is measurement error that is presumed to be small in norm. The most plausible guess for x is

  $$\widehat{x} = \underset{z}{\text{argmin}}\, \|Az - y\|.$$

- Geometric interpretation:Consider the subspace $\mathcal{A} = \mathcal{R}(A) \subset \mathbb{R}^m$ and a poiont $b \in \mathbb{R}^m$. A projection of b onto $\mathcal{A}$ in the norm $\|\cdot\|$ is any point in $\mathcal{A}$ that is closest to b.

The choice of norm leads to various interesting problems:

- Weighted norm approximation: If we define the W norm as $\|z\|_W = \|Wz\|$, then the problem argmin $\|Ax - b\|_W$ = argmin $\|W(Ax - b)\|$ is called a weighted norm approximation problem, where W is a weighting matrix.

- Least-squares: Note that if we take $\|\cdot\|_2$, then the problem is the famous least-squares approximation problem, which has a closed form solution.

- Chebyshev/Minimax Approximation: Take $\|\cdot\|_\infty$. The approximation problem can be cast as a linear program.

- Sum of absolute residuals: Taking $\|\cdot\|_1$, we obtain a robust estimator (for reasons that will be clear soon).

5.2. Regularization

This is a common scalarization method used to solve bi-criterion problems. One form is to minimize the weighted sum of the objectives:

$$\text{argmin} \quad \|Ax - b\| + \gamma\|x\|,$$

where $\gamma \in \mathbb{R}^+$ traces out the optimal trade-off curve as it varies. Another common method is

$$\text{argmin} \quad \|Ax - b\|_2^2 + \delta\|x\|^2$$

where $\delta > 0$ varies.

5.2.1. Tikhonov regularization

With Eucliden norms, this is known as Tikhonov Regularization:

$$\text{argmin} \quad \|Ax - b\|_2^2 + \delta\|x\|_2^2 = x^\intercal(A^\intercal A + \delta I)x - 2b^\intercal Ax + b^\intercal b.$$

This has an analytical solution $x = (A^\intercal A + \delta I)^{-1}A^\intercal b$. Note that since $A^\intercal A + \delta I \succ 0$ for any $\delta > 0$, the regularized problem requires no rank or dimension assumptions.

5.2.2. Smoothing regularization

Consider $\|Dx\|$ in place of x where D represents an approximate differentiation or second-order differentiation operation. In other words, $\|Dx\|$ represents a measure of the smoothness of x.

5.2.3. ℓ_1-norm regularization

Regularization with ℓ_1-norm can be used a heuristic for finding a sparse solution. For example, if we consider

$$\text{argmin} \quad \|Ax - b\|_2 + \gamma\|x\|_1$$

varying γ traces out an approximation of the optimal trade-off curve between $\|Ax - b\|_2$ and the sparsity of x. This can be recast and solved as an SOCP.

5.3. Reconstruction, smoothing, and de-noising

In reconstruction problems, we start with a signal $x \in \mathbb{R}^n$, where the coefficients x_i correspond to the value of some function of time, evaluated at evenly spaced points. It is usually assumed that $x_i \approx x_{i+1}$. The signal x is corrupted by an additive noise v: $x_c = x + v$. The goal is to form an estimate $\widehat{x}$ of the original signal x given x_c.

A simple formulation of the reconstruction problem is the bi-criterion problem

$$\text{argmin} \quad (\|\widehat{x} - x_c\|_2, \varphi(\widehat{x})).$$

The function $\varphi : \mathbb{R}^n \to \mathbb{R}$ is convex, and is called the regularization function or smoothing objective.

For example, the simplest reconstruction method uses the quadratic smoothing function

$$\varphi_{\text{quad}}(x) = \sum_{i=1}^{n-1}(x_{i+1} - x_i)^2 = \|Dx\|_2^2,$$

where $D \in \mathbb{R}^{(n-1)\times n}$ is the bidiagonal matrix

$$D = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -1 & 1 \end{bmatrix}$$

The solution to the reconstruction problem can be solved and efficiently computed as

$$\widehat{x} = (I + \delta D^\intercal D)^{-1} x_c.$$

Another example is total variation reconstruction, which is more useful for rapidly varying signals. This is based on the function

$$\varphi_{tv}(\widehat{x}) = \sum_{i=1}^{n-1} |\widehat{x}_{i+1} - \widehat{x}_i| = \|D\widehat{x}\|_1.$$

Like the quadratic smoothing measure, this assigns large values to rapidly varying signals. However, it assigns relatively less penalty to large values of $|x_{i+1} - x_i|$.

5.4. Robust approximation

Let $A = \overline{A} + U$ where $A \in \mathbb{R}^{m\times n}$, $\overline{A}$ is a constant matrix and U is a random matrix with mean zero. The stochastic robust approximation problem is to minimize the expectation of $\|Ax - b\|$,

$$\text{argmin} \quad \mathbb{E}\|Ax - b\|.$$

Although this is a convex optimization problem, this is usually not tractable because it is very difficult to evaluate the objective or its derivatives.

This is possible to solve when A only takes finitely many values: $\Pr(A = A_i) = p_i$, $i = 1, \dots, k$, where $A_i \in \mathbb{R}^{m\times n}$, $\mathbf{1}^\intercal p = 1$, $p \succeq 0$. In this case, we have the sum of norms problem

$$\text{argmin} \quad \sum_{i=1}^{k} p_i \|A_i x - b\|,$$

or equivalently

$$\begin{aligned} &\text{argmin} && p^\intercal t, \\ &\text{s.t.} && \|A_i x - b\| \le t_i \quad \forall i \end{aligned}$$

If the norm is the Euclidean norm, the problem is a SOCP. If it is ℓ_1, ℓ_∞, it can be expressed as an LP. Some variations on the problem are tractable, for example, the *stochastic robust least-squares problem*

$$\text{argmin} \quad \mathbb{E}\|Ax - b\|_2^2.$$

This is because the objective can be expressed as

$$\begin{aligned} \mathbb{E}\|Ax - b\|_2^2 &= \mathbb{E}(\overline{A}x - bUx)^\intercal(\overline{A}x - b + Ux) \\ &= (\overline{A}x - b)^\intercal(\overline{A}x - b) + \mathbb{E}x^\intercal U^\intercal U x \\ &= \|\overline{A}x - b\|_2^2 + x^\intercal P x, \end{aligned}$$

where $P = \mathbb{E}U^\intercal U$. Therefore, this has the form of a regularized least-squares problem:

$$\text{argmin} \quad \|\overline{A}x - b\|_2^2 + \|P^{1/2}x\|_2^2,$$

which has solution

$$x = (\overline{A}^\intercal \overline{A} + P)^{-1}\overline{A}^\intercal b.$$

This aligns with our intuition about regularization: when A is subject to variation, the vector Ax has more variation the larger x is, increasing the average value of $\|Ax - b\|_2$ by Jensen's inequality. So we balance making $\overline{A}x - b$ small with the desire to keep x small. Conversely, we can also interpret the Tikhonov regularized least-squares problem as a robust least-squares problem, taking account the variation in A.

Another interesting variation is considering a worst-case approach: Let $A \in \mathcal{A} \in \mathbb{R}^{m\times n}$ describe the uncertainty of A, which is assume is nonempty and bounded. Define the worst-case error of a candidate approximation solution $x \in \mathbb{R}^n$ as

$$e(x) = \sup_{A\in\mathcal{A}} \{\|Ax - b\|\}.$$

The tractability of such a problem is highly dependent on $\mathcal{A}$, despite the fact that it is always a convex optimiation problem.

- Finite set: if $\mathcal{A} = \{A_1, \ldots, A_k\}$, then we have the problem

  $$\text{argmin} \quad \max_{i=1,\ldots,k} \|A_i x - b\|.$$

  This problem is equivalent to the one where take the polyhedral set $\mathcal{A} = \text{conv}\{A_1, \ldots, A_k\}$, and can be cast in epigraph form. As before, if it is in Euclidean norm, this is an SOCP. If it is ℓ_1, ℓ_∞, then it can be expressed as an LP.

- Norm bound error: we have that $\mathcal{A}$ is a norm ball: $\mathcal{A} = \{\overline{A} + U : \|U\| \le a\}$. The expression for $e(x)$ can be simplified in several cases. If we take the Euclidean norm on $\mathbb{R}^n$ and teh corresponding induced norm on $\mathbb{R}^{m\times n}$, then if $\overline{A}x - b \neq 0$ and $x \neq 0$, then the supremum is attained for $U = auv^\intercal$ with

  $$u = \frac{\overline{A}x - b}{\|\overline{A}x - b\|_2}, \quad v = \frac{x}{\|x\|_2}.$$

  The corresponding worst case error is $e(x) = \|\overline{A}x - b\|_2 + a\|x\|_2$, which corresponds to a regularized norm problem.

- Uncertainty ellipsoids: consider the following characterization as an ellipsoid of possible values for each row:

  $$\mathcal{A} = \{[a_1, \ldots, a_m]^\intercal : a_i \in \mathcal{E}_i, i = 1, \ldots, m\},$$

  $$\mathcal{E}_i = \{\overline{a}_i = P_i u : \|u\|_2 \le 1\},$$

  where P_i describes the variation in a_i. We allow P_i to have a nontrivial nullspace, so that we can model the situation when the variation in a_i is restricted by a subspace. Now, note that we have

  $$\sup_{a_i\in\mathcal{E}_i} = |a_i^\intercal x - b_i| = |\overline{a}_i^\intercal x - b_i| + \|P_i^\intercal x\|_2.$$

  This result allows to solve several robust approximation problems.

5.5. Solutions to selected problems

Exercise 5.1 (6.1)**.** Show that for the log barrier penalty function, if $\|u\|_\infty < a$, then

$$\|u\|_2^2 \leq \sum_{i=1}^m \varphi(u_i) \leq \frac{\varphi(\|u\|_\infty)}{\|u\|_\infty^2}\|u\|_2^2.$$

Proof. Note that

$$\sum_{i=1}^m \varphi(u_i) = \sum_{i=1}^m \sum_{k=1}^\infty a^2 \left(\frac{u_i}{a}\right)$$

from which both inequalities immediately follow. □

Exercise 5.2 (6.4)**.** Differentiable approximation of ℓ_1-norm problem.

Proof. (a) Note that the derivative of the approximate problem at $\widehat{x}$ is zero giving

$$\sum_{i=1}^m \varphi'(\widehat{r}_i)a_i = \sum_{i=1}^m \widehat{r}_i(\widehat{r}_i^{\,2} + \varepsilon)^{-1/2}a_i = 0.$$

Now, the dual of the ℓ_1-norm problem is

$$\begin{aligned} \text{argmax} \quad & \sum_{i=1}^m b_i\lambda_i \\ & |\lambda_i| \leq 1 \\ & \sum_{i=1}^m \lambda_i a_i = 0. \end{aligned}$$

From the dual-feasibility of $\lambda_i = \widehat{r}_i(\widehat{r}_i^{\,2} + \varepsilon)^{-1/2}$, we obtain the inequality

$$p^* \geq \sum_{i=1}^m -b_i\lambda_i = \sum_{i=1}^m (a_i^\intercal \widehat{x} - b_i)\lambda_i = \sum_{i=1}^m \widehat{r}_i^{\,2}(\widehat{r}_i^{\,2} + \varepsilon)^{-1/2}.$$

□

Proof. (b) With part (a) in hand, note that

$$p^* + \sum_{i=1}^m |\widehat{r}_i| \left(1 - \frac{|\widehat{r}_i|}{(\widehat{r}_i^2 + \varepsilon)^{1/2}}\right) \geq \sum_{i=1}^m |r_i| = \|A\widehat{x} - b\|_1.$$

□

Exercise 5.3 (6.6)**.** Duals of some penalty function approximation problems.

Proof. First, note that the Lagrangian is given by

$$L(x, r, \lambda) = \sum_{i=1}^m \varphi(r_i) + \lambda^\intercal(Ax - b - r),$$

which is only bounded if $\lambda \in \mathcal{N}(A^\intercal)$, giving the dual problem:

$$g(\lambda) = \begin{cases} -b^\intercal\lambda + \sum_{i=1}^m \inf_{r_i}(\varphi(r_i) - \lambda_i r_i) & A^\intercal\lambda = 0, \\ -\infty, \text{otherwise} \end{cases}$$

But note that

$$\inf_{r_i}(\varphi(r_i) - \lambda_i r_i) = -\sup_{r_i}(\lambda_i r_i - \varphi(r_i)) = -\varphi^*(\lambda_i).$$

It follows that the general dual problem can be expressed as

$$\begin{aligned} \text{argmax} \quad & -b^\intercal \lambda - \sum_{i=1}^{m} \varphi^*(\lambda_i) \\ & A^\intercal \lambda = 0. \end{aligned}$$

□

6. Unconstrained minimization

6.1. Introduction

We finally discuss algorithms for solving optimization problems, starting with the unconstrained problem:

$$\operatorname{argmin} f(x)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is convex and twice differentiable. We assume the problem is solvable and we denote the optimal value as p^*. It is clear that a necessary and sufficient condition for x^* to be optimal is $\nabla f(x^*) = 0$, so solving the problem is equivalent to finding a solution to the gradient problem, which is a set of n equations in n variables $x_1, \ldots, x_n$. Sometimes, the problem can be solved analyticcally, but most of the time, it requires an iterative algorithm. By this, we mean an algorith that computes $x^0, x^1, \cdots \in \operatorname{dom} f$ with $f(x^k) \to p^*$ as $k \to \infty$. The algorithm is terminated when $f(x^k) - p^* \leq \varepsilon$, for some fixed tolerance, $\varepsilon > 0$.

Furthermore, we assume that the starting point $x^0 \in \operatorname{dom} f$ and the sublevel set $C_{f(x^0)}$ is closed. Such a condition is satisfied for all $x^0 \in \operatorname{dom} f$ if f is a closed.

6.2. Strong convexity

Definition 6.1 (Strong convexity)**.** A function $f \in C^2(\mathbb{R}^n; \mathbb{R})$ is strongly convex on S if there exists $m > 0$ such that $\nabla^2 f(x) \succeq mI$, for all $x \in S$.

From strong convexity, we can derive interesting consequences.

Proposition 6.2

For $x, y \in S$, if f is m-strongly convex, then we have

$$f(y) \geq f(x) + \nabla f(x)^\intercal (y - x) + \frac{m}{2}\|y - x\|_2^2.$$

Proof. By Taylor's theorem, we have

$$f(y) = f(x) + \nabla f(x)^\intercal (y - x) + \frac{1}{2}(y - x)^\intercal \nabla^2 f(z)(y - z)$$

for some z on the line segment $[x, y]$. Note that strong convexity implies that the last term is at least $(m/2)\|y - x\|_2^2$. □

Proposition 6.3

For any $x \in S$,

$$p^* \geq f(x) - \frac{\|\nabla f(x)\|_2^2}{2m}.$$

Proof. Note that the inequality from the previous proposition is a convex quadratic function of y. Setting the gradient with respect to y equal to zero, we find that $\tilde{y} = x - (1/m)\nabla f(x)$ minimizes the righthand side.

Therefore,

$$\begin{aligned} f(y) &\geq f(x) + \nabla f(x)^\intercal (y - x) + \frac{m}{2}\|y - x\|_2^2 \\ &\geq f(y) + \nabla f(x)^\intercal (\tilde{y} - x) + \frac{m}{2}\|\tilde{y} - x\|_2^2 \\ &= f(x) - \frac{\|\nabla f(x)\|_2^2}{2m}. \end{aligned}$$

Since this holds for all $y \in S$, the inequality holds. □

Remark 6.4. Note that we can also interpret the above inequality as a condition for suboptimality which generalizes the basic optimality condition:

$$\|\nabla f(x)\|_2 \leq (2m\varepsilon)^{1/2} \implies f(x) - p^* \leq \varepsilon.$$

Proposition 6.5

For any $x \in S$ and optimal point x^*, we have

$$\|x - x^*\|_2 \leq \frac{2}{m}\|\nabla f(x)\|_2.$$

Proof. Take $y = x^*$ to obtain

$$\begin{aligned} p^* = f(x^*) &\geq f(x) + \nabla f(x)^\intercal (x^* - x) + \frac{m}{2}\|x^* - x\|_2^2 \\ &\geq f(x) - \|\nabla f(x)\|_2 \|x^* - x\|_2 + \frac{m}{2}\|x^* - x\|_2^2 \end{aligned}$$

where we applied the Cauchy-Schwarz inequality. One obtains the desired result by noting the fact that $p^* \leq f(x)$ and rearranging the inequality. □

Proposition 6.6 (Upper bound on $\nabla^2 f(x)$)

There exists a constant M such that for all $x, y \in S$,

$$\nabla^2 f(x) \preceq MI,$$

$$f(y) \leq f(x) + \nabla f(x)^\intercal (y - x) + \frac{M}{2}\|y - x\|_2^2,$$

$$p^* \leq f(x) - \frac{1}{2M}\|\nabla f(x)\|_2^2.$$

Proof. The inequality from Proposition 6.2 implies that the sublevel sets contained in S are bounded, so S is bounded. Therefore, the maximum eigenvalue of $\nabla^2 f(x)$ is also bounded above on S, since it is a continuous function of x on S. From this, each of the results easily follow. □

6.2.1. Condition number of sublevel sets

From the strong convexity inequality $mI \preceq \nabla^2 f(x) \preceq MI$ for all $x \in S$, the ratio $\kappa = M/m$ is an upper bound on the condition number of $\nabla^2 f(x)$, which is the ratio of its largest and smallest eigenvalues.

Define the width of a convex set $C \subseteq \mathbb{R}^n$ in direction q, where $\|q\|_2 = 1$ as

$$W(C, q) = \sup_{z \in C} q^\intercal z - \inf_{z \in C} q^\intercal z.$$

Then, the minimum and maximum width are defined as

$$W_{\min} = \inf_{\|q\|_2 = 1} W(C, q), \quad W_{\max} = \sup_{\|q\|_2 = 1} W(C, q).$$

The condition number of the convex set C is then defined as

$$\text{cond}(C) = \frac{W_{\max}^2}{W_{\min}^2}.$$

Now, suppose f satisfies $mI \preceq \nabla^2 f(x) \preceq MI$ for all $x \in S$, we will derive a bound on $\text{cond}(C_\alpha)$, where $p^* < \alpha \leq f(x^0)$.

Note that with $x = x^*$, we have

$$p^* + (M/2)\|y - x^*\|_2^2 \geq f(y) \geq p^* + (m/2)\|y - x^*\|_2^2.$$

This implies that $B_{\text{m}} \subseteq C_\alpha \subseteq B_{\text{M}}$, where

$$B_k = \{y : \|y - x^*\|_2 \leq (2(\alpha - p^*)/k)^{1/2}\}.$$

This gives an upper bound on the condition number of C_α,

$$\text{cond}(C_\alpha) \leq M/m.$$

We can also gives a geometric interpretation of the condition number of the Hessian at the optimum as follows. From the Taylor series expansion of f around x^*,

$$f(y) \approx p^* + \frac{1}{2}(y - x^*)^\intercal \nabla^2 f(x^*)(y - x^*),$$

for α close to p^*, we have that

$$C_\alpha = \{y : (y - x^*)^\intercal \nabla^2 f(x^*)(y - x^*) \leq 2(\alpha - p^*)\}.$$

Therefore,

$$\lim_{\alpha \to p^*} \text{cond}(C_\alpha) = \kappa(\nabla^2 f(x^*)).$$

6.3. Descent methods

In this section, we produce a minimizing sequence $x^{(k)}$ where

$$x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x^{(k)},$$

and $t^{(k)} > 0$ (except when $x^{(k)}$ is optimal). All the methods we describe are called descent methods, which means that

$$f(x^{(k+1)}) < f(x^{(k)}),$$

except when $x^{(k)}$ is optimal. This implies that for all k, we have

$$x^{(k)} \in S \subset \operatorname{dom} f.$$

From convexity, we know that $\nabla f(x^{(k)})^\intercal (y - x^{(k)}) \geq 0$ implies that $f(y) \geq f(x^{(k)})$, so the search direction in a descent method must satisfy

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

The general descent method is then described as follows:

1. Take a starting point $x \in \operatorname{dom} f$.
2. repeat until the stopping condition is satisfied:
   a) Determine a descent direction Δx.
   b) Line search: choose a step size $t > 0$.
   c) Update: $x \leftarrow x + t\Delta x$.

The line search method has two main variants:

- Exact line search: t is chosen to minimize f along the ray $\{x + t\Delta x : t \geq 0\}$:

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

  This is used when the cost of the minimiation problem with one variable is low compared to the cost of computing the search direction.

- Backtracking line search: t is chosen to approximately minimize or just reduce f enough. The backtracking method depends on two constants $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$ and is as follows:
  1. Take a descent direction Δx for f at $x \in \operatorname{dom} f$.
  2. Set $t = 1$.
  3. while $f(x + t\Delta x) > f(x) + \alpha t \nabla f(x)^\intercal \Delta x$, set $t \leftarrow \beta t$.

Note that since Δx is a descent direction, for small enough t, we have

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

which shows that the backtracking line search eventually terminates. The constant α represents the fraction of decrease in f predicted by linear extrapolation that we will accept.

6.4. Gradient descent method

A natural choice for $\Delta x = -\nabla f(x)$, which is called the gradient descent algorithm. The stopping criterion is usually of the form $\|\nabla f(x)\|_2 \leq \eta$.

6.4.1. Convergence analysis

We present a basic convergence analysis. Take $x^+ = x + t\Delta x$ for $x^{(k+1)} = x^{(k)} + t^{(k)}\Delta x^{(k)}$. Suppose f is strongly convex on S, so there exist m, M with $mI \preceq \nabla^2 f(x) \preceq MI$ for all $x \in S$. Define $\tilde{f} : \mathbb{R} \to \mathbb{R}$ by $\tilde{f}(t) = f(x - t\nabla f(x))$. We only consider t with $x - t\nabla f(x) \in S$. If we use the inequality from Proposition 6.6, we obtain

$$\tilde{f}(t) \leq f(x) - t\|\nabla f(x)\|_2^2 + \frac{Mt^2}{2}\|\nabla f(x)\|_2^2.$$

Suppose we use exact line search. We obtain t_e, the exact step length that minimizes $\tilde{f}$. The righthand side is quadratic, minimized by $t = 1/M$, giving

$$f(x^+) = \tilde{f}(t_e) \leq f(x) - \frac{1}{2M}\|\nabla f(x)\|_2^2.$$

If we subtract p^* from both sides, we have

$$f(x^+) - p^* \leq f(x) - p^* - \frac{1}{2M}\|\nabla f(x)\|_2^2.$$

We can combine this with $\|\nabla f(x)\|_2^2 \geq 2m(f(x) - p^*)$ to conclude that

$$f(x^+) - p^* \leq (1 - m/M)(f(x) - p^*).$$

Applying this recursively, we have

$$f(x^{(k)}) - p^* \leq c^k(f(x^{(0)}) - p^*)$$

with $c := 1 - m/M < 1$, which implies that $f(x^{(k)})$ converges to p^* as $k \to \infty$. In particular, we must have $f(x^{(k)}) - p^* \leq \varepsilon$ after at most

$$\frac{\log((f(x^{(0)}) - p^*)/\varepsilon)}{\log(1/c)}$$

iterations of the gradient method with exact line search.

For backtracking line search, we first show that the backtracking exit condition is satisfied whenever $t \in [0, 1/M]$. Note that $0 \leq t \leq 1/M$ implies that $-t + \frac{Mt^2}{2} \leq -t/2$ from convexity, so it follows that

$$\begin{aligned}\tilde{f}(t) &\leq f(x) - t\|\nabla f(x)\|_2^2 + \frac{Mt^2}{2}\|\nabla f(x)\|_2^2 \\ &\leq f(x) - (t/2)\|\nabla f(x)\|_2^2 \\ &\leq f(x) - \alpha t\|\nabla f(x)\|_2^2,\end{aligned}$$

since $\alpha < 1/2$. Therefore, the backtracking line search terminates with either $t = 1$ or $t \geq \beta/M$.

- In the first case, we have

$$f(x^+) \leq f(x) - \alpha\|\nabla f(x)\|_2^2.$$

- In the second case, we have

$$f(x^+) \leq f(x) - (\beta\alpha/M)\|\nabla f(x)\|_2^2.$$

It follows that

$$f(x^+) \leq f(x) - \min\{\alpha, \beta\alpha/M\}\|\nabla f(x)\|_2^2.$$

Defining $c := 1 - \min\{2m\alpha, 2\beta\alpha m/M\} < 1$ and proceeding with the same analysis as with exact line search, we obtain

$$f(x^+) - p^* \leq c^k(f(x^{(0)}) - p^*)$$

which gives the same linear rate of convergence.

Some numerical results show the following:

1. The gradient method exhibits approximately linear convergence.
2. The choice of backtracking parameters α, β have a noticeable but not dramatic effect on the convergence. Exact line search can improve the convergence, but the effect is not large.
3. The convergence rate depends greatly on the condition number of the Hessian, or the sublevel sets.

6.5. Steepest descent method

Note that the first-order Taylor expansion of $f(x + v)$ around x is

$$f(x + v) \approx \widehat{f}(x + v) = f(x) + \nabla f(x)^\intercal v,$$

where the second term is called the directional derivative of f at x in direction v. We now address the problem of choosing v so that the directional derivative is as negative as possible, and we need to normalize by the length of v to make the question sensible.

Definition 6.7 (Normalized steepest descent direction)**.** Let $\|\cdot\|$ be any norm on $\mathbb{R}^m$. We define a normalized steepest descent direction as

$$\Delta x_{nsd} = \text{argmin}\{\nabla f(x)^\intercal v : \|v\| = 1\}.$$

We can also consider an unnormalized steepest descent step as

$$\Delta x_{sd} = \|\nabla f(x)\|_* \Delta x_{nsd},$$

where $\|\cdot\|_*$ is the dual norm. It follows that

$$\nabla f(x)^\intercal \Delta x_{sd} = \|\nabla f(x)\|_* \nabla f(x)^\intercal \Delta x_{nsd} = -\|\nabla f(x)\|_*^2.$$

This is interesting because it allows us to define steepest descent methods based on a choice of norm with analysis based on the general method. See the text for various examples of norms and a proof of linear convergence.

6.6. Newton's method

For $x \in \text{dom}\, f$, the Newton step is given by

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

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

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

unless $\nabla f(x) = 0$, so the Newton step is a descent direction (unless x is optimal). There are many interesting ways to motivate the Newton step:

- Minimizer of second-order approximation: the second-order Taylor approximation of f at x is
$$\widehat{f}(x+v) = f(x) + \nabla f(x)^\intercal v + \frac{1}{2} v^\intercal \nabla^2 f(x) v,$$
which is a convex quadratic function of v that is minimized at $v = \Delta x_{nt}$.

- Steepest descent direction in Hessian norm: Note that for the quadratic norm defined by the Hessian
$$\|u\|_{\nabla^2 f(x)} = (u^\intercal \nabla^2 f(x) u)^{1/2},$$
the Newton step is also the steepest descent direction at x. The quadratic norm P converges very rapidly when the Hessian, after the associated change of coordinates, has a small condition number. So the choice of norm defined by the Hessian is very well motivated.

- Solution of linearized optimality condition: if we linearize the optimality condition $\nabla f(x^*) = 0$ near x, we obtain
$$\nabla f(x+v) \approx \nabla f(x) + \nabla^2 f(x) v = 0,$$
with solution $v = \Delta x_{nt}$.

One interesting feature of the Newton step is that it is independent of linear changes of coordinates. We can prove this as follows: suppose $T \in \mathbb{R}^{n \times n}$ is nonsingular and define $\tilde{f}(y) = f(Ty)$. Then,

$$\nabla \tilde{f}(y) = T^\intercal \nabla f(x), \qquad \nabla^2 \tilde{f}(y) = T^\intercal \nabla^2 f(x) T$$

where $x = Ty$. It follows that for $\tilde{f}$ at y, the Newton step is given by

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

It follows that

$$x + \Delta x_{nt} = T(y + \Delta y_{nt}).$$

Definition 6.8 (Newton decrement)**.** Define the quantity

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

which is called the Newton decrement at x.

We note some useful properties of the Newton decrement, which will be a key tool in the analysis of the method and as a stopping criterion.

If we let $\widehat{f}$ be a second-order approximation of f at x, then note that

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

We can also express the decrement as

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

Furthermore, the constant $-\lambda(x)^2 = \nabla f(x)^\intercal \Delta x_{nt}$ is the one used in a backtracking line search, and can be interpreted as the directional derivative of f in the direction of the Newton step:

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

Finally, the Newton decrement is also affine invariant. With the Newton decrement in hand, we state Newton's method:

- given a starting point $x \in \operatorname{dom} f$, tolerance $\varepsilon > 0$
- repeat:
  1. Compute the Newton step and decrement.
  2. quit if $\lambda^2/2 \leq \varepsilon$
  3. Choose step size t by backtracking line search
  4. update $x \leftarrow x + t\Delta x_{nt}$

6.6.1. Convergence analysis

We assume that $f \in C^2(\mathbb{R}^n; \mathbb{R})$ and (m, M)-strongly convex. We also assume that the Hessian of f is L-Lipschitz continuous on S. There are two main phases of the algorithm - we will show that there exist $\eta \in (0, m^2/L]$ and $\gamma > 0$ such that

- Damped phase: If $\|\nabla f(x^{(k)})\|_2 \geq \eta$ then

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

Proof. Strong convexity implies that $\nabla^2 f(x) \preceq MI$ on S, so

$$\begin{aligned} f(x + t\Delta x_{nt}) &\leq f(x) + t\nabla f(x)^\intercal \Delta x_{nt} + \frac{M\|\Delta x_{nt}\|_2^2}{2} t^2 \\ &\leq f(x) - t\lambda(x)^2 + \frac{M}{2m} t^2 \lambda(x)^2. \end{aligned}$$

Then, note that $\widehat{t} = m/M$ satisfies the exit condition of the line search:

$$f(x + \widehat{t}\Delta x_{nt}) \leq f(x) - \frac{m}{2M}\lambda(x)^2 \leq f(x) - \alpha\widehat{t}\lambda(x)^2.$$

Therefore, the line search returns a step size $t \geq \beta m/M$, resulting in a decrease of the objective function

$$\begin{aligned} f(x^+) - f(x) &\leq -\alpha t \lambda(x)^2 \\ &\leq -\alpha\beta \frac{m}{M}\lambda(x)^2 \\ &\leq -\alpha\beta \frac{m}{M^2}\|\nabla f(x)\|_2^2 \\ &\leq -\alpha\beta\eta^2 \frac{m}{M^2}. \end{aligned}$$

Therefore, the desired inequality is satisfied with $\gamma = \alpha\beta\eta^2 \frac{m}{M^2}$. □

- Quadratic phase: If $\|\nabla f(x^{(k)})\|_2 < \eta$, then the backtracking line search selects $t^{(k)} = 1$ and
$$\frac{L}{2m^2}\|\nabla f(x^{(k+1)})\|_2 \leq \left(\frac{L}{2m^2}\|\nabla f(x^{(k)})\|_2\right)^2.$$

Proof. First, we show that the backtracking line search selects unit steps provided
$$\eta \leq 3(1-2\alpha)\frac{m^2}{L}.$$
By the Lipschitz condition, for $t \geq 0$ we have
$$\|\nabla^2 f(x + t\Delta x_{nt}) - \nabla^2 f(x)\|_2 \leq t\|\Delta x_{nt}\|_2,$$
so it follows that
$$|\Delta x_{nt}^\intercal \left(\nabla^2 f(x + t\Delta x_{nt}) - \nabla^2 f(x)\right) \Delta x_{nt}| \leq tL\|\Delta x_{nt}\|_2^3.$$
If we define $\tilde{f} = f(x + t\Delta x_{nt})$, the above inequality is equivalent to
$$|\tilde{f}''(t) - \tilde{f}''(0)| \leq tL\|\Delta x_{nt}\|_2^3.$$
Now, we establish an upper bound on $\tilde{f}(t)$. Note that
$$\tilde{f}''(t) \leq \tilde{f}''(0) + tL\|\Delta x_{nt}\|_2^3 \leq \lambda(x)^2 + t\frac{L}{m^{3/2}}\lambda(x)^3.$$
Integrating the inequality, we obtain
$$\tilde{f}'(t) \leq \tilde{f}'(0) + t\lambda(x)^2 + t^2\frac{L}{2m^{3/2}}\lambda(x)^3 = -\lambda(x)^2 + t\lambda(x)^2 + t^2\frac{L}{2m^{3/2}}\lambda(x)^3.$$
Integrating another time, we obtain
$$\tilde{f}(x) \leq \tilde{f}(0) - t\lambda(x)^2 + t^2\frac{1}{2}\lambda(x)^2 + t^3\frac{L}{6m^{3/2}}\lambda(x)^3.$$
Finally, taking $t = 0$, we obtain
$$f(x + \Delta x_{nt}) \leq f(x) - \frac{1}{2}\lambda(x)^2 + \frac{L}{6m^{3/2}}\lambda(x)^3.$$
If $\|\nabla f(x)\|_2 \leq \eta \leq 3(1-2\alpha)m^2/L$, then by strong convexity, we have
$$\lambda(x) \leq 3(1-2\alpha)m^{3/2}/L$$
so we obtain
$$f(x + \Delta x_{nt}) \leq f(x) - \lambda(x)^2\left(\frac{1}{2} - \frac{L\lambda(x)}{6m^{3/2}}\right) \leq f(x) + \alpha\nabla f(x)^\intercal x_{nt}.$$
Finally, if we apply the Lipschitz condition, note that
$$\begin{aligned}
\|\nabla f(x^+)\|_2 &= \|\nabla f(x + \Delta x_{nt}) - \nabla f(x) - \nabla^2 f(x)\Delta x_{nt}\|_2 \\
&= \left\|\int_0^1 \left(\nabla^2 f(x + t\Delta x_{nt}) - \nabla^2 f(x)\right)\Delta x_{nt}\, dt\right\| \\
&\leq \frac{L}{2}\|\Delta x_{nt}\|_2^2 \\
&= \frac{L}{2}\|\nabla^2 f(x)^{-1}\nabla f(x)\|_2^2 \\
&\leq \frac{L}{2m^2}\|\nabla f(x)\|_2^2.
\end{aligned}$$
So in particular, taking $\eta = \min 1, 3(1-2\alpha)\frac{m^2}{L}$ gives the desired conclusion. □

First, we consider the second condition. Note that if $\|\nabla f(x^{(k)})\|_2 < \eta$ and $\eta \leq m^2/L$, it follows that we also have $\|\nabla f(x^{(k+1)})\|_2 < \eta$, so the condition holds for all $l \geq k$. Applying the inequality recursively, we have

$$\frac{L}{2m^2}\|\nabla f(x^{(l)})\|_2 \leq \left(\frac{L}{2m^2}\|\nabla f(x^{(k)})\|_2\right)^{2^{l-k}} \leq 2^{-2^{l-k}}$$

Therefore,

$$f(x^{(l)}) - p^* \leq \frac{1}{2m}\|\nabla f(x^{(l)})\|_2^2 \leq \frac{2m^3}{L^2}2^{-2^{l-k+1}},$$

which corresponds to a quadratic convergence rate.

Now, note that since f decreases by at least γ at every iteration, the number of damped Newton steps cannot exceed

$$f(x^{(0)} - p^*)/\gamma$$

since otherwise, f would be less than p^*. Furthermore, the number of iterations untill $f(x) - p^* \leq \varepsilon$ in the quadratic phase is no more than $\lg\lg(\varepsilon_0/\varepsilon)$ iterations, where $\varepsilon_0 = 2m^3/L^2$. Therefore, the total number of steps is bounded by

$$\frac{f(x^{(0)}) - p^*}{\gamma} + \lg\lg(\varepsilon_0/\varepsilon)$$

Some general properties of Newton's method that can be seen in practice are as follows:

- Convergence of Newton's method is rapid in general, and quadratic near x^*.
- Newton's method is affine invariant and completely insensitive to the choice of coordinates or the condition number of the sublevel sets of the objective.
- Newton's method scales well with problem size.
- Good performance of Newton's method is not dependent on the choice of parameters. In contrast, the choice of norm for steepest descent plays a critical role in its performance.

The main disadvantage of Newton's method is the cost of forming and storing the Hessian, and the cost of computing the Newton step, which requires solving a set of linear equations. Sometimes, it is possible to exploit problem structure to substantially reduce the cost of computing the Newton step(sparse factorizations, band matrix structure, etc.)

Another family of alternatives to Newton's method for unconstrained problems are called quasi-Newton methods, which require less computational effort to form the search direction, but share some of the strong advantages of Newton methods, such as local rapid convergence. This will be covered more extensively in *Part III, Numerical Optimization.*

6.7. Solutions to selected problems

Exercise 6.9 (9.1)**.** Consider the problem argmin $f(x) = (1/2)x^\intercal Px + q^\intercal x + r$ where $P \in \mathcal{S}^n$. Show that if $P \succeq 0$ but the optimality condition $Px^* = -q$ doesn't have a solution, then the problem is unbounded below.

Proof. We have $q \notin \mathcal{R}(P)$, so we have express $q = \tilde{q} + q_\perp$ where $q_\perp$ is the orthogonal projection of $\tilde{q}$ onto $\mathcal{R}(P)$ and $q_\perp$ is nonzero. Then for $x = tq_\perp$, note that

$$f(x) = (1/2)q_\perp^\intercal Pq_\perp + tq_\perp^\intercal q_\perp + r = tq_\perp^\intercal q_\perp + r$$

which is unbounded below. □

Exercise 6.10 (9.5)**.** Suppose f is strongly convex with $mI \preceq \nabla^2 f(x) \preceq MI$. Let Δx be a descent direction at x. Show that the backtracking stopping condition holds for

$$0 < t \leq -\frac{\nabla f(x)^\intercal \Delta x}{M\|\Delta x\|_2^2}.$$

Use this to give an upper bound on the number of backtracking iterations.

Proof. Use the strong convexity upper bound and check condition is satisfied. The final result holds from setting $t_0 \leq 1$ and iterating until $\beta^k t_0$ falls in the desired range, where we can find an upper bound on k by taking logs. □

Exercise 6.11 (9.7)**.** Let Δx_{nsd} and Δx_{sd} be normalized and unnormalized steepest descent directions for the norm $\|\cdot\|$. Prove the following identities:

1. $\nabla f(x)^\intercal \Delta x_{nsd} = -\|\nabla f(x)\|_*$.
2. $\nabla f(x)^\intercal \Delta x_{sd} = -\|\nabla f(x)\|_*^2$.
3. $\Delta x_{sd} = \operatorname{argmin}_v (\nabla f(x)^\intercal v + (1/2)\|v\|^2)$.

Proof. For 1, note that

$$-\|\nabla f(x)\| = -\sup_{\|v\|=1}\{|\nabla f(x)^\intercal v|\} = \inf_{\|v\|=1}\{\nabla f(x)^\intercal v\},$$

where we note that fact that v is a descent direction so $\nabla f(x)^\intercal v < 0$. 2 immediately follows from the definition.

For 3, we define $v = tw$ where $t \geq 0$ and $\|w\| = 1$ is fixed. Optimizing over t and w separately gives $t = \|\nabla f(x)\|_*$ and $w = \Delta x_{nsd}$, which gives the result. □

Exercise 6.12 (9.9)**.** Show that the Newton decrement satisfies

$$\lambda(x) = \sup_{v^\intercal \nabla^2 f(x) v = 1} (-v^\intercal \nabla f(x)) = \sup_{v \neq 0} \frac{-v^\intercal \nabla f(x)}{(v^\intercal \nabla^2 f(x) v)^{1/2}}.$$

Proof. Define the change of variables $w = (\nabla^2 f(x))^{1/2} v$ so that $w^\intercal w = v^\intercal \nabla^2 f(x) v$. Now, note that

$$\begin{aligned}
\sup_{v^\intercal \nabla^2 f(x) v = 1} (-v^\intercal \nabla f(x)) &= \sup_{w^\intercal w = 1} \left(([\nabla^2 f(x)]^{-1/2} w)^\intercal \nabla f(x)\right) \\
&= \sup_{\|w\|_2 = 1} \left(w^\intercal (\nabla^2 f(x))^{-1/2} \nabla f(x)\right) \\
&= \sup_{\|w\|_2 = 1} \left\langle w, (\nabla^2 f(x))^{-1/2} \nabla f(x)\right\rangle \\
&= \left\|(\nabla^2 f(x))^{-1/2} \nabla f(x)\right\|_2 \\
&= \left(\nabla f(x)^\intercal \nabla^2 f(x)^{-1} \nabla f(x)\right)^{1/2} \\
&= \lambda(x).
\end{aligned}$$

□

Exercise 6.13 (9.11)**.** Suppose $\varphi : \mathbb{R} \to \mathbb{R}$ is increasing and convex, $f : \mathbb{R}^n \to \mathbb{R}$ is convex, so $g(x) = \varphi(f(x))$ is convex. Furthermore, assume f, g are twice differentiable. Compare the gradient method and Newton's method, applied to f and g. How are the search directions related? How are the methods related if an exact line search is used?

Proof. The first part just follows from the chain rule - we see that the gradient of g is a positive multiple of the gradient of f, the with exact line search we choose the same step size. Things could vary with the backtracking line search depending on the constant.

For Newton's method, note that the iterate is given by

$$\begin{aligned}
\nabla x_{nt}^g &= -\left(\varphi''(x)\nabla f(x)\nabla f(x)^\intercal + \varphi'(x)\nabla^2 f(x)\right)^{-1}\nabla f(x) \\
&= \left(I + \sum_{k\geq 2}\left(\frac{\varphi''(f(x))}{\varphi'(f(x))}\Delta x_{nt}\nabla f(x)^\intercal\right)^k\right)\Delta x_{nt} \\
&= \Delta x_{nt}\left(1 + \sum_{k\geq 2}\left(\frac{\varphi''(f(x))}{\varphi'(f(x))}\right)^k(\nabla f(x)^\intercal \Delta x_{nt})^k\right) \\
&= \Delta x_{nt}\left(1 + \frac{\left(\frac{\varphi''(f(x))}{\varphi'(f(x))}\nabla f(x)^\intercal \Delta x_{nt}\right)^2}{1 - \left(\frac{\varphi''(f(x))}{\varphi'(f(x))}\nabla f(x)^\intercal \Delta x_{nt}\right)}\right)
\end{aligned}$$

which is a positive multiple of Δx_{nt}. □

Remark 6.14. Note that the second equality follows from careful manipulations after applying the Woodbury formula.

Exercise 6.15 (9.12)**.** In this problem, we introduce the trust region Newton method. If $\nabla^2 f(x)$ is singular, the Newton step $\Delta x_{nt} = -\nabla^2 f(x)^{-1}\nabla f(x)$ is not well defined. Instead, we can define a search direction Δx_{tr} as the solution of

$$\text{argmin} \quad (1/2)v^\intercal H v + g^\intercal v \qquad \text{s.t. } \|v\|_2 \leq \gamma,$$

where $H = \nabla^2 f(x)$, $g = \nabla f(x)$, $\gamma > 0$. The point $x + \Delta x_{tr}$ minimizes the second-order approximation of f and x, subject to the constraint that $\|(x + \Delta x_{tr}) - x\|_2 \leq \gamma$.

The set $\{v : \|v\|_2 \leq \gamma\}$ is called the trust region. The parameter γ reflects our confidence in the second-order model. Show that Δx_{tr} minimizes

$$(1/2)v^\intercal H v + g^\intercal v + \widehat{\beta}\|v\|_2^2$$

for some $\widehat{\beta}$.

Proof. The objective and constraint is differentiable, so we take the dual and apply the KKT conditions, where β corresponds to the dual constraint. They are easy to derive based on casework on whether or not H is singular. □

7. Equality constrained minimization

We describe methods for solving a convex optimization problem with equality constraints:

$$\begin{aligned} \text{argmin} \quad & f(x) \\ & Ax = b, \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is convex and twice differentiable, $A \in \mathbb{R}^{p\times n}$ with $\operatorname{rank} A = p < n$. These assumptions mean that there are fewer equality constraints than variables, and that the equality constraints are independent. We will assume that x^* exists with corresponding optimal value p^*.

Recall that $x^* \in \operatorname{dom} f$ is optimal if and only if there exists $v^* \in \mathbb{R}^p$ such that

$$Ax^* = b, \qquad \nabla f(x^*) + A^\intercal v^* = 0$$

so solving the ECM problem is equivalent to finding a solution of the corresponding KKT conditions. The primal feasibility equations are $Ax^* = b$, which is a linear system. The dual feasibility equations are $\nabla f(x^*) + A^\intercal v^* = 0$, which is in general, nonlinear.

- Note that we could eliminate the equality constraints to obtain an equivalent unconstrained problem. So we could apply the methods of the previous section.
- Another approach is the solve the dual problem using an unconstrained minimization method and use the dual solution to obtain a primal solution.
- We also have extensions of Newton's method that directly handle the equality constraints. This is preferable, because dualized problems often destroy special problem structure (for example sparsity).

7.1. Equality constrained convex quadratic minimization

In the case of quadratic objectives, we can solve the primal and dual feasibility conditions analytically. Consider the problem $f(x) = (1/2)x^\intercal Px + q^\intercal x + r$ subject to the equality $Ax = b$, where $P = \mathcal{S}^n_+$ and $A \in \mathbb{R}^{p\times n}$.

We can write the optimality conditions as

$$\begin{bmatrix} P & A^\intercal \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ v^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}.$$

This is called the KKT system for the problem and the coefficient matrix is called the KKT matrix.

When the KKT matrix is nonsingular, there is a unique optimal primal-dual pair (x^*, v^*). If not, but the KKT system is solvable, then any solution yields an optimal pair. IF the KKT system is not solvable, then the quadratic problem is unbounded below or infeasible. Indeed, in this case, there exist $v \in \mathbb{R}^n$, $w \in \mathbb{R}^p$ such that

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

If we let $\widehat{x}$ be any feasible point, then $x = \widehat{x} + tv$ is feasible for all t, and

$$\begin{aligned} f(x) &= f(\widehat{x}) + t(v^\intercal P\widehat{x} + q^\intercal v) + (1/2)t^2 v^\intercal Pv \\ &= f(\widehat{x}) + t(-\widehat{x}^\intercal A^\intercal w + q^\intercal v) - (1/2)t^2 w^\intercal Av \\ &= f(\widehat{x}) + t(-b^\intercal w + q^\intercal v) \xrightarrow{t\to\infty} -\infty. \end{aligned}$$

Note that there are several conditions that are equivalent to the nonsingularity of the KKT matrix:

- $\mathcal{N}(P) \cap \mathcal{N}(A) = \{0\}$.
- $Ax = 0, x \neq 0$ implies that $x^\intercal Px > 0$ or P is positive definite on $\mathcal{N}(A)$.
- $F^\intercal PF \succ 0$ where $F \in \mathbb{R}^{n\times(n-p)}$ is a matrix for which $\mathcal{R}(F) = \mathcal{N}(A)$.

As a special case, $P \succ 0$ implies that the KKT matrix is nonsingular.

7.2. Eliminating equality constraints

As mentioned before, one general approach to solving the equality constrained problem is to eliminate the equality constraint and use unconstrained methods. First, we find a matrix $F \in \mathbb{R}^{n\times(n-p)}$ so that $\mathcal{R}(F) = \mathcal{N}(A)$. The coresponding reduced problem is

$$\text{argmin} \quad \tilde{f}(z) = f(Fz + \widehat{x}).$$

Remark 7.1. Note that there are many possible choices for the elimination matrix F. But they all form equivalent problems, essentially through changes of coordinate systems.

Note that we can also construct an optimal dual variable v^* as

$$v^* = -(AA^\intercal)^{-1}A\nabla f(x^*).$$

It is easy to verify that the dual feasibility condition holds.

Remark 7.2. It is interesting to see a "least-squares" type of solution, but I suppose the intuition follows from the fact that the dualized problem is a linear approximation with the lower bound property.

7.3. Newton's method with equality constraints

Now, we describe an extension of Newton's method to include equality constraints. The biggest differences are as follows:

- The initial point must be feasible.
- The definition of the Newton step is modified to account for the equality constraints. In particular, the direction should be feasible: $A\Delta x_{nt} = 0$.

In order to derive the Newton step, we replace the objective with its second-order Taylor approximation:

$$\begin{aligned} \text{argmin} \quad & \widehat{f}(x+v) = f(x) + \nabla f(x)^\intercal v + (1/2)v^\intercal \nabla^2 f(x) v \\ & A(x+v) = b. \end{aligned}$$

This is a convex quadratic minimization problem with equality constraints, so we can solve this analytically, as in Section 7.1. The Newton step is characterized by

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

where w is the associated dual variable for the quadratic problem. Note that Newton step is only defined at points for which the KKT matrix is nonsingular. The algorithm and analysis proceed similarly as before. It is interesting to note that the Newton decrement stays the same as the unconstrained version.

7.4. Infeasible-start Newton method

Note that the above method is a feasible descent method, in that we rely on the assumption that $x \in \operatorname{dom} f$ and $Ax = b$, and all the future iterates are feasible (unless they are optimal).

Instead, suppose we have $x \in \operatorname{dom} f$ but we do not assume it to be feasible. Our goal is to find a step Δx so that $x + \Delta x$ satisfies (at least approximatly) the optimality conditions. To do this, substite $x + \Delta x$ for x^* and w for v^* and use the first-order approximation:

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

in order to obtain

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

$$\begin{bmatrix} \nabla^2 f(x) & A^\intercal \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{nt} \\ w \end{bmatrix} = - \begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix}.$$

Note that the equations are the same as the ones that define the Newton step at a feasible point, except the second block component of the righthand side contains the residual $Ax - b$. Of course, this vanishes when x is feasible. When we refer to Δ_{nt} in the future for equality constrained problems, it will refer to the above one in order to remove that assumption that we start at a feasible point.

7.4.1. Primal-dual interpretation

Define the function $r : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n \times \mathbb{R}^p$ as

$$r(x, \nu) = (r_{dual}(x, \nu), r_{pri}(x, \nu)),$$

where

$$r_{dual}(x, \nu) = \nabla f(x) + A^\intercal \nu, \quad r_{pri}(x, \nu) = Ax - b$$

are the dual and primal residuals respectively. The optimality conditions are given by $r(x^*, \nu^*) = 0$. The first-order Taylor approximation of r is

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

where $Dr(y) \in \mathbb{R}^{(n+p)\times(n+p)}$ is the derivative of r evaluated at y.

Now, we define the primal-dual Newton step as

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

It is easy to show that the primal-dual Newton step is related to the original Newton step as

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

A nice property of r is that the norm of the residual decreases in the Newton direction:

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

so it follows that

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

This is contrary to the usual Newton direction, where we have

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

which is not necessarily negative, unless x is feasible.

We also have a full step feasibility property: recall that the Newton step has the property that $A(x+\Delta x_{nt}) = b$, which implies that if x is feasibility, future iterates are also feasible.

We can analyze the effect of a damped step on the equality constraint residual r_{pri}. With a step length $t \in [0,1]$, the next iterate is $x + t\Delta x_{nt}$, so we have

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

It follows that

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

where $r^{(i)} = Ax^{(i)} - b$. This shows that the residual at each step is in the direction of the initial primal residual and is scaled down at every step. Once a full step is taken, all future iterates are primal feasible.

7.4.2. Algorithm

With the residual in hand, we state the algorithm:

- Given $x \in \operatorname{dom} f, \nu$, tolerance $\varepsilon > 0$, $\alpha \in (0, 1/2), \beta \in (0,1)$.
- repeat until $Ax = b$ and $\|r(x,\nu)\|_2 \leq \varepsilon$.
  1. Compute primal and dual Newton steps $\Delta x_{nt}, \Delta\nu_{nt}$.
  2. Backtracking line search on $\|r\|_2$:
     a) $t := 1$
     b) while $\|r(x+t\Delta x_{nt}, \nu + t\Delta\nu_{nt})\|_2 > (1-\alpha t)\|r(x,\nu)\|_2$, $t \leftarrow \beta t$
  3. Update: $x \leftarrow x + t\Delta x_{nt}, \quad \nu \leftarrow \nu + t\Delta\nu_{nt}$.

The convergence analysis proceeds very similarly to the original Newton method.

7.5. Convex-concave games

Suppose $r : \mathbb{R}^N \to \mathbb{R}^n$ is differentiable, the derivative satisfies a Lipschitz condition on S, and $\|Dr(x)^{-1}\|_2$ is bounded on S, where

$$S = \{x \in \operatorname{dom} r : \|r(x)\|_2 \leq \|r(x^{(0)})\|\}$$

is a closed set. Then the infeasible start Newton method, starting at $x^{(0)}$, converges to a solution of $r(x) = 0$. But the general framework can also be applied in other settings - we apply it to solving convex-concave games.

An unconstrained game on $\mathbb{R}^p \times \mathbb{R}^q$ is defined by the payoff function $f : \mathbb{R}^{p+q} \to \mathbb{R}$, where player 1 chooses a value $u \in \mathbb{R}^p$ and player 2 chooses a value $v \in \mathbb{R}^q$. Based on the

choises, player 1 pays player 2 the amount $f(u, v)$. The goal of player 1 is to minimize the payment while player 2 wants to maximize it.

If player 1 moves first, then player 2 knows the choice so they will choose v to maximize $f(u, v)$ which results in a payoff $\sup_v f(u, v)$. If player 1 assumes player 2 will make this choice, they should choose u to minimize $\sup_v f(u, v)$, resulting in the overall payoff from player 1 to player 2 as

$$\inf_u \sup_v f(u, v).$$

On the other hand, if player 2 makes the first choice, then we have the reverse strategy

$$\sup_v \inf_u f(u, v).$$

The minimax inequality (or weak duality) says that the first strategy is always greater than or equal to the second one, and the difference between the two is referred to as the advantage afforded to the player who makes the second move.

We say that (u^*, v^*) is a solution of the game if it satisfies a saddle-point property:

$$f(u^*, v) \leq f(u^*, v^*) \leq f(u, v^*).$$

If a solution exists, then we see that there is no advantage to making the second move since $f(u^*, v^*)$ is a common value of both payoffs (we showed this in a previous exercise).

The game is called convex-concave if for each v, $f(u, v)$ is a convex function of u and for each u, $f(u, , v)$ is a concave function of v. When f is differentiable, a saddle-point is characterized by $\nabla f(u^*, v^*) = 0$.

7.5.1. Solution via infeasible start Newton method

Define the residual as

$$r(u, v) = \nabla f(u, v) = \begin{bmatrix} \nabla_u f(u, v) \\ \nabla_v f(u, v) \end{bmatrix},$$

and apply the infeasible start Newton method. The convergence is guaranteed provided that $Dr = \nabla^2 f$ has a boundedd inverse and satisfies the Lipschitz condition on S. There is also a simple analog of the strong convexity condition: we say that a game with payoff f is strongly convex-concave if for some $m > 0$, $\nabla^2_{uu} f(u, v) \succeq mI$ and $\nabla^2_{vv} f(u, v) \preceq -mI$ for all $(u, v) \in S$. Unsurprisingly, the strong convex-concave condition implies the bounded inverse condition.

7.6. Solutions to selected problems

Exercise 7.3 (10.1)**.** Recall the KKT matrix given by

$$K = \begin{bmatrix} P & A^\intercal \\ A & 0 \end{bmatrix}$$

where $P \in \mathcal{S}_+^n$, $A \in \mathbb{R}^{p\times n}$ where $\operatorname{rank} A = p < n$.

(a) Show that the following statements are equivalent:

1. The KKT matrix is nonsingular.
2. $\mathcal{N}(P) \cap \mathcal{N}(A) = \{0\}$.
3. $Ax = 0, x \neq 0 \implies x^\intercal P x > 0$.
4. $F^\intercal P F \succ 0$ where $F \in \mathbb{R}^{n\times(n-p)}$ is a matrix for which $\mathcal{R}(F) = \mathcal{N}(A)$.
5. $P + A^\intercal Q A \succ 0$ for some $Q \succeq 0$.

(b) Show that if the KKT matrix is nonsingular, then it has exactly n positive and p negative eigenvalues.

Proof. First, note that if the KKT matrix is nonsingular, then $KX = 0$ implies that $X = 0$. But note that $KX = 0$ is equivalent to $PX_1 + A^\intercal X_2 = AX_1 = 0$, and we can take $X_2 = 0$ without loss of generality. It follows that $\mathcal{N}(P) \cap \mathcal{N}(A) = \{0\}$.

For 2 implies 3, if $Ax = 0$ and $x \neq 0$, then if $x^\intercal P x = 0$, then note that writing $P = B^\intercal B$ gives

$$x^\intercal B^\intercal B x = \|Bx\|_2^2 = 0,$$

which implies that $Bx = 0$, so $B^\intercal B x = Px = 0$, which implies that $x \in \mathcal{N}(P) \cap \mathcal{N}(A) = \{0\}$, a contradiction.

For 3 implies 4, note that $\dim \mathcal{N}(A) = n - p$ by the rank-nullity theorem, so using 3, we can construct a basis for $\mathcal{N}(A)$ which defines the matrix F which satisfies $F^\intercal P F \succ 0$ by construction.

For 4 implies 5,

$$x^\intercal (P + A^\intercal A) x = x^\intercal P x + x^\intercal A^\intercal A x = x^\intercal P x + \|Ax\|_2^2.$$

If $x \notin \mathcal{N}(A)$, then we are done. Otherwise, we can take $x = Fz$ gives

$$z^\intercal F^\intercal P F z + z^\intercal F^\intercal A^\intercal A F z = z^\intercal F^\intercal P F z > 0.$$

Finally, if we have 5 but the KKT matrix is singular, then we can find x, z not both zero so that $Px + A^\intercal z = 0$, $Ax = 0$, which implies that $x^\intercal P x + x^\intercal A^\intercal z = x^\intercal P x = 0$. But this implies that

$$x^\intercal (P + A^\intercal Q A) x = x^\intercal P x + x^\intercal A^\intercal Q A x = 0$$

so we must have $x = 0$. But then $A^\intercal z = 0$, which contradicts the fact that $\operatorname{rank} A = p$. □

Exercise 7.4 (10.3)**.** In this problem, we explore Newton's method for solving the dual of the equality constrained minimization problem. We assume that $f \in C^2$, $\nabla^2 f(x) \succ 0$ for all $x \in \operatorname{dom} f$ and that for each $v \in \mathbb{R}^p$, the Lagrangian $L(x, \nu) = f(x) + \nu^\intercal (Ax - b)$ has a unique minimizer $x(\nu)$.

(a) Show that the dual function g is twice differentiable, and find an expression for the Newton step evaluated at ν.

Proof. This essentially follows from the properties of the Legendre transform. We obtain $\nabla g = -b + A\nabla f^*(-A^\intercal \nu) = -b + Ax(\nu)$, and $\nabla^2 g = A\nabla^2 f^*(-A^\intercal \nu)A^\intercal = A\nabla^2 f^*(x(\nu))^{-1}A^\intercal$. From this we can easily derive the Newton decrement.

Part (b) is less trivial and follows from the block matrix inversion formula and fact that the 2-norm is an operator norm, allowing us to obtain that $\|A\nabla^2 f(x)^{-1}A^{-1}\| \geq K$, from which we can easily derive the conclusion. □

Exercise 7.5 (10.6)**.** Show that the Newton decrement satisfies

$$f(x) - \inf\{\widehat{f}(x+v) | A(x+v) = b\} = \lambda(x)^2/2.$$

Proof. Note that by definition of the Newton step,

$$\begin{aligned} f(x) - \inf\{\widehat{f}(x+v) | A(x+v) = b\} &= f(x) - \widehat{f}(x + \Delta x_{nt}) \\ &= -\nabla f(x)^\intercal \Delta x_{nt} - (1/2)\Delta x_{nt}^\intercal \nabla^2 f(x) \Delta x_{nt} \\ &= \lambda(x)^2/2 \end{aligned}$$

□

Part II.
Numerical Optimization

This will contain notes from *Numerical Optimization* by Nocedal and Wright, as well as lecture notes from STAT 31020 at UChicago.

Part III.
Online Convex Optimization

Part IV.
Riemannian Optimization