Chapter 1

Introduction

Optimization problems arise naturally in many computer vision and machine learning applications that estimate model parameters from input images, videos, range sensors, or training data. These model parameters can represent pixels of a noise-free image, the motion of the objects on a scene, a surface scanned with a range sensor, or the filter of neural network. To find the model parameters that best fit the observed data and satisfy the constraints of the physical world, we define an optimization problem that penalizes deviation between the observed data and the behavior of the model. Finding a solution of the resulting optimization problem is far from trivial.

Most optimization problems are unsolvable. Among the solvable ones, convex problems form a large subset that comes with useful mathematical properties and can be solved efficiently with numerical algorithms that exploit these properties. Most commercial packages for optimization, however, are designed for problems with differentiable objective functions that poorly fit computer vision and machine learning models. Vision and machine learning problems are characterized, not by their differentiablity, but by their use of large amounts of data and by objective functions that combine data-fidelity terms and regularization penalties. The goal of this course is to present techniques that exploit the structure of this problems to develop efficient algorithms for a large set of computer vision and machine learning problems.

In this course, we will not discuss how to design models for a particular vision or machine learning problem. The models will be given to us and we will only try to understand the properties of each model in terms of their computational cost and algorithmic choices. This is important because very often we have to choose between a *good* model, which we cannot solve, and a *reasonable* model which can be solved efficiently. To distinguish between the two, it is necessary to be aware of some theory that explains what we can and what we cannot do in optimization. Convexity plays a key role on this discussion.

This first chapter is a summary of Chapter 1 of *Introductory Lectures on Convex Optimization*, by Nesterov.

1.1 Why Convex Optimization

Let us start by describing our optimization problem. Let $u \in \mathbb{R}^n$ be an n-dimensional real vector, $C \subset \mathbb{R}^n$ be a subset of $\mathbb{R}^n$, and E be a real-valued functions of u. We study different variants of the following general minimization problem:

$$\hat{u} \in \arg\min_{u \in C} E(u) \tag{1.1}$$

The function $E : \mathbb{R}^n \to \mathbb{R}$ is the objective function, while the set C is the feasible set. We consider a minimization problem by convention, but we can also consider a maximization problem with $-E$ as objective function.

There is a natural classification of the types of minimization problems that we will study: unconstrained problems where $C = \mathbb{R}^n$, smooth problems where E is differentiable, and non-smooth problems where E is not differentiable. We also distinguish two different types of solutions to the minimization problem.

Definition $u^\star$ is a **global solution** of

$$\hat{u} \in \arg\min_{u \in C} E(u)$$

if $E(u^\star) \geq E(u)$ for all $u \in C$.

Definition $u^\star$ is a **local solution** of

$$\hat{u} \in \arg\min_{u \in C} E(u)$$

if there exists a $r > 0$ such that

$$E(u^\star) \geq E(u) \quad \forall u \in C,\ \|u - u^\star\| < r.$$

Local minima are easier to find that global ones. For instance, given an estimate of the the minimizer u^0, we can create a sequence $u_0, u_1, u_2, \ldots = \{u^k\}_{k=1}^\infty$ that decreases the value of the energy at each step. Formally, we say that these type of optimization methods create a relaxation sequence $\{E(u^k)\}_{k=1}^\infty$ that satisfies $E(u^{k+1}) \leq E(u^k)$ and always improves the initial value of the objective function. If E is bounded below on $\mathbb{R}^n$, then the sequence $\{E(u^k)\}_{k=1}^\infty$ converges to a local minimum. Let us formalize what we mean by convergence.

Definition We say that a sequence $\{a^k\}_{k=1}^\infty \subset \mathbb{R}^n$ converges to $\hat{a} \in \mathbb{R}^n$ if for all $\epsilon > 0$ there exists an $k_0 \in \mathbb{N}$ such that

$$\|a^k - \hat{a}\| < \epsilon \quad \forall k \geq k_0.$$

To implement the idea of relaxation we use another fundamental principle of numerical analysis, approximation. In an approximation strategy, we replace the original objective function E by a simplified objective function that is close to the original. When the objective function is differentiable, the approximation is usually local and determined by its Taylor expansion at the current estimate. For instance, let $E(u)$ be differentiable at u^0, then for $u \in \mathbb{R}^n$, we have

$$E(u) = E(u^0) + \langle \nabla E(u^0), u - u^0 \rangle + o(\|u - u^0\|) \ \text{ where } \ \lim_{r \to 0} \frac{o(r)}{r} = 0.$$

The function

$$E(u; u^0) = E(u^0) + \langle \nabla E(u^0), u - u^0 \rangle$$

is a linear approximation of E in a neighborhood of u^0. Given an initial estimate of the minimizer u^0, we can then use this linear approximation to reduce the value of $E(u)$ in a neighborhood of u^0 at each time step. In particular we can decide to iteratively step in the direction of maximum descent with a constant stepsize τ as follows:

$$\begin{aligned} u^1 &= u^0 - \tau \nabla E(u^0) \\ u^2 &= u^1 - \tau \nabla E(u^1) \\ &\cdots \\ u^{k+1} &= u^k - \tau \nabla E(u^k). \end{aligned}$$

This gives us a very simple algorithm know as gradient descent. We will see in this course that, under certain conditions, the algorithm creates a relaxation sequence that decreases the value of the objective function and converges to a point $\hat{u} \in \mathbb{R}^n$. This point then satisfies $\hat{u} = \hat{u} - \tau \nabla E(\hat{u}) \Rightarrow \nabla E(\hat{u}) = 0$. This is a necessary condition for optimality, as the next theorem shows.

Theorem 1. First-order Optimality Condition. *Let $u^\star$ be a local minimum of differentiable function $E(u)$. Then $\nabla E(u^\star) = 0$.*

Proof. Since $u^\star$ is a local minimum of $E(u)$, then there exists $r > 0$ such that for all v with $\|v - u^\star\| \leq r$, we have $E(v) \geq E(u^\star)$. Since E is differentiable, this implies that

$$E(v) = E(u^\star) + \langle \nabla E(u^\star), v - u^\star \rangle + o(\|v - u^\star\|) \geq E(u^\star).$$

Thus, for all $s = v - u^\star$ we have $\langle \nabla E(u^\star), s \rangle \geq 0$. If we consider the directions s and $-s$, we get $\nabla E(u^\star) = 0$. □

Note that we have proved only a necessary condition of a local minimum. The points satisfying this condition are called the stationary points of function. In order to see that such points are not always the local minima, it is enough to look at function $E(u) = u^3$. The optimality condition $E'(u) = 3u^2 = 0$ suggests that 0 should be a local minimum, even though $E'(u) > 0$ for all u and the function is always decreasing and cannot have a minimum at 0. The point 0 is in fact a stationary point, not a maximum or minimum.

To discern between local minima and stationary points of a function, let us introduce the second-order approximation. Let function $E(u)$ be twice differentiable with Hessian $\nabla^2 E(u)$ at u. Then

$$E(v) = E(u) + \langle \nabla E(u), v - u \rangle + \frac{1}{2} \langle \nabla^2 E(u)(v - u), v - u \rangle + o(\|v - u\|^2).$$

The function

$$E(v; u) = E(u) + \langle \nabla E(u), v - u \rangle + \frac{1}{2} \langle \nabla^2 E(u)(v - u), v - u \rangle$$

is the quadratic or second-order approximation of function E at u. The Hessian is a symmetric matrix that can be seen as a derivative of the vector function ∇E. As a result, using a linear approximation to each component of ∇E, we have

$$\nabla E(v) = \nabla E(u) + \nabla^2 E(u)(v - u) + o(\|v - u\|).$$

Using the second-order approximation, we can write down the second-order optimality conditions.

Theorem 2. Second-order Optimality Condition *Let $u^\star$ be a local minimum of twice differentiahte function $E(u)$. Then $\nabla E(u^\star) = 0$ and $\nabla^2 E(u^\star)$ is symmetric and positive semi-definite, that we denote by $\nabla^2 E(u^\star) \succeq 0$.*

Proof. Since $u^\star$ is a local minimum of function E, there exists $r > 0$ such that

$$E(u) \geq E(u^\star) \quad \forall u \quad \text{with} \quad \|u - u^\star\| < r.$$

The first order optimality condition gives us $\nabla E(u^\star) = 0$ and, as a result

$$E(u) = E(u^\star) + \frac{1}{2} \langle \nabla^2 E(u^\star)(v - u^\star), v - u^\star \rangle + o(\|v - u^\star\|^2) \geq E(u^\star).$$

Thus, $\langle \nabla^2 E(u^\star)(v - u^\star), v - u^\star \rangle \geq 0$. Letting $s = v - u^\star$ we have $\langle \nabla^2 E(u^\star) s, s \rangle \geq 0$, which implies positive semi-definiteness of $\nabla^2 E(u^\star)$. □

This second-order characteristic of a local minimum is also sufficient.

Theorem 3. *Let function $E(u)$ be twice differentiable on $\mathbb{R}^n$ and let $u^\star$ satisfy $\nabla E(u^\star) = 0$ and $\nabla^2 E(u^\star) \succ 0$. Then $u^\star$ is a strict local minimum of E.*

Proof. In a small neighborhood of $u^\star$, $E(u)$ can be represented as

$$E(u) = E(u^\star) + \frac{1}{2}\langle \nabla^2 E(u^\star)(u-u^\star), u-u^\star\rangle + o(\|u-u^\star\|^2).$$

Since $\lim_{r\to 0}\frac{o(r)}{r} = 0$, there exists a value $\bar{r}$ such that for all $r \in [0,\bar{r}]$ we have

$$|o(r)| \leq \frac{r}{4}\lambda_1,$$

where $\lambda_1 > 0$ is the smallest eigenvalue of matrix $\nabla^2 E(u^\star)$. As $\nabla^2 E(u^\star)$ is symmetric and positive definite, it has positive eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n > 0$ and orthogonal eigenvectors $q_1, q_2, \ldots q_n$, such that $\nabla^2 E(u^\star) = \sum_{1\leq i\leq n} \lambda_i q_i q_i^T$ and $\|q_i^T v\| = \|v\|$ for all $v \in \mathbb{R}^n$. As a result,

$$\begin{aligned}
E(u) &= E(u^\star) + \frac{1}{2}\langle \nabla^2 E(u^\star)(u-u^\star), u-u^\star\rangle + o(\|u-u^\star\|^2) \\
&= E(u^\star) + \frac{1}{2}\sum_{1\leq i\leq n} \lambda_i \langle q_i q_i^T (u-u^\star), u-u^\star\rangle + o(\|u-u^\star\|^2) \\
&= E(u^\star) + \frac{1}{2}\sum_{1\leq i\leq n} \lambda_i \langle q_i^T (u-u^\star), q_i^T (u-u^\star)\rangle + o(\|u-u^\star\|^2) \\
&= E(u^\star) + \frac{1}{2}\sum_{1\leq i\leq n} \lambda_i \|q_i^T (u-u^\star)\|^2 + o(\|u-u^\star\|^2) \\
&= E(u^\star) + \frac{1}{2}\sum_{1\leq i\leq n} \lambda_i \|u-u^\star\|^2 + o(\|u-u^\star\|^2) \\
&\geq E(u^\star) + \frac{\lambda_1}{2}\|u-u^\star\|^2 + o(\|u-u^\star\|^2) \\
&\geq E(u^\star) + \frac{\lambda_1}{4}\|u-u^\star\|^2 \geq E(u^\star).
\end{aligned} \tag{1.2}$$

□

For general optimization problems, we thus require second-order differentiablity to formulate necessary and sufficient optimality conditions. The optima described by these conditions is, moreover, only local. This is quite disappointing because most applications in computer vision and machine learning have objective functions that are not differentiable and these general optimality conditions are meaningless. Even in the rare cases where second-order derivatives exists, computing the Hessian is not feasible because the size of the problem is too large. For these reasons, we resort to the field of convex optimization. In convex optimization optimality conditions are not only necessary but sufficient and the objective function does not need to be differentiable to find them.

1.2 Characterization of continuous and smooth functions

Assuming only differentiability of the objective function we cannot get many reasonable properties of minimization processes. We usually have to impose some additional assumptions on the magnitude of the derivatives. In optimization these kind of assumptions are presented in the form of a Lipschitz condition for a derivative of certain order. Among them, we will make heavy use of Lipschitz continuity and L-smoothness.

Let us first define Lipschitz continuity.

Definition A function $E : \mathbb{R}^n \to \mathbb{R}^m$ is **Lipschitz continuous** with *Lipschitz constant* L if for all $u,\ v \in \mathrm{dom}(E)$

$$\|E(u) - E(v)\|_2 \leq L\|u-v\|_2$$

A function is **locally Lipschitz continuous** if for every $u \in \mathrm{dom}(E)$ there exists $\epsilon > 0$ such that $f_{|B(\epsilon,u)}$ is Lipschitz continuous

If the a function is differentiable, then we can compute the Lipchitz constant of its gradient operator to measure its smoothness. This results in the definition of L-smoothness.

Definition ***L*-smooth function** A differentiable function $E\colon \mathbb{R}^n \to \mathbb{R}$ is L-smooth is

$$\|\nabla E(u) - \nabla E(v)\| \leq L\|u - v\| \quad \forall u, v \in \mathbb{R}^n.$$

If a function is twice-continusouly differentiable, a sufficient condition for L-smoothness is the following:

Lemma 4. *A twice-continuously differentiable function E is L-smooth if and only if $\|\nabla^2 E(u)\| \leq L \quad \forall u \in \mathbb{R}^n$.*

Proof. Let us first prove that any twice-continuously differentiable function with bounded Hessian is L-smooth. Given any $u, v \in \mathbb{R}^n$, we have the componentwise inequality

$$\begin{aligned}
\nabla E(v) &= \nabla E(u) + \nabla E(v) - \nabla E(u) \\
&= \nabla E(u) + \int_0^1 \nabla^2 E(u + \tau(v-u))(v-u)\mathrm{d}\tau \\
&= \nabla E(u) + \left(\int_0^1 \nabla^2 E(u + \tau(v-u))\mathrm{d}\tau\right)(v-u)
\end{aligned}$$

Re arranging terms and using Cauchy-Schwarz inequality,

$$\begin{aligned}
\|\nabla E(v) - \nabla E(u)\| &= \left\|\left(\int_0^1 \nabla^2 E(u + \tau(v-u))\mathrm{d}\tau\right)(v-u)\right\| \\
&\leq \left\|\int_0^1 \nabla^2 E(u + \tau(v-u))\mathrm{d}\tau\right\| \|v-u\| \\
&\leq \int_0^1 \|\nabla^2 E(u + \tau(v-u))\|\mathrm{d}\tau \|v-u\| \leq L\|v-u\|.
\end{aligned}$$

Let us now prove the other direction, that is, that a twice-continuously differentiable function that is L-smooth has bounded Hessian. As E is wice-continuously differentiable, we have

$$\left\|\left(\int_0^\alpha \nabla^2 E(u + \tau s)\mathrm{d}\tau\right)s\right\| = \|\nabla E(u + \alpha s) - \nabla E(u)\| \leq \alpha L\|s\|$$

Dividing this inequality by α and tending $\alpha \to 0$ we obtain $\|\nabla^2 E(u)\| \leq L$. □

The next statement is important for the geometric interpretation of L-smooth functions.

Lemma 5. *If E is L-smooth, then for any $u, v \in \mathbb{R}^n$*

$$|E(v) - E(u) - \langle \nabla E(u), v - u\rangle| \leq \frac{L}{2}\|v - u\|^2.$$

Proof. For all $u, v \in \mathbb{R}^n$, we have

$$\begin{aligned}
E(v) &= E(u) + \int_0^1 \langle \nabla E(u + \tau(v-u)), v - u\rangle \mathrm{d}\tau \\
&= E(u) + \langle \nabla E(u), v - u\rangle + \int_0^1 \langle \nabla E(u + \tau(v-u)) - \nabla E(u), v - u\rangle \mathrm{d}\tau
\end{aligned}$$

Re-arranging terms and using Cauchy-Schwarz inequality, we get

$$\begin{aligned}
|E(v) - E(u) - \langle \nabla E(u), v - u \rangle| &= \left| \int_0^1 \langle \nabla E(u + \tau(v-u)) - \nabla E(u), v - u \rangle \mathrm{d}\tau \right| \\
&\leq \int_0^1 |\langle \nabla E(u + \tau(v-u)) - \nabla E(u), v - u \rangle| \mathrm{d}\tau \\
&\leq \int_0^1 \|\nabla E(u + \tau(v-u)) - \nabla E(u)\| \|v - u\| \mathrm{d}\tau \\
&\leq \int_0^1 \tau L \|v - u\|^2 \mathrm{d}\tau \\
&= \frac{1}{2} L \|v - u\|^2
\end{aligned}$$

□

Geometrically, we can draw the following picture. Given a differentiable L-smooth function E and $u_0 \in \mathbb{R}^n$, we can define two quadratic functions

$$\begin{aligned}
\phi_1(u) &= E(u_0) + \langle \nabla E(u_0), u - u_0 \rangle - \frac{L}{2} \|u - u_0\|^2 \\
\phi_2(u) &= E(u_0) + \langle \nabla E(u_0), u - u_0 \rangle + \frac{L}{2} \|u - u_0\|^2
\end{aligned}$$

that upper and lower bound the function

$$\phi_1(u) \leq E(u) \leq \phi_2(u) \quad \forall u \in \mathbb{R}^n.$$

Chapter 2

Convex Analysis

2.1 Convex Optimization

We start this section discussing the unconstrained minimization problem

$$\min_{u \in \mathbb{R}^n} E(u). \tag{2.1}$$

In the general situation we cannot do too much: even when the function is smooth, the gradient method converges only to a stationary point of function E. To make the problem tractable we introduce a key assumption on the kind of functions E that we minimize.

We call for the following property: for any E differentiable, a point $\hat{u}$ is a global solution of

$$\min_{u \in \mathbb{R}^n} \mathbb{E}(u)$$

if and only if $\nabla E(\hat{u}) = 0$. That is, the first-order optimality condition is necessary and sufficient. We will see that convex functions come with this guarantee.

Definition A function $E\colon \mathbb{R}^n \to \mathbf{R}$ is a **convex function** if and only if for any $u, v \in \mathbb{R}^n$ and $\theta \in [0, 1]$

$$E(\theta u + (1-\theta)v) \leq \theta E(u) + (1-\theta)E(v).$$

E is strictly convex if the inequality is strict for all $\theta \in (0, 1)$, $v \neq u$.

The definition of convex functions implicitly assumes that it is possible to evaluate the function at any point of the segment

$$[u, v] = \{z = \theta u + (1-\theta)v \colon 0 \leq \theta \leq 1\}.$$

As a result, it is natural to consider a set that contains the whole segment between any two points in the set. Such sets are called convex.

Definition The set C is a **convex set** if for any $u, v \in C$ and $\theta \in [0, 1]$, $\theta u + (1-\theta)v \in C$.

We can then include this notion in the definition of convex functions with restricted domain.

Definition The **domain** of a function $E\colon \mathbb{R}^n \to \mathbb{R}$ is the set

$$\mathrm{dom}(E) = \{u \in \mathbb{R}^n \colon E(u) < \infty\}$$

We can now extend the definition of convexity to functions that can take infinity values. Formally, they are know as **extended real-valued functions**.

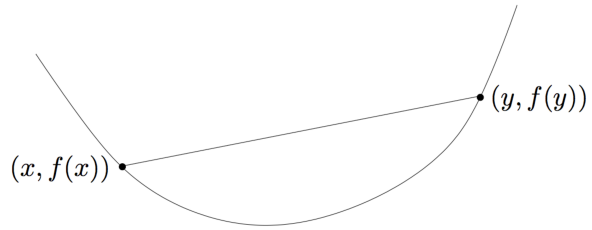


Fig. 2.1: Graph of a convex function f: the chord between two graph points lies above the graph of the function. Image source: Boyd, Vandenberghe, *Convex optimization theory*,2004.

Definition The extended real-valued function $E : \mathbb{R}^n \to \overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ is **convex** if

- its domain $\mathrm{dom}(E)$ is a convex set.
- For all $u, v \in \mathrm{dom}(E)$ and all $\theta \in [0, 1]$ it holds that

$$E(\theta u + (1 - \theta)v) \leq \theta E(u) + (1 - \theta)E(v).$$

E is **strictly convex** if the inequality is strict for all $\theta \in (0, 1)$, $v \neq u$.

In terms of the graph of the function, we can characterize convexity as follows: for any two points $(u, E(u))$ and $(v, E(v))$ in the graph of a convex function E, the chord between the two points lies above the graph of the function between these points, that is

$$\underbrace{E(\theta u + (1 - \theta)v)}_{\text{graph of the function in } [u,\, v]} \quad \leq \quad \underbrace{\theta E(u) + (1 - \theta)E(v)}_{\text{chord between } (u, E(u)) \text{ and } (v, E(v))} \quad .$$

This is illustrated in Figure **??** .

In the following we assume that the domain of E is not empty or E is proper.

Definition Function $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ is **proper** if its domain is not empty.

Now that we know what a convex set and a convex function is, we can define a convex minimization problem as an optimization problem of the form

$$\hat{u} \in \arg\min_{u \in C} E(u), \tag{2.2}$$

where C is a convex set and E is a convex function. To write such a problem in our familiar unconstrained optimization form, we define the **extended real-valued function** $\tilde{E}$ by introducing the constraint $u \in C$ into the domain of the original energy function E as follows:

$$\tilde{E} : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\} \qquad \tilde{E}(u) = \begin{cases} E(u) & \text{if } u \in C, \\ \infty & \text{else.} \end{cases}$$

We can then re-write

$$\hat{u} \in \arg\min_{u \in C} E(u)$$

as

$$\hat{u} \in \arg\min_{u \in \mathbb{R}^n} \tilde{E}(u).$$

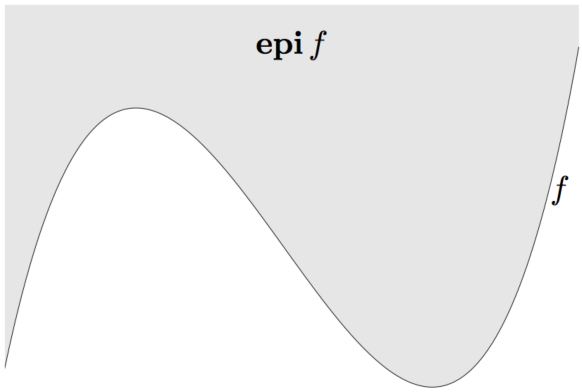


Fig. 2.2: Epigraph of a function f shaded. The lower boundary, in a darker shade, is the graph of the function. Image source: Boyd, Vandenberghe, *Convex optimization theory*,2004.

2.2 Convex Sets

We have already seen examples of convex sets as the domain of convex functions. We will show now this connection explicitly.

Lemma 6. *If E is a convex function, then for any $\beta \in \mathbb{R}$, its level set $\{z : E(z) \leq \beta\}$ is either convex or empty.*

Proof. Let $u, v \in \text{dom}(E)$ with $E(u) \leq \beta$ and $E(v) \leq \beta$, by convexity of E we have $\theta u + (1-\theta)v \in \text{dom}(E)$ and

$$E(\theta u + (1-\theta)v) \leq \theta E(u) + (1-\theta)E(v) \leq \theta\beta + (1-\theta)\beta = \beta.$$

We have just seen that if there are $u, v \in \{z : E(z) \leq \beta\}$, then $\theta u + (1-\theta)v \in \{z : E(z) \leq \beta\}$, that is, if the level set $\{z : E(z) \leq \beta\}$ is not empty, then it is convex. □

Lemma 7. *Let E be a convex function, then its **epigraph** $epi(E) = \{(u, \beta) : E(u) \leq \beta\}$ is a convex set.*

Proof. Let $p = (u, \alpha), q = (v, \beta) \in \text{epi}(E)$, that is $u, v \in \text{dom}(E)$ with $E(u) \leq \alpha$ and $E(v) \leq \beta$. As E is convex, its domain is convex and we have $\theta u + (1-\theta)v \in \text{dom}(E)$. At the same time, by convexity of E

$$E(\theta u + (1-\theta)v) \leq \theta E(u) + (1-\theta)E(v) \leq \theta\alpha + (1-\theta)\beta.$$

As a result

$$\begin{aligned} (\theta u + (1-\theta)v, \theta\alpha + (1-\theta)\beta) &\in \text{epi}(E) \\ \theta(u, \alpha) + (1-\theta)(v, \beta) &\in \text{epi}(E) \\ \theta q + (1-\theta)q &\in \text{epi}(E), \end{aligned} \tag{2.3}$$

which shows that $\text{epi}(E)$ is convex because given $p, q \in \text{epi}(E)$ we have seen that $\theta q + (1-\theta)q \in \text{epi}(E)$. □

Figure **??** illustrates the epigraph of a one-dimensional function.

To determine if a set is convex, a few properties are useful.

Lemma 8. *Let $C \subset \mathbb{R}^n, D \subset \mathbb{R}^m$ be convex sets and $\mathscr{A} : \mathbb{R}^n \to \mathbb{R}^m$ be a linear operator, then the following sets are convex:*

- *Intersection $C \cap D$.*

- *Sum* $C + D = \{u = x + y \ : \ x \in C,\ y \in D\}$ *if* $n = m$.
- *Affine image* $\mathscr{A}(C) = \{u \in \mathbb{R}^m \ : \ u = \mathscr{A}(x),\ x \in C\}$.
- *Inverse affine image* $\mathscr{A}^{-1}(D) = \{v \in \mathbb{R}^n \ : \ \mathscr{A}(v) \in D\}$.

Proof. Left as exercise. □

As a result of the previous lemma, the following sets are convex

- Half-space $\{u \in \mathbb{R}^n \ : \ \langle a, u\rangle \leq \beta\}$ is convex as a sublevel set of a linear function, which is a convex function.
- Polytope $\{u \in \mathbb{R}^n \ : \ \langle a_i, u\rangle \leq b_i\}$ is convex as an intersection of convex sets (half-spaces).
- Ellipsoid $\{u \in \mathbb{R}^n \ : \ \langle Au, u\rangle \leq 1 \text{ with } A \succeq 0\}$ is convex as a sub-level set of the convex function $\langle Au, u\rangle$.

The next theorem make the connection between convex sets and convex functions explicit.

Theorem 9. Convexity and Epigraphs. *A proper function* $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ *is convex if and only if its epigraph is convex.*

Proof. We have already seen one direction, the other is part of an exercise sheet. □

2.3 Convex Functions

In order to determine if a function is convex, it is useful to know some equivalent definitions of convexity.

Lemma 10. Jensen's Inequality. *For any convex function* E, $u_1, \ldots, u_m \in dom(E)$ *and coefficients* $\theta_1, \ldots, \theta_m \geq 0$ *such that* $\sum_{i=1}^m \theta_i = 1$ *it holds*

$$E(\sum_{i=1}^m \theta_i u_i) \leq \sum_{i=1}^m \theta_i E(u_i).$$

Proof. By induction on m. The case $m = 2$ is a result of the definition and the general case is part of an exercise sheet. □

Corollary 11. *Let* $\Delta = \text{Conv}\{u_1, \ldots, u_m\}$ *be the convex hull of* $u_1, \ldots, u_m$ *that is defined by*

$$\text{Conv}\{u_1, \ldots, u_m\} = \{u \ : \ u = \sum_{i=1}^m \theta_i u_i, \quad \sum_{i=1}^m \theta_i = 1, \quad \theta_1, \ldots, \theta_m \geq 0\}.$$

Then a consequence of Jensen's inequality is

$$\max_{u \in \Delta} E(u) = \max_{1 \leq i \leq m} E(u_i).$$

Lemma 12. *Function* $E : C \to \mathbb{R}$ *is convex if and only if* C *is convex and for all* $u, v \in C$, $\beta \geq 0$ *such that* $u + \beta(u - v) \in C$ *it holds that*

$$E(u + \beta(u - v)) \geq E(u) + \beta(E(u) - E(v)).$$

Proof. Let E be convex, we first prove the alternative definition. Given $\beta > 0$ define $\theta = \frac{\beta}{\beta+1} \in (0,1]$ and $x = u + \beta(u - v)$ such that

$$u = \frac{1}{1+\beta}(x + \beta v) = (1-\theta)x + \theta v.$$

Now, by convexity of E, we have

$$E(u) \leq (1-\theta)E(x) + \theta E(v) = \frac{1}{1+\beta}E(u + \beta(u-v)) + \frac{\beta}{1+\beta}E(v) \tag{2.4}$$

and multiplying both sides by $(1+\beta)$ and re-arranging we have the alternate definition:

$$\begin{aligned}(1+\beta)E(u) - \beta E(v) &\leq E(u + \beta(u-v)) \\ E(u) + \beta(E(u) - E(v)) &\leq E(u + \beta(u-v)).\end{aligned} \tag{2.5}$$

Let us now prove that this alternative definition implies convexity. Given any $u, v \in \mathrm{dom}(E)$, $\theta \in (0,1]$, define $\beta = \frac{1-\theta}{\theta}$ and $x = \theta u + (1-\theta)v$ such that

$$u = \frac{1}{\theta}x - \frac{1-\theta}{\theta}v = \left(1 + \frac{1-\theta}{\theta}\right)x - \frac{1-\theta}{\theta}v = (1+\beta)x - \beta v = x + \beta(x - v).$$

The inequality of the alternate definition can be expressed in terms of $E(u)$ as follows

$$\begin{aligned}E(u) &= E(x + \beta(x-v)) \geq E(x) + \beta[E(x) - E(v)] \\ E(u) &\geq (1+\beta)E(x) - \beta E(v) = \frac{1}{\theta}E(x) - \frac{1-\theta}{\theta}E(v) \\ \theta E(u) &+ (1-\theta)E(v) \geq E(\theta u + (1-\theta)v),\end{aligned} \tag{2.6}$$

which implies convexity of E. □

Theorem 13. Monotonicity of the gradient *Let $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and continuously differentiable, then E is convex if and only if for any $u, v \in dom(E)$*

$$E(v) \geq E(u) + \langle \nabla E(u), v - u \rangle. \tag{2.7}$$

Proof. Given $u, v \in \mathrm{dom}(E)$, and $\theta \in [0,1]$, let $u_\theta = \theta u + (1-\theta)v$. If E is continuously differentiable and satisfies the theorem's inequality, we have

$$\begin{aligned}E(v) &\geq E(u_\theta) + \langle \nabla E(u_\theta), v - u_\theta \rangle = E(u_\theta) + \theta \langle \nabla E(u_\theta), v - u \rangle \\ E(u) &\geq E(u_\theta) + \langle \nabla E(u_\theta), u - u_\theta \rangle = E(u_\theta) - (1-\theta)\langle \nabla E(u_\theta), v - u \rangle.\end{aligned}$$

Multiplying the first inequality by $1-\theta$, the second by θ, and adding the results, we get the inequality that defines a convex function $\theta E(u) + (1-\theta)E(v) \geq E(u_\theta) = E(\theta u + (1-\theta)v)$.

We now prove that a convex and continuously differentiable function satisfies the theorem's inequality. Given $u, v \in \mathrm{dom}(E)$ and $\theta \in [0,1)$, by convexity of E we have

$$\begin{aligned}E(u_\theta) = E(\theta u + (1-\theta)v) &\leq \theta E(u) + (1-\theta)E(v) \\ E(u_\theta) - \theta E(u) &\leq (1-\theta)E(v)\end{aligned} \tag{2.8}$$

multiplying both sides by $\frac{1}{1-\theta}$ we obtain

$$E(v) \geq \frac{1}{1-\theta}[E(u_\theta) - \theta E(u)] = E(u) + \frac{1}{1-\theta}[E(u_\theta) - \theta E(u) - (1-\theta)E(u)] = E(u) + \frac{1}{1-\theta}[E(u_\theta) - E(u)]. \tag{2.9}$$

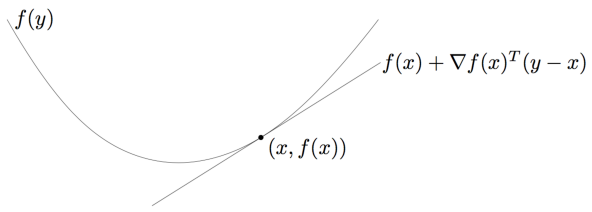


Fig. 2.3: If f is convex and differentiable, then $f(x) + \nabla f(x)^T(y - x)$, as a function of x, is a global underestimator of the function f.Image source: Boyd, Vandenberghe, *Convex optimization theory*,2004.

That is, given $u, v \in \text{dom}(E)$ for any $\theta \in [0, 1)$, we have

$$E(v) \geq E(u) + \frac{1}{1-\theta}[E(\theta u + (1-\theta)v) - E(u)]. \tag{2.10}$$

As E is differentiable, the limit when θ tends to 1 exists and we obtain $E(v) \geq E(u) + \langle \nabla E(u), v - u \rangle$. □

The affine function of v given by $E(u) + \langle \nabla E(u), v - u \rangle$ is the first-order Taylor approximation of E near u. The inequality (2.7) states that for a convex function, the first-order Taylor approximation is in fact a global underestimator of the function. Conversely, if the first-order Taylor approximation of a function is always a global underestimator of the function, then the function is convex. See Figure **??**.

The monotonicity of gradients is a key property because it shows that from local information about a convex function (its value and derivative at a point) we can derive global information (a global underestimator of the function). This explains some of the remarkable properties of convex functions for optimization.

2.3.1 Necessary and Sufficient Optimality Conditions

Theorem 14. *Let $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex. Any local minimum of E is global.*

Proof. Let $u^\star$ be a global minimum of E and $\bar{u}$ a local minimum that is not global, that is, $E(u^\star) < E(\bar{u})$. By definition of local minimum, there exists an $\epsilon > 0$ such that $E(v) \geq E(\bar{u})$ for any $v \in \text{dom}(E)$ with $\|\bar{u} - v\| < \epsilon$. As $u^\star, \bar{u} \in \text{dom}(E)$ convex, $\theta\bar{u} + (1-\theta)u^\star \in \text{dom}(E)$ and

$$E(\theta\bar{u} + (1-\theta)u^\star) \leq \theta E(\bar{u}) + (1-\theta)E(u^\star) < E(\bar{u})$$

As θ tends to 1, $\|\theta\bar{u} + (1-\theta)u^\star - \bar{u}\| < \epsilon$ and this contradicts the definition of $\bar{u}$ as local minimum. □

When a convex function is differentiable, we can prove that first-order optimality conditions are sufficient.

Theorem 15. *If $E \colon \mathbb{R}^n \to \mathbb{R}$ is a convex continuously differentiable function with $\nabla E(u^\star) = 0$ then $u^\star$ is the global minimum of $E(x)$.*

Proof. As $\nabla E(u^\star) = 0$, the inequality $E(v) \geq E(u^\star) + \langle \nabla E(u^\star), v - u^\star \rangle \ \forall v \in \text{dom}(E)$ gives us the condition $E(v) \geq E(u^\star)$ that characterizes a global minimum. □

When the objective function is two-times differentiable, we can also characterize convexity in terms of the Hessian.

Theorem 16. *Two times continuously differentiable function $E \colon \mathbb{R}^n \to \mathbb{R}$ is convex if and only for any $u \in \mathbb{R}^n$ we have $\nabla^2 E(u) \succeq 0$.*

Proof. This is part of an exercise sheet. □

As a result, for any matrix A symmetric and positive semi-definite, the quadratic function $E(u) = \alpha + \langle a, u\rangle + \langle u, Au\rangle$ is convex because $\nabla^2 E(u) = A \succeq 0$.

2.3.2 Analytic Properties of Convex Functions

The behavior of convex functions at the boundary of their domain can be out of control. To prevent this case, we restrict our analysis to convex and closed functions.

Definition A convex function is **closed** if its epigraph is closed.

Lemma 17. *If E is convex and closed, all its level sets are closed.*

Proof. For each β, the level-set $\{u : E(u) = \beta\} = \text{epi}(E) \cap \{(x,t) : t = \beta\}$ can be described as the intersection of the epigraph of E, which is closed and convex, and the closed and convex set $\{(x,t) : t = \beta\}$. As the intersection of closed sets is closed, we have seen that any level set of E is closed. □

Function $E(u) = \|u\|$, where $\|\cdot\|$ is any norm, is closed and convex as a result of the triangle inequality and homogeneity properties that define any norm:

$$\|\theta u + (1-\theta)v\| \leq \|\theta u\| + \|(1-\theta)v\| = |\theta|\|u\| + |1-\theta|\|v\| = \theta\|u\| + (1-\theta)\|\|v\|$$

The norms more common in computer vision and machine learning are the ℓ_p norms:

$$\|u\|_p = \left(\sum_{i=1}^{n} |u_i|^p\right)^{\frac{1}{p}} \quad u \in \mathbb{R}^n.$$

- the Euclidean norm: $\|u\|_2 = \sqrt{\sum_{i=1}^n u_i^2}$.
- the non-differentiable ℓ_1 norm $\|u\|_1 = \sum_{i=1}^n |u_i|$.
- the ℓ_∞ norm $\|u\|_\infty = \max_{1\leq i\leq n} \ |u_i|$.

As a result of the convexity of the norm, we can prove that any ball of radius r centered at any point $u \in \mathbb{R}^n$,

$$B_p(u, r) = \{v \in \mathbf{R^n} : \|v - u\|_p \leq r\},$$

is a convex set. By default, $\|u\| = \|u\|_2$ denotes the Euclidean norm in $\mathbb{R}^n$.

If E is convex function, its domain $\text{dom}(E)$ closed, and E is continuous then E is closed. The converse is not true, a closed convex function is not necessarily continuous. Consider the following examples:

- $E(u) = \frac{1}{u}$ is convex in its domain $\text{dom}(E) = \mathbb{R}_{++} = \{u \in \mathbb{R} \ : u > 0\}$. The domain $\text{dom}(E)$ is thus open even though the function is closed because its epigraph $\{(u,t) \in \mathbb{R} \times \mathbb{R}_{++} : \frac{1}{t} \leq u\}$ is closed.
- the function

$$E(x,y) = \begin{cases} 0 & \text{if } \ x^2 + y^2 < 1 \\ \phi(x,y) & \text{if } \ x^2 + y^2 = 1 \end{cases}$$

  has a closed domain, the unit ball, and is convex for any $\phi(x,y) > 0$ defined on the unit circle, the boundary of the function domain. Imposing that the function is closed, which implies $\phi(x,y) = 0$, ensures that the function is well-behaved also on the boundary of its domain.

The behavior of convex function at the boundary of their domain can be disappointing, but their behavior in their interior is very simple.

Theorem 18. *Let function $E\colon C \subset \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex, then E is locally bounded in the interior of its domain intdom(E).*

Proof. Let us choose $\epsilon > 0$ such that $u \pm \epsilon e_i \in \text{int dom}(E) \;\; i = 1, \ldots, n$, where e_i is the i-th coordinate vector of $\mathbb{R}^n$ and define $\hat{\epsilon} = \frac{\epsilon}{\sqrt{n}}$. A simple drawing show us that

$$B(u, \hat{\epsilon}) \subset \Delta = \text{Conv}\{u \pm \epsilon e_i \;\; i = 1, \ldots, n\}.$$

From the corollary to Jensen's inequality, we find a local bound M to E

$$\max_{v \in B(u,\hat{\epsilon})} \; E(v) \leq \max_{v \in \Delta} \; E(v) \leq \max_{1 \leq i \leq n} \; E(u \pm \epsilon e_i) = M.$$

□

Theorem 19. Continuity of Convex Functions *If $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is convex, then E is locally Lipschitz and hence continuous on int$(dom(E))$.*

Proof. Let us first recall the definition of Lipschitz continuity. A function $E : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz continuous with *Lipschitz constant* L if for all $u, \; v \in \text{dom}(E)$

$$\|E(u) - E(v)\|_2 \leq L\|u - v\|_2$$

A function is **locally Lipschitz continuous** if for every $u \in \text{dom}(E)$ there exists $\epsilon > 0$ such that $f_{|B(\epsilon,u)}$ is Lipschitz continuous.

Now, let $B(u_0, \epsilon) \subset \text{dom}(E)$ and $M = \sup_{u \in B(u_0,\epsilon)} \; E(u) < \infty$.
Consider $v \in B(u_0, \epsilon)$, $v \neq u_0$ like in Figure **??** and define

$$\alpha = \frac{1}{\epsilon}\|v - u_0\| \qquad\qquad z = u_0 + \frac{1}{\alpha}(v - u_0)$$

It is clear that $\|z - u_0\| = \epsilon$, $\alpha \leq 1$, and $v = \alpha z + (1 - \alpha)u_0$. By convexity of E then

$$E(v) \leq \alpha E(z) + (1-\alpha)E(u_0) = E(u_0) + \alpha(E(z) - E(u_0)) \leq E(u_0) + \alpha(M - E(u_0)) = E(u_0) + \frac{M - E(u_0)}{\epsilon}\|v - u_0\|$$

Now define $y = u_0 + \frac{1}{\alpha}(u_0 - v)$ with $\|y - u_0\| = \epsilon$ and $v = u_0 + \alpha(u_0 - y)$. Using the alternative definition of convex functions, we have

$$E(v) \geq E(u_0) + \alpha(E(u_0) - E(y)) = E(u_0) - \alpha(E(y) - E(u_0)) \geq E(u_0) - \alpha(M - E(u_0)) = E(u_0) - \frac{M - E(u_0)}{\epsilon}\|v - u_0\|$$

As a result of the two inequalities

$$|E(v) - E(u_0)| \leq \frac{M - E(u_0)}{\epsilon}\|v - u_0\|,$$

which shows that E is locally Lipschitz continuous at u_0. □

2.3.3 Examples of Convex Functions

The next statements significantly increases our possibilities of constructing convex functions.

Lemma 20. *Given a closed convex function ϕ and a linear operator $\mathscr{A} : \mathbb{R}^m \to \mathbb{R}^n$, then $E(u) = \phi(\mathscr{A}(u))$ is closed and convex with*

$$dom(E) = \{u \in \mathbb{R}^m \;\; : \mathscr{A}(u) \in dom(\phi)\}.$$

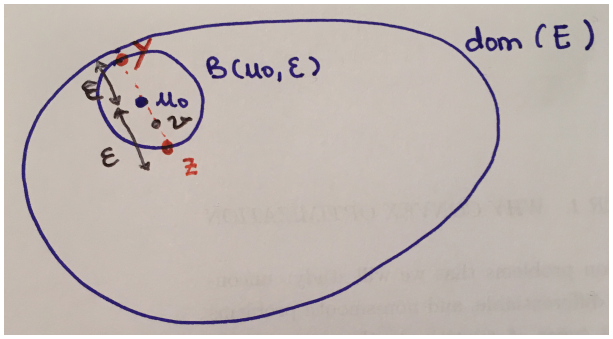


Fig. 2.4: Illustration accompanying the proof that a convex function is locally Lipschitz.

Proof. Let $\mathscr{A}(u) = Au + b = x \in \mathrm{dom}(\phi)$ and $\mathscr{A}(v) = Av + b = y \in \mathrm{dom}(\phi)$, then by convexity of ϕ for any $\theta \in [0,1]$ we have $\theta x + (1-\theta)y \in \mathrm{dom}(\phi)$ and

$$E[\theta u+(1-\theta)v] = \phi[\mathscr{A}(\theta u+(1-\theta)v)] = \phi[\theta(Au+b)+(1-\theta)(Av+b)] \leq \theta\phi(Au+b)+(1-\theta)\phi(Av+b) = \theta E(u)+(1-\theta)E(v).$$

This proves convexity of E. The closedness of its epigraph follows from continuity of the linear operator $\mathscr{A}$. □

Lemma 21. *Given two convex function E_1, E_2 and $\alpha_1, \alpha_2 > 0$, then $E = \alpha_1 E_1 + \alpha_2 E_2$ is convex with $dom(E) = dom(E_1) \cap dom E_2$.*

Proof. Let $u, v \in \mathrm{dom}(E_1) \cap \mathrm{dom} E_2$ and $\theta \in [0,1]$, by convexity of each E_1, E_2 we have

$$\begin{aligned} E(\theta u + (1-\theta)v) &= \alpha_1 E_1(\theta u + (1-\theta)v) + \alpha_2 E_2(\theta u + (1-\theta)v) \\ &\leq \alpha_1\theta E_1(u) + \alpha_1(1-\theta)E(v) + \alpha_2\theta E_1(u) + \alpha_2(1-\theta)E(v) \\ &\leq \theta[\alpha_1 E_1(u) + \alpha_2 E_2(u)] + (1-\theta)[\alpha_1 E_1(v) + \alpha_2 E_2(v) = \theta E(u) + (1-\theta)E(v). \quad (2.11) \end{aligned}$$

This proves the convexity of E. □

Taking into account that the following 1-dimensional functions are convex:

$$\begin{aligned} E(u) &= \exp(u) \\ E(u) &= |u|^p \quad p > 1 \\ E(u) &= |x| - \log(1+|x|) \end{aligned}$$

the previous lemmas imply that the following multi-dimensional functions are convex:

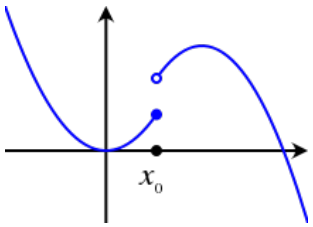
$$\begin{aligned} E(u) &= \sum_{i=1}^{n} \exp(\alpha + \langle u, a_i \rangle) \\ E(u) &= |\langle u, a_i \rangle - b_i|^p \quad p > 1 \end{aligned}$$

Lemma 22. *Given two closed and convex function E_1, E_2, then $E(u) = \max\{E_1(u), E_2(u)\}$ is closed and convex with $dom(E) = dom(E_1) \cup dom(E_2)$.*

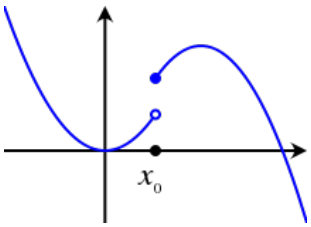
Proof. The epigraph is closed and convex because it is the intersection of two closed convex sets

$$\mathrm{epi}(E) = \{(u,t) \,:\, u \in \mathrm{dom}(E_1) \cap \mathrm{dom}(E_2),\ E_1(u) \leq t,\ E_2(u) \leq t\} = \mathrm{epi}(E_1) \cap \mathrm{epi}(E_2).$$

□



Example of a lower semi-continuous function.



Example of a function that is not lower semi-continuous at x_0.

Fig. 2.5: The function on the left is lower semi-continuous because any sequence, approaching x_0 from the left or the right, verifies $\liminf_{x\to x_0} E(x) \geq E(x_0)$. This is not satisfies by the function on the right for any sequence approaching x_0 from the right.

We have an even more general result.

Theorem 23. *Let D be some set, not necessarily convex or finite dimensional, and*

$$E(u) = \sup_{y\in D} \ \phi(u,y)$$

such that ϕ is closed and convex in u for all $y \in D$, then E is closed and convex with

$$dom(E) = \{u \in \cap_{y\in D} dom(\phi(\cdot,y)) \ : \ \exists \gamma \in \mathbb{R} \ s.t. \ \phi(u,y) \leq \gamma \ \forall y \in D\}.$$

Proof. We first show the definition of the domain. If u belongs to $\{u \in \cap_{y\in D}\text{dom}(\phi(\cdot,y)) \ : \ \exists\gamma \in \mathbb{R}$ s.t. $\phi(u,y) \leq \gamma \ \ \forall y \in D\}$, then $E(u) < \infty$ and $u \in \text{dom}(E)$. If u does not belong to this set, then there exists a sequence $\{y_k\}_{k=1}^{\infty}$ such that $\phi(u,y_k) \to \infty$ and u does not belong to $\text{dom}(E)$.
To show that E is closed and convex, we will show that its epigraph is closed and convex. We start noting that $(u,t) \in \text{epi}(E)$ if and only if for all $y \in D$ we have $u \in \text{dom}(\phi(\cdot,y))$ and $\phi(u,y) \leq t$. As a results $\text{epi}(E) = \cap_{y\in D}\text{epi}(\phi(\cdot,y))$ is closed and convex as the intersection of closed and convex sets. □

As a result of this lemma, the function $E^*(y) = \sup_{u\in\text{dom}(E)} \ \langle u,y\rangle - E(u)$ is convex for any E. We will use this function to define the convex conjugate.

2.4 Existence and Uniqueness of Minimizers

It only makes sense to try to solve an optimization problem if it has a solution. Specially, if the solution is the limit of a relaxation sequence that is computed through costly iterative algorithms that might never converge. To show that a convex problem has a minimizer, we will see that it satisfies the necessary conditions to frame the problem in the more general framework of lower semi-continuous functions. This section explains the tools from this framework that we will use.

Definition Lower semi-continuity. A function $E : \mathbb{R}^n \to \mathbb{R}$ is lower semi-continuous (l.s.c.), if for all u it holds that

$$\liminf_{v\to u} E(v) \geq E(u).$$

The picture to keep in mind of a lower semi-continuous function is Figure **??**.

Theorem 24. *Let $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ be l.s.c. and let there exist an α such that the sublevelset*

$$S_\alpha = \{u \in \mathbb{R}^n \mid E(u) \leq \alpha\}$$

is nonempty and bounded, then there exists

$$\hat{u} \in \arg\min_u E(u).$$

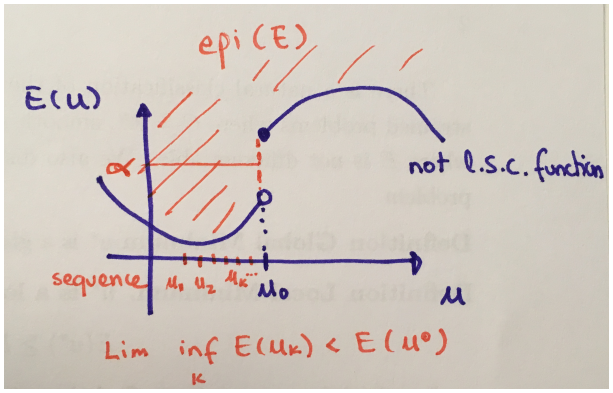


Fig. 2.6: Illustration accompanying the proof that lower semi-continuity and closedness are equivalent.

Proof. Remember that the infimum is the largest lower bound on all possible values of $E(u)$ and consider a sequence $(u_k)_{k=1}^{\infty}$ such that $E(u_k) \to \inf_u E(u)$.
We distinguish two cases: For $\alpha = \inf_u E(u)$ the non-emptyness of S_α yields the assertion. For $\alpha > \inf_u E(u)$ it holds that for some sufficiently large k_0 on, we will have $u_k \in S_\alpha$. Since S_α is bounded there exists a convergent subsequence $u_{k_l} \to \bar{u}$. Due to the lower semi-continuity we find

$$\inf_u E(u) = \lim_{k\to\infty} E(u_k) = \lim_{l\to\infty} E(u_{k_l}) \geq E(\bar{u}).$$

Since by definition $\inf_u E(u) \leq E(\bar{u})$ we obtain equality and hence there exists $\bar{u} \in \operatorname{argmin}_u E(u)$. □

Theorem 25. Equivalence of l.s.c. and closedness. *For $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ the following two statements are equivalent*

- *E is lower semi-continuous (l.s.c.).*
- *E is closed (its epigraph is closed).*

Proof. Let E be closed and assume that E is not l.s.c. See Figure **??**. Then there exists a point u^0 and a sequence $\{u_k\}_{k=1}^{\infty}$ with $\lim_k u_k = u^0$ such that

$$\liminf_k E(u_k) < E(u^0).$$

In particular, there exists $\alpha \in \mathbb{R}$ and a subsequence $\{u_{k_l}\}_{l=1}^{\infty}$ such that

$$E(u_{k_l}) \leq \alpha < E(u^0) \quad \forall k \tag{2.12}$$

Obviously, $(u_{k_l}, \alpha) \in \operatorname{epi}(E)$ for all k_l and $(u_{k_l}, \alpha) \to (u^0, \alpha)$, but according to (2.12) $(u^0, \alpha) \notin \operatorname{epi}(E)$, which contradicts the closedness of E.
To prove the other direction of the claim, let E be l.s.c. and assume that E is not closed. Then there exists a sequence $(u_k, \alpha_k) \in \operatorname{epi}(E)$ with $(u_k, \alpha_k) \to (u^0, \alpha^0) \notin \operatorname{epi}(E)$. We find

$$\liminf_k E(u_k) \leq \lim_k \alpha_k = \alpha^0 < E(u^0).$$

On the other hand, due to E being l.s.c. we have $E(u^0) \leq \liminf_k E(u_k)$, which is a contradiction. □

2.4.1 Existence of Minimizers of Convex Functions

Definition Coercivity. A function $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is called coercive if $E(v_n) \to \infty$ for all sequences $\{v_n\}_{n=1}^{\infty}$ with $\|v_n\| \to \infty$.

It is easy to proof that coercivity implies existence of a bounded sublevelset by contradiction. We have now all the tools to prove existence of minimizers of convex functions.

Theorem 26. Existence of a Minimizer *Let $E : \mathbb{R}^n \to \mathbb{R}$ be convex and coercive, then an element $\hat{u} \in \arg\min_u E(u)$ exists.*

Proof. As $\mathrm{dom}(E) = \mathbb{R}^n$ and E convex, E is Lipschitz continuous, and thus continuous. At the same time, as E is coercive, there exists a non-empty bounded sublevelset, and we can apply the theorem on the existence of minimizers for lower semi-continuous functions to prove existence of a minimizer. □

Theorem 27. Uniqueness. *If $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ is strictly convex, then there exists at most one local minimum which is the unique global minimum.*

Proof. Assume there are two global minima u, v with $u \neq v$, $E(u) = E(v)$, then any $\theta \in [0, 1]$ we have

$$E(\theta u + (1 - \theta)v) < \theta E(u) + (1 - \theta)E(v),$$

which contradicts the definition of $u,\ v$ as global minima. □

2.5 Subdifferentials

Up to now we were describing properties of convex functions in terms of function values or their gradients. When the function is not differentiable, we need to define a direction that acts as the gradient of differentiable convex functions in the inequality known as *monotonicity of the gradients*

$$E(v) \geq E(u) + \langle \nabla E(u), v - u \rangle \quad \forall u, v \in \mathrm{dom}(E).$$

It is easy to see that such directions are defined by supporting hyperplanes to the graph of E at the point $E(u)$. In general, we define supporting hyperplanes to convex sets.

2.5.1 Supporting Hyperplanes

Definition Let C be a set. We say that hyperplane

$$\mathscr{H}(g, \gamma) = \{u \in \mathbb{R}^n \ : \ \langle g, u \rangle = \gamma, \ \ g \neq 0\}$$

is supporting to C if any $u \in C$ satisfies $\langle g, u \rangle \leq \gamma$.
We say that the hyperplane $\mathscr{H}(g, \gamma)$ separates a point u_0 from C if

$$\langle g, u \rangle \leq \gamma \leq \langle g, u_0 \rangle \quad \forall u \in C.$$

This is illustrated in Figure 2.7.

The next separation theorem deals with boundary points of convex sets.

Theorem 28. Supporting Hyperplane Theorem *Let C be a closed convex set and u_0 in the boundary of C. Then there exists a hyperplane $\mathscr{H}(g, \gamma)$ supporting to C and passing through u_0.*

Proof. See Boyd and Vandenberghe, *Convex Optimization Theory*, pp 50–51. □

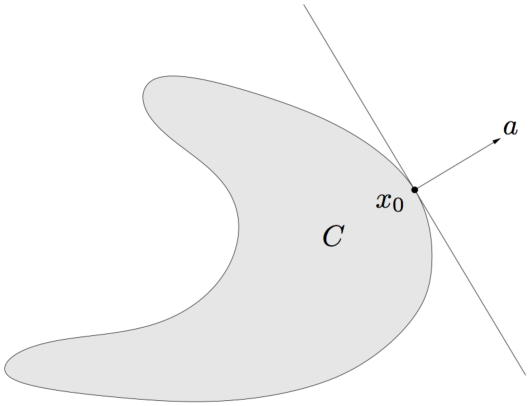


Fig. 2.7: The hyperplane $\{x \ : \ a^T x = a^T x_0\}$ supports C at x_0. Image source: Boyd, Vandenberghe, *Convex optimization theory*,2004.

2.5.2 The Subdifferential

We now have all the tools to introduce the notion of subdifferential.

Definition Let $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex, the **subdifferential** of E at u is the set

$$\partial E(u) = \{p \in \mathbb{R}^n \mid E(v) - E(u) - \langle p, v - u\rangle \geq 0, \ \forall v \in \mathbb{R}^n\}$$

- Elements of $\partial E(u)$ are called subgradients.
- If $\partial E(u) \neq \emptyset$, we call E subdifferentiable at u.
- By convention, $\partial E(u) = \emptyset$ for $u \notin \mathrm{dom}(E)$.

Defining the subdifferential ∂E as a set is necessary because subgradients are not unique. Consider for example a function as friendly-looking as the absolute value at zero:

$$\forall g \in [-1, 1], \quad E(u) = |u| \geq gu = E(0) + \langle g, u - 0\rangle$$

As a result, the subdifferential at 0 contains the interval $\partial E(0) = [-1, 1]$. In general $\partial E(u)$ is a set. Form its definition as a set of linear constraints, we can easily see that it is closed and convex. In the case of the absolute value at zero, the subdifferential is the interval $[-1, 1]$.

2.5.3 Subdifferentiablity and Convexity

The subdifferentiability of a function is important because it implies its convexity.

Theorem 29. *If for any $u \in dom(E)$ the subdifferential $\partial E(u)$ is non-empty, then E is a convex function.*

Proof. Given $u, v \in \mathrm{dom}(E)$, and $\theta \in [0, 1]$, let $u_\theta = \theta u + (1 - \theta)v$. As the subdifferential $\partial E(u_\theta)$ is non-empty, we can pick $g \in \partial E(u_\theta)$ satisfying

$$E(v) \geq E(u_\theta) + \langle g, v - u_\theta\rangle = E(u_\theta) + \theta\langle g, v - u\rangle$$
$$E(u) \geq E(u_\theta) + \langle g, u - u_\theta\rangle = E(u_\theta) - (1 - \theta)\langle g, v - u\rangle.$$

Multiplying the first inequality by $1 - \theta$, the second by θ, and adding the results, we get the inequality that defines a convex function $\theta E(u) + (1 - \theta)E(v) \geq E(u_\theta) = E(\theta u + (1 - \theta)v)$. □

The converse statement is also true. Before we can prove that, however, we need to define where subdifferentials exists. In the same way that the gradient of a differentiable function is only defined for points in the interior of the domain, the subdifferential of a proper convex function is always defined for points in the relative interior of its domain.

The relative interior of a set is a refinement of the concept of the interior that is useful when dealing with low-dimensional sets embedded in higher-dimensional spaces. Intuitively, the relative interior of a set contains all points that are not on the "edge" of the set, relative to the smallest subspace in which this set lies. When the set is convex, the definition takes the following simple form:

Definition The **relative interior** of a convex set C is defined as

$$\mathrm{ri}(C) := \{x \in C \mid \forall y \in C,\ \exists \lambda > 1,\ \text{s.t. } \lambda x + (1-\lambda)y \in C\}$$

As mentioned earlier, the subdifferentiability of convex functions can be guaranteed for points that are not necessarily in the interior of the domain, but that are in its relative interior. To better understand this difference, consider the line segment $I = [-1, 1]$ as a convex subset of the Euclidean plane $I \subset \mathbb{R}^2$. The interior of I is empty with the Euclidean topology of $\mathbb{R}^2$, but its relative interior is the open line segment $\mathrm{ri}(I) = (0, 1)$.

One key property of the relative interior is that it is not empty for convex sets.

Theorem 30. *Let C be a non-empty convex set, then ri(C) is not empty.*

Theorem 31. *If E is a closed convex function and $u \in$ int$(\mathrm{dom}(E))$, then $\partial E(u)$ is a non-empty bounded set.*

Proof. Note that the point $(E(u), u)$ belongs to the boundary of $\mathrm{epi}(E)$, which is convex. As a result, there exists a hyperplane $\mathscr{H} = (g, \gamma)$ supporting to $\mathrm{epi}(E)$ at $(E(u), u)$:

$$\gamma\tau + \langle g, u\rangle \leq \gamma E(u) + \langle g, u\rangle \quad \forall (u, \tau) \in \mathrm{epi}(E)$$

Without loss of generality, we can assume $\|g\|^2 + \gamma^2 = 1$. We can determine the sign of γ by checking the inequality for any point in the epigraph. In particular for any $\tau \geq E(u)$, we have $(u, \tau) \in \mathrm{epi}(E)$ that results in $\gamma > 0$.

To find a subgradient $p \in \partial E(u)$, we will use that a convex function is locally Liptschitz in the interior of its domain. That is, there is some $\epsilon > 0, M > 0$ such that $B(u, \epsilon) \subset \mathrm{dom}(E)$ and

$$E(v) - E(u) \leq M\|v - u\| \quad \forall v \in B(u, \epsilon)$$

For any v from this ball, the supporting hyperplane equation reads

$$\langle g, v - u\rangle \leq \gamma(E(v) - E(u)) \leq \gamma M\|v - u\|$$

In particular, if we choose $v = u + \epsilon g$ we get $\|g\|^2 \leq M\gamma\|d\|$. Plugging now the condition $\|g\|^2 + \gamma^2 = 1$ we get

$$\gamma \geq \frac{1}{\sqrt{1 + M^2}}.$$

If we choose $p = \frac{g}{\gamma}$ we obtain

$$E(v) \geq E(u) + \langle p, v - u\rangle \quad \forall v \in \mathrm{dom}(E)$$

and p is a subgradient of E at u. Finally, to show that the subdifferential is bounded we assume that $p \neq 0$ and consider the point $v = u + \epsilon\frac{p}{\|p\|}$ such that

$$\epsilon\|p\| = \langle p, v - u\rangle \leq E(v) - E(u) \leq M\|v - u\| = M\epsilon$$

Thus, $\partial E(u)$ is bounded by M. □

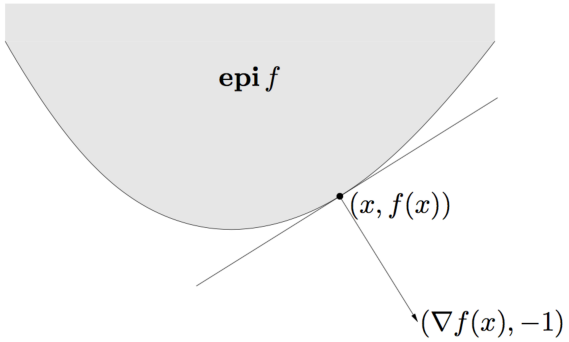


Fig. 2.8: For a convex differentiable function f, the vector $(\nabla f(x), -1)$ defines a supporting hyperplane to the epigraph of f at x. Image source: Boyd, Vandenberghe, *Convex optimization theory*,2004.

The conditions of this theorem cannot be relaxed. For instance, the function $E(u) = -\sqrt{u}$ is convex and closed in its domain $\{u\colon u \geq 0\}$, but its subdifferential does not exists at the only point (0) that is not in its interior. This is just another reminder that considering the interior of the domain for convex functions is important.

To conclude this section, let us point out to the property of the subgradients that makes it important for optimization.

Theorem 32. Optimality Condition. $0 \in \partial E(\hat{u})$ *if and only if* $\hat{u} \in \arg\min_{u\in\mathbb{R}^n} E(u)$.

Proof. If $0 \in \partial E(\hat{u})$, by definition of the subgradient

$$E(u) \geq E(\hat{u}) + \langle 0, u - \hat{u}\rangle = E(\hat{u}) \quad \forall u \in \mathrm{dom}(E)$$

and we conclude that $\hat{u}$ is a minimizer of E. On the other hand, if $E(u) \geq E(\hat{u})$ for all $u \in \mathrm{dom}(E)$, then 0 satisfies the condition of subgradient of E at $\hat{u}$. □

2.5.4 Alternative Definitions of Subgradients

The supporting hyperplane theorem appears on the proof of the "subdifferentiability" theorem because subgradients can be interpreted in terms of supporting hyperplanes.

Theorem 33. Geometric interpretation of Subgradients. *Any subgradient* $p \in \partial E(u)$ *represents a non-vertical supporting hyperplane to epi*(E) *at* $(u, E(u))$.

Proof. Let $p \in \partial E(u)$. Then, by definition of subgradient,

$$\begin{aligned} &E(v) - E(u) - \langle p, v - u\rangle \geq 0 && \forall v \in \mathbb{R}^n \\ &\alpha - E(u) - \langle p, v - u\rangle \geq 0 && \forall (v, \alpha) \in \mathrm{epi}(E) \\ &\left\langle \begin{bmatrix} -p \\ 1 \end{bmatrix}, \begin{bmatrix} v \\ \alpha \end{bmatrix} - \begin{bmatrix} u \\ E(u) \end{bmatrix} \right\rangle \geq 0 && \forall (v, \alpha) \in \mathrm{epi}(E). \end{aligned}$$

As a result, the non-vertical hyperplane $\mathscr{H} = (g, \gamma)$ with $g = (-p, 1)$ and $\gamma = \langle p, u\rangle - E(u)$ supports epi(E) at $(u, E(u))$. □

Apart from this geometric interpretation, it is useful to compute the subdifferential of a differentiable function to understand why it is a generalization of the gradient. The next theorem does that.

Theorem 34. Subdifferential of of Differentiable Functions. *Let the convex function $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be differentiable at $u \in int(dom(E))$. Then*

$$\partial E(u) = \{\nabla E(u)\}.$$

Proof. The subdifferential $\partial E(u)$ of some convex E at $u \in \mathrm{dom}(E)$ is given as

$$\{p \in \mathbb{R}^n : E(z) - E(u) - \langle p, z - u\rangle \geq 0, \forall\, z \in \mathrm{dom}(E)\}.$$

Since $u \in \mathrm{int}(\mathrm{dom}(E))$, we find that for all $v \in \mathbb{R}^n$, $z = u \pm \epsilon v \in \mathrm{dom}(E)$ for ϵ small enough. Therefore, it holds that

$$E(u + \epsilon v) \geq E(u) + \epsilon\langle p, v\rangle, \quad E(u - \epsilon v) \geq E(u) - \epsilon\langle p, v\rangle,$$

for all $v \in \mathbb{R}^n$ and ϵ small enough. This implies that

$$\lim_{\epsilon\to 0} \frac{E(u + \epsilon v) - E(u)}{\epsilon} \geq \langle p, v\rangle, \quad \lim_{\epsilon\to 0} \frac{E(u) - E(u - \epsilon v)}{\epsilon} \leq \langle p, v\rangle,$$

which means

$$\langle \nabla E(u), v\rangle \geq \langle p, v\rangle, \quad \langle \nabla E(u), v\rangle \leq \langle p, v\rangle,$$

i.e.

$$\langle \nabla E(u) - p, v\rangle = 0$$

for all $v \in \mathbb{R}^n$. For the particular choice of $v := \nabla E(u) - p$ we find $p = \nabla E(u)$. The above concludes the proof if we can show that $\partial E(u)$ is non-empty, which follows from the Theorem on Subdifferentiability. □

As a result, we have that for a convex differentiable function E, the vector $(\nabla E(u), -1)$ defines a supporting hyperplane to the epigraph of E at u. This is illustrated in Figure **??**.

2.5.5 Subdifferential Rules

Now that we understand where subdifferentials exists, we can learn the rules that guide their computation.

Theorem 35. Sum Rule. *Let E_1, E_2 be convex functions, then $\partial(E_1 + E_2)(u) = \partial E_1(u) + \partial E_2(u)$ for all $u \in ri(dom(E_1)) \cap ri(dom(E_2))$.*

Proof. See Nesterov, *Introductory Lectures on Convex Optimization*, Lemma. 3.1.9. □

Theorem 36. Chain Rule *Given the linear operator $A \in \mathbb{R}^{m\times n}$ and the convex function $E : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$, then $\partial(E \circ A)(u) = A^*\partial E(Au)$ for all $u \in ri(dom(E)) \cap range(A)$.*

Proof. See Nesterov, *Introductory Lectures on Convex Optimization*, Nesterov, Lemma. 3.1.8. □

Chapter 3

Fixed-Point Iterations

Convex optimization problems come in so many shapes and sizes that the algorithms developed to solve them form a zoo. Each algorithm exploits a particular feature of the convexity of the objective function or the constrain set to find the solution of the problem. As a result, we traditionally also analyze the convergence of each algorithm and its properties in a case by case manner.

It is possible to interpret many of these algorithms as fixed-point iterations in a unified manner and analyze their convergence with the same approach. To do so, we first need to formulate the optimization problem as finding a zero of a monotone operator. This problem is converted into the problem of finding a fixed point of a function and solved by the fixed point iteration algorithm. Different choices of the monotone operator and fixed point function result in different well-known algorithms.

This new view on many classic algorithms provides a convenient strategy to analyze their convergence with a single approach. The price to pay, however, is an additional level of abstraction that might at first seem disconnected from intuitive algorithms like gradient descent. Be patient, and read on.

The material of this chapter is taken mostly from: Ryu and Boyd, *Primer on Monotone Operator Methods*, 2016.

3.1 Nonexpansive mappings and contractions

Definition A function $F : \mathbb{R}^n \to \mathbb{R}^n$ is a **contraction** if it is Lipschitz continuous with constant $L < 1$, that is,

$$\|F(x) - F(y)\| \leq L\|x - y\| \quad \forall x,\ y \in \mathrm{dom}(F).$$

When $L = 1$, we say that F is a **nonexpansive** operator.

In other words, mapping a pair of points by a contraction reduces the distance between them; mapping them by a nonexpansive operator does not increase the distance between them. See Figure 3.1.

Intuitively, it is useful to keep in mind an exemplary contraction and an exemplary nonexpansive operator. You can think of a contraction as a "zoom-out" that reduces the distance between two points, and think of a nonexpansive operator as a rotation of the coordinate plane. It is then only natural to see that the combination of two zoom-outs (contractions) is still a zoom-out, while the combination of a zoom-out and a rotation is also a zoom-out. The following lemma describes this.

Lemma 37. *Convex combinations as well as compositions of nonexpansive operators are nonexpansive.*

Proof. If $F_1 : \mathbb{R}^n \to \mathbb{R}^n$ has Lipschitz constant L_1 and $F_2 : \mathbb{R}^n \to \mathbb{R}^n$ has Lipschitz constant L_2, then the composition F_2F_1 has Lipschitz constant L_2L_1. Indeed, let $x,\ y \in \mathrm{dom}(F_1)$ such that $F_1x,\ F_1y \in \mathrm{dom}(F_2)$

(a) Contraction Mapping (b) Nonexpansive Mapping

Fig. 3.1: Illustration of a contractive and a nonexpansive mapping F on two points. Source: Ryu and Boyd, *Primer on Monotone Operator Methods*, 2016

then

$$\|F_2F_1x - F_2F_1y\| \leq L_2\|F_1x - F_1y\| \leq L_2L_1\|x - y\|$$

As a result, the composition of nonexpansive operators is nonexpansive, and the composition of a contraction and a nonexpansive operator is a contraction.

Similarly, if $\alpha_1, \alpha_2 \in \mathbb{R}$, then $\alpha_1F_1 + \alpha_2F_2$ has Lipschitz constant $|\alpha_1|L_1 + |\alpha_2|L_2$. Indeed, let $x,\ y \in \mathrm{dom}(F_1) \cap \mathrm{dom}(F_2)$, then

$$\begin{aligned}\|(\alpha_1F_1 + \alpha_2F_2)x - (\alpha_1F_1 + \alpha_2F_2)y\| &\leq \|\alpha_1F_1x - \alpha_1F_1y\| + \|\alpha_2F_2x - \alpha_2F_2y\| \\ &\leq |\alpha_1|L_1\|x - y\| + |\alpha_2|L_2\|x - y\| \\ &\leq (|\alpha_1|L_1 + |\alpha_2|L_2)\|x - y\|. \end{aligned} \tag{3.1}$$

As a result, a weighted average of nonexpansive operators $\theta F_1 + (1-\theta)F_2$ with $\theta \in [0,1]$ is also nonexpansive. If one of them is a contraction and $\theta \in (0,1)$, then the weighted average is a contraction. □

Contractions are important for us because they have a single fixed point and we can use this property to design iterative algorithms that converge to it.

Theorem 38. *If F is nonexpansive and $dom(F) = \mathbb{R}^n$, then its set of fixed points*

$$\{x \in dom(F) \ :\ x = F(x)\}$$

is closed and convex. If F is a contraction and $dom(F) = \mathbb{R}^n$, its has exactly one fixed point.

Proof. Let $F\colon \mathbb{R}^n \to \mathbb{R}$ be nonexpansive and denote by X the set of its fixed points. Note that we can also define $X = (I-F)^{-1}(\{0\})$, where I is the identity function. From this definition, X is closed because it is the preimage of a continuous function $(F - I)$ on a closed set $(\{0\})$. To show that it is convex, let $x, y \in \mathrm{dom}(F)$ and $\theta \in [0,1]$ and define $z = \theta x + (1-\theta)y$. We will show that $z \in X$. As F is nonexpansive

$$\begin{aligned}\|Fz - x\| = \|Fz - Fx\| \leq \|z - x\| = (1-\theta)\|x - y\| \\ \|Fz - y\| = \|Fz - Fy\| \leq \|z - y\| = \theta\|x - y\| \\ \|x - y\| \leq \|Fz - x\| + \|Fz - y\| \leq \|x - y\|\end{aligned}$$

The last triangle inequality tells us that Fz is on the line segment between x and y. In particular, as $\|Fz - y\| = \theta\|x - y\|$, we have $Fz = \theta x + (1-\theta)y = z$ and z is a fixed point of F.

Let $x,\ y \in X$ be again fixed points of F and F be now a contraction with Lipschitz constant $L < 1$, then

$$\|x - y\| = \|Fx - Fy\| \leq L\|x - y\|.$$

This is a contradiction unless $x = y$, which implies that there is a single fixed-point. □

There are few examples of contractions that are common in convex optimization. Most of the time we work with nonexpansive operators. Among them, a particular type called averaged operator, is specially useful and common.

Definition An operator G is **averaged** if $G = (1-\alpha)I + \alpha R$ for some $\alpha \in (0,1)$ and nonexpansive R.

G is nonexpansive because it is a convex combination of nonexpansive operators (the identity I is nonexpansive). Moreover, it is easy to see that G has the same fixed points as R.

$$u^* = Ru^* \;\Leftrightarrow\; (1-\alpha)u^* + \alpha u^* = (1-\alpha)u^* + \alpha R u^* \;\Leftrightarrow\; u^* = \left[(1-\alpha)I + \alpha R\right] u^* = Gu^* \tag{3.2}$$

We will use this property to design algorithms that find fixed points of nonexpansive operators R and are parametrized by $\alpha \in (0,1)$.

Properties of Averaged Operators

Lemma 39. *If a function $G : \mathbb{R}^n \to \mathbb{R}^n$ is averaged with respect to $\alpha \in (0,1)$, then it is also averaged with respect to any other parameter $\tilde{\alpha} \in (0, \alpha)$.*

Proof. Since G is averaged with respect to α there exists a nonexpansive operator R such that $G = \alpha I + (1-\alpha)R$. We find

$$\begin{aligned} G &= \alpha I + (1-\alpha)R \\ &= \tilde{\alpha} I + (\alpha - \tilde{\alpha})I + (1-\alpha)R \\ &= \tilde{\alpha} I + (1-\tilde{\alpha}) \underbrace{\left(\frac{\alpha - \tilde{\alpha}}{1-\tilde{\alpha}} I + \frac{1-\alpha}{1-\tilde{\alpha}} R \right)}_{=:\tilde{R}}. \end{aligned}$$

And $\tilde{R}$ is still nonexpansive because

$$\begin{aligned} \|\tilde{R}(u) - \tilde{R}(v)\| &\leq \frac{\alpha - \tilde{\alpha}}{1-\tilde{\alpha}} \|u - v\| + \frac{1-\alpha}{1-\tilde{\alpha}} \|R(u) - R(v)\| \\ &\leq \frac{\alpha - \tilde{\alpha}}{1-\tilde{\alpha}} \|u - v\| + \frac{1-\alpha}{1-\tilde{\alpha}} \|u - v\| \\ &= \|u - v\|. \end{aligned}$$

□

Lemma 40. *If $G_1 : \mathbb{R}^n \to \mathbb{R}^n$ and $G_2 : \mathbb{R}^n \to \mathbb{R}^n$ are averaged, then $G_2 \circ G_1$ is also averaged.*

Proof. Let $G_1 = \alpha_1 I + (1-\alpha_1)R_1$ and $G_2 = \alpha_2 I + (1-\alpha_2)R_2$ for nonexpansive operators R_1 and R_2. Then

$$\begin{aligned} G_2(G_1)(u) &= \alpha_2 G_1(u) + (1-\alpha_2) R_2(G_1(u)) \\ &= \alpha_1\alpha_2 u + \alpha_2(1-\alpha_1)R_1(u) + (1-\alpha_2)R_2(G_1(u)) \\ &= \alpha_1\alpha_2 u + (1-\alpha_1\alpha_2)\left(\frac{\alpha_2(1-\alpha_1)}{1-\alpha_1\alpha_2} R_1(u) + \frac{(1-\alpha_2)}{1-\alpha_1\alpha_2} R_2(G_1(u)) \right). \end{aligned}$$

Since the concatenation of nonexpansive operators is nonexpansive, and convex combinations of nonexpansive operators are nonexpansive, we conclude that $G_2 \circ G_1$ is averaged. □

It is possible to determine if an operator is averaged without explicitly finding its decomposition into a convex combination of the identity and a nonexpansive operator. We do so through the notion of firmly nonexpansive operators.

Definition A function $G : \mathbb{R}^n \to \mathbb{R}^n$ is called **firmly nonexpansive**, if for all $u, v \in \mathbb{R}^n$ it holds that

$$\|G(u) - G(v)\|_2^2 \leq \langle G(u) - G(v), u - v\rangle.$$

Lemma 41. *A function $G : \mathbb{R}^n \to \mathbb{R}^n$ is firmly nonexpansive if and only if G is averaged with $\alpha = \frac{1}{2}$.*

Proof. First, let G be averaged with $\alpha = 1/2$, i.e., $G = \frac{1}{2}I + \frac{1}{2}R$ for some nonexpansive operator $R = 2G - I$. As R is nonexpansive, we have

$$\begin{aligned}\|u - v\|_2^2 \geq \|R(u) - R(v)\|_2^2 &= \|2G(u) - 2G(v) - (u - v)\|_2^2 \\ &= 4\|G(u) - G(v)\|_2^2 - 4\langle G(u) - G(v), u - v\rangle + \|u - v\|_2^2,\end{aligned}$$

which implies $\langle G(u) - G(v), u - v\rangle \geq \|G(u) - G(v)\|_2^2$ and shows that G is firmly nonexpansive.

Second, let G be firmly nonexpansive and define $R = 2G - I$, then

$$\begin{aligned}\|R(u) - R(v)\|_2^2 &= \|2G(u) - 2G(v) - (u - v)\|_2^2 \\ &= 4\|G(u) - G(v)\|_2^2 - 4\langle G(u) - G(v), u - v\rangle + \|u - v\|_2^2 \\ &\leq \|u - v\|_2^2,\end{aligned}$$

which shows that R is nonexpansive, i.e., $G = \frac{1}{2}I + \frac{1}{2}R$ is averaged. □

3.2 Fixed-point Iterations

We are now ready to discuss the main algorithm of this chapter.

Definition Let $G\colon \mathbb{R}^n \to \mathbb{R}$, and $u^0 \in \mathbb{R}^n$ be a starting point, the **fixed-point or Picard iteration** is

$$u^{k+1} = G(u^k).$$

As the name suggests, the fixed-point iteration is used to find a fixed point $\hat{u}$ of G. Using this iteration to solve an optimization problem involves two steps: 1) find a suitable G whose fixed points are solutions to the problem at hand, 2) show that the iteration converges to a fixed point. For this second step, we show two simple conditions that guarantee convergence.

Theorem 42. Banach fixed-point theorem. *If the update rule $G : \mathbb{R}^n \to \mathbb{R}^n$ is a contraction with Lipschitz constant $L < 1$, then the fixed-point iteration converges to the unique fixed-point $\hat{u}$ of G with*

$$\|u^k - \hat{u}\| \leq L^k \|u^0 - \hat{u}\|.$$

Theorem 43. Krasnosel'skii-Mann Theorem. *If the operator $G : \mathbb{R}^n \to \mathbb{R}^n$ is averaged and has a fixed-point, then the iteration*

$$u^{k+1} = G(u^k)$$

converges to a fixed point of G for any starting point $u^0 \in \mathbb{R}^n$.

Proof. We'll make use of the identity

$$\|(1 - \theta)a + \theta b\|^2 = (1 - \theta)\|a\|^2 + \theta\|b\|^2 - \theta(1 - \theta)\|a - b\|^2,$$

which holds for any $\theta \in \mathbb{R}$, $a, b \in \mathbb{R}^n$. It can be verified by expanding both sides as a quadratic function of θ. The first two terms correspond to the definition of convexity for function $\|\cdot\|^2$, the third one improves this bound.

Because G is averaged, there exists a non-expansive mapping $T : \mathbb{R}^n \to \mathbb{R}^n$ such that $G = (1-\theta)I+\theta T$. Recall that T has the same fixed points as F. We consider the fixed point iteration

$$u^{k+1} = G(u^k) = (1-\theta)u^k + \theta T u^k.$$

Let U be the (nonempty) set of fixed-points of G and $u^* \in U$, we have then $G(u^*) = u^*$ and

$$\begin{aligned}
\|u^{k+1} - u^*\|^2 &= \|(1-\theta)(u^k - u^*) + \theta(Tu^k - u^*)\|^2 \\
&= (1-\theta)\|u^k - u^*\|^2 + \theta\|Tu^k - u^*\|^2 - \theta(1-\theta)\|Tu^k - u^k\|^2 \\
&= (1-\theta)\|u^k - u^*\|^2 + \theta\|Tu^k - Tu^*\|^2 - \theta(1-\theta)\|Tu^k - u^k\|^2 \\
&\leq (1-\theta)\|u^k - u^*\|^2 + \theta\|u^k - u^*\|^2 - \theta(1-\theta)\|Tu^k - u^k\|^2 \\
&= \|u^k - u^*\|^2 - \theta(1-\theta)\|Tu^k - u^k\|^2
\end{aligned} \tag{3.3}$$

This shows that the distance to the solution set decreases at each step. We call this property Fejèr monotonicity.

Applying the inequality k times yields

$$\|u^{k+1} - u^*\|^2 \leq \|u^0 - u^*\|^2 - \theta(1-\theta)\sum_{j=0}^{k}\|Tu^j - u^j\|^2$$

and hence

$$\sum_{j=0}^{k}\|Tu^j - u^j\|^2 \leq \frac{\|u^0 - u^*\|^2 - \|u^{k+1} - u^*\|^2}{\theta(1-\theta)} \leq \frac{\|u^0 - u^*\|^2}{\theta(1-\theta)}.$$

As the upper bound does not depend on k, the series of non-negative terms remains bounded as $k \to \infty$ and we conclude that $\|Tu^k - u^k\| \to 0$ as $k \to \infty$.

From that we can also estimate a convergence rate of the fixed-point residual:

$$\min_{j=0\ldots k}\|Tu^j - u^j\|^2 \leq \frac{\|u^0 - u^*\|^2}{(k+1)\theta(1-\theta)},$$

Since the iterates $\{u^k\}_{k=1}^{\infty}$ lie in the compact set

$$C = \left\{v \mid \|v - u^*\| \leq \|u^0 - u^*\|\right\},$$

there exists at least one subsequence $\{u^{k_l}\}_{l=1}^{\infty}$ which converges to some point $\widehat{u}$.

Since $Tu^{k_l} - u^{k_l} \to 0$, we have $Gu^{k_l} - u^{k_l} = (G-I)u^{k_l} \to 0$ and, as $G-I$ is Lipschitz continuous, we have that $G\widehat{u} = \widehat{u}$ and the subsequence converges to a point in $\widehat{u} \in U$.

As (3.3) holds for any point from $u^* \in U$, we can apply it to the point $\widehat{u}$ our subsequence converges to. We know that for the iterates of the original sequence the distance to this point is monotonically decreasing,

$$\|u^{k+1} - \widehat{u}\| \leq \|u^k - \widehat{u}\|.$$

Since a subsequence $\{u^{k_l}\}_{l=1}^{\infty}$ of $\{u^k\}_{k=1}^{\infty}$ is converging to $\widehat{u}$, and $\|u^k - \widehat{u}\|$ is monotonically decreasing, we have convergence of the entire sequence to $\widehat{u}$. □

3.3 Gradient Descent as an Averaged Operator

Given a differentiable convex function $E\colon \mathbb{R}^n \to \mathbb{R}$, consider the problem

$$u \in \arg\min_{u\in\mathbb{R}^n} E(u).$$

The first-order optimality conditions of the problem characterize the solution $u^\star$ by

$$\nabla E(u^\star) = 0 \iff u^\star = (I - \tau \nabla E)u^\star$$

for any $\tau \neq 0$. The fixed-point iteration for this setup is

$$u^{k+1} = u^k - \tau \nabla E(u^k).$$

This algorithm is called **gradient descent** with a constant step size $\tau > 0$. To guarantee convergence of this fixed-point iteration, we need to determine under which conditions $(I - \tau \nabla E)$ is a contraction or an averaged operator. To this purpose, we will use the following result.

Theorem 44. Baillon-Haddad theorem. *A continuously differentiable convex function $E : \mathbb{R}^n \to \mathbb{R}$ is L-smooth if and only if $\frac{1}{L}\nabla E$ is firmly nonexpansive, i.e.*

$$\langle \nabla E(u) - \nabla E(v), u - v \rangle \geq \frac{1}{L} \|\nabla E(u) - \nabla E(v)\|_2^2$$

for all $u, v \in \mathbb{R}^n$.

Proof. Let $u, v \in \mathrm{dom}(E)$, we first show that if E is firmly non-expansive, then E is L-smooth. By combining Cauchy-Schwarz inequality with the inequality that defines a firmly non-expansive function, we have

$$\|\nabla E(u) - \nabla E(v)\| \|u - v\| \geq \langle \nabla E(u) - \nabla E(v), u - v \rangle \geq \frac{1}{L} \|\nabla E(u) - \nabla E(v)\|^2.$$

Dividing the inequality above by $\|\nabla E(u) - \nabla E(v)\|$, when $\nabla E(u) \neq \nabla E(v)$, we obtain that inequality that defines L-smoothness.

$$\|\nabla E(u) - \nabla E(v)\| \geq L \|u - v\|$$

and E^{-1} has Lipschitz constant $\frac{1}{L}$.

To prove the other direction, assume that E is L-smooth,

$$\begin{aligned}
0 &\geq \|\nabla E(u) - \nabla E(v) - L(u - v)\|^2 \\
&= \|\nabla E(u) - \nabla E(v)\|^2 + L^2 \|u - v\|^2 - 2L \langle \nabla E(u) - \nabla E(v), u - v \rangle \\
&\geq \|\nabla E(u) - \nabla E(v)\|^2 + \|\nabla E(u) - \nabla E(v)\|^2 - 2L \langle \nabla E(u) - \nabla E(v), u - v \rangle \\
&\geq 2\|\nabla E(u) - \nabla E(v)\|^2 - 2L \langle \nabla E(u) - \nabla E(v), u - v \rangle.
\end{aligned}$$

Rearranging the terms, we obtain

$$\begin{aligned}
\langle \nabla E(u) - \nabla E(v), u - v \rangle &\geq \frac{1}{L} \|\nabla E(u) - \nabla E(v)\|^2 \\
\langle \frac{1}{L}\nabla E(u) - \frac{1}{L}\nabla E(v), u - v \rangle &\geq \|\frac{1}{L}\nabla E(u) - \frac{1}{L}\nabla E(v)\|^2,
\end{aligned}$$

that is, $\frac{1}{L}\nabla E$ is firmly non-expansive.

□

We can now determine the conditions under which gradient descent with a constant step size converges.

Theorem 45. *If $E : \mathbb{R}^n \to \mathbb{R}$ has a minimizer, is convex, and L-smooth, then the gradient descent iteration*

$$u^{k+1} = u^k - \tau \nabla E(u^k)$$

with constant step size $\tau \in (0, \frac{2}{L})$ converges to a minimizer of E.

Proof. We will show that the fixed-point operator of gradient descent $G(u) = u - \tau\nabla E(u)$ is averaged.

By Baillon-Haddad theorem, we know that $\frac{1}{L}\nabla E$ is firmly non-expansive, or equivalently, averaged with $\alpha = 1/2$. Let $\frac{1}{L}\nabla E = \frac{1}{2}(I + T)$ for a non-expansive T, it holds

$$G(u) = u - \tau L\frac{1}{L}\nabla E(u) = u - \tau L(\frac{1}{2}u + \frac{1}{2}Tu) = \left(1 - \frac{L\tau}{2}\right)u + \frac{L\tau}{2}(-Tu)$$

It is clear that $-T$ is non-expansive because T is non-expansive. Consequently G is averaged for $\frac{L\tau}{2} \in (0, 1)$, that is $\tau \in (0, \frac{2}{L})$. □

To achieve a linear convergence rate, we need a fixed-point iteration defined by a contraction. In the case of gradient descent, this requires the function to be a little more than convex, it has to be strongly convex.

Definition A function $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ is called *strongly convex* with constant m or m-strongly convex if $E(u) - \frac{m}{2}\|u\|_2^2$ is still convex.

To determine if a function is strongly convex, we will use the following equivalent definition.

Theorem 46. *For a continuously differentiable function E, the following are equivalent:*

1. $E(u) - \frac{m}{2}\|u\|^2$ *is convex*
2. $E(v) \geq E(u) + \langle\nabla E(u), v - u\rangle + \frac{m}{2}\|v - u\|^2$
3. $\langle\nabla E(u) - \nabla E(v), u - v\rangle \geq m\|v - u\|^2$
4. *If E is twice-continuously differentiable, $\nabla^2 E(x) \succeq mI$.*

Proof. See *Ryu, Boyd, A Primer on Monotone Operator Methods, Appendix A* □

We are now in a position to determine when the fixed-point update of the gradient descent algorithm defines a contraction and achieves a linear convergence rate.

Theorem 47. *If $E : \mathbb{R}^n \to \mathbb{R}$ is strongly convex with parameter m and strongly smooth with parameter L, then the gradient descent iteration with constant step size $\tau \in (0, \frac{2}{L})$ converges to the unique minimizer $u^\star$ with geometric convergence rate*

$$\|u^k - u^\star\| \leq c^k\|u^0 - u^\star\|.$$

Proof. We will show that $(I - \tau\nabla E)$ is Lipschitz with parameter $c = \max\{|1 - \tau m|, |1 - \tau L|\}$. To simplify the proof, we will assume that E is twice continuously differentiable although the result is still true without this assumption. If E is twice continuously differentiable, we have

- $D(I - \tau\nabla E) = I_n - \tau\nabla^2 E$, where I_n is the identity matrix
- m strong convexity is equivalent to $\nabla^2 E \succeq mI_n$
- L-smoothness corresponds to $\nabla^2 E \preceq LI_n$

Putting these together, we have

$$\begin{aligned}(1 - \tau L)I_n \preceq D(I - \tau\nabla E) &\preceq (1 - \tau m)I_n \\ \|D(I - \tau\nabla E)\| &\leq \max\{|1 - \tau m|, |1 - \tau L|\} \\ (I - \tau\nabla E) \text{ has Lipscitz constant } c &= \max\{|1 - \tau m|, |1 - \tau L|\}. \end{aligned} \tag{3.4}$$

As a result, $(I - \tau\nabla E)$ is a contraction for $\tau \in (0, \frac{2}{L})$ and the fixed-point iteration converges to the unique fixed point of the contraction with the geometric rate c^k by Banach fixed-point theorem.

□

3.4 Projected Gradient Descent

Definition Projection For a (nonempty) closed convex set $C \subset \mathbb{R}^n$,

$$\pi_C(v) = \operatorname*{argmin}_{u \in C} \|u - v\|_2^2$$

is called the projection of v onto the set C.

In plain English, the projection of a point v onto C is the point in C that is closest v. As a result, if $v \in C$ then $\pi_C(v) = v$ and the reverse if also true: $\pi_C(v) = v$ if and only if $v \in C$.

As the projection is defined in terms of a minimization problem, it is natural to wonder if the optimization problem has a solution and whether this solution is unique. The convexity of the set C, gives us a positive answer.

Theorem 48. Existence and Uniqueness of the Projection *For any (nonempty) closed convex set $C \subset \mathbb{R}^n$ and any v the projection $\pi_C(v)$ exists and is single valued.*

Proof. To show that $\pi_C(v)$ is not empty, we define

$$E(u) = \begin{cases} \|u - v\|^2 & \text{if } u \in C \\ \infty & \text{otherwise} \end{cases}.$$

As C is not empty, we can pick $v_0 \in C$ and define the sublevel set $S_{E(v_0)} = \{u \in \mathbb{R}^n : \ E(u) \leq E(v_0)\}$. As $v_0 \in S_{E(v_0)}$, the sublevel set is not empty. It is also bounded because any $u \in S_{E(v_0)}$, satisfies

$$\|u\| \leq \|u - v\| + \|v\| = \sqrt{E(u)} + \|v\| \leq \sqrt{E(v_0)} + \|v\|,$$

where v is here the fixed point where we evaluate $\pi_C(v)$.

We also have that $\operatorname{epi}(E)$ is closed because C is closed, and we have already seen that closed functions are l.s.c. As a result, we can use Theorem 24 to prove existence of a minimizer of $E(u)$.

The minimizer is unique because $E(u)$ is strictly convex as a result of the convexity of C and the strict convexity of $\|u - v\|^2$. □

Because the minimizer is unique, although $\pi_C(v)$ is by definition a set, we usually identify $\pi_C(v)$ with the single element in the set.

It is useful to know the form of the projection operator for some common sets. We derived in the class the projection for the following sets:

- $C = \{u \in \mathbb{R}^n \mid \|u\|_2 \leq 1\}$
- $C = \{u \in \mathbb{R}^n \mid \|u\|_\infty := \max_i |u_i| \leq 1\}$
- $C = \{u \in \mathbb{R}^n \mid u_i \in [a, b]\}$
- $C = \{u \in \mathbb{R}^n \mid u_i \geq a\}$
- $C = \{u \in \mathbb{R}^n \mid \|u\|_1 = \sum_i |u_i|\}$

Projection operators are necessary in optimization to solve problems subject to a closed convex constraint set C

$$u^* \in \arg\min_{u \in C} E(u), \tag{3.5}$$

When the objective function E is also convex, the optimization problem is convex but we do not know how to solve it yet. The projected gradient algorithm is the first step towards this goal.

The projected gradient descent algorithm builds on gradient descent to find a solution of (3.5) when E is convex and L-smooth. To this goal, let us look at the gradient descent update rule

$$u^{k+1} = u^k - \tau \nabla E(u^k).$$

The problem with this update for solving problem (3.5) is that, even if $u^k \in C$, the update u^{k+1} might lie outside the feasible set C. Gradient projection solves this by simply projecting every iteration back to the feasible set with $u^{k+1} = \pi_C(u^k - \tau \nabla E(u^k))$.

Definition Gradient Projection Algorithm Let $C \subset \mathbb{R}^n$ be a nonempty closed convex set and let $E : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable. Then, for $u^0 \in C$

$$u^{k+1} = \pi_C(u^k - \tau \nabla E(u^k))$$

is called the *gradient projection* algorithm.

Similar to gradient descent, we can write the gradient projection algorithm as a fixed point iteration of an operator

$$G(u) = \pi_C(u - \tau \nabla E(u))$$

to analyze its convergence. In particular, we need to determine under which conditions G is an averaged operator or a contraction to use Banach or Krasnosel'skii-Mann theorems to prove its convergence.

From the analysis of gradient descent, we know that if E is L-smooth and $\tau \in (0, \frac{2}{L})$ the operator

$$G_1(u) = u - \tau \nabla E(u)$$

is averaged. If we know recollect that the composition of averaged operators is also averaged, we only need to show that the projection π_C is averaged. In fact, we will see that is it firmly nonexpansive, and therefore, it is averaged with $\alpha = \frac{1}{2}$.

Theorem 49. *The projection π_C onto a nonempty closed convex set $C \subset \mathbb{R}^n$ is firmly nonexpansive, i.e. it meets*

$$\langle u - v, \pi_C(u) - \pi_C(v) \rangle \geq \|\pi_C(u) - \pi_C(v)\|^2 \qquad \forall u, v \in \mathbb{R}^n.$$

Proof. Let δ_C be the indicator function of the convex set C, which is defined as

$$\delta_C(v) = \begin{cases} 0 & \text{if } v \in C \\ \infty & \text{otherwise} \end{cases}.$$

As C is not empty, closed, and convex, δ_C is proper, closed, and convex. We can now write

$$\pi_C(u) = \underset{z \in \mathbb{R}^n}{\operatorname{argmin}} \ \delta_C(z) + \|z - u\|_2^2$$

From the optimality conditions of the optimization problem we have

$$u - \pi_C(u) \in \partial \delta_C(\pi_C(u))$$
$$v - \pi_C(v) \in \partial \delta_C(\pi_C(v)).$$

At the same time, recall the definition of the subgradient of function E

$$E(z) - E(x) \geq \langle p, z - x \rangle \ \ \forall z \ \ p \in \partial E(x).$$

If we apply this inequality with $E = \delta$ at the points $x = \pi_C(u)$ and $x = \pi_C(v)$, we have

$$\begin{aligned}\delta_C(z) - \delta_C(\pi_C(u)) &\geq \langle u - \pi_C(u), z - \pi_C(u)\rangle \quad \forall z \\ \delta_C(z) - \delta_C(\pi_C(v)) &\geq \langle v - \pi_C(v), z - \pi_C(v)\rangle \quad \forall z.\end{aligned} \tag{3.6}$$

If we choose $z = \pi_C(v)$ for the first inequality and $z = \pi_C(u)$ for the second and consider that $\delta_C(\pi_C(u)) = \delta_C(\pi_C(v)) = 0$ we have

$$\begin{aligned}0 &\geq \langle u - \pi_C(u), \pi_C(v) - \pi_C(u)\rangle \\ 0 &\geq \langle v - \pi_C(v), \pi_C(u) - \pi_C(v)\rangle.\end{aligned} \tag{3.7}$$

Adding both inequalities, we obtain

$$\begin{aligned}0 &\geq \langle u - \pi_C(u) + \pi_C(v) - v, \pi_C(v) - \pi_C(u)\rangle \\ 0 &\geq \langle u - v, \pi_C(v) - \pi_C(u)\rangle + \|\pi_C(v) - \pi_C(u)\|^2 \\ \langle u - v, \pi_C(u) - \pi_C(v)\rangle &\geq \|\pi_C(u) - \pi_C(v)\|^2,\end{aligned}$$

which shows that the projection is firmly nonexpansive. □

We can now state the main convergence result of projected gradient algorithm.

Theorem 50. *For an L-smooth energy E that has a minimizer and a choice $\tau \in (0, \frac{2}{L})$ the gradient projection converges to a solution of*

$$u^* \in \arg\min_{u \in C} E(u) \tag{3.8}$$

with convergence rate $\mathcal{O}(1/k)$.

The convergence rate $\mathcal{O}(1/k)$ of the vanilla projected gradient algorithm is suboptimal, but it can be improved to $\mathcal{O}(1/k^2)$ with acceleration techniques that introduce an extrapolation step exploiting the L-smoothness of E . We can also improve this rate if our objective function is m-strongly convex.

Theorem 51. *For E being L-smooth and m-strongly convex and $\tau \in (0, \frac{2}{L})$ the gradient projection algorithm converges to the (unique) global minimizer u^* with $E(u^k) - E(u^*) \in \mathcal{O}(c^k)$ with $c < 1$.*

Proof. Recall that the composition of a non-expansive operator with a contraction is a contraction. As a result, whenever $G_1(u) = u - \tau\nabla E(u)$ is a contraction, the gradient projection operator $\pi_C(G_1(u))$ is a contraction and we can use Banach fixed-point theorem to prove convergence with a linear rate . □

3.5 Proximal Gradient Descent

Definition Proximal Operator Given a closed, proper, convex function $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, the mapping $\mathsf{prox}_E : \mathbb{R}^n \to \mathbb{R}^n$ defined as

$$\mathsf{prox}_E(v) := \operatorname*{argmin}_{u \in \mathbb{R}^n} \; E(u) + \frac{1}{2}\|u - v\|^2$$

is called the proximal operator or proximal mapping of E.

The proximal operator is a generalization of the projection. Indeed, given a nonempty, closed, convex set C, the projection π_C is the proximal operator of the indicator function

$$\delta_C(v) = \begin{cases} 0 & \text{if } v \in C \\ \infty & \text{otherwise} \end{cases}.$$

Many of the properties of the projection are inherited by the proximal operator.

Lemma 52. *Given a closed, proper, and convex function $E\colon \mathbb{R}^n \to \bar{\mathbb{R}}$ and $u \in \mathbb{R}^n$ there exists a unique proximal $\mathsf{prox}_E(u)$.*

Proof. We have already proved the existence of minimizers of convex functions that are coercive, that it, functions that have bounded sublevel sets. In the case of the proximal operator, it is easy to see that $E(u) + \frac{1}{2}\|u - v\|^2$ is convex if E is convex and that it has bounded sublevel sets.

Uniqueness is a consequence of the strong convexity of $(1/2)\|u - v\|^2$, which implies strong convexity of $E(u) + (1/2)\|u - v\|^2$ □

Similarly, the proximal operator is also firmly nonexpansive, as the next theorem shows.

Theorem 53. *The proximal operator $prox_E$ for a closed, proper, convex function E is firmly nonexpansive, that is, it satisfies*

$$\langle u - v, prox_E(u) - prox_E(v)\rangle \geq \|prox_E(u) - prox_E(v)\|^2 \quad \forall u,\ v \in \mathbb{R}^n.$$

Proof. Let $x = \mathsf{prox}_E(u)$ and $y = \mathsf{prox}_E(v)$, the optimality conditions then

$$\begin{aligned} x = \min_z\ E(z) + 0.5\|z - u\|^2 &\Rightarrow u - x \in \partial E(x) \\ x = \min_z\ E(z) + 0.5\|z - v\|^2 &\Rightarrow v - y \in \partial E(y). \end{aligned} \tag{3.9}$$

At the same time, by definition of the subgradient we have

$$\begin{aligned} E(z) - E(x) \geq \langle \partial E(x), z - x\rangle = \langle u - x, z - x\rangle \quad \forall z \\ E(z) - E(y) \geq \langle \partial E(y), z - y\rangle = \langle v - y, z - y\rangle \quad \forall z, \end{aligned} \tag{3.10}$$

where in the second step of the inequalities we have used that $u - x \in \partial E(x)$ and $v - y \in \partial E(y)$. If we now choose $z = y$ for the first inequality and $z = x$ for the second, we have

$$\begin{aligned} E(y) - E(x) &\geq \langle u - x, y - x\rangle \\ E(x) - E(y) &\geq \langle v - y, x - y\rangle \end{aligned}$$

adding both inequalities we have

$$\begin{aligned} 0 &\geq \langle u - x + y - v, y - x\rangle \\ 0 &\geq \langle u - v, y - x\rangle + \|y - x\|^2 \\ \langle u - v, x - y\rangle &\geq \|x - y\|^2. \\ \langle u - v, \mathsf{prox}_E(u) - \mathsf{prox}_E(v)\rangle &\geq \|\mathsf{prox}_E(u) - \mathsf{prox}_E(v)\|^2 \end{aligned}$$

which shows that the proximal mapping of a convex function is firmly non-expansive. □

For many common convex objective functions, the proximal operators is simple to compute and has a closed-form expression.

- Quadratic functions

$$f(x) = \frac{1}{2}\|Au - b\|^2, \quad \mathsf{prox}_{\tau f}(v) = (I + \tau A^T A)^{-1}(v - \tau b)$$

- Euclidean norm

$$f(x) = \|x\|, \quad \mathsf{prox}_{\tau f}(v) = \begin{cases} (1 - \tau/\|v\|)v & \text{if } \|v\| \geq \tau \\ 0 & \text{otherwise.} \end{cases}$$

- ℓ_1-norm (cf. exercise sheet 3)

$$f(x) = \|x\|_1, \quad \left(\mathsf{prox}_{\tau f}(v)\right)_i = \begin{cases} v_i + \tau & \text{if } v_i < -\tau \\ 0 & \text{if } |v_i| \leq \tau \\ v_i - \tau & \text{if } v_i > \tau. \end{cases}$$

In the same way as we used the projection to generalize gradient descent to constrained minimization problems, we use the proximal operator to generalize gradient descent to optimization problems of the form

$$E(u) = E_1(u) + E_2(u),$$

where both E_1 and E_2 are proper, closed, and convex and satisfy

- $E_1 : \mathbb{R}^n \to \mathbb{R}$ is L-smooth.
- $E_2 : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ has an easy-to-evaluate proximal operator.

In this case, we can generalize the projected gradient algorithm by taking gradient descent steps on E_1 and proximal steps on E_2. This strategy is known as proximal gradient method.

Definition Proximal Gradient Method For a closed, proper, convex function $E_1, E_2 : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, where E_1 is differentiable, and given an initial point $u^0 \in \mathbb{R}^n$ and a step size τ, the algorithm

$$u^{k+1} = \mathsf{prox}_{\tau E_2}\left(u^k - \tau \nabla E_1(u^k)\right), \qquad k = 0, 1, 2, \ldots,$$

is called the *proximal gradient method* or *forward-backward splitting*.

For a constant E_2, the proximal gradient method reduces to gradient descent, while for $E_2 = \delta_C$ it reduces to projected gradient descent. The case we have not seen, constant E_1, results in the *proximal point algorithm*.

Intuitively, the *proximal gradient method* does an alternate minimization. The first step decreases the value of E_1 with a gradient descent update

$$u^{k+\frac{1}{2}} = u^k - \tau \nabla E_1(u^k)$$

while the second step

$$u^{k+1} = \min_u E_2(u) + \frac{1}{2}\|u - u^{k+\frac{1}{2}}\|^2$$

finds a minimizer of $E_2(u)$ that is no far from $u^{k+\frac{1}{2}}$. The term $\frac{1}{2}\|u - u^{k+\frac{1}{2}}\|^2$ in the proximal map, then ensures that this second step does not revert any progress made in the minimization of E_1 by the gradient descent step.

We can again analyze the convergence of the proximal gradient method as the fixed-point iteration of the operator

$$G(u) = \mathsf{prox}_{\tau E_2}(u - \tau \nabla E_1(u)).$$

We do so by determining the conditions under which G is averaged or a contraction.

Theorem 54. *For closed, proper, convex functions E_1 and E_2, with E_1 L-smooth and having a minimizer u^* of $E(u) = E_1(u) + E_2(u)$, the proximal gradient method with constant step size $\tau \in (0, \frac{2}{L})$ converges to u^* with rate $E(u^k) - E(u^*) \in \mathcal{O}(1/k)$.*

Proof. The operator G is averaged for $\tau \in (0, \frac{2}{L})$ because it is the composition of two averaged operators, the gradient-descent operator $G_1(u) = u - \tau E_1(u)$ and the proximal operator $\mathsf{prox}_{\tau E_2}$. Indeed, we have already seen that if E_1 is L-smooth closed, proper, convex, then G_1 is averaged for $\tau \in (0, \frac{2}{L})$. At the same time, $\mathsf{prox}_{\tau E_2}$ is averaged with $\alpha = \frac{1}{2}$ because it is firmly nonexpansive. The convergence rate results from particularizing the analysis of the convergence rate of fixed-point iterations of an averaged operator to the proximal gradient. □

As we have done with gradient descent and the projected gradient algorithms, we can obtain linear convergence with some additional assumptions on the objective function. To this goal, we need to determine under which conditions the proximal operator is not only nonexpansive but a contraction.

Theorem 55. *If the proper, closed function E is m-strongly convex, then $prox_{\tau E} : \mathbb{R}^n \to \mathbb{R}^n$ is a contraction.*

As the composition of a contraction with a nonexpansive mapping is a contraction, $\mathsf{prox}_{\tau E_2}(u - \tau \nabla E_1)$ is a contraction whenever $(u - \tau \nabla E_1)$ or $\mathsf{prox}\tau E$ are a contraction. This condition translates into strong convexity of E_1 and E_2.

Corollary 56. *If E_1 is L-smooth, $\tau \in (0, \frac{2}{L})$, and either E_1 or E_2 is strongly convex, then the proximal gradient method converges linearly, i.e., $\|u^k - u^*\|_2^2 \in \mathcal{O}(c^k)$ for some $c < 1$.*

Proof. This is an immediate result of Banach fixed-point iteration theorem. □

Chapter 4

Duality

Functional transforms can shed light into an optimization problem by presenting it from a different perspective. In convex analysis, this new perspective is the result of duality. There are many kinds of duality in mathematics. Even in optimization, some classes of problems have much stronger duality theorems than others.

In general, a dual problem is another problem formulated with the data of the original problem that tells something about the original problem. In nonlinear optimization the dual gives lower (or upper) bounds for the original problem. In convex optimization, the results are much stronger: every convex problem has a dual version, and their solutions are related by a transform.

We will start by introducing the convex conjugate of a function $E(u)$, which is convex, and then show how conjugating twice a convex functions give us the original function reformulated as a maximization problem over a dual variable p. As a result, the minimization problem over u can be reformulated into a saddle-point problem

$$\min_u E(u) \iff \min_u \max_p \mathscr{L}(u, p).$$

For the moment this is only and act of faith, but we will see that under certain conditions we can additionally swap min and max

$$\min_u E(u) \iff \min_u \max_p \mathscr{L}(u, p) \iff \max_p \min_u \mathscr{L}(u, p).$$

If we can now analytically solve the interior minimization problem

$$u^\star = \operatorname*{argmin}_u \mathscr{L}(u, p)$$

then we have found a closed-form formulation of the dual problem

$$\min_u E(u) \iff \max_p \mathscr{L}(u^\star, p).$$

The dual problem usually has complementary properties to the primal (the original) one. For instance, if the primal problem is the composition of affine transforms with simple convex functions, the dual tends to have a larger number of optimization variables but has simpler proximal operators or gradients. Duality is also useful when the primal problem is not differentable but is m-strongly convex, because its dual is $\frac{1}{m}$-smooth and can be solved with simple gradient descent. Duality increases the set of problems that we can solve with the algorithms we have studied so far, or simply let us do it more efficiently, with lower memory requirements, or operation counts. This forms the first part of this chapter.

When looking at dual problem does not help us solve the primal one better, we look at the saddle point problem to try to combine the best of the primal and the dual worlds. This results in the primal-dual algorithms that we will cover in the second part of this chapter.

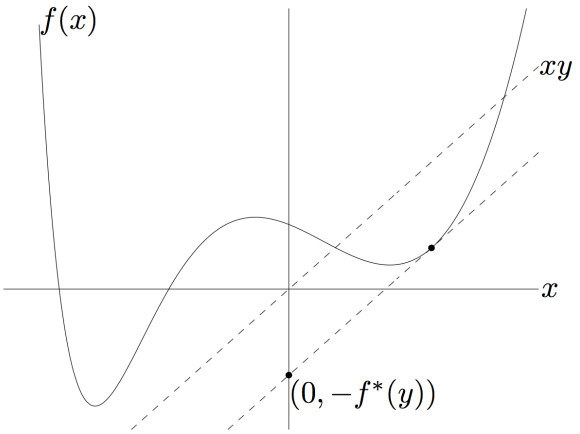


Fig. 4.1: Illustration of convex conjugate of a one-dimensional function $f\colon \mathbb{R} \mapsto \mathbb{R}$ evaluated at y. $f^*(y)$ is the maximum gap between the linear function yx and $f(x)$, as shown by the dashed line. If f is differentiable, this occurs at a point x where $f'(x) = y$. Source: Boyd, and Vandenberghe. *Convex optimization theory.*2004

4.1 Convex Conjugate

Definition Let $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be any function, not necessarily convex, we define its *convex conjugate* to be

$$E^*(p) = \sup_{u \in \mathbb{R}^n} \left[\langle u, p\rangle - E(u)\right].$$

The domain of the conjugate function consists of $p \in \mathbb{R}^n$ for which the supremum is finite, i.e., for which the difference $\langle u, p\rangle - E(u)$ is bounded above on $\mathrm{dom}(E)$.

Figure 4.1 provides a first geometric interpretation of the conjugate. Given a one-dimensional function $E\colon \mathbb{R} \mapsto \mathbb{R}$ and $p \in \mathbb{R}$, the conjugate function $E^*(p)$ is the maximum gap between the linear function pu and $E(u)$.

As the name suggests, the convex conjugate of a function is convex, even when the function is not. In fact, it is also closed.

Lemma 57. ***Convexity of the Convex Conjugate*** *The convex conjugate*

$$E^*(p) = \sup_{u \in \mathbb{R}^n} \left(\langle u, p\rangle - E(u)\right).$$

of any proper function $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is convex and closed.

Proof. The convexity of E^* results from its definition as a supremum. We can prove it, by proving the convexity of its epigraph. Let $(p, \alpha) \in \mathrm{epi}(E^*)$, $(q, \beta) \in \mathrm{epi}(E^*)$, by definition of epigraph we have

$$\alpha \geq \sup_{u \in \mathbb{R}^n} \left[\langle u, p\rangle - E(u)\right] \tag{4.1}$$

$$\beta \geq \sup_{u \in \mathbb{R}^n} \left[\langle u, q\rangle - E(u)\right]. \tag{4.2}$$

Multiplying the first inequality by $\theta \in [0, 1]$ and the second by $(1 - \theta)$ and adding them up, we obtain

$$\begin{aligned}
\theta\alpha + (1-\theta)\beta &\geq \sup_{u \in \mathbb{R}^n} \left[\langle u, \theta p\rangle - \theta E(u)\right] + \sup_{u \in \mathbb{R}^n} \left[\langle u, (1-\theta)q\rangle - (1-\theta)E(u)\right] \\
&\geq \sup_{u \in \mathbb{R}^n} \left[\langle u, \theta p\rangle - \theta E(u) + \langle u, (1-\theta)q\rangle - (1-\theta)E(u)\right] \\
&\geq \sup_{u \in \mathbb{R}^n} \left[\langle u, \theta p + (1-\theta)q\rangle - E(u)\right]
\end{aligned}$$

which shows that $\theta(p,\alpha) + (1-\theta)(q,\beta) \in \text{epi}(E^*)$. To show that it is closed, we will show that its epigraph is the intersection of an aribraty number of closed convex sets $\text{epi}(E^*) = \{(p,\alpha) \in \mathbb{R}^n \times \mathbb{R} : \langle u,p\rangle - E(u) \leq \alpha \ \forall u\} = \cap_u \text{epi}E_u$, where $E_u(\cdot) = \langle u, \cdot\rangle - E(u)$ is continuous and thus closed. Since each $\text{epi}E_u$ is closed, and any arbitrary intersection of closed sets is closed, $\text{epi}(E^*)$ is closed. □

As we have seen in the previous proof,

$$\begin{aligned}(p,\alpha) \in \text{epi}(E^*) &\iff \alpha \geq \sup_{u\in\mathbb{R}^n} [\langle u,p\rangle - E(u)] \\ &\iff E(u) \geq \langle u,p\rangle - \alpha \ \ \forall u \in \mathbb{R}^n. \end{aligned} \tag{4.3}$$

This means that the points of the epigraph of E^* parameterize the affine functions minorizing E. In other words, if the affine function $l(u) = \langle p,u\rangle - \alpha$ minorizes E, then the affine function $m(u) = \langle p,u\rangle - E^*(p)$ is the largest affine minorizer and satisfies

$$l(u) \leq m(u) \leq E(u).$$

We will mostly study convex conjugates of convex functions, and refer to them as conjugates for simplicity. Some of them are classic examples that is useful to know instead of derive every single time.

- If $E(u) = \frac{1}{2}\|u\|^2$, the optimality conditions of

  $$\sup_u \ \langle p,u\rangle - \frac{1}{2}\|u\|^2$$

  show that the supremum is attained at $\hat{u} = p$, where $\langle p,\hat{u}\rangle - \frac{1}{2}\|\hat{u}\|^2 = \frac{1}{2}\|p\|^2$. This yields $E^*(p) = \frac{1}{2}\|p\|^2$.

- If E is the affine transform $E(u) = \langle a,u\rangle + b$, the function of p

  $$\sup_{u\in\mathbb{R}^n} \langle p,u\rangle - \langle a,u\rangle - b$$

  is bounded if and only if $p = a$, in which case it is constant. Therefore the domain of the conjugate function E^* is the singleton $\{a\}$, and

  $$E^*(p) = \begin{cases} -b & \text{if } \ p = a. \\ \infty & \text{otherwise.} \end{cases}$$

- If $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$, the convex conjugate of $E(u) = \|u\|$ is $E^*(p) = \begin{cases} 0 & \text{if } \|p\|_* \leq 1, \\ \infty & \text{else.} \end{cases}$

  Recall the definition of the dual norm: given $p \in \mathbb{R}^n$

  $$\|p\|_* = \sup\{\langle p,u\rangle \ : \ u \in \mathbb{R}^n, \ \|u\| \leq 1\}.$$

  As a result, if $\|p\|_* > 1$, there exists $x \in \mathbb{R}^n$ with $\|x\| < 1$ such that $\langle x,p\rangle > 1$ and $\langle p,x\rangle - \|x\| > 0$. Now, define $z = tx$, we have

  $$\sup_{u\in\mathbb{R}^n} \ \langle p,u\rangle - \|u\| \geq \sup_t \ \langle p,tx\rangle - \|tx\| = \sup_{t>0} \ t[\langle p,x\rangle - \|x\|] = \infty.$$

  Conversely, if $\|p\| \leq 1$, we have $\langle p, \frac{u}{\|u\|}\rangle \leq 1$ for all u, which implies $\langle p,u\rangle \leq \|u\|$ for all u. Therefore $u = 0$ is the value that maximizes $\langle p,u\rangle - \|u\|$ with maximum value 0.

  In particular, we have

  - The conjugate of $E(u) = \|u\|_2$ is $E^*(p) = \begin{cases} 0 & \text{if } \|p\|_2 \leq 1, \\ \infty & \text{else.} \end{cases}$
  - The conjugate of $E(u) = \|u\|_1$ is $E^*(p) = \begin{cases} 0 & \text{if } \|p\|_\infty \leq 1, \\ \infty & \text{else.} \end{cases}$

  - The conjugate of $E(u) = \|u\|_\infty$ is $E^*(p) = \begin{cases} 0 & \text{if } \|p\|_1 \leq 1, \\ \infty & \text{else.} \end{cases}$

- The convex conjugate of the indicator function of the unit ball $E(u) = \begin{cases} 0 & \text{if } \|u\| \leq 1, \\ \infty & \text{else.} \end{cases}$ is the dual norm $E^*(p) = \|p\|_*$. Indeed

$$\sup_{u \in \mathbb{R}^n} \langle p, u \rangle - E(u) = \sup_{\|u\| \leq 1} \langle p, u \rangle = \|p\|_*.$$

  In particular, we have

  - $E(u) = \begin{cases} 0 & \text{if } \|u\|_2 \leq 1, \\ \infty & \text{else.} \end{cases}$ leads to $E^*(p) = \|p\|_2$.
  - $E(u) = \begin{cases} 0 & \text{if } \|u\|_\infty \leq 1, \\ \infty & \text{else.} \end{cases}$ leads to $E^*(p) = \|p\|_1$.
  - $E(u) = \begin{cases} 0 & \text{if } \|u\|_1 \leq 1, \\ \infty & \text{else.} \end{cases}$ leads to $E^*(p) = \|p\|_\infty$.

Now that we know some basic conjugates, it is useful to investigate how the conjugation affects some basic operations like linear composition, scaling or affine transforms to increase our repertoire of conjugates.

- **conjugation reverses inequalities**: if $E_1(u) \leq E_2(u) \;\; \forall u \in \mathbf{R}^n$, then $E_1^*(p) \leq E_2^*(p) \;\; \forall p \in \mathbb{R}^n$. This follows immediately from the definition of conjugate function

$$E_1^*(p) = \sup_{u \in \mathbb{R}^n} \langle u, p \rangle - E_1(u) \geq \sup_{u \in \mathbb{R}^n} \langle u, p \rangle - E_2(u) = E_2^*(p). \tag{4.4}$$

- **Scalar multiplication** :If $E(u) = \alpha \tilde{E}(u)$

$$E^*(p) = \sup_u \langle p, u \rangle - \alpha \tilde{E}(u) = \alpha \sup_u \langle \frac{p}{\alpha}, u \rangle - \tilde{E}(u) = \alpha \tilde{E}^*(p/\alpha).$$

- **Separable sum**: If $E(u_1, u_2) = E_1(u_1) + E_2(u_2)$

$$\begin{aligned} E^*(p) &= \sup_{p=(p_1,p_2)} \langle p_1, u_1 \rangle + \langle p_2, u_2 \rangle - E_1(u_1) - E_2(u_2) \\ &= \sup_{p_1} \langle p_1, u_1 \rangle - E(u_1) + \sup_{p_2} \langle p_2, u_2 \rangle - E(u_2) \\ &= E_1^*(p_1) + E_2^*(p_2). \end{aligned} \tag{4.5}$$

- **Sum rule**: If E_1, E_2 are closed, convex, proper and $E(u) = E_1(u) + E_2(u)$

$$\begin{aligned} E^*(p) &= \sup_p \langle p, u \rangle - E_1(u) - E_2(u) \\ &= \sup_{p=p_1+p_2} \langle p_1, u \rangle - E_1(u) + \langle p_2, u \rangle - E_2(u) \\ &= \inf_{p=p_1+p_2} \sup_{p_1} \langle p_1, u \rangle - E_1(u) + \sup_{p_2} \langle p_2, u \rangle - E_2(u) \\ &= \inf_{p=p_1+p_2} E_1^*(p_1) + E_2^*(p_2). \end{aligned} \tag{4.6}$$

  Where we have used that

$$\sup_{p=p_1+p_2} F(p_1, p_2) = \inf_{p=p_1+p_2} \sup_{p_1} \sup_{p_2} F(p_1, p_2)$$

- **Translation**: If $E(u) = \tilde{E}(u - b)$

$$\begin{aligned} E^*(p) &= \sup_u \langle p, u\rangle - \tilde{E}(u-b) = \sup_u \langle p, b\rangle + \langle p, u-b\rangle - \tilde{E}(u-b) \\ &= \sup_{u-b} \langle p, b\rangle + \langle p, u-b\rangle - \tilde{E}(u-b) \\ &= \langle p, b\rangle + \tilde{E}^*(p). \end{aligned} \tag{4.7}$$

- **Additional affine functions**: If $E(u) = \tilde{E}(u) + \langle b, u\rangle + a$

$$\begin{aligned} E^*(p) &= \sup_u \langle p, u\rangle - \tilde{E}(u) - \langle b, u\rangle - a = \sup_u -a + \langle p-b, u\rangle - \tilde{E}(u) \\ &= -a + \tilde{E}^*(p-b). \end{aligned} \tag{4.8}$$

The conjugate of a differentiable function E is also called the Legendre transform of E. In this case, it takes a particular form because if E is convex and differentiable, with $\mathrm{dom}(E) = \mathbb{R}^n$, any maximizer u^* of $\langle p, u\rangle - E(u)$ satisfies $p = \nabla E(u^*)$, and conversely, if u^* satisfies $p = \nabla E(u^*)$, then u^* maximizes $\langle p, u\rangle - E(u)$. Therefore, if $p = \nabla E(u^*)$, we have

$$E^*(p) = \langle \nabla E(u^*), u^*\rangle - E(u^*).$$

This allows us to determine $E^*(p)$ for any p for which we can solve the gradient equation $p = \nabla E(v)$ for v. We can express this another way. Let $v \in \mathbb{R}^n$ be arbitrary and define $p = \nabla E(v)$. Then we have

$$E^*(p) = \langle \nabla E(v), v\rangle - E(v).$$

4.2 Duality Theorems

Theorem 58 (**Fenchel-Young Inequality**). *Let E be proper, convex and closed, $u \in dom(E) \subset \mathbb{R}^n$, and $p \in \mathbb{R}^n$, then*

$$E(u) + E^*(p) \geq \langle u, p\rangle.$$

Equality holds if and only if $p \in \partial E(u)$.

Proof. The inequality follows immediately from the definition of the conjugate

$$E(u) + E^*(p) = E(u) + \sup_v \langle v, p\rangle - E(v) \geq E(u) + \langle u, p\rangle - E(u) = \langle u, p\rangle.$$

To show the equality statement, we will show the remaining inequality

$$E(u) + E^*(p) \leq \langle u, p\rangle,$$

or, in other words,

$$E(u) + \langle p, z\rangle - E(z) \leq \langle u, p\rangle, \ \forall z.$$

Rewritten, the above is nothing but

$$E(z) - E(u) - \langle p, z-u\rangle \geq 0, \ \forall z,$$

which is simply the definition of the subgradient $p \in \partial E(u)$. □

Theorem 59 (Biconjugate). *Let $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be proper, convex and closed, then $E^{**} = E$.*

Proof. We'll show an incomplete proof that only considers the relative interior but gives already a quick intuition of why the statement makes sense. For the full proof, please check Rockafellar's book *Convex Analysis*, Theorem 12.2.

First of all, note that $E^{**}(u) \leq E(u)$ because

$$E^{**}(u) = \sup_p \langle p, u\rangle - E^*(p) = \sup_p \langle p, u\rangle - \sup_v[\langle p, v\rangle - E(v)] \leq \sup_p \langle p, u\rangle - [\langle p, u\rangle - E(u)] = E(u).$$

If E is subdifferentiable at u, let $q \in \partial E(u)$. Fenchel-Young inequality tells us that $E(u) + E^*(q) = \langle u, q\rangle$ and, by definition of the supremum, we have

$$E^{**}(u) = \sup_p \langle p, u\rangle - E^*(p) \geq \langle q, u\rangle - E^*(q) = E(u).$$

Combining both inequalities, we have $E^{**}(u) = E(u)$. □

From the previous proof, we have $E^{**}(u) \leq E(u)$ for any E. This provides a function E^{**} that minorizes everywhere the original function E, is convex, and lower semi-continuous (closed). In fact, we have

Lemma 60. *E^{**} is the largest convex lower semi-continuous envelope of E.*

Proof. To see this, recall:

- conjugation reverses inequalities: if $E \leq F$ then $E^* \geq F^*$.
- The conjugate function is always convex and lower semi-continuous (or closed).
- If E is proper, convex, and lower semi-continuous, then $E^{**} = E$.

From the above, it follows that if $\hat{E}$ is a lower semi-continuous convex function such that $\hat{E} \leq E$, then $\hat{E}^{**} = \hat{E} \leq E^{**}$. □

There are also interesting connections between the subdifferentials of E and E^*.

Lemma 61. Subgradient of convex conjugate *Let E be proper, convex and closed, then the following two conditions are equivalent:*

- $p \in \partial E(u)$
- $u \in \partial E^*(p)$

Proof. Let $p \in \partial E(u)$, by the Fenchel-Young Inequality we know that

$$E(u) + E^*(p) = \langle u, p\rangle.$$

On the other hand, $E = E^{**}$ such that

$$E^{**}(u) + E^*(p) = \langle u, p\rangle,$$

and the Fenchel-Young Inequality tells us that $u \in \partial E^*(p)$. Similarly, $u \in \partial E^*(p)$ implies $p \in \partial E(u)$. □

This let us prove an important property, that the convex conjugate of a strongly convex function is smooth. This is important because it means that we can minimize non-differentiable strongly convex problems with a descent technique through its conjugate.

Theorem 62. Conjugation of strongly convex functions *If $E : \mathbb{R}^n \to \bar{\mathbb{R}}$ is proper, closed and m-strongly convex, then E^* is proper, closed, convex and $1/m$-smooth.*

Proof. For a proper, closed, strongly convex function, the supremum in the definition of E^* is attained. To see this, let us re-write

$$\max_u \langle u, p\rangle - E(u) = -\min_u \; E(u) - \langle u, p\rangle$$

Now the convexity of E let us substitute the minimum by an infimum

$$\max_u \langle u, p\rangle - E(u) = -\inf_u \; E(u) - \langle u, p\rangle = \sup_u \langle u, p\rangle - E(u) = E^*(p).$$

The strong convexity of E also means that the minimum is unique. The optimality condition immediately yields that this minimum is attained for $p \in \partial E(u)$, i.e. for $u \in \partial E^*(p)$. Since the optimal u is unique, the subdifferential $\partial E^*(p)$ is single valued for all p, which yields the differentiability of E^*.

As E is m-strongly convex, $E - \frac{m}{2}\|\cdot\|^2$ is convex. As a result

$$\langle u - v, p - q\rangle \geq m\|u - v\|^2 \qquad \forall p \in \partial E(u), q \in \partial E(v),$$

or in other words

$$\langle \nabla E^*(p) - \nabla E^*(q), p - q\rangle \geq m\|\nabla E^*(p) - \nabla E^*(q)\|^2 \qquad \forall p, q.$$

Now, by Cauchy-Schwarz inequality we have

$$\|\nabla E^*(p) - \nabla E^*(q)\|\|p - q\| \geq \langle \nabla E^*(p) - \nabla E^*(q), p - q\rangle \geq m\|\nabla E^*(p) - \nabla E^*(q)\|^2 \qquad \forall p, q$$

which implies $\frac{1}{m}$-smoothness of E^* as

$$\|\nabla E^*(p) - \nabla E^*(q)\| \leq \frac{1}{m}\|p - q\| \;\; \forall p, q.$$

□

This give us the first hint on how to reformulate a convex problem into a friendlier version: simply changing E by its biconjugate E^{**} and checking if it is *in some way simpler* to solve.

Theorem 63 (Fenchel's Duality[1])**.** *Let $G : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ and $F : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ be proper, closed, convex functions and let there exist a $u \in ri(dom(G))$ such that $Ku \in ri(dom(F))$. Then*

	$\inf_u$	$G(u) + F(Ku)$	*"Primal"*
$=$	$\inf_u \sup_q$	$G(u) + \langle q, Ku\rangle - F^*(q)$	*"Saddle point"*
$=$	$\sup_q \inf_u$	$G(u) + \langle q, Ku\rangle - F^*(q)$	*"Saddle point"*
$=$	$\sup_q$	$-G^*(-K^*q) - F^*(q)$	*"Dual"*

Proof. Partial proof: Let us assume a minimum is attained at some $\hat{u}$. Our assumptions let us apply the sum rule of the subgradient to compute the optimality conditions of the primal problem

$$\hat{u} \in \operatorname*{argmin}_u \; G(u) + F(Ku) \tag{4.9}$$

$$0 \in q + K^*\hat{p} \qquad q \in \partial G(\hat{u}) \qquad \hat{p} \in \partial F(K\hat{u}) \tag{4.10}$$

$$q \in -K^*\hat{p} \qquad \hat{u} \in \partial G^*(q) \qquad K\hat{u} \in \partial F^*(\hat{p}) \tag{4.11}$$

$$\hat{u} \in \partial G^*(-K^*\hat{p}) \qquad K\hat{u} \in \partial F^*(\hat{p}) \tag{4.12}$$

[1] C.f. Rockafellar, *Convex Analysis*, Section 31

If we know take $\hat{u} \in \partial G^*(-K^*\hat{p})$ and $K\hat{u} \in \partial F^*(\hat{p})$ and use it to write

$$0 = K\hat{u} - K\hat{u} \in K\partial G^*(-K^*\hat{p}) - \partial F^*(\hat{p})$$

we see that $\hat{p}$ satisfies the optimality conditions of the dual problem $\hat{p} \in \arg\max_p \ -G^*(-K^*p) - F^*(p)$. Moreover, the optimal solution pair $(\hat{u}, \hat{p})$ satisfies

$$K\hat{u} \in \partial F^*(\hat{p}) \qquad -K^*\hat{p} \in \partial G(\hat{u}) \tag{4.13}$$

$$K\hat{u} \in \partial F^*(\hat{p}) \qquad \hat{u} \in \partial G^*(-K^*\hat{p}) \tag{4.14}$$

$$\hat{p} \in \partial F(K\hat{u}) \qquad \hat{u} \in \partial G^*(-K^*\hat{p}) \tag{4.15}$$

$$\hat{p} \in \partial F(K\hat{u}) \qquad -K^*\hat{p} \in \partial G(\hat{u}). \tag{4.16}$$

□

This immediately give us the following alternative characterizations of the solution of our problem.

Corollary 64. *Let the assumptions from Fenchel's Duality Theorem hold. If there exists a pair $(u, q) \in \mathbb{R}^n \times \mathbb{R}^n$ such that one of the following four equivalent conditions are met*

1. $-K^*q \in \partial G(u), \quad q \in \partial F(Ku),$
2. $-K^*q \in \partial G(u), \quad Ku \in \partial F^*(q),$
3. $u \in \partial G^*(-K^*q), \quad q \in \partial F(Ku),$
4. $u \in \partial G^*(-K^*q), \quad Ku \in \partial F^*(q),$

Then u solves the primal and q solves the dual optimization problem.

This corollary tells us how to obtain a solution of the primal problem from a solution of the dual.

4.3 Applications of Duality in Vision

We covered in class how to solve problem with the descent algorithms that we know by looking at their dual.

The classic TV denoising model minimizes a combination of least-squares data term $G(u) = \frac{1}{2}\|u-f\|^2$ that penalizes the deviations of u to the input image f and the TV regularizer $F(Ku) = \alpha\|Ku\|_{2,1}$ that penalizes non-smooth images in terms of the discrete gradient operator K. To find the dual formulation of the problem, we substitute $F(Ku)$ by its biconjugate

$$F^{**}(Ku) = \max_q \ \langle Ku, q > -\iota_{\|\cdot\|_{2,\infty}\leq 1}(q)$$

to obtain

$$\begin{aligned}\min_u \frac{1}{2}\|u-f\|_2^2 + \alpha\|Ku\|_{2,1} &= \min_u \max_{\|q\|_{2,\infty}\leq 1} \frac{1}{2}\|u-f\|_2^2 + \alpha\langle Ku, q\rangle \\ &= \max_{\|q\|_{2,\infty}\leq 1} \min_u \frac{1}{2}\|u-f\|_2^2 + \alpha\langle Ku, q\rangle\end{aligned}$$

Now the inner minimization problem obtains its optimum at $u = f - \alpha K^*q$ and the remaining problem in q becomes

$$\begin{aligned}\max_{\|q\|_{2,\infty}\leq 1} \frac{1}{2}\|f - \alpha K^*q - f\|_2^2 + \alpha\langle K(f - \alpha K^*q), q\rangle &= \max_{\|q\|_{2,\infty}\leq 1} \frac{1}{2}\|\alpha K^*q\|_2^2 + \alpha\langle Kf, q\rangle - \|\alpha D^*q\|_2^2 \\ &= \max_{\|q\|_{2,\infty}\leq 1} -\frac{1}{2}\|\alpha K^*q\|_2^2 + \alpha\langle Kf, q\rangle \\ &= \max_{\|q\|_{2,\infty}\leq 1} -\frac{1}{2}\|\alpha K^*q - f\|_2^2.\end{aligned}$$

Since we prefer minimizations over maximizations, we write

$$\begin{aligned}
\hat{q} &= \operatorname*{argmax}_{\|q\|_{2,\infty}\leq 1} -\frac{1}{2}\|\alpha K^* q - f\|_2^2 \\
\hat{q} &= \operatorname*{argmin}_{\|q\|_{2,\infty}\leq 1} \frac{1}{2}\left\|K^* q - \frac{f}{\alpha}\right\|_2^2 \\
\hat{q} &= \operatorname*{argmin}_{q\in C} \frac{1}{2}\left\|K^* q - \frac{f}{\alpha}\right\|_2^2 \text{ where } C = \{q \in \mathbb{R}^{nm\times 2c} \mid \|q\|_{2,\infty} \leq 1\}
\end{aligned}$$

This is a minimization of a convex, proper, closed, L-smooth function over a convex set C that we can solve with gradient projection as follows:

$$q^{k+1} = \pi_C\left(q^k - \tau D\Big(K^* q^k - \frac{f}{\alpha}\Big)\right).$$

Working with the dual of a convex problem does not solve all of our problems and a simple modification to the TV denoising problem is enough to render our trick useless. Consider TV-ℓ^1 denoising model

$$\begin{aligned}
&\inf_u \ \|u - f\|_1 + \alpha\|Du\|_{2,1} \\
=&\inf_u \sup_q \ \|u - f\|_1 + \alpha\langle q, Du\rangle - \delta_{\|\cdot\|_{2,\infty}\leq 1}(q) \\
=&\sup_q \inf_u \ \|u - f\|_1 + \alpha\langle q, Du\rangle - \delta_{\|\cdot\|_{2,\infty}\leq 1}(q) \\
=&\sup_q \left(-\sup_u \ \langle -\alpha D^* q, u\rangle - \|u - f\|_1\right) - \delta_{\|\cdot\|_{2,\infty}\leq 1}(q) \\
=&\sup_q \ \langle \alpha D^* q, f\rangle - \delta_{\|\cdot\|_\infty\leq 1}(-\alpha D^* q) - \delta_{\|\cdot\|_{2,\infty}\leq 1}(q)
\end{aligned}$$

The problem did not become easier because neither the primal nor the dual have simple proximal operators or satisfy all the smoothness assumptions that descent algorithms require. To overcome this situations, we will work directly with the saddle-point formulation of the problem.

4.4 Primal-Dual Algorithms

Primal-dual techniques solve minimization problems of the form

$$\min_u \ G(u) + F(Ku) \tag{4.17}$$

by solving the equivalent saddle-point problem

$$\min_u \max_p \ G(u) + \langle p, Ku\rangle - F^*(p). \tag{4.18}$$

The algorithm that we will study uses an interpretation of the proximal mapping as implicit gradient descent. This generalization results from the optimality conditions of the proximal update

$$\begin{aligned}
u^{k+1} &= \mathsf{prox}_{\tau E}(u^k) \\
u^{k+1} &= \operatorname*{argmin}_u \tau E(u) + \frac{1}{2}\|u - u^k\|^2 \\
0 &\in \tau\partial E(u^{k+1}) + u^{k+1} - u^k \\
u^{k+1} &\in u^k - \tau\partial E(u^{k+1}).
\end{aligned} \tag{4.19}$$

If E is differentiable, we obtain an update rule

$$u^{k+1} = u^k - \tau\nabla E(u^{k+1})$$

that coincides to an implicit discretization of gradient descent, where the gradient is computed at the next iterate. The **proximal-point algorithm** uses this interpretation to generalize gradient descent to non-differentiable functions with a simple proximal operator.

We will use this algorithm to solve the saddle-point problem (4.18). The energy that we want to minimize over u and maximize over p is called the Lagrangian

$$\mathscr{L}(u,p) := G(u) + \langle p, Ku\rangle - F^*(p). \tag{4.20}$$

If we alternate implicit ascent steps in p with implicit descent steps in u, we obtain the following update rules:

$$\begin{aligned} p^{k+1} &= \mathsf{prox}_{-\sigma\mathscr{L}(u^k,\cdot)}(p^k) \\ u^{k+1} &= \mathsf{prox}_{\tau\mathscr{L}(\cdot,p^{k+1})}(u^k) \end{aligned}$$

We can exploit the structure of the Lagrangian, inherited from the decomposition of the energy as the sum of convex functions $G(u) + F(Ku)$, to formulate the dual update in terms of the conjugate of F as follows:

$$\begin{aligned} p^{k+1} =& \mathsf{prox}_{-\sigma\mathscr{L}(u^k,\cdot)}(p^k), \\ =& \operatorname*{argmin}_p \ \frac{1}{2}\|p-p^k\|^2 + \sigma F^*(p) - \sigma\langle Ku^k, p\rangle \\ =& \operatorname*{argmin}_p \ \frac{1}{2}\|p-p^k-\sigma Ku^k\|^2 + \sigma F^*(p) \\ =& \mathsf{prox}_{\sigma F^*}(p^k + \sigma Ku^k). \end{aligned}$$

Similarly, we can formulate the primal update in terms of the function G as follows:

$$\begin{aligned} u^{k+1} =& \mathsf{prox}_{\tau\mathscr{L}(\cdot,p^{k+1})}(u^k), \\ =& \operatorname*{argmin}_u \ \frac{1}{2}\|u-u^k\|^2 + \tau G(u) + \tau\langle Ku, p^{k+1}\rangle \\ =& \operatorname*{argmin}_u \ \frac{1}{2}\|u-u^k+\tau K^*p^{k+1}\|^2 + \tau G(u) \\ =& \mathsf{prox}_{\tau G}(u^k - \tau K^*p^{k+1}). \end{aligned}$$

This results in a very temping update rule

$$\begin{aligned} p^{k+1} &= \mathsf{prox}_{\sigma F^*}(p^k + \sigma Ku^k), \\ u^{k+1} &= \mathsf{prox}_{\tau G}(u^k - \tau K^*p^{k+1}) \end{aligned}$$

for a primal-dual algorithm. However, decoupling the updates of the primal and dual variables violates the optimality conditions of the saddle-point problem. To overcome this limitation, we need to substitute u^k in the dual update by an extrapolated variable $\bar{u}$. The result is the primal-dual hybrid gradient method.

Definition We will call the iteration

$$\begin{aligned} p^{k+1} &= \mathsf{prox}_{\sigma F^*}(p^k + \sigma K\bar{u}^k), \\ u^{k+1} &= \mathsf{prox}_{\tau G}(u^k - \tau K^*p^{k+1}), \\ \bar{u}^{k+1} &= 2u^{k+1} - u^k. \end{aligned} \tag{PDHG}$$

the **Primal-Dual Hybrid Gradient (PDHG) Method**.

As we will see, the Primal-Dual Hybrid Gradient method converges if $\tau\sigma < \frac{1}{\|K\|^2}$. It is actually easy to show that it converges to a minimizer of the original energy because a fixed-point $(u^\star, p^\star, \bar{u}^\star)$ of PDHG satisfies:

$$\begin{aligned}
\bar{u}^\star &= u^\star, \\
p^\star &= \ \mathsf{prox}_{\sigma F^*}(p^\star + \sigma K u^\star) \\
&= \operatorname*{argmin}_u \sigma F^*(p) + \frac{1}{2}\|p - p^\star - \sigma K u^\star\|^2 \iff 0 \in \sigma \partial F^*(p^\star) - \sigma K u^\star, \\
u^\star &= \ \mathsf{prox}_{\tau G}(u^\star - \tau K^* p^\star) \\
&= \operatorname*{argmin}_u \tau G(u) + \frac{1}{2}\|u - u^\star + \tau K^* p^\star\|^2 \iff 0 \in \tau \partial G(u^\star) + \tau K^* p^\star.
\end{aligned} \tag{4.21}$$

Combining these expressions we obtain the optimality conditions of Corollary 64

$$-K^* p^\star \in \partial G(u^\star) \qquad\qquad K\bar{u}^\star = \partial F^*(p^\star). \tag{4.22}$$

To analyze the convergence of the algorithm, it is useful to modify the order of the primal and dual updates to

$$\begin{aligned}
u^{k+1} &= \mathsf{prox}_{\tau G}(u^k - \tau K^* p^k) \\
\bar{u}^{k+1} &= 2u^{k+1} - u^k \\
p^{k+1} &= \mathsf{prox}_{\sigma F^*}(p^k + \sigma K \bar{u}^{k+1}).
\end{aligned} \tag{4.23}$$

and expand the proximal operators in terms of subgradients. For the primal variable, we have

$$\begin{aligned}
u^{k+1} = \mathsf{prox}_{\tau G}(u^k - \tau K^* p^k) &= \operatorname{argmin}\ \tau G(u) + \frac{1}{2}\|u - u^k + \tau K^* p^k\|^2 \\
0 &\in \partial G(u^{k+1}) + \frac{1}{\tau}(u^{k+1} - u^k) + K^* p^k \\
0 &\in \partial G(u^{k+1}) - K^* p^{k+1} + \frac{1}{\tau}(u^{k+1} - u^k) + K^*(p^{k+1} - p^k).
\end{aligned} \tag{4.24}$$

Similarly, for the dual variable

$$\begin{aligned}
p^{k+1} = \mathsf{prox}_{\sigma F^*}(p^k + \sigma K(2u^{k+1} - u^k)\ ) &= \operatorname{argmin}\ \sigma F^*(p) + \frac{1}{2}\|p - p^k - \sigma K(2u^{k+1} - u^k)\|^2 \\
0 &\in \partial F^*(p^{k+1}) + \frac{1}{\sigma}(p^{k+1} - p^k) - K(2u^{k+1} - u^k) \\
0 &\in \partial F^*(p^{k+1}) - K u^{k+1} + \frac{1}{\sigma}(p^{k+1} - p^k) - K(u^{k+1} - u^k).
\end{aligned} \tag{4.25}$$

If we combine both expressions into a single update rule and define a set-valued operator T and matrix M as follows

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \underbrace{\begin{pmatrix} \partial G & K^T \\ -K & \partial F^* \end{pmatrix}}_{T} \begin{pmatrix} u^{k+1} \\ p^{k+1} \end{pmatrix} + \underbrace{\begin{pmatrix} \frac{1}{\tau} I & -K^T \\ -K & \frac{1}{\sigma} I \end{pmatrix}}_{M} \begin{pmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{pmatrix}, \tag{4.26}$$

we can write the update rule of PDHG as the customized proximal-point algorithm

$$\begin{aligned}
z^{k+1} &= (M + T)^{-1}(M z^k) \\
0 &\in T(z^{k+1}) + M(z^{k+1} - z^k).
\end{aligned} \tag{4.27}$$

The name originates from its similarity with the proximal point algorithm

$$\begin{aligned}
u^{k+1} &= prox_E(u^k) = (I + \tau \partial E)^{-1}(u^k) \\
0 &\in \partial E(u^{k+1}) + \frac{1}{\tau}(u^{k+1} - u^k).
\end{aligned} \tag{4.28}$$

We will build on this algorithm to analyze the convergence of PDHG. In particular, as we showed convergence of the proximal point algorithm by showing that $prox_E = (I+\tau\partial E)^{-1}$ is firmly nonexpansive, thus averaged, and the Picard iteration converges; we will show convergence of PDHG by showing that T is maximal monotone, thus $(I+T)^{-1}$ is averaged, and the Picard iteration converges. Before we can do this, we need to define set-valued and monotone operators.

4.5 Monotone Operators

The mapping of a set-valued operator defines a relation. In monotone operator theory, a relation, a point-to-set mapping, a multi-valued function, and a correspondence all refer to the same concept.

Definition A relation or set-valued operator T on $\mathbb{R}^n$ is a subset of $\mathbb{R}^n \times \mathbb{R}^n$.

For simplicity we write $T(x)$ and Tx to mean the set $\{y \ : (x,y) \in T\}$. This is an abuse of notation that is useful when T is a function and $T(x) = y$ is the singleton set $T(x) = \{y\}$.

For instance, the subdifferential of a function $E\colon \mathbb{R}^n \to \bar{\mathbb{R}}$ defines the subdifferential relation

$$\partial E = \{(u,g) \in \mathrm{dom}(E) \times \mathbb{R}^n \ : \ E(v) \geq E(u) + \langle g, v-u\rangle \ \ \forall v \in \mathbb{R}^n\}.$$

The set $\partial E(u)$ is a well-defined closed convex set at any point $u \in \mathrm{dom}(E)$, but it can be empty. We have already seen that when E is convex, $\partial E(u) \neq \emptyset$ for any $u \in \mathrm{ri}(\mathrm{dom}(E))$.

We can extend many notions for functions to relations. For example, if T and S are relations, we define the domain $\mathrm{dom}(T)$, composition TS and the sum $T+S$ as

$$\begin{aligned}
\mathrm{dom}(T) &= \{x \in \mathbb{R}^n \ : \ T(x) \neq \emptyset\} \\
TS &= \{(x,z) \ : \ \exists y \ \text{ s.t. } \ (x,y) \in S, (y,z) \in T\} \\
T+S &= \{(x, y+z) \ : \ (x,y) \in T, \ (x,z) \in S\}.
\end{aligned}$$

More interesting for us are the inverse relation and the zeros of a relation.

Definition Inverse Relation. The inverse relation of T is defined as

$$T^{-1} = \{(x,y) \ : \ (y,x) \in T\}.$$

This always exists, even when T is a function that is not one-to-one. The inverse relation is not quite an inverse in the usual sense, as we can have $T^{-1}T \neq I$ like in the case of the zero function. In fact, we only have $T^{-1}Tx = x$ when T^{-1} is a function and $x \in \mathrm{dom}(T)$.

Definition Zeros of a Relation. When $0 \in T(x)$, we say that x is a zero of T. The zero set of a relation T is

$$T^{-1}(\{0\}) = T^{-1}(0) = \{x \ : \ (x,0) \in T\}.$$

Zeros and inverses are important for us when applied to the subdifferential. The **inverse of subdifferential** $(\partial E)^{-1}$ is defined as

$$\begin{aligned}
(u,\hat{u}) \in (\partial E)^{-1} &\iff (\hat{u}, u) \in \partial E \\
&\iff u \in \partial E(\hat{u}) \\
&\iff 0 \in \partial E(\hat{u}) - u \\
&\iff \hat{u} \in \arg\min_v E(v) + \langle u, v\rangle
\end{aligned}$$

Definition A relation T on $\mathbb{R}^n$ is **monotone** if it satisfies

$$\langle u - v, x - y \rangle \geq 0 \quad \forall (x,u),\ (y,v) \in T$$

In multi-valued function notation, monotonicity reads

$$\langle Tx - Ty, x - y \rangle \geq 0 \quad \forall x,\ y \in \mathrm{dom}(T),$$

where the left-hand side is a subset of $\mathbb{R}$ and the inequality means that it lies in $\mathbb{R}_+$.

Definition The relation T is **maximal monotone** if there is no monotone operator that properly contains it as a subset of $\mathbb{R}^n \times \mathbb{R}^n$.

In other words, if the monotone operator T is not maximal, then there is $(x,u) \notin T$ such that $T \cup \{(x,u)\}$ is monotone. We already know some examples of (maximally) monotone operators.

Lemma 65. *$E\colon \mathbb{R}^n \to \bar{\mathbb{R}}$, then ∂E is a monotone operator. If E is closed convex and proper then ∂E is maximal monotone.*

Proof. To prove monotonicity, we add the inequalities that define the subdifferential

$$E(y) \geq E(x) + \langle \partial E(x), y - x \rangle$$
$$E(x) \geq E(y) + \langle \partial E(y), x - y \rangle$$

to obtain $\langle \partial E(x) - \partial E(y), x - y \rangle \geq 0$. This holds even when E is not convex. When E is convex, we need to show that ∂E is maximal, i.e., for any $(x,p) \notin \partial E$ there exists $(y,q) \in \partial E$ such that

$$\langle x - y, p - q \rangle < 0.$$

because $\partial E \cup (x,p)$ is not monotone. Let $y = \arg\min_z E(z) + \frac{1}{2}\|z - (x+p)\|^2$, then

$$0 \in \partial E(y) + y - x - p$$
$$x - y \in q - p \ \text{ for } \ q \in \partial E(y).$$

As $(x,p) \notin \partial E$, either $x \neq y$ or $p \neq q$ and we have

$$\langle x - y, p - q \rangle = -\|x - y\|^2 = -\|p - q\|^2 < 0.$$

□

A relation on $\mathbb{R}$ is monotone if it is a curve in $\mathbb{R}^2$ that is always nondecreasing (it can have flat and vertical portions). If it is a continuous curve with no end points, then it is maximal monotone.

Lemma 66. *A continuous monotone function $F\colon \mathbb{R}^n \to \mathbb{R}$ with $dom(F) = \mathbb{R}^n$ is maximal.*

Proof. Assume by contradiction that there is a pair $(y,v) \notin F$ such that

$$\langle v - F(x), y - x \rangle \geq 0 \ \ \forall x \in \mathrm{dom}(F) = \mathbb{R}^n.$$

Given any $z \in \mathbb{R}^n$, the previous inequality should hold true for $x = y - t(z - y)$ for all $t > 0$, that is

$$\langle v - F(y - t(z-y)), t(z-y) \rangle \geq 0$$
$$\langle v - F(y - t(z-y)), z - y \rangle \geq 0. \tag{4.29}$$

As F is continuous, when $t \to 0$ we have

$$\langle v - F(y), z - y \rangle \geq 0.$$

As this should hold for an arbitrary z, we have $v = F(y)$ and we obtain the contradiction $(y,v) \in F$. □

Definition The **resolvent** of a relation T on $\mathbb{R}^n$ is

$$R_T = (I + \alpha T)^{-1},$$

where $\alpha \in \mathbb{R}$. The **Cayley operator**, reflection operator, or reflected resolvent of T is defined as

$$C_T = 2R_T - I,$$

where I is the identity function.

Lemma 67. *If T is maximal monotone, then the resolvent $R_T = (I + \alpha T)^{-1}$ with $\alpha > 0$ and the Caley operator $C_T = 2R_T - I$ are nonexpansive functions.*

Proof. We show first that R_T is nonexpansive. Suppose $(x, u) \in R_T$ and $(y, v) \in R_T$. By definition of R_T, we have

$$(x, u) \in R_T \tag{4.30}$$
$$(u, x) \in (I + \alpha T) \tag{4.31}$$
$$x \in u + \alpha T(u). \tag{4.32}$$

Similarly $y \in v + \alpha T(v)$ and we can combine both into

$$x - y \in u - v + \alpha(T(u) - T(v)). \tag{4.33}$$

Multiplying both sides by $u - v$ and using the monotonicity of T, we get

$$\langle u - v, x - y\rangle = \|u - v\|^2 + \alpha\langle u - v, T(u) - T(v)\rangle \tag{4.34}$$
$$\langle u - v, x - y\rangle \geq \|u - v\|^2. \tag{4.35}$$

Now we apply Cauchy-Schwarz inequality and divide by $\|u - v\|$ to get

$$\|x - y\| \geq \|u - v\| = \|R_T(x) - R_T(y)\| \tag{4.36}$$

which shows that R_T is nonexpansive.

Next, we show that $C_T = 2R_T - I$ is also nonexpansive. Using the inequality (4.35), we get

$$\begin{aligned} \|C_T(x) - C_T(y)\|^2 &= \|2(u - v) - (x - y)\|^2 \\ &= 4\|u - v\|^2 - 4\langle x - y, u - v\rangle + \|x - y\|^2 \\ &\leq \|x - y\|^2, \end{aligned} \tag{4.37}$$

which shows that C_T is nonexpansive. □

The following properties describe how we can combine monotone operators without loosing the monotonicity.

Lemma 68. *If F and G are (maximal) monotone and $dom(F) \cap dom(G) \neq \emptyset$, then $F + G$ is (maximal) monotone.*

Theorem 69. Convergence proximal point algorithm *Let T be a maximal monotone operator, and let there exist a z such that $0 \in T(z)$. Then the (generalized) proximal point algorithm*

$$\begin{aligned} z^{k+1} &= (T + I)^{-1}(z^k) \\ 0 &\in T(z^{k+1}) + z^{k+1} - z^k \end{aligned} \tag{4.38}$$

converges to a point $\tilde{z}$ with $0 \in T(\tilde{z})$.

Proof. From Lemma 67 we know that if T is maximal monotone, then the resolvent $R_T = (T+I)^{-1}$ and the Caley operator $C_T = 2R_T - I$ are nonexpansize. Since $R_T = \frac{1}{2}I + \frac{1}{2}C_T$, the resolvent $R_T$is an averaged operator and the generalized proximal point algorithm

$$z^{k+1} = R_T(z^k) \tag{4.39}$$

is a fixed-point iteration of an averaged operator that converges by Krasnoselskii-Mann Theorem. □

To apply this theorem to the PDHG algorithm

$$0 \in T(z^{k+1}) + Mz^{k+1} - Mz^k, \tag{4.40}$$

we need to factor the matrix M. If M is symmetric positive definite, we can decompose it as $M = L^T L$ and define the operator $\tilde{T} = L^{-T}TL^{-1}$ whose fixed-point iteration

$$\begin{aligned}
z^{k+1} &= (\tilde{T}+I)^{-1}(z^k) \\
0 &\in \tilde{T}(z^{k+1}) + z^{k+1} - z^k \\
0 &\in L^{-T}TL^{-1}(z^{k+1}) + z^{k+1} - z^k \\
0 &\in TL^{-1}(z^{k+1}) + L^T(z^{k+1} - z^k) \\
0 &\in T(L^{-1}z^{k+1}) + L^TL(L^{-1}z^{k+1} - L^{-1}z^k) \\
0 &\in T(x^{k+1}) + L^TL(x^{k+1} - x^k) \\
0 &\in T(x^{k+1}) + M(x^{k+1} - x^k),
\end{aligned} \tag{4.41}$$

coincides with the PDHG algorithm update. As a result, we can show convergence of the PDHG algorithm by determining the conditions under which the set-valued operator $\tilde{T}$ is maximal monotone, and the matrix M is symmetric positive definite.

The next lemma shows that if T is maximal monotone, then $\tilde{T}$ is also maximal monotone.

Lemma 70. *If T is (maximal) monotone, then $L^{-T}TL^{-1}$ is (maximal) monotone, too.*

Proof. To show that $L^{-T}TL^{-1}$ is monotone, we need to show the inequality

$$\langle q_u - q_v, u - v\rangle \geq 0$$

holds for all u, v, $q_u \in L^{-T}TL^{-1}(u)$ and $q_v \in L^{-T}TL^{-1}(v)$. That is, for all u, v, $p_u \in TL^{-1}(u)$ and $p_v \in TL^{-1}(v)$

$$\begin{aligned}
\langle L^{-T}p_u - L^{-T}p_v, u - v\rangle &\geq 0 \\
\langle p_u - p_v, L^{-1}u - L^{-1}v\rangle &\geq 0.
\end{aligned}$$

If we now define $\bar{u} = L^{-1}u$, $\bar{v} = L^{-1}v$ we have the following inequality

$$\langle p_u - p_v, \bar{u} - \bar{v}\rangle \geq 0$$

for all $\bar{u}$, $\bar{v}$, $p_u \in T(\bar{u})$ and $p_v \in T(\bar{v})$ which is a result of the monotonicity of T. The maximal monotonicity can be proven easily by contradiction. □

We now need to show that T is maximal monotone.

Lemma 71. *Let $G : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ and $F : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ be proper, closed, convex functions and matrix $K \in \mathbb{R}^{m\times n}$, then the set valued operator*

$$T = \begin{pmatrix} \partial G & K^T \\ -K & \partial F^* \end{pmatrix} = \begin{pmatrix} \partial G & 0 \\ 0 & \partial F^* \end{pmatrix} + \begin{pmatrix} 0 & K^T \\ -K & 0 \end{pmatrix}.$$

is maximal monotone

Proof. Let $x = (u, p) \in T \in \mathbb{R}^n \times \mathbb{R}^m$ and $y = (v, q) \in T \in \mathbb{R}^n \times \mathbb{R}^m$, by bilinearity of the scalar product we have

$$\begin{aligned}
\langle x - y, Tx - Ty\rangle &= \langle \begin{pmatrix} u - v \\ p - q \end{pmatrix}, \begin{pmatrix} \partial G(u) + K^\top p \\ -Ku + \partial F^\star(p) \end{pmatrix} - \begin{pmatrix} \partial G(v) + K^\top q \\ -Kv + \partial F^\star(q) \end{pmatrix} \rangle \\
&= \langle \begin{pmatrix} u - v \\ p - q \end{pmatrix}, \begin{pmatrix} \partial G(u) - \partial G(v) + K^\top (p - q) \\ -K(u - v) + \partial F^\star(p) - \partial F^\star(q) \end{pmatrix} \rangle \\
&= \langle u - v, \partial G(u) - \partial G(v) + K^\top (p - q)\rangle + \langle p - q, -K(u - v) + \partial F^\star(p) - \partial F^\star(q)\rangle \\
&= \langle u - v, \partial G(u) - \partial G(v)\rangle + \langle u - v, K^\top (p - q)\rangle - \langle p - q, K(u - v)\rangle + \langle p - q, \partial F^\star(p) - \partial F^\star(q)\rangle \\
&= \langle u - v, \partial G(u) - \partial G(v)\rangle + \langle p - q, \partial F^\star(p) - \partial F^\star(q)\rangle. \qquad (4.42)
\end{aligned}$$

As F and G are proper, closed, and convex, their subdifferentials are maximal monotone operators (Lemma 67), that is,

$$\begin{aligned}
\langle u - v, \partial G(u) - \partial G(v)\rangle &\geq 0 \\
\langle p - q, \partial F^\star(p) - \partial F^\star(q)\rangle &\geq 0. \qquad (4.43)
\end{aligned}$$

Adding both inequalities we prove that T is monotone because

$$\langle x - y, Tx - Ty\rangle = \langle u - v, \partial G(u) - \partial G(v)\rangle + \langle p - q, \partial F^\star(p) - \partial F^\star(q)\rangle \qquad \geq 0. \qquad (4.44)$$

To show that T is maximal, let us write is as the sum of maximal monotone operators: the subdifferential of convex, proper, closed functions G and F^*, that are maximal monotone by Lemma 67, and an affine a function that is mononote by Lemma 70.

$$T = \begin{pmatrix} \partial G & K^T \\ -K & \partial F^* \end{pmatrix} = \begin{pmatrix} \partial G & 0 \\ 0 & \partial F^* \end{pmatrix} + \begin{pmatrix} 0 & K^T \\ -K & 0 \end{pmatrix}. \qquad (4.45)$$

□

By construction M is already symmetric, if $\tau\sigma < \frac{1}{\|K\|^2}$ it is also positive definite. We can thus summarize the previous result in the following theorem.

Theorem 72. Convergence PDHG *Let F and G be proper, closed, and convex, and assume that the problem*

$$\min_{u \in \mathbb{R}^n} \; G(u) + F(Ku) \qquad (4.46)$$

has a solution and there exists $u \in ri(G)$ such that $Ku \in ri(F)$. Then, the operator

$$T = \begin{pmatrix} \partial G & K^T \\ -K & \partial F^* \end{pmatrix}$$

is maximally monotone. For $\tau\sigma < \frac{1}{\|K\|^2}$ the matrix

$$M = \begin{pmatrix} \frac{1}{\tau} I & -K^T \\ -K & \frac{1}{\sigma} I \end{pmatrix}$$

in the PDHG algorithm is positive definite and PDHG converges.

4.6 Applications of Primal-Dual Algorithms in Vision

Let K being a discretization of the multichannel gradient operator, we will use the saddle-point formulation to solve the following problems:

The classic ROF denoising model minimizes a least-squares data term $G(u) = \frac{1}{2}\|u - f\|^2$ and the TV regularizer $F(Ku) = \alpha\|Ku\|_{2,1}$ as follows:

$$\min_u P(u) = \min_u \frac{1}{2}\|u - f\|^2 + \alpha\|Ku\|_{2,1} = \min_u \max_p \frac{1}{2}\|u - f\|^2 + \langle Ku, p\rangle - \iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p).$$

As the proximal steps of both G and F^* are simple, the PDHG updates

$$\begin{aligned}
p^{k+1} &= \operatorname*{argmin}_p \frac{1}{2}\|p - (p^k + \sigma K\bar{u}^k)\|^2 + \sigma\iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p) \\
u^{k+1} &= \operatorname*{argmin}_u \frac{1}{2}\|u - (u^k - \tau K^* p^{k+1})\|^2 + \frac{\tau}{2}\|u - f\|^2 = \frac{u^k - \tau K^* p^{k+1} + \tau f}{1+\tau} \\
\bar{u}^{k+1} &= 2u^{k+1} - u^k.
\end{aligned}$$

are extremely simple to code. They are also efficient and result in an algorithm that parallelizes well.

The TV-L^1 denoising problem keeps the TV regularizer $F(Ku) = \alpha\|Ku\|_{2,1}$ but changes the norm used for the data term G to

$$\min_u P(u) = \min_u \|u - f\|_1 + \alpha\|Ku\|_{2,1} = \min_u \max_p \frac{1}{2}\|u - f\|_1 + \langle Ku, p\rangle - \iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p).$$

as a result, only the proximal step of the primal variable appearing in G changes. In particular, we have

$$\begin{aligned}
p^{k+1} &= \operatorname*{argmin}_p \frac{1}{2}\|p - (p^k + \sigma K\bar{u}^k)\|^2 + \sigma\iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p) \\
u^{k+1} &= \operatorname*{argmin}_u \frac{1}{2}\|u - (u^k - \tau K^* p^{k+1})\|^2 + \tau\|u - f\|_1 \\
\Rightarrow u_i^{k+1} &= f_i + \operatorname{sign}(v_i^k)\max(|v_i^k| - \tau, 0) \\
\bar{u}^{k+1} &= 2u^{k+1} - u^k.
\end{aligned}$$

The TV deblurring problem is more interesting because it admits multiple formulations as a saddle-point problem. Given A the convolution operator with a blur kernel, the TV deblurring problem reads

$$\min_u P(u) = \min_u \frac{1}{2}\|Au - f\|^2 + \alpha\|Ku\|_{2,1}.$$

To derive the first PDHG formulation of the problem, we substitute the TV regularizer by its biconjugate as we have done before, that is

$$\min_u P(u) = \min_u \max_p \frac{1}{2}\|Au - f\|^2 + \langle Ku, p\rangle - \iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p).$$

and obtain the following PDHG updates

$$\begin{aligned}
p^{k+1} &= \operatorname*{argmin}_p \frac{1}{2}\|p - (p^k + \sigma K\bar{u}^k)\|^2 + \sigma\iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p) \\
u^{k+1} &= \operatorname*{argmin}_u \frac{1}{2}\|u - (u^k - \tau K^* p^{k+1})\|^2 + \frac{\tau}{2}\|Au - f\|^2 \\
&= (I + \tau A^* A)^{-1}(u^k - \tau K^* p^{k+1} + \tau f) \\
\bar{u}^{k+1} &= 2u^{k+1} - u^k.
\end{aligned}$$

In the second case, we introduce an additional dual variable q for the blurred image Au and substitute the data term $\frac{1}{2}\|Au - f\|^2$ by its biconjugate:

$$\begin{aligned}\min_u P(u) &= \min_u \max_{p,q} \langle Au - f, q\rangle - \frac{1}{2}\|q\|^2 + \langle Ku, p\rangle - \iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p) \\ &= \min_u \max_{p,q} \left\langle \begin{pmatrix} A \\ K \end{pmatrix} u, \begin{pmatrix} q \\ p \end{pmatrix} \right\rangle - \langle f, q\rangle - \frac{1}{2}\|q\|^2 - \iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p).\end{aligned}$$

Now we have

$$F^*(p,q) = \langle f, q\rangle + \frac{1}{2}\|q\|^2 + \iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p) \qquad G(u) = 0 \qquad \tilde{K} = \begin{pmatrix} A \\ K \end{pmatrix}$$

and the PDHG updates are simpler because $F^*(p,q)$ decomposes into independent minimizations in p and q but involve an additional variable

$$\begin{aligned} q^{k+1} &= \operatorname*{argmin}_q \frac{1}{2}\|q - (q^k + \sigma A\bar{u}^k)\|^2 + \sigma\langle f, q\rangle + \frac{\sigma}{2}\|q\|^2 \\ p^{k+1} &= \operatorname*{argmin}_p \frac{1}{2}\|p - (p^k + \sigma K\bar{u}^k)\|^2 + \sigma\iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p) \\ u^{k+1} &= u^k - \tau K^* p^{k+1} - \tau A^* q^{k+1} \\ \bar{u}^{k+1} &= 2u^{k+1} - u^k. \end{aligned}$$

This is a common situation: different formulations of the dual or saddle-point problem exist for the same primal problem depending on how the terms in the objective functional are split, and the primal and dual variables are defined. Usually this different formulations result in primal-dual updates with different computational or memory cost. A common case finds two different formulations that trade off memory for computations. The right choice is a matter of the resources and constraints allocated to the problem.

4.7 ADMM: Alternating Direction Method of Multipliers

The optimality conditions of the saddle point formulation pf the problem

$$\begin{aligned} &\min_u \; G(u) + F(Ku) \\ &\min_u \max_p \; G(u) + \langle Ku, p\rangle - F^*(p) \end{aligned} \tag{4.47}$$

are

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \underbrace{\begin{pmatrix} \partial G & K^T \\ -K & \partial F^* \end{pmatrix}}_{T} \underbrace{\begin{pmatrix} \hat{u} \\ \hat{p} \end{pmatrix}}_{\hat{z}} \tag{4.48}$$

$$0 \in T(\hat{z}). \tag{4.49}$$

We have used the operator T and the variable z to write them more compactly as

$$0 \in T(\hat{z}). \tag{4.50}$$

In the previous section, we developed an iterative algorithm

$$0 \in T(z^{k+1}) + M(z^{k+1} - z^k) \quad M \succ 0 \tag{4.51}$$

to find $\hat{z}$ with $0 \in T(\hat{z})$ as a fixed point of the iteration. That is,

$$\hat{z} \text{ is a fixed point of } 0 \in T(z^{k+1}) + M(z^{k+1} - z^k) \iff 0 \in T(\hat{z})$$

In terms of $(\hat{u}, \hat{p}) = \hat{z}$ we have

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial G & K^T \\ -K & \partial F^* \end{pmatrix} \begin{pmatrix} u^{k+1} \\ p^{K+1} \end{pmatrix} + \overbrace{\begin{pmatrix} M_1 & M_3 \\ M_4 & M_2 \end{pmatrix}}^{=:M} \begin{pmatrix} u^{k+1} - u^k \\ p^{K+1} - p^k \end{pmatrix}$$

If we now expand the update rule of each component

$$\begin{aligned} 0 &\in \partial G(u^{k+1}) + K^T p^{k+1} & &+ M_1(u^{k+1} - u^k) + M_3(p^{k+1} - p^k) \\ 0 &\in -K u^{k+1} + \partial F^*(p^{k+1}) & &+ M_4(u^{k+1} - u^k) + M_2(p^{k+1} - p^k) \end{aligned}$$

we realize that we need to set $M_3 = -K^T$ or $M_4 = K$ to have sequential updates. Indeed, for $M_3 = -K^T$, for instance, the update in u^{k+1}

$$0 \in \partial G(u^{k+1}) + M_1(u^{k+1} - u^k) + K^T p^k$$

does not depend on p^{k+1} and we can first compute u^{k+1} from $u^k,\ p^k$ and then compute p^{k+1} that depends on $u^k,\ u^{k+1},\ p^k$. Similarly, if $M_4 = K$, the update of p^{k+1} does not depend on u^{k+1} and we first update the variable p and then u.

The matrix M is constrained to be symmetric and positive definite to guarantee that the iteration update is an averaged operator and the algorithm converges. For a symmetric M, we need to constraint $M_3 = (M_4)^T$, while there are many options to achieve positive definiteness. PDHG achieves it with $M_1 = \frac{1}{\tau} I$, $M_2 = \frac{1}{\sigma} I$ for $\frac{1}{\sigma\tau} \leq \|K\|^2$. The Alternating Direction Method of Multipliers (ADMM) proposes another one.

ADMM defines the iteration

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial G & K^T \\ -K & \partial F^* \end{pmatrix} \begin{pmatrix} u^{k+1} \\ p^{K+1} \end{pmatrix} + \begin{pmatrix} \frac{1}{\lambda} I & -K^T \\ -K & \lambda K K^T \end{pmatrix} \begin{pmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{pmatrix}. \tag{4.52}$$

The resulting M is only positive semi-definite, but we can use other properties of fixed-point iterations of averaged operators to show convergence of ADMM.

If we decompose this equation component by component, in u we have

$$\begin{aligned} 0 &\in \partial G(u^{k+1}) + \frac{1}{\lambda}(u^{k+1} - u^k) + K^T p^k \\ 0 &\in \lambda \partial G(u^{k+1}) + u^{k+1} - (u^k - \lambda K^T p^k) \\ u^{k+1} &= \operatorname*{argmin}_u \ \lambda G(u) + \frac{1}{2}\|u - (u^k - \lambda K^T p^k)\|^2 \\ u^{k+1} &= \mathsf{prox}_{\lambda G}(u^k - \lambda K^T p^k), \end{aligned}$$

which requires a proximal step to update the primal variable, like PDHG. In variable p, however, the sub-problem defined by the iteration

$$\begin{aligned} 0 &\in \partial F^*(p^{k+1}) + \lambda K K^T (p^{k+1} - p^k) - K(2u^{k+1} - u^k) \\ p^{k+1} &= \operatorname*{argmin}_p \ F^*(p) + \frac{\lambda}{2} \left\| K^T p - K^T p^k - \frac{1}{\lambda} K(2u^{k+1} - u^k) \right\|^2, \end{aligned}$$

is more difficult than the proximal step of PDHG and we need a special structure of K or F^* to solve it. This step usually involves solving a linear system of equations.

Indeed, we can reformulate our generic problem

$$\min_u G(u) + F(Ku)$$

in terms of additional variables v, d as follows:

$$\min_{u,v,d} G(v) + F(d), \qquad \text{s.t.} \underbrace{\begin{pmatrix} I & -I & 0 \\ K & 0 & -I \end{pmatrix}}_{\hat{K}} \underbrace{\begin{pmatrix} u \\ v \\ d \end{pmatrix}}_{\hat{u}} = 0$$

$$\min_{\hat{u}=(u,v,d)} \underbrace{G(v) + F(d)}_{\hat{G}} + \hat{F}(K\hat{u}) \quad \hat{F}(z) = \begin{cases} 0 & \text{if } z = 0 \\ \infty & \text{otherwise} \end{cases}$$

$$\min_{\hat{u}} \hat{G}(\hat{u}) + \hat{F}(K\hat{u}) \quad \hat{F}(z) = \begin{cases} 0 & \text{if } z = 0 \\ \infty & \text{otherwise} \end{cases}$$

$$\min_{\hat{u}} \hat{G}(\hat{u}) - \hat{F}^*(p) + \langle K\hat{u}, p \rangle$$

where $F^* = \sup_z \langle z, p \rangle - F(z) = \langle 0, p \rangle - F(0) = 0$ and the solution of the ADMM subproblem in p becomes a linear system.

ADMM is often derived from a different perspective. In this perspective, the above ADMM is the classical algorithm applied to the dual formulation of the problem. The primal version is

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial G & K^T \\ -K & \partial F^* \end{pmatrix} \begin{pmatrix} u^{k+1} \\ p^{K+1} \end{pmatrix} + \underbrace{\begin{pmatrix} \lambda K^T K & K^T \\ K & \frac{1}{\lambda} I \end{pmatrix}}_{=:M} \begin{pmatrix} u^{k+1} - u^k \\ p^{K+1} - p^k \end{pmatrix}.$$

and requires G to be sufficiently simple in order to solve the update equations, i.e.

$$p^{k+1} = \mathsf{prox}_{\lambda F^*}(p^k + \lambda K u^k) \tag{4.53}$$

$$u^{k+1} = \arg\min_u \frac{\lambda}{2} \left\| Ku - Ku^k + \frac{1}{\lambda}(2p^{k+1} - p^k) \right\|^2 + G(u) \tag{4.54}$$

Detailed convergence rate of ADMM is still an active field of research. Whether or not ADMM is faster than PDHG and its variants largely depends on how efficient the non-prox step can be computed. In fact, it often even depends on the architecture you are computing on. The general tendency is

- PDHG is better parallelizable → GPU.
- ADMM makes more progress per iteration → CPU.

4.8 Stopping Criteria

To define a meaningful stopping criterion for the customized proximal point algorithm

$$0 \in \begin{bmatrix} \partial G & K^T \\ -K & \partial F^* \end{bmatrix} \begin{bmatrix} u^{k+1} \\ p^{k+1} \end{bmatrix} + \overbrace{\begin{bmatrix} M_1 & -K^T \\ -K & M_2 \end{bmatrix}}^{M \succ 0} \begin{bmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{bmatrix} \tag{4.55}$$

it is necessary to keep in mind that we use this iteration to find a point $(\hat{u}, \hat{p})$ that satisfies the optimality conditions

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial G & K^T \\ -K & \partial F^* \end{pmatrix} \begin{pmatrix} \hat{u} \\ \hat{p} \end{pmatrix}$$

of the saddle point problem

$$\min_u \max_p \; G(u) + \langle Ku, p\rangle - F^*(p).$$

As a result, the natural questions that our stopping criteria has to satisfy are the following:

- At convergence, $-K^T p^{k+1}$ should be an element of $\partial G(u^{k+1})$.
- At convergence, Ku^{k+1} should be an element of $\partial F^*(p^{k+1})$.

From the update (4.55), we see that we can measure the distance between $-K^T p^{k+1}$ and the set $\partial G(u^{k+1})$ as well as the distance between Ku^{k+1} and the set $\partial F^*(p^{k+1})$ by measuring how close (u^{k+1}, p^{k+1}) is to (u^k, p^k) with the distance induced by the symmetric positive definite matrix M. This define the following primal and dual residuals::

$$\begin{aligned} r_p^{k+1} &= M_2(p^{k+1} - p^k) - K(u^{k+1} - u^k) \\ r_d^{k+1} &= M_1(u^{k+1} - u^k) - K^T(p^{k+1} - p^k) \end{aligned}$$

We consider our algorithm to be convergent if $\|r_d^{k+1}\|^2 + \|r_p^{k+1}\|^2 \to 0$, because it implies

$$\operatorname{dist}(-K^T p^{k+1}, \partial G(u^{k+1})) \to 0, \qquad \operatorname{dist}(Ku^{k+1}, \partial F^*(p^{k+1})) \to 0.$$

Determining sensible upper bounds on the residuals to guarantee convergence require some though. The simplest option of iterating until

$$\|r_d^{k+1}\| \le \epsilon \qquad \|r_p^{k+1}\| \le \epsilon \tag{4.56}$$

is unfair if $u^k \in \mathbb{R}^n$ and $p^k \in \mathbb{R}^m$ have very different sizes ($n >> m$). A better option is to scale the residuals to a unit size and use

$$\|r_d^{k+1}\| \le \sqrt{n}\,\epsilon \qquad \|r_p^{k+1}\| \le \sqrt{m}\,\epsilon. \tag{4.57}$$

This option is unfair if the primal and dual variables have very different scales, for instance if u represent pixel values in the RGB spectrum or is a probability in the unit interval. We account for this possibility by combining absolute and relative error criteria into

$$\|r_d^{k+1}\| \le \sqrt{n}\,\epsilon^{abs} + \sigma_d \epsilon^{rel} \tag{4.58}$$

$$\|r_p^{k+1}\| \le \sqrt{m}\,\epsilon^{abs} + \sigma_p \cdot \epsilon^{rel}, \tag{4.59}$$

where the scale factors $\sigma_p,\ \sigma_d$ account for the scaling of the primal and dual residuals.

Recall that the primal residual

$$r_p^{k+1} = M_2(p^{k+1} - p^k) - K(u^{k+1} - u^k)$$

measures how far Ku^{k+1} is away from a particular element in $\partial F^*(p^{k+1})$. As a result, it scales with the magnitude of elements in $\partial F^*(p^{k+1})$.

$$\begin{aligned} & 0 \in \partial F^*(p^{k+1}) - Ku^{k+1} + r_p^{k+1} \\ \Rightarrow & 0 \in \partial F^*(p^{k+1}) - K^T(2u^{k+1} - u^k) + M_2(p^{k+1} - p^k). \\ \Rightarrow & \underbrace{M_2(p^k - p^{k+1}) + K^T(2u^{k+1} - u^k)}_{=: z^{k+1}} \in \partial F^*(p^{k+1}) \end{aligned}$$

and we can use the magnitude of z to define the scaling factor of the primal residual $\sigma_s = \|z^{k+1}\|$.

Similarly, the dual residual

$$r_d^{k+1} = M_1(u^{k+1} - u^k) - K^T(p^{k+1} - p^k)$$

measures how far $-K^T p^{k+1}$ is away from a particular element in $\partial G(u^{k+1})$, and scales with the magnitude of elements in $\partial G(u^{k+1})$.

$$\begin{aligned} & 0 \in \partial G(u^{k+1}) + K^T p^{k+1} + r_d^{k+1}. \\ \Rightarrow\ & 0 \in \partial G(u^{k+1}) + K^T p^k + M_1(u^{k+1} - u^k) \\ \Rightarrow\ & \underbrace{M_1(u^k - u^{k+1}) - K^T p^k}_{=:v^{k+1}} \in \partial G(u^{k+1}) \end{aligned}$$

that we can measure with the magnitude of the variable v. The result in the following stopping criterion:

$$\begin{aligned} \|r_p^{k+1}\| &\leq \sqrt{m}\ \epsilon^{abs} + \|z^{k+1}\| \cdot \epsilon^{rel}, \\ \|r_d^{k+1}\| &\leq \sqrt{n}\ \epsilon^{abs} + \|v^{k+1}\| \cdot \epsilon^{rel}. \end{aligned}$$

In our previous considerations, we have only rewritten the known fact that PDHG and ADMM generate iterates $(u^{k+1}, p^{k+1}, v^{k+1}, z^{k+1})$ with

$$v^{k+1} \in \partial G(u^{k+1}), \qquad z^{k+1} \in \partial F^*(p^{k+1}).$$

and achieve convergence when

$$\| \underbrace{z^{k+1} - Ku^{k+1}}_{=r_p^{k+1}} \| \to 0 \ \text{ and } \| \underbrace{v^{k+1} + K^T p^{k+1}}_{=r_d^{k+1}} \| \to 0!$$

to find elements that satisfy the optimality conditions.

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial G & K^T \\ -K & \partial F^* \end{pmatrix} \begin{pmatrix} \hat{u} \\ \hat{p} \end{pmatrix}.$$

We can also use this concept to accelerate convergence by balancing the primal and dual residuals. The main idea is to keep the residual balanced and increase the sppen of the dual update whenever the dual residual is much larger than the primal one, or the reverse.