MATH3027: Optimization 2022

Week 7: Convex Optimization

Prof. Richard Wilkinson
Please send any comments or mistakes to r.d.wilkinson@nottingham.ac.uk

Having studied convex sets and functions, we are now in a position to study convex optimization problems, that is, the optimization of convex functions constrained to convex sets. Convex problems, like linear programmes, can be solved quickly and reliably for very large problems. The difficulty is often determining whether or not a given problem is convex.

This week we introduce the concept of stationarity as a characterization of optimality in convex problems, and then we will study projection operators and propose an algorithm for convex optimization, known as the projected gradient method.

Four Important Theorems for Convex Functions

We begin by stating four important results for convex functions. The first relates to the continuity of convex functions. In particular, convex functions are continuous at interior points of their domain:

Theorem (Continuity of Convex Functions)**.** *Let $f : C \longrightarrow \mathbb{R}$ be a convex function defined over a convex set $C \subseteq \mathbb{R}^n$. Let $\mathbf{x}_0 \in \mathrm{int}(C)$. Then there exist $\varepsilon > 0$ and $L > 0$ such that $B\left[\mathbf{x}_0, \varepsilon\right] \subseteq C$ and*

$$|f(\mathbf{x}) - f(\mathbf{x}_0)| \leq L \|\mathbf{x} - \mathbf{x}_0\| \text{ for any } \mathbf{x} \in B\left[\mathbf{x}_0, \varepsilon\right]$$

In addition, all the directional derivatives of a convex function exist at the interior of the domain.

Theorem (Existence of Directional Derivatives of Convex Functions). *Let $f : C \longrightarrow \mathbb{R}$ be a convex function over the convex set $C \subseteq \mathbb{R}^n$. Let $\mathbf{x} \in \operatorname{int}(C)$. Then for any $\mathbf{d} \neq 0$, the directional derivative $f'(\mathbf{x}; \mathbf{d})$ exists.*

The last two theorems relate to the problem of **maximizing** a non-constant convex function over a convex set. As we'll soon see, *minimizing* a convex function over a convex set is in some sense relatively easy, whereas in contrast, maximizing a convex function is hard as the stationarity condition does not hold.

Theorem (No Maximum Inside the Convex Set). *Let $f : C \longrightarrow \mathbb{R}$ be convex and non-constant over the nonempty convex set $C \subseteq \mathbb{R}^n$. Then f does not attain a maximum at a point in $\operatorname{int}(C)$.*

Finally, we state that the maximum of convex function over compact convex sets can be found at the extreme points of the set.

Theorem (Maximum of a Convex Function Over a Compact Convex Set). *Let $f : C \longrightarrow \mathbb{R}$ be convex over the nonempty convex and compact set $C \subseteq \mathbb{R}^n$. Then there exists at least one maximizer of f over C that is an extreme point of C.*

Proof. Let $\mathbf{x}^*$ be a maximizer of f over C. If $\mathbf{x}^*$ is an extreme point of C, then the result is established. Otherwise, by Krein-Milman, $C = \operatorname{conv}(\operatorname{ext}(C))$ implies the existence of $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \in \operatorname{ext}(C)$ such that

$$\mathbf{x}^* = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i$$

for $\boldsymbol{\lambda} \in \Delta_k$. By the convexity of f,

$$f(\mathbf{x}^*) \leq \sum_{i=1}^{k} \lambda_i f(\mathbf{x}_i)$$

or equivalently

$$\sum_{i=1}^{k} \lambda_i \left(f(\mathbf{x}_i) - f(\mathbf{x}^*) \right) \geq 0$$

as $\lambda_i \geq 0$. Since $\mathbf{x}^*$ is a maximizer of f over C, we have $f(\mathbf{x}_i) \leq f(\mathbf{x}^*)$ for all $i = 1, \dots, k$. This implies that $f(\mathbf{x}_i) = f(\mathbf{x}^*)$. Consequently, the extreme points $\mathbf{x}_1, \dots, \mathbf{x}_k$ are all maximizers of f over C. □

Convex Optimization Problems

A **convex optimization** problem consists of minimizing a convex function $f(\mathbf{x})$ over a convex set C:

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{x} \in C \end{array} \tag{CVX}$$

A more explicit way of writing this is in a functional form:

$$\begin{array}{lll} \min & f(\mathbf{x}) & \\ \text{s.t.} & g_i(\mathbf{x}) \leq 0, & i = 1, 2, \ldots, m \\ & h_j(\mathbf{x}) = 0, & j = 1, 2, \ldots, p, \end{array}$$

where $f, g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ are convex functions, and $h_1, h_2, \ldots, h_p : \mathbb{R}^m \to \mathbb{R}$ are affine functions.

A very important feature of convex optimization problems is that local minima are global minima!

Theorem (Local minima are global minima in convex problems.)**.** *Let $f : C \to \mathbb{R}$ be a convex function defined on the convex set $C \subseteq \mathbb{R}^n$. Let $\mathbf{x}^* \in C$ be a local minimum of f over C. Then x^* is a global minimum of f over C.*

Proof. Assume $\mathbf{x}^*$ is a local minimum of f over C. This implies that there exists $r > 0$ such that $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for any $\mathbf{x} \in C \cap B[\mathbf{x}^*, r]$. Let $\mathbf{x}^* \neq \mathbf{y} \in C$. We will show that $f(\mathbf{y}) \geq f(\mathbf{x}^*)$. Let $\lambda \in (0, 1)$ be such that $\mathbf{x}^* + \lambda(\mathbf{y} - \mathbf{x}^*) \in B[\mathbf{x}^*, r]$. Since $\mathbf{x}^* + \lambda(\mathbf{y} - \mathbf{x}^*) \in B[\mathbf{x}^*, r]$, it follows that $f(\mathbf{x}^*) \leq f(\mathbf{x}^* + \lambda(\mathbf{y} - \mathbf{x}^*))$ and hence

$$\begin{aligned} f(\mathbf{x}^*) &\leq f(\mathbf{x}^* + \lambda(\mathbf{y} - \mathbf{x}^*)) \\ &= f((1-\lambda)\mathbf{x}^* + \lambda \mathbf{y}) \\ &\leq (1-\lambda) f(\mathbf{x}^*) + \lambda f(\mathbf{y}). \end{aligned}$$

by Jensen's inequality. The desired inequality $f(\mathbf{x}^*) \leq f(\mathbf{y})$ follows. □

A small variation of the proof of the last theorem yields the following.

Theorem. *Let $f : C \to \mathbb{R}$ be a strictly convex function defined on the convex set C. Let $x^* \in C$ be a local minimum of f over C. Then x^* is a strict global minimum of f over C.*

Another important and easily deduced property of convex problems is that the set of optimal solutions is also convex.

Theorem. *Let $f : C \to \mathbb{R}$ be a convex function defined over the convex set $C \subseteq \mathbb{R}^n$. Then the set of optimal solutions of the problem*

$$\min\{f(\mathbf{x}) : \mathbf{x} \in C\}$$

is convex. If, in addition, f is strictly convex over C, then there exists at most one optimal solution of the problem.

Finally, note that maximizing a concave function over a convex set is also a convex optimization problem.

Examples:

- Linear Programming

$$(\mathbf{LP}) : \quad \begin{array}{ll} \min & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{B}\mathbf{x} = \mathbf{g} \end{array}$$

  (LP) is a convex optimization problem (constraints and objective function are linear / affine and hence convex). It is also equivalent to a problem of maximizing a convex (linear) function subject to a convex constraints set. Hence, if the feasible set C is compact and nonempty, then by the theorems from last week, **there exists at least one optimal solution which is an extreme point, or equivalently, a basic feasible solution**.

- Convex Quadratic Problems consist of minimizing a convex quadratic function subject to affine constraints. The general form is

$$\begin{array}{ll} \min & \mathbf{x}^\top \mathbf{Q}\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{c} \end{array}$$

  $\mathbf{Q} \in \mathbb{R}^{n \times n}$ is positive semidefinite, $\mathbf{b} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{c} \in \mathbb{R}^m$. Convex QP problems frequently occur in statistics and machine learning.

Optimization over a Convex Set and Stationarity

We will consider the constrained optimization problem given by

$$\min_{\mathbf{x}} \quad \{f(\mathbf{x}) : \mathbf{x} \in C\}, \tag{P}$$

where C is a closed convex subset of $\mathbb{R}^n$, and f is continuously differentiable over C, not necessarily convex.

When we looked at unconstrained optimization, we showed that a necessary condition for $\mathbf{x}^*$ to be a local optima is that $\nabla f(\mathbf{x}^*) = 0$. We called points for which the gradient was zero *stationary points*. For constrained optimizations problems such as (P), we need an alternative definition of what it means to be stationary. Instead of defining stationarity solely in terms of the function f, we have to consider stationary points of the problem (P). Once we've defined this new concept of stationarity for convex problems, we will then prove similar results as in the unconstrained case, namely that stationarity is a necessary condition for optimality, and that under certain conditions (convexity of f) it is also a sufficient condition for optimality.

Definition (Stationarity)**.** *Let f be a continuously differentiable function over a closed and convex set C. Then $\mathbf{x}^*$ is called a stationary point of* (P) *if*

$$\nabla f\left(\mathbf{x}^*\right)^\top \left(\mathbf{x}-\mathbf{x}^*\right) \geq 0\,, \qquad \textit{for any } \mathbf{x} \in C\,.$$

Intuition: this condition says that if $\mathbf{x}^*$ is stationary, then there are no feasible descent directions of f at $\mathbf{x}^*$, i.e., there is no direction $\mathbf{d}$ such that the directional derivative $f'(\mathbf{x}^*, \mathbf{d}) = \nabla f(\mathbf{x}^*)^\top \mathbf{d} < 0$ and $\mathbf{x}^* + \mathbf{d} \in C$. This suggests that stationarity is a necessary condition for $\mathbf{x}^*$ to be a local minima of (P).

Theorem (Stationarity as a Necessary Optimality Condition)**.** *Let f be a continuously differentiable function over a nonempty closed convex set C, and let $\mathbf{x}^*$ be a local minimum of* (P) *Then $\mathbf{x}^*$ is a stationary point of* (P).

Proof. Let $\mathbf{x}^*$ be a local minimum of (P), and assume in contradiction that $\mathbf{x}^*$ is not a stationary point of (P). This implies that there exists $\mathbf{x} \in C$ such that

$$\nabla f\left(\mathbf{x}^*\right)^\top \left(\mathbf{x}-\mathbf{x}^*\right) < 0\,.$$

Thus, $f'\left(\mathbf{x}^*; \mathbf{d}\right) < 0$, where $\mathbf{d} = \mathbf{x} - \mathbf{x}^*$. Therefore, there exists $\varepsilon \in (0,1)$ such that $f\left(\mathbf{x}^* + t\mathbf{d}\right) < f\left(\mathbf{x}^*\right)$, $\forall t \in (0, \varepsilon)$. Finally, since $\mathbf{x}^* + t\mathbf{d} = (1-t)\mathbf{x}^* + t\mathbf{x} \in C$, $\forall t \in (0, \varepsilon)$, we conclude that $\mathbf{x}^*$ is not a local optimum point of (**P**). This contradicts out initial assumption. □

Examples of Stationarity Conditions

- For unconstrained problems, i.e., where $C = \mathbb{R}^n$, we can show that the new concept of stationarity is equivalent to the stationarity condition we studied previously for unconstrained problems (i.e. that $\nabla f(\mathbf{x}^*) = 0$).

  Firstly, assume that $\mathbf{x}^*$ is a stationary point of (P), i.e., that

  $$\nabla f\left(\mathbf{x}^*\right)^\top \left(\mathbf{x}-\mathbf{x}^*\right) \geq 0 \quad \forall \mathbf{x} \in \mathbb{R}^n\,.$$

  Plugging $\mathbf{x} = \mathbf{x}^* - \nabla f\left(\mathbf{x}^*\right)$ in the above implies

  $$-\left\|\nabla f\left(\mathbf{x}^*\right)\right\|^2 \geq 0$$

  and thus $\nabla f\left(\mathbf{x}^*\right) = 0$.

  Conversely, if $\nabla f\left(\mathbf{x}^*\right) = 0$, then obviously the inequality in the new definition of stationarity is statisified. We've thus proved that when $C = \mathbb{R}^n$

  $$\nabla f\left(\mathbf{x}^*\right) = 0 \text{ if and only if } \nabla f\left(\mathbf{x}^*\right)^\top \left(\mathbf{x}-\mathbf{x}^*\right) \geq 0 \quad \forall \mathbf{x} \in \mathbb{R}^n\,.$$

- For $C = \mathbb{R}^n_+$, $\mathbf{x}^*$ is a stationary point iff

$$\nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{x} \in \mathbb{R}^n_+ .$$

  This is equivalent to

$$\nabla f(\mathbf{x}^*)^\top \mathbf{x} - \nabla f(\mathbf{x}^*)^\top \mathbf{x}^* \geq 0 \text{ for all } \mathbf{x} \geq 0.$$

  Noting that

$$\mathbf{a}^\top \mathbf{x} + b \geq 0 \text{ for all } \mathbf{x} \geq 0 \text{ if and only if } \mathbf{a} \geq 0 \text{ and } b > 0$$

  we can see that this is equivalent to $\nabla f(\mathbf{x}^*) \geq 0$ and $\nabla f(\mathbf{x}^*)^\top \mathbf{x}^* \leq 0$. This is then equivalent to

$$\nabla f(\mathbf{x}^*) \geq 0 \qquad \text{and} \qquad x_i^* \frac{\partial f}{\partial x_i}(\mathbf{x}^*) = 0, \quad i = 1, 2, \ldots, n,$$

  which can be summarized as

$$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) \begin{cases} = 0 & x_i^* > 0 \\ \geq 0 & x_i^* = 0. \end{cases}$$

  If x_i^* is in the interior of C then $\mathbf{x}^*$ is only stationary if the partial derivative $\frac{\partial f}{\partial x_i}(\mathbf{x}^*)$ is zero, but if x_i^* is on the boundary of C (i.e. $x_i^* = 0$) then we only require the partial derivative to be non-negative[1].

The following table summarizes some important stationarity conditions:

Feasible Set	Explicit Stationarity Condition
$\mathbb{R}^n$	$\nabla f(\mathbf{x}^*) = \mathbf{0}$
$\mathbb{R}^n_+$	$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) \begin{cases} = 0 & x_i^* > 0 \\ \geq 0 & x_i^* = 0 \end{cases}$
$\{\mathbf{x} \in \mathbb{R}^n : \mathbf{e}^\top \mathbf{x} = 1\}$	$\frac{\partial f}{\partial x_1}(\mathbf{x}^*) = \ldots = \frac{\partial f}{\partial x_n}(\mathbf{x}^*)$
$B[\mathbf{0}, 1]$	$\nabla f(\mathbf{x}^*) = \mathbf{0}$ or $\lVert \mathbf{x}^* \rVert = 1$ and $\exists \lambda \leq 0 : \nabla f(\mathbf{x}^*) = \lambda \mathbf{x}^*$

where $\mathbf{e} = \begin{pmatrix} 1 & 1 & \ldots & 1 \end{pmatrix}^\top$.

We've shown for constrained optimization problems where the feasible set C is convex, that stationarity is a necessary condition for a point to be a local minima. For convex problems, i.e., when the objective f is also a convex function, stationarity is also a sufficient condition.

[1] If it was negative, it would suggest that moving away from $x_i = 0$ to $x_i > 0$ would yield a smaller value of f.

Theorem (Stationarity in Convex Optimization). *Let f be a continuously differentiable convex function over a nonempty closed and convex set $C \subseteq \mathbb{R}^n$. Then $\mathbf{x}^*$ is a stationary point of* (P) *iff $\mathbf{x}^*$ is an optimal solution of* (P)*.*

Proof. If $\mathbf{x}^*$ is an optimal solution of (P), then we already showed that it is a stationary point of (P). Assume that $\mathbf{x}^*$ is a stationary point of (P). Let $\mathbf{x} \in C$. Then

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq f(\mathbf{x}^*)$$

establishing the optimality of $\mathbf{x}^*$. The first inequality above follows from the gradient inequality for convex functions (Equation **??**), and the second from the definition of stationarity. □

The Orthogonal Projection Operator

We now introduce the orthogonal projection operator $P_C(\mathbf{x})$, which returns the point in C closest to $\mathbf{x}$.

Definition (Orthogonal Projection). *Given a nonempty closed convex set C, the orthogonal projection operator $P_C : \mathbb{R}^n \to C$ is defined by*

$$P_C(\mathbf{x}) = \operatorname*{argmin}_{\mathbf{y} \in C} \left\{ \|\mathbf{y} - \mathbf{x}\|^2 \right\} .$$

Note that as $f(\mathbf{y}) = \|\mathbf{y} - \mathbf{x}\|^2$ is a strictly convex function of $\mathbf{y}$, finding the orthogonal projection operator requires us to solve a convex optimization problem. Because the feasible set C is convex, we can prove that this problem has a unique optimal solution.

Theorem (The First Projection Theorem). *Let $C \subseteq \mathbb{R}^n$ be a nonempty closed and convex set. Then for any $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection $P_C(\mathbf{x})$ exists and is unique.*

Proof. $f(\mathbf{y}) = \|\mathbf{y} - \mathbf{x}\|^2$ is a coercive function, hence the minimum is attained in C, and moreover, $f(\mathbf{y})$ is strictly convex[2] hence there is only one optimal solution. □

[2] It is a quadratic function with positive definite matrix

$$f(\mathbf{y}) = (\mathbf{y} - \mathbf{x})^\top \mathbf{I}(\mathbf{y} - \mathbf{x})$$

.

Examples of Orthogonal Projections: In general, computing the orthogonal projection operator $P_C(\mathbf{x})$ can be difficult, but in some cases it can be computed explicitly.

- For $C = \mathbb{R}^n$ we have $P_C(\mathbf{x}) = \mathbf{x}$.
- For $C = \mathbb{R}^n_+$, we need to solve

$$\begin{aligned} \min_{\mathbf{y}} ||\mathbf{y} - \mathbf{x}|| &= \textstyle\sum (y_i - x_i)^2 \\ \text{s.t. } y_i &\geq 0 \text{ for all } i \end{aligned}$$

  The objective is a sum, and the constraints are separable, and so we need to just solve the problems

$$\begin{aligned} \min_{y_i} \quad & (y_i - x_i)^2 \\ \text{s.t. } y_i \quad & \geq 0. \end{aligned}$$

  These are solved at

$$y_i^* = \max(x_i, 0).$$

  Thus we have

$$P_{\mathbb{R}^n_+}(\mathbf{x}) = [\mathbf{x}]_+$$

  where $[\mathbf{x}]_+ = (\max\{x_1, 0\}, \max\{x_2, 0\}, \ldots, \max\{x_n, 0\})^\top$.
- For the closed ball in $\mathbb{R}^n$, $C = B[0, r]$, we can show that

$$P_{B[0,r]} = \begin{cases} \mathbf{x} & ||\mathbf{x}|| \leq r \\ r\frac{\mathbf{x}}{||\mathbf{x}||} & ||\mathbf{x}|| > r \end{cases}$$

There is an important geometric characterization of the projection operator. It says that for $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in C$, the angle between $\mathbf{x} - P_C(\mathbf{x})$ and $\mathbf{y} - P_C(\mathbf{x})$ is greater than or equal to 90°.

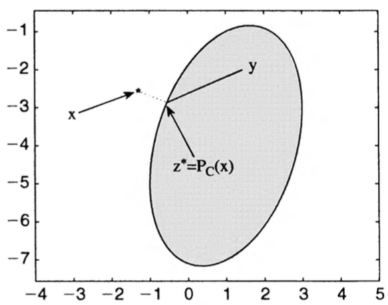

Figure 1: The orthogonal projection operator onto a convex set C. If $\mathbf{z}^* = P_C(\mathbf{x})$, then the angle between $\mathbf{x} - \mathbf{z}^*$ and $\mathbf{y} - \mathbf{z}^*$ for $\mathbf{y} \in C$ is always greater than 90°.

Theorem (The Second Projection Theorem)**.** *Let C be a nonempty closed convex set and let $\mathbf{x} \in \mathbb{R}^n$. Then $\mathbf{z} = P_C(\mathbf{x})$ if and only if*

$$(\mathbf{x} - \mathbf{z})^\top (\mathbf{y} - \mathbf{z}) \leq 0, \quad \textit{for any } \mathbf{y} \in C$$

Proof. $P_C(\mathbf{x})$ is the solution of the convex optimization problem

$$\begin{array}{ll} \min_{\mathbf{y}} & f(\mathbf{y}) = ||\mathbf{y} - \mathbf{x}||^2 \\ \text{s.t.} & \mathbf{y} \in C. \end{array}$$

Thus $\mathbf{z} = P_C(\mathbf{x})$ if and only if it is a stationary point, i.e.,

$$\nabla f(\mathbf{z})^\top (\mathbf{y} - \mathbf{z}) \geq 0 \text{ for all } \mathbf{y} \in C.$$

But $\nabla f(\mathbf{z}) = 2(\mathbf{z} - \mathbf{x})$ giving the result as desired. □

Finally, an important result in convex optimization is the representation of stationarity using the orthogonal projection operator.

Theorem (Representation of Stationarity via the Orthogonal Projection Operator)**.** *Let f be a continuously differentiable function over the nonempty closed convex set C, and let $s > 0$. Then x^* is a stationary point of* (P) *if and only if*

$$\mathbf{x}^* = P_C\left(\mathbf{x}^* - s\nabla f\left(\mathbf{x}^*\right)\right) .$$

Proof. By the second projection theorem, $\mathbf{x}^* = P_C\left(\mathbf{x}^* - s\nabla f\left(\mathbf{x}^*\right)\right)$ iff

$$\left(\mathbf{x}^* - s\nabla f\left(\mathbf{x}^*\right) - \mathbf{x}^*\right)^\top \left(\mathbf{y} - \mathbf{x}^*\right) \leq 0 \text{ for any } \mathbf{y} \in C\,,$$

which is equivalent to

$$\nabla f\left(\mathbf{x}^*\right)^\top \left(\mathbf{y} - \mathbf{x}^*\right) \geq 0 \text{ for any } \mathbf{y} \in C\,,$$

namely, the definition of stationarity. □

The Gradient Projection Method

The result that x^* is a stationary point of (P) if and only if

$$\mathbf{x}^* = P_C\left(\mathbf{x}^* - s\nabla f\left(\mathbf{x}^*\right)\right) .$$

suggests we can find stationary points by solving this fixed-point problem. The gradient projection method is a gradient descent type algorithm that uses the orthogonal projection operator to solve this problem:

Algorithm 1: The Gradient Projection Method

Initialization: A tolerance parameter $\varepsilon > 0$ and $\mathbf{x}^0 \in C$.
General Step: for any $k = 0, 1, 2, \ldots$ execute the following steps:
1 Pick a stepsize t^k by a line search procedure.
2 Set $\mathbf{x}^{k+1} = P_C\left(\mathbf{x}^k - t^k\nabla f\left(\mathbf{x}^k\right)\right)$.
3 If $\left\|\mathbf{x}^k - \mathbf{x}^{k+1}\right\| \leq \varepsilon$, then STOP and $\mathbf{x}^{k+1}$ is the output.

Another way of writing this is to split step 2 into two steps:

2a $\mathbf{y}^{k+1} = \mathbf{x}^k - t^k \nabla f\left(\mathbf{x}^k\right)$

2b $\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}\in C} ||\mathbf{x} - \mathbf{y}^{k+1}||$

Note that in the unconstrained case $P_C(\mathbf{x}) = \mathbf{x}$ and so the algorithm simplifies to

$$\mathbf{x}^{k+1} = \mathbf{x}^k - t^k \nabla f\left(\mathbf{x}^k\right),$$

i.e., the gradient descent method we studied at the start of the module. In the constrained case, the algorithm essentially does the usual gradient descent step $\mathbf{x}^k - t^k \nabla f\left(\mathbf{x}^k\right)$ and then projects this point back into C using the orthogonal projection P_C if it lies outside of C.

As before, there are several strategies for choosing the stepsizes t^k. When $f \in C_L^{1,1}$, we can choose t^k to be constant and equal to $\frac{1}{L}$. An alternative is to use backtracking.

Before describing backtracking, lets first define the gradient mapping to be

$$G_M(\mathbf{x}) = M\left[\mathbf{x} - P_C\left(\mathbf{x} - \frac{1}{M}\nabla f(\mathbf{x})\right)\right],$$

where $M > 0$ is a positive constant. In the unconstrained case $G_M(\mathbf{x}) = \nabla f(\mathbf{x})$, so the gradient mapping is an extension of the usual gradient operation. By the previous theorem, $G_M(\mathbf{x}) = \mathbf{0}$ if and only if $\mathbf{x}$ is a stationary point of (P). This means that we can consider $||G_M(\mathbf{x})||^2$ to be an optimality measure, i.e., for the stationary points $||G_M(\mathbf{x})||^2 = 0$. Otherwise $||G_M(\mathbf{x})||^2 > 0$.

The constrained version of the backtracking rule we studied in the unconstrained case, gives the following algorithm:

Algorithm 2: The Gradient Projection Method with Backtracking

---

Initialization: A tolerance parameter $\varepsilon > 0$ and $\mathbf{x}^0 \in C$. Parameters $s > 0$, $\alpha \in (0,1)$, and $\beta \in (0,1)$.

General Step: for any $k = 0, 1, 2, \ldots$ execute the following steps:

Pick $t^k = s$.

While $f(\mathbf{x}^k) - f(P_C(\mathbf{x}^k - t^k\nabla f(\mathbf{x}^k))) < \alpha t^k \left\|G_{\frac{1}{t^k}}(\mathbf{x}^k)\right\|^2$, set $t^k := \beta t^k$.

Set $\mathbf{x}^{k+1} = P_C\left(\mathbf{x}^k - t^k\nabla f\left(\mathbf{x}^k\right)\right)$.

If $\left\|\mathbf{x}^k - \mathbf{x}^{k+1}\right\| \le \varepsilon$, then STOP and $\mathbf{x}^{k+1}$ is the output.

Finally, if we make some assumptions about f then it is possible to prove (but beyond the scope of the module) that the gradient projection method will converge.

Theorem (Convergence of the Gradient Projection Method)**.** *Let $\left\{\mathbf{x}^k\right\}$ be the sequence generated by the gradient projection method for solving problem* (P) *with either a constant stepsize $\bar{t} \in \left(0, \frac{2}{L}\right)$, where L is a Lipschitz constant of ∇f (i.e. $f \in C_L^{1,1}(C)$) or a backtracking stepsize strategy. Assume that f is bounded below. Then:*

1. *The sequence $\left\{f\left(\mathbf{x}^k\right)\right\}$ is nonincreasing.*

2. $G_d\left(\mathbf{x}^k\right) \to 0$ *as* $k \to \infty$, *where*

$$d = \begin{cases} 1/\bar{t} & \textit{constant stepsize} \\ 1/s & \textit{backtracking.} \end{cases}$$

Checklist

The idea of this checklist is to help you to self-evaluate your progress and understanding of the subject, and to give you some guidance on where to focus. If you can tick all the boxes it means you're doing alright, otherwise you need to study a bit more, grab a book, watch the videos, or seek help from classmates, the lecturers, or the demonstrators. Try to fill as many gaps as quickly as possible.

And remember to do the exercises!

Learning Outcome	Check
I understand the links between convexity, minimizers, and maximizers.	
I can identify a convex optimization problem.	
I understand stationarity in convex constrained optimization as a generalization of the stationarity concept studied in Week 3.	
I can compute stationarity conditions for different constraints.	
I understand the link between stationarity, orthogonal projection, and projected gradient descent.	

Exercises

1. Let $g : S \to \mathbb{R}$ be a function defined over $S \subseteq \mathbb{R}^n$. The **level set** of g with level a, is

$$\mathrm{Lev}(g, a) = \{\mathbf{x} : g(\mathbf{x}) \le a\}.$$

   Show that the level sets of convex functions are convex.

2. Consider the functional form for optimization problems:

$$\begin{array}{lll} \min & f(\mathbf{x}) & \\ \text{s.t.} & g_i(\mathbf{x}) \le 0, & i = 1, 2, \dots, m \\ & h_j(\mathbf{x}) = 0, & j = 1, 2, \dots, p, \end{array}$$

   where $f, g_1, \dots, g_m : \mathbb{R}^n \to \mathbb{R}$ are convex functions, and $h_1, h_2, \dots, h_p : \mathbb{R}^m \to \mathbb{R}$ are affine functions.

   Show that this functional form of the problem does fit into the general formulation (CVX), i.e., show that

$$C = \{\mathbf{x} : g_i(\mathbf{x}) \le 0, h_j(\mathbf{x}) = 0\}$$

   is a convex set.

3. Prove that for convex optimization problems $\min\{f(\mathbf{x}) : \mathbf{x} \in C\}$ where f and C are convex, the set of optimal solutions is also convex. In addition, prove that if f is strictly convex over C, then there exists at most one optimal solution of the problem.

4. Show that

$$\begin{array}{ll} \min & -2x_1 + x_2 \\ \text{s.t.} & x_1^2 + x_2^2 \le 3 \end{array}$$

   is a convex optimization problem. Show that

$$\begin{array}{ll} \min & x_1^2 - x_2 \\ \text{s.t.} & x_1^2 + x_2^2 = 3 \end{array}$$

   is a non-convex problem.

5. **Maximization of convex functions:** Find the maximum of the following convex maximization problems:

   a)
$$\max\{\mathbf{x}^\top \mathbf{Q}\mathbf{x} : ||\mathbf{x}||_\infty \le 1\}$$

   where $\mathbf{Q} \succeq 0$.

   b)
$$\max\{\, ||\mathbf{A}\mathbf{x}||_1 : ||\mathbf{x}||_1 \le 1\}$$

   where $\mathbf{A} \in \mathbb{R}^{m \times n}$.

6. Solve

$$\begin{array}{ll} \max_{\mathbf{x}} & 2x_1 + 3x_2 - x_3 \\ & x_1 + x_2 + x_3 = 6 \\ \text{such that} & \quad x_2 + x_4 = 3 \\ & x_1, x_2, x_3, x_4 \geq 0 \end{array}$$

What is the minimum of this problem?

7. Solve

$$\begin{array}{ll} \min_{\mathbf{x}} & x_1 + x_2 \\ & 2x_1 + x_2 \leq 2 \\ \text{such that} & x_2 + x_3 \leq 3 \\ & \quad x_1, x_2, x_3 \geq 0 \end{array}$$

Hint: you can introduce new variables $x_4, x_5 \geq 0$ to convert the inequality constraints into equality constraints.

8. Consider the convex optimization problem

$$\min\{f(\mathbf{x}) : \mathbf{e}^\top \mathbf{x} = 1\}$$

where f is a convex function. Show that the stationarity condition for this problem is equivalent to

$$\frac{\partial f}{\partial x_1}(\mathbf{x}^*) = \ldots = \frac{\partial f}{\partial x_n}(\mathbf{x}^*)$$

9. Let C be a closed and convex set. Prove that the distance function from the set $d_C(\mathbf{x}) \equiv \inf_{\mathbf{y} \in C} \|\mathbf{x} - \mathbf{y}\|$ is a convex function.

10. A box is a subset of $\mathbb{R}^n$ of the form

$$B = [\ell_1, u_1] \times [\ell_2, u_2] \times \cdots \times [\ell_n, u_n] = \{\mathbf{x} \in \mathbb{R}^n : \ell_i \leq x_i \leq u_i\} \ ,$$

where $\ell_i \leq u_i$ for all $i = 1, 2, \ldots, n$. Show that the orthogonal projection for this set is

$$[P_B(\mathbf{x})]_i = \begin{cases} u_i & x_i \geq u_i \\ x_i & \ell_i < x_i < u_i \\ \ell_i & x_i \leq \ell_i \end{cases}$$