EEOR E6616: Convex Optimisation Spring 2024

Convex Optimisation

Prof. Cédric Josz *Scribe: Ekene Ezeunala*

Contents

§1. Lecture 01—19th January, 2024

Logistics. This is a course on convex optimisation, taught by Prof. Cédric Josz. Prerequisites for this course are previous courses in undergraduate linear algebra and real analysis—it is not necessary to have taken a course in optimisation before. All the exams for this course will be held in class and will be done closed book. The course grade will be composed from 50% of the midterm exam and 50% of the final exam—there will be no homeworks or projects.
The textbooks for the course are [BV04], [Bec17], [N+18], and [Roc15].

§1.1. Introduction

In the most fundamental sense, convex optimisation is about finding the points $x^* \in X$ such that $f(x^*) \leqslant f(x)$ for all other $x \in X$, where X is something called a *convex set*, $f : X \to \overline{\mathbb{R}}$ is a *convex function*, and $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ is the extended real line.

> **Example 1.1.** If $X = (0, \infty)$ and $f : X \to \overline{\mathbb{R}}$ is given by $f(x) = 1/x$, then although X is convex, there is no such x^* that minimises f.

This example suggests that we might be looking to do something a bit more general; in full generality, we are interested in finding the greatest lower bound of $\{f(x) : x \in X\}$ which we recall is denoted $\inf\{f(x) : x \in X\}$. If this infimum is attained as in the first sentence above, we write $\min\{f(x) : x \in X\}$.

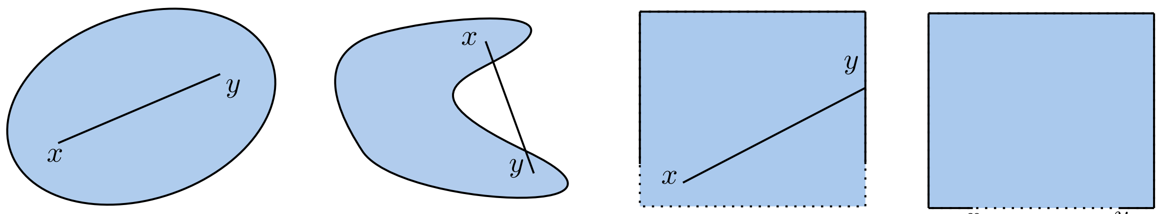
For cases like that of the example above, it will often be more convenuent to consider the function $f : \mathbb{R} \to \overline{\mathbb{R}}$ defined as

$$x \mapsto \begin{cases} 1/x & \text{if } x \in (0, \infty) \\ \infty & \text{otherwise.} \end{cases}$$

This notion will be more useful when we consider more intricate operations over convex functions, et cetera. Let's now define some of the terms we will be frequently referring back to throughout the course.

Definition 1.2 (Convex set)**.** *Let E be a vector space. A subset $C \subseteq E$ is said to be* convex *if for all $x, y \in C$ and for all $t \in [0, 1]$, we have $(1 - t)x + ty \in C$.*

Note that this is a very general idea; we didn't even need to define a metric or a norm or anything of the sort on E or C to establish this definition; we just have a vector space, we have a subset of that vector space, and we have a linear combination of two elements of that subset. The idea of a convex set beckons to some notion of "stability" for a set:



convex set   nonconvex set   convex set   nonconvex set

Fact 1.3 (Union and intersection of convex sets)**.** *The union of convex sets is not necessarily convex, but the intersection of convex sets is convex.*

Proof. For the union, one may simply consider the union of two disjoint convex sets. For the intersection, let $C_1, C_2 \subseteq E$ be convex sets. Then for all $x, y \in C_1 \cap C_2$ and for all $t \in [0,1]$, we have

$$\begin{aligned}(1-t)x + ty \in C_1 \cap C_2 &\iff (1-t)x + ty \in C_1 \text{ and } (1-t)x + ty \in C_2 \\ &\iff x \in C_1 \text{ or } x \in C_2 \text{ and } y \in C_1 \text{ or } y \in C_2 \\ &\iff (1-t)x + ty \in C_1 \text{ and } (1-t)x + ty \in C_2 \\ &\iff (1-t)x + ty \in C_1 \cap C_2.\end{aligned}$$ □

Question: Can we replace the criterion $t \in [0,1]$ with $t \in 1/2$ in the definition of a convex set? The answer is no, and we can see this by considering the set of rationals $\mathbb{Q}$.

Having defined a convex set, we may now talk in greater detail about what a convex function is. We will first define a few terms that will be useful in the definition of a convex function and throughout the course.

Definition 1.4 (Domain, epigraph, strict epigraph)**.** *Given $f : \mathbb{E} \to \overline{\mathbb{R}}$, the* domain *of f is the set*

$$\operatorname{dom} f = \{x \in \mathbb{E} : f(x) < \infty\},$$

and the epigraph *of f is the set*

$$\operatorname{epi} f = \{(x, \alpha) \in \mathbb{E} \times \mathbb{R} : f(x) \leqslant \alpha\}.$$

The strict epigraph *of f is the epigraph without the equality constraint:*

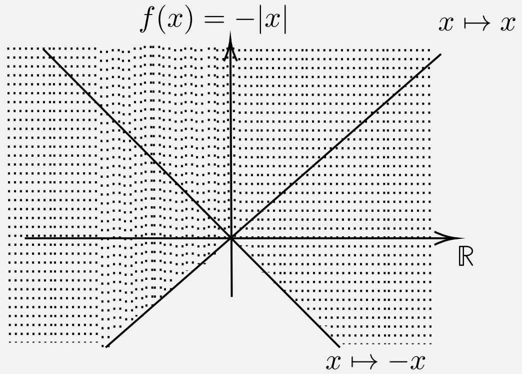
$$\operatorname{epi}_s f = \{(x, \alpha) \in \mathbb{E} \times \mathbb{R} : f(x) < \alpha\}.$$

Given the above notation, we may now properly define a convex function:

Definition 1.5 (Convex function)**.** *A function $f : E \to \overline{\mathbb{R}}$ is said to be* convex *if its epigraph* $\operatorname{epi} f$ *is a convex set.*

Example 1.6.

1. The infimum of two convex functions is not convex. To see this, choose two functions $x \mapsto x$ and $x \mapsto -x$ over $\mathbb{R}$. Then the mapping $x \mapsto \inf\{x, -x\}$ which is the same as the mapping $x \mapsto -|x|$ is not convex.



Clearly the epigraph of $x \mapsto -|x|$ (as shaded) is not convex, so $x \mapsto -|x|$ is not convex.

2. On the other hand, the supremum of convex functions is itself convex, that is, given a family of functions $\{f_i\}_{i\in I}$, the mapping $x \mapsto \sup_{i\in I} f_i(x)$ is convex. To see this, note that

$$\begin{aligned}\operatorname{epi} f &= \{(x,\alpha) \in E\times\mathbb{R} : f(x) \leqslant \alpha\} = \{(x,\alpha)\in E\times\mathbb{R} : \sup_{i\in I} f_i(x) \leqslant \alpha\} \\ &= \{(x,\alpha)\in E\times\mathbb{R} : \forall i\in I, f_i(x)\leqslant\alpha\} = \bigcap_{i\in I}\{(x,\alpha)\in E\times\mathbb{R} : f_i(x)\leqslant\alpha\} \\ &= \bigcap_{i\in I}\operatorname{epi} f_i,\end{aligned}$$

and since the intersection of convex sets is convex, we have that $\operatorname{epi} f$ is convex, so f is convex.

One example we will see for how the epigraph is a great perspective to take on a convex function comes in form of the *marginal function*—which we will see more about—of a convex function, defined in the following way for vector spaces E and F:

$$\begin{cases}\varphi : E\times F \to \overline{\mathbb{R}} \\ f(x) = \inf_{y\in F}\varphi(x,y),\end{cases}$$

for all $x\in E$ and for $f : E\to\overline{\mathbb{R}}$. This function shows up in an abundance of situations, especially in the areas of constrained optimisation and first-order methods, and the epigraph will be very useful for thinking about the function.

Definition 1.7 (Proper convex function)**.** *A convex function $f : E\to\overline{\mathbb{R}}$ is said to be* proper *if $\operatorname{dom} f \neq \emptyset$ (i.e. $\exists x\in E$ such that $f(x) < \infty$ and $f(x) > -\infty$ for all $x\in E$). The set of proper convex functions is denoted* $\operatorname{conv}(E)$.

Let us equip the vector space E with a suitable topology, the usual norm topology $\|\cdot\|_E$ on $\mathbb{R}^n$ for example. Then we may define the following:

Definition 1.8 (Closed function)**.** *A function $f : E\to\overline{\mathbb{R}}$ is said to be* closed *if its epigraph* $\operatorname{epi} f$ *is a closed set. The set of closed proper convex functions is denoted* $\overline{\operatorname{conv}}(E)$.

Let us now see our first result:

Proposition 1.9 (Characterisation of convex functions)**.** *The domain of a convex function is convex. Furthermore, a function $f : E\to\overline{\mathbb{R}}$ is convex if and only if for all $x,y\in\operatorname{dom} f$ and for all $t\in(0,1)$, we have*

$$f\left((1-t)x+ty\right) \leqslant (1-t)f(x)+tf(y). \tag{1}$$

Proof. Assume that $f : E\to\overline{\mathbb{R}}$ is convex. Then for all $x,y\in\operatorname{dom} f$ and for all $t\in(0,1)$, and real numbers $\alpha,\beta\in\mathbb{R}$ such that $f(x)<\alpha$ and $f(y)<\beta$. Note that these α,β exist since $x,y\in\operatorname{dom} f$. Then we have that (x,α) and (y,β) belong to $\operatorname{epi} f$, which is convex by assumption (since f is convex), so we have that $(1-t)(x,\alpha)+t(y,\beta)=((1-t)x+ty,(1-t)\alpha+t\beta)$, which belongs to $\operatorname{epi} f$ since $\operatorname{epi} f$ is convex. Thus we have that

$$f\left((1-t)x+ty\right)\leqslant(1-t)f(x)+tf(y)<\infty.$$

In other words, $(1-t)x+ty\in\operatorname{dom} f$, so $\operatorname{dom} f$ is convex.

For the second statement, since α and β are arbitrary, it follows that with $\alpha=f(x)$ and $\beta=f(y)$, we have that

$$f\left((1-t)x+ty\right)\leqslant(1-t)f(x)+tf(y),$$

for all $x,y\in\operatorname{dom} f$ and for all $t\in(0,1)$, so f is convex. Conversely, assume that (1) holds; we will show that f is convex by showing that its epigraph is convex. Let $(x,\alpha),(y,\beta)\in\operatorname{epi} f$ and let $t\in(0,1)$. Then we have that $x,y\in\operatorname{dom} f$, so $t\in(0,1)$ and

$$\begin{aligned}f\left((1-t)x+ty\right)&\leqslant(1-t)f(x)+tf(y)\\&\leqslant(1-t)\alpha+t\beta,\end{aligned}$$

so $(1-t)x+ty\in\operatorname{epi} f$, and thus $\operatorname{epi} f$ is convex, so f is convex by definition. □

Remark 1.10. *A similar construction from set to function is used for nonconvex functions. For example, we say that a* set $S\subseteq\mathbb{R}^n$ is semialgebraic *if*

$$S=\bigcup_{i=1}^{m}\{x\in\mathbb{R}^n:p_{j_1}(x)=\cdots=p_{j_i}(x)=0,q_{k_1}(x),\ldots,q_{k_i}(x)>0\},$$

for polynomials $p_{j_1},\ldots,p_{j_i},q_{k_1},\ldots,q_{k_i}\in\mathbb{R}[X]$*. Then a* function $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ is semialgebraic *if its epigraph* $\operatorname{epi} f$ *is a semialgebraic subset of* $\mathbb{R}^{n+1}$*. Indeed semialgebraic functions share parallel properties with convex functions—for example, the domain of a semialgebraic function is itself semialgebraic.*

Example 1.11. 1. Consider the parabola f defined as $x\mapsto x^2$ for $x\in\mathbb{R}$. This function is convex and semialgebraic. Its epigraph is

$$\begin{aligned}\operatorname{epi} f&=\{(x,\alpha)\in\mathbb{R}\times\mathbb{R}:x^2\leqslant\alpha\}\\&=\{(x,\alpha)\in\mathbb{R}^2:\alpha-x^2>0\}\cup\{(x,\alpha)\in\mathbb{R}^2:\alpha-x^2=0\},\end{aligned}$$

Therefore the epigraph is a finite union of a strict system of polynomials, and thus is semialgebraic. Consequently, $x\mapsto x^2$ is semialgebraic.

2. Consider the distorted parabola g defined by $x\mapsto\sqrt{|x|}$. This function is not convex, but it is semialgebraic. Its epigraph is

$$\begin{aligned}\operatorname{epi} g&=\{(x,\alpha)\in\mathbb{R}\times\mathbb{R}:\sqrt{|x|}\leqslant\alpha\}\\&=\{(x,\alpha)\in\mathbb{R}^2:x\leqslant\alpha^2,x\geqslant-\alpha^2,\alpha\geqslant0\},\end{aligned}$$

which is semialgebraic.

Semialgebraic functions, unlike convex functions, can have a "kink" in their graph, as we see in the example of g. This is because the epigraph of a semialgebraic function is not necessarily convex. As we will show, however, all convex functions are locally Lipschitz continuous, and thus have no kinks.

§1.2. Why study convex optimisation?

The study of convex optimisation involves a lot of the beautiful theory of optimisation, but beyond this, there are so many applications of convex optimisation that it is hard to list them all. Here are a few:

1. **Linear regression.** In this area we are interested in the problem of finding a line that best fits a set of data points. This is a problem that is very common in statistics and machine learning. The problem can be formulated as follows: given a set of data points $\{(x_i, y_i)\}_{i=1}^n \subseteq \mathbb{R}^2$, we want to find the line $y = ax + b$ that minimises the sum of the squared errors written as $\sum_{i=1}^n (y_i - ax_i - b)^2$. This problem can be formulated as an unconstrained optimisation problem:
$$\inf \left\{ \|Ax - b\|_2^2 : x \in \mathbb{R}^n \right\}.$$
This is a very popular convex problem, and we fully understand its solutions as prescribed by the Moore-Penrose pseudoinverse. We will see this problem again in the course.

2. **Ridge regression.** This is a variant of linear regression that is used to prevent overfitting. The problem is formulated as follows: given a set of data points $\{(x_i, y_i)\}_{i=1}^n \subseteq \mathbb{R}^2$, we want to find the line $y = ax+b$ that minimises the sum of the squared errors written as $\sum_{i=1}^n (y_i - ax_i - b)^2$ plus a regularisation term $\lambda(a^2 + b^2)$, where $\lambda > 0$ is a parameter that controls the amount of regularisation. This problem can be formulated as an unconstrained optimisation problem:
$$\inf \left\{ \|Ax - b\|_2^2 + \lambda\|x\|_2^2 : x \in \mathbb{R}^n \right\}.$$
This is also a very popular convex problem, and we fully understand its solutions as prescribed by the Moore-Penrose pseudoinverse.

3. **Support vector machines.** This is a very popular machine learning algorithm that is used for classification. The problem is formulated as follows: given a set of labelled data points $\{(x_i, y_i)\}_{i=1}^n \subseteq \mathbb{R}^p \times \{-1, +1\}$, we want to find the line $y = ax + b$ that minimises the margin between the line and the data points. This problem can be formulated as a constrained optimisation problem:
$$\inf \left\{ \frac{1}{n} \sum_{i=1}^n \xi_i + \lambda\|\omega\|_2^2 : (\xi, \omega) \in X \right\},$$
where $X \coloneqq \{(\xi, \omega) \in \mathbb{R}^n \times \mathbb{R}^p : y_i^\top (w^\top x_i - b) \geqslant 1 - \xi_i \text{ and } \forall\, i \in [n], \xi_i \geqslant 0\}$.

4. **Control theory.** This is a very popular area of mathematics that is used in engineering and economics. The problem is formulated as follows: given a system of differential equations $\dot{x} = f(x, u)$, where $x \in \mathbb{R}^n$ is the state of the system and $u \in \mathbb{R}^m$ is the control, we want to find the control u that minimises the cost function $\int_0^T \ell(x(t), u(t))\, \mathrm{d}t$, where $\ell : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is the running cost and $T > 0$ is the time horizon. For us we will typically be interested in conic optimisation problems, for example, finding a stabilising controller leads us to the following problem:
$$\inf \left\{ \langle C, X \rangle : X \in A \cap S_+^n \right\},$$

where $A \coloneqq \{X \in S^n : \langle A_i, X\rangle + \langle X, A_i\rangle \leqslant 0, \forall\, i \in [m]\}$ is an affine subspace of S^n and S^n_+ is the cone of positive semidefinite matrices.

5. **Portfolio optimisation.** This is a very popular area of mathematics that is used in finance. The problem is formulated as follows: given a set of assets $\{x_i\}_{i=1}^n \subseteq \mathbb{R}^n$, we want to find the portfolio $x \in \mathbb{R}^n$ that maximises the expected return $\mathbb{E}[x^\top x_i]$ subject to the constraint that the portfolio has a certain risk, which is measured by the variance $\operatorname{Var}[x^\top x_i]$. This problem can be formulated as a constrained optimisation problem:
$$\sup\left\{\sum_{i=1}^n \mu_i x_i : \sum_{i=1}^n x_i = 1, \sum_{i=1}^n \sigma_i^2 x_i \leqslant \sigma^2, x \in \mathbb{R}^n\right\},$$
where μ_i is the expected return of asset i and σ_i^2 is the variance of asset i.

6. **Neural networks.** Another popular example for optimisation where convexity may not be present in the objective but convex optimisation shows up in the analysis are problems of the form
$$\inf\left\{f\left(W_1, \ldots, W_\ell, b_1, \ldots, b_\ell\right) : W_i \in \mathbb{R}^{n_i \times m_i}, b_i \in \mathbb{R}\right\},$$
where W is a matrix of weights, b is a vector containing the bias terms, and f takes the form
$$f\left(W_1, \ldots, W_\ell, b_1, \ldots, b_\ell\right) = \frac{1}{n}\sum_{i=1}^n \left(W_\ell \sigma\left(W_{\ell-1}\sigma\left(\cdots \sigma\left(W_1 x_i + b_1\right)\cdots\right) + b_{\ell-1}\right) + b_\ell - y_i\right),$$
with training data $\{(x_i, y_i)\}_{i=1}^n \subseteq \mathbb{R}^p \times \mathbb{R}^q$ and σ is a nonlinear activation function. This function f, while nonconvex, is still locally Lipschitz as a composition of all of these applications of the activation σ to the affine forms. Thus $\forall\, a \in \mathbb{R}^n$, there are $r, L > 0$ such that for all $x, y \in B(a, r)$, we have that $|f(x) - f(y)| \leqslant L\|x - y\|_2$, and so we can define a subderivative $\partial f(x) \rightrightarrows \mathbb{R}^n$ as the set-valued map that goes from $\mathbb{R}^n$ to the set of nonempty closed subsets of $\mathbb{R}^n$. We have that ∂f is convex for all $x \in \mathbb{R}^n$ and that $\partial f(x) \subseteq \partial f(y) + L\|x - y\|_2$ for all $x, y \in \mathbb{R}^n$. This is a very important result that we will frequently use for the analyses.

7. **Convergence rates for semialgebraic optimisation.** Let $x^* \in \mathbb{R}^n$ be a limit point of a sequence $\{x_k\}_{k=1}^\infty \subseteq \mathbb{R}^n$ and let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a semialgebraic function such that $\partial f = 0$ at x^*. Then we know that there exists a convex function $\psi : [0, \infty) \to \mathbb{R}$ and $r > 0$ such that although $\|\nabla f\| \to 0$ as $x \to x^*$ for $f \in C^1$, we can postcompose f with ψ so as to bound it away from 0:
$$\|\nabla(\psi \circ f)(x)\| \geqslant 1 \quad \text{for all} \quad x \in B(x^*, r) \quad \text{such that} \quad f(x) = f(x^*).$$
A useful application of the above arises when we try to establish linear convergence rates for the analyses of many convex optimisation algorithms.

8. **Compressed sensing.** This is a very popular area of mathematics that is used in signal processing. The problem is formulated as follows: given a set of measurements $y \in \mathbb{R}^m$ and a sensing matrix $A \in \mathbb{R}^{m \times n}$, we want to find the signal $x \in \mathbb{R}^n$ that minimises the ℓ_1-norm $\|x\|_1$ subject to the constraint that $Ax = y$. This problem can be formulated as a constrained optimisation problem:
$$\inf\left\{\|x\|_1 : Ax = y, x \in \mathbb{R}^n\right\}.$$
This has a very close relationship to the restricted isometry property [CT05], which has made its way into low-rank matrix recovery, matrix completion, and many other areas of mathematics.

9. **Principal component analysis (PCA).** This is yet another area where convex optimisation shines, particularly in the area of robust PCA where we are interested in taking some data matrix $M \in \mathbb{R}^{m\times n}$ and decomposing it as $M = L_0 + S_0$ where L_0 is a low-rank matrix and S_0 is a sparse matrix. This problem can be formulated as a constrained optimisation problem:

$$\inf\left\{\|L\|_* + \lambda\|S\|_1 : L + S = M, L \in \mathbb{R}^{m\times n}, S \in \mathbb{R}^{m\times n}\right\},$$

where $\|\cdot\|_*$ is the nuclear norm and $\|\cdot\|_1$ is the entry-wise ℓ_1-norm. This problem is very popular in the area of machine learning.

10. **Polynomial optimisation.** In this area we are interested in the problem of finding the global minimum of a polynomial $f : \mathbb{R}^n \to \mathbb{R}$ over a semialgebraic set $X \subseteq \mathbb{R}^n$. This problem can be formulated as a constrained optimisation problem:

$$\inf\left\{f(x) : x \in X\right\},$$

where $X = \{x \in \mathbb{R}^n : p_1(x) = \cdots = 0 \text{ and } p_{k+1}(x), \ldots, p_m(x) \geqslant 0\}$ is a closed basic semialgebraic set, where $p_1, \ldots, p_m \in \mathbb{R}[X]$ are polynomials in n variables. Although we can't take unions freely for this kind of semialgebraic set, we still have a degree of generality which captures as special cases problems in linear optimisation, quadratic programming, integer optimisation (since the constraint $x_i \in \{0,1\}$ is the same as $x_i^2 - x_i = 0$ over $\mathbb{R}$), and many other areas of mathematics.

One can solve polynomial optimisation problems to global optimality using a sequence of convex optimisation problems—this is a remarkable fact! Consider the minimiser of a function

$$\inf_X f = \inf_{\substack{\mu\in\mathcal{M}(X)\\ \operatorname{supp}(X)\neq\emptyset}} \left\{\int f \,\mathrm{d}\mu : \mu \geqslant 0, \int \mathrm{d}\mu = 1\right\}.$$

The point of view we will be taking here is that we can take this problem which is in finite dimension (over $\mathbb{R}^n$), "lift" it to an infinite-dimensional problem (over $\mathcal{M}(X)$), and discretise the formulation of the problem to get finer and finer approximations. Then for all $x \in X$,

$$\inf_\mu \int f \,\mathrm{d}\delta_{x_*} = f(x_*) \leqslant f(x).$$

This is a very powerful result that we will see in the course.

11. **Optimality conditions for nonconvex and convex problems.** These problems often crucially rely on convex cones, as we will see in the course. An example is the following optimisation problem

$$\inf\{xy - x^2\} \quad \text{subject to} \quad \begin{cases} x^2 + y^2 \leqslant 1 \\ 3xy \geqslant 1. \end{cases}$$

§1.3. Examples of convex sets

Let E, E_1, E_2, and F denote Euclidean spaces, that is, finite-dimensional real-valued vector spaces equipped with an inner product $\langle\cdot,\cdot\rangle$ and the induced norm $\|\cdot\|$. Here is a nonexhaustive list of convex sets we will come back to from time to time:

1. **Hyperplane.** A hyperplane of $(E, \langle\cdot,\cdot\rangle)$ is a set of the form
$$H(\xi,\alpha) \coloneqq \{x \in E : \langle \xi, x\rangle = \alpha\},$$
where $\xi \in E \setminus \{0\}$ and $\alpha \in \mathbb{R}$. This is a convex subset of E as one can readily verify via the definition of convexity.

2. **Halfspace.** A closed halfspace of E is a set of one of the following forms:
$$H^+(\xi,\alpha) \coloneqq \{x \in E : \langle \xi, x\rangle \geqslant \alpha\},$$
$$H^-(\xi,\alpha) \coloneqq \{x \in E : \langle \xi, x\rangle \leqslant \alpha\},$$
where $\xi \in E \setminus \{0\}$ and $\alpha \in \mathbb{R}$. We say that the halfspaces are open if we replace the inequalities with strict inequalities. These are all convex subsets of E as one can readily verify via the definition of convexity.

3. **Orthant.** The nonnegative orthant is the set of points with nonnegative components, *i.e.*,
$$\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x_i \geq 0, i = 1, \ldots, n\} = \{x \in \mathbb{R}^n \mid x \succeq 0\}.$$
For example, the cone of positive semidefinite matrices is a nonnegative orthant. The orthant is a convex subset of E.

4. **Minkowski sum of convex sets.** Given two convex sets $C_1, C_2 \subseteq E$, we define the Minkowski sum of C_1 and C_2 as
$$C_1 + C_2 \coloneqq \{x_1 + x_2 : x_1 \in C_1, x_2 \in C_2\}.$$
This is a convex subset of E. Likewise, the product αC of a convex set $C \subseteq E$ and a positive scalar $\alpha > 0$ is a convex subset of E.

5. **Affine subspace.** An affine subspace of E is a set of the form
$$A(x_0, \xi) \coloneqq \{x \in E : x = x_0 + \xi\},$$
where $x_0 \in E$ and $\xi \in E \setminus \{0\}$. This is a convex subset of E.

6. **Intersection of convex sets.** Given a family of convex sets $\{C_i\}_{i\in I}$, we define the intersection of the family as
$$\bigcap_{i\in I} C_i \coloneqq \{x \in E : \forall\, i \in I, x \in C_i\}.$$
This is a convex subset of E. Like we showed, the union of a family of convex sets is not necessarily convex.

7. **Image of linear map.** Let $A : E \to F$ be a linear map. Then the image $A(C)$ (resp. the preimage $A^{-1}(C)$) of A for a convex set $C \subseteq E$ (resp. $C \subseteq F$) is convex.

8. **Cross product of convex sets.** Given two convex sets $C_1 \subseteq E_1$ and $C_2 \subseteq E_2$, we define the cross product of C_1 and C_2 as
$$C_1 \times C_2 \coloneqq \{(x_1, x_2) \in E_1 \times E_2 : x_1 \in C_1, x_2 \in C_2\}.$$
This is a convex subset of $E_1 \times E_2$ as one can readily verify via the definition of convexity.

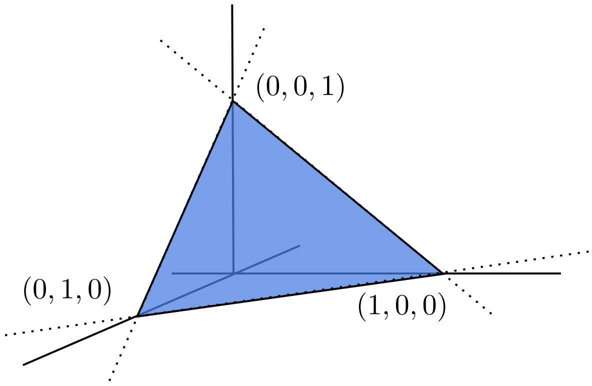
9. **Convex polyhedron.** A convex polyhedron of E is a set of the form $\{x \in E : Ax \leqslant b\}$, where $A : E \to \mathbb{R}^m$ is a linear map, $b \in \mathbb{R}^m$, and $Ax \leqslant b$ means $(Ax)_i \leqslant b_i$ for all $i \in [m]$. This is an intersection of a finite number of closed halfspaces and thus is convex.

§2. Lecture 02—26th January, 2024

Last time we introduced some early notions in convex optimisation. Today we will explore even more, starting with more examples of convex sets.

§2.1. Further examples of convex sets

10. **Simplex.** The simplex $\Delta_n = \{x \in \mathbb{R}^n : x \succeq 0, \mathbb{1}^\top x = 1\}$ is a convex subset of $\mathbb{R}^n$, as it is an intersection of an affine subspace and a positive orthant. The n-dimensional simplex is the convex hull of the standard basis vectors in $\mathbb{R}^n$ (we will see what the convex hull is later on).



11. **Set of positive (semi)definite matrices.** Let S^n be the vector space of symmetric matrices of size $n \times n$, with dimension $\binom{n}{2}$. Then the sets

$$S^n_+ \coloneqq \{A \in S^n : A \succeq 0\},$$
$$S^n_{++} \coloneqq \{A \in S^n : A \succ 0\},$$

that is, the sets S^n_+ and S^n_{++} of positive semidefinite and positive definite matrices respectively, are convex. Indeed, if $A, B \in S^n_+$, then for all $t \in [0,1]$ and for all $\nu \in \mathbb{R}^n$, we have that

$$\nu^\top \left((1-t)A + tB\right)\nu = (1-t)\nu^\top A\nu + t\nu^\top B\nu \geqslant 0,$$

so $(1-t)A + tB \in S^n_+$. The same argument applies for S^n_{++}.

12. **Intersection of affine space and PSD matrix.** If A is an affine subspace of S^n_+, then $S^n_+ \cap A$ is convex, as an intersection of convex sets. As an example, consider the set

$$A = \{X \in S^n : L(X) = b\},$$

where $L : S^n \to \mathbb{R}^m$ is a linear map and $b \in \mathbb{R}^m$. Then $S^n_+ \cap A$ is convex if $X \succeq 0$ implies that $L(X) = b$. As anptother example, consider the set

$$A = \left\{A_0 + \sum_{i=1}^m \alpha_i A_i : \alpha \in \mathbb{R}^m\right\},$$

where $A_0, \ldots, A_m \in S^n$. Then $S^n_+ \cap A \ni X = A_0 + \sum_{i=1}^m \alpha_i A_i \succeq 0$ is convex. These are called linear matrix inequalities (LMIs) and are very important in control theory.

§2.2. Geometric aspects of convex sets

Let E be an Euclidean space. We will now explore some geometric aspects of convex sets.

§2.2.1. Affine hull

An affine set can be understood as a set that is "flat" in the sense that it contains all of its line segments, a translation of a linear subspace. An affine set is something like $\{x\} + L$, where $x \in E$ and L is a linear subspace of E. The affine hull of a set is the smallest affine set that contains the set. Before we define the affine hull more formally, we present the following proposition:

Proposition 2.1. *Let $A \subset A$ and $A \neq \varnothing$. Then the following are equivalent:*

(i) A is an affine subspace.

(ii) For all $x, y \in A$ and for all $t \in \mathbb{R}$, we have that $(1-t)x + ty \in A$.

(iii) The Minkowski difference $A - A$ is a linear subspace of E, and $A + (A - A) = A$.

Proof. We will show that $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (i)$.

- $(i) \Rightarrow (ii)$. Let A be an affine subspace. Then there exists a linear subspace L and a vector x_0 such that $A = x_0 + L$. Let $x, y \in A$ and let $t \in \mathbb{R}$. Then $x = x_0 + u$ and $y = x_0 + v$ for some $u, v \in L$, so
$$(1-t)x + ty = (1-t)(x_0 + u) + t(x_0 + v) = x_0 + (1-t)u + tv \in A,$$
so (ii) holds.
- $(ii) \Rightarrow (iii)$. Let A be a set such that for all $x, y \in A$ and for all $t \in \mathbb{R}$, we have that $(1-t)x + ty \in A$. Then for all $x, y \in A$ and for all $t \in \mathbb{R}$, we have that
$$x - y = (1-t)x + ty - x \in A,$$
so $A - A \subseteq A$. Similarly, we have that $A - A \subseteq A$, so $A - A = A$. Let $x \in A$ and let $y \in A - A$. Then $y = x - z$ for some $z \in A$, so $x = y + z \in A + A$, so $A \subseteq A + A$. Conversely, let $x \in A + A$. Then $x = y + z$ for some $y, z \in A$, so $x - y = z \in A$, so $A + A \subseteq A$, so $A + A = A$. Thus $A - A$ is a linear subspace of E.
- $(iii) \Rightarrow (i)$. Let A be a set such that the Minkowski difference $A - A$ is a linear subspace of E, and $A + (A - A) = A$. Then for all $x, y \in A$ and for all $t \in \mathbb{R}$, we have that $x - y \in A - A$, so $(1-t)x + ty = x - (1-t)x + ty \in A - A$, so $(1-t)x + ty \in A$, and so it follows that A is an affine subspace. □

We now define the affine hull of a set.

Definition 2.2 (Affine hull)**.** *The* affine hull *of a subset $P \subset E$ is the smallest affine set containing P which has the form*
$$\operatorname{aff} P = \bigcap_{\substack{A \text{ is an affine} \\ \text{subset of } E \\ \text{containing } P}} A.$$

Definition 2.3 (Affine combination)**.** *An* affine combination *of P is an element $x \in E$ of the form $x = \sum_{i=1}^{m} t_i x_i$, where $m \in \mathbb{N}^* = \{1, 2, \ldots\}$, $t = (t_1, \ldots, t_m) \in \mathbb{R}^m$ and $x_i \in P$ such that $\mathbb{1}^\top t = 1$.*

The following proposition provides a clearer connection for these objects and how they all relate.

Proposition 2.4. *(i) A set is an affine subspace if and only if it contains all affine combinations of its elements.*

(ii) If $P \subset E$, then $\operatorname{aff} P$ is the set of affine combinations of elements in P, that is,

$$\operatorname{aff} P := \left\{ \sum_{i=1}^{m} t_i x_i : m \in \mathbb{N}^*, t \in \mathbb{R}^m, x_i \in P, \mathbb{1}^\top t = 1 \right\}.$$

Proof. We prove each part separately.

(i) The proof for the reverse direction is immediate given the (ii) $\implies$ (i) part of Proposition 2.1. For the forward direction, let A be an affine subspace; we will prove the result by induction on the number of elements in the affine combination. If $m = 1$, then the result is immediate. Suppose $m > 1$. Define the event $\mathcal{H}_m$ as the event that A contains all affine combinations of m of its elements. Then $\mathcal{H}_2$ is true by the (ii) $\implies$ (iii) $\implies$ (i) part of Proposition 2.1. Suppose that $\mathcal{H}_m$ is true; we will prove the result for $\mathcal{H}_{m+1}$. Consider an affine combination with $m + 1$ elements of A. Then

$$\begin{aligned}
x &= \sum_{i=1}^{m+1} t_i && \text{for } \sum_{i=1}^{m+1} t_i = 1 \text{ and } x_i \in A \\
&= t_1 x_1 + \sum_{i=2}^{m+1} t_i x_i \\
&= \underbrace{t_1 x_1 + (1 - t_1) \underbrace{\sum_{i=2}^{m+1} \frac{t_i}{1 - t_i} x_i}_{\in A \text{ by } \mathcal{H}_m}}_{\in A \text{ by } \mathcal{H}_2} && \text{since } t_i \neq 1.
\end{aligned}$$

Note that the $t_i \neq 1$ condition is satisfied because $\sum_{i=1}^{m+1} t_i = 1$ and therefore without loss of generality, it is impossible for all $t_i = 1$. Thus $\mathcal{H}_m \implies \mathcal{H}_{m+1}$, so by induction, $\mathcal{H}_m$ is true for all $m \in \mathbb{N}^*$.

(ii) Define $X = \left\{\sum_{i=1}^{m} t_i x_i : m \in \mathbb{N}^*, t \in \mathbb{R}^m, x_i \in P, \mathbb{1}^\top t = 1\right\}$. We will show that $\operatorname{aff} P = X$ by showing that $\operatorname{aff} P \subseteq X$ and $X \subseteq \operatorname{aff} P$. First we know that $\operatorname{aff} P \subseteq X$, since $P \subseteq X$ (simply by taking $m = 1$), and X is affine, so that $\operatorname{aff} P \subseteq X$. Next we show that $X \subseteq \operatorname{aff} P$ via the following series of inclusions:

$$\begin{aligned}
X &= \left\{ \sum_{i=1}^{m} t_i x_i : m \in \mathbb{N}^*, t \in \mathbb{R}^m, \mathbb{1}^\top t = 1, x_i \in P, \right\} \\
&\subseteq \left\{ \sum_{i=1}^{m} t_i x_i : m \in \mathbb{N}^*, t \in \mathbb{R}^m, \mathbb{1}^\top t = 1, x_i \in \operatorname{aff} P \right\} && \text{since } P \subseteq \operatorname{aff} P \\
&\subseteq \operatorname{aff} P && \text{as } \operatorname{aff} P \text{ is affine and by Prop. 2.4(i).}
\end{aligned}$$

Thus the results hold. $\square$

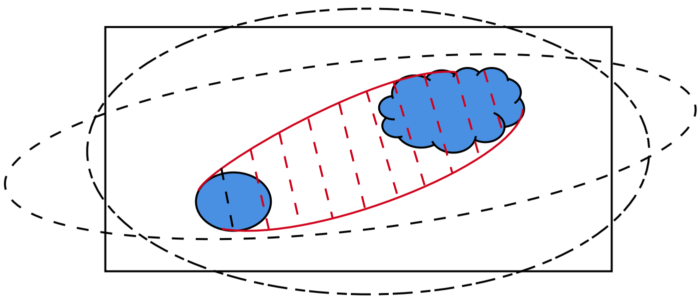
§2.2.2. Convex hull

The convex hull of a set is the smallest convex set containing the set, and is defined as follows:

Definition 2.5 (Convex hull)**.** *The* convex hull *of a subset P of E is the smallest convex set of E containing P, which has the form*

$$\operatorname{co} P = \bigcap_{\substack{A \text{ is an convex} \\ \text{subset of } E \\ \text{containing } P}} A.$$

Definition 2.6 (Convex combination)**.** *A* convex combination *of P is an element $x \in E$ of the form $x = \sum_{i=1}^m t_i x_i$, where $m \in \mathbb{N}^*$, $t = (t_1, \ldots, t_m) \in \Delta_m = \{\alpha \in \mathbb{R}^m : \mathbb{1}^\top \alpha = 1, \alpha \succeq 0\}$ and $x_i \in P$ such that $\mathbb{1}^\top t = 1$ and $t \geqslant 0$. The set of all convex combinations of P is denoted by* $\operatorname{conv} P$.



In the figure above, the blue region is the subset P, and the red region is the convex hull $\operatorname{co} P$.

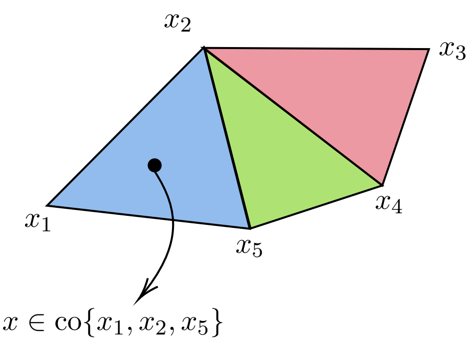
Proposition 2.7. *(i) A set S is convex iff it contains all convex combinations of its elements.*

(ii) If $P \subset E$, then $\operatorname{co} P$ is the set of convex combinations of elements in P, that is, $\operatorname{co} P := \left\{\sum_{i=1}^m t_i x_i : m \in \mathbb{N}^*, t \in \Delta_m, x_i \in P\right\}$.

Proof. Exercise. (See Appendix A for a proof.) $\square$

§2.3. Fundamental property of the convex hull: Carathéodory's theorem

On $\mathbb{R}^2$, consider a point x in the convex hull of x_1, x_2, x_5.



Here $P = \{x_1, \ldots, x_5\}$, and by Proposition 2.7, $x \in \operatorname{co} P$ if and only if x is a convex combination of $x_1, \ldots, x_5$. Note that here $\operatorname{co} P = \{t_1 x_1 + \ldots + t_5 x_5 : t_1 + \ldots + t_5 = 1, t_1, \ldots, t_5 \geqslant 0\}$. Carathéodory's theorem states that if $x \in \operatorname{co} P$, then x is a convex combination of at most $n+1$ points in P, where n is the dimension of the space. In the figure above, x is a convex combination of x_1, x_2, x_5, and so x is a convex combination of at most 3 points in P.

Theorem 2.1 (Carathéodory's theorem, [Car07])**.** *Let P be a subset of a vector space E of dimension n. Then any element of $\operatorname{co} P$ can be written as a convex combination of at most $n+1$ elements of P.*

Proof. We will show a weaker version: that any element of $\operatorname{co} P$ can be written as a convex combination of $n+1$ elements of P. Let $x \in \operatorname{co} P$ be a convex combination of $m > n+1$ elements $x_1, \ldots, x_m \in P$, that is,

$$x = \sum_{i=1}^{m} t_i x_i, \quad t \in \Delta_m, \quad t_1, \ldots, t_m > 0.$$

It suffices to show that x can be written as a convex combination of $m-1$ elements among $x_1, \ldots, x_m$:

$$\underbrace{\begin{pmatrix} x_1 \\ 1 \end{pmatrix}, \begin{pmatrix} x_2 \\ 1 \end{pmatrix}, \ldots, \begin{pmatrix} x_m \\ 1 \end{pmatrix}}_{>n+1} \in \underbrace{E \times \mathbb{R}}_{\dim(E\times\mathbb{R})=n+1}.$$

This the vectors are linearly dependent, *i.e.*, there exist $(\alpha_1, \ldots, \alpha_m) \neq (0, \ldots, 0)$ such that

$$\alpha_1 \begin{pmatrix} x_1 \\ 1 \end{pmatrix} + \ldots + \alpha_m \begin{pmatrix} x_m \\ 1 \end{pmatrix} = 0,$$

that is to say,

$$\begin{cases} \alpha_1 x_1 + \ldots + \alpha_m x_m = 0, \\ \alpha_1 + \ldots + \alpha_m = 0. \end{cases}$$

Since $\alpha \neq \mathbf{0}$, we can assume without loss of generality that there exists an index k such that $\alpha_k \neq 0$. Then we can write

$$x_k = -\sum_{\substack{1 \leqslant i \leqslant m \\ i \neq k}} \frac{\alpha_i}{\alpha_k} \cdot x_i$$

$$x = \sum_{\substack{1 \leqslant i \leqslant m \\ i \neq k}} \left(t_i - t_k \cdot \frac{\alpha_i}{\alpha_k} \right) \cdot x_i.$$

Observe now that

$$\sum_{\substack{1 \leqslant i \leqslant m \\ i \neq k}} \left(t_i - t_k \cdot \frac{\alpha_i}{\alpha_k} \right) = \sum_{\substack{1 \leqslant i \leqslant m \\ i \neq k}} t_i - t_k \sum_{\substack{1 \leqslant i \leqslant m \\ i \neq k}} \frac{\alpha_i}{\alpha_k}$$
$$= 1 - t_k - t_k \cdot \frac{1}{\alpha_k} (-\alpha_k) = 1.$$

All that is left to prove is that $t_i - t_k \cdot \frac{\alpha_i}{\alpha_k} \geqslant 0$ for all $i \neq k$. This condition is equivalent to, recalling that $\alpha_k > 0$ for $t_1, \ldots, t_m$,

$$\frac{\alpha_k}{t_k} \geqslant \frac{\alpha_i}{t_i}, \quad \text{for all } i \neq k.$$

This holds if we choose $k \in \arg\max\{\alpha_i/t_i : 1 \leqslant i \leqslant m\}$, which indeed yields $\alpha_k > 0$. Thus x can be written as a convex combination of $m-1$ elements of P. By induction, we can conclude that any element of $\operatorname{co} P$ can be written as a convex combination of at most $n+1$ elements of P. $\square$

§2.3.1. Relation to linear programming

Recall that a linear program (LP) is a problem of the form: given an input of the following m inequality constraints:

$$K \subseteq \mathbb{R}^n = \begin{cases} a_{11}x_1 + \ldots + a_{1n}x_n \leqslant b_1, \\ \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n \leqslant b_m, \end{cases}$$

where every $a_{ij}, b_j \in \mathbb{Q}$, for $i \in [m]$ and $j \in [n]$. The goal is to determine whether there exists a solution, *i.e.* $K \neq \varnothing$, or to maximise $c^\top x$ for some $c \in \mathbb{Q}^n$. If $K \neq \varnothing$, we output a point in $\mathbb{Q}^n$ in K; if $K = \varnothing$, we output a proof that it is empty.

Intermission: the Fourier-Motzkin elimination. If some nonnegative linear combinations of the m input constraints give us $0 \geqslant 1$, then clearly the LP is infeasible. So we define the following:

Definition 2.8 (LP unsatisfiability)**.** *A proof of unsatisfiability is m multipliers $\lambda_1, \ldots, \lambda_m \geqslant 0$ such that the sum over $i \in [m]$ of λ_i times the ith constraint is equal to $0 \geqslant 1$, i.e.,*

$$\begin{cases} \lambda_1 a_{1i} + \ldots + \lambda_m a_{mi} = n \quad \textit{for all } i \in [n], \\ \lambda_1 b_1 + \ldots + \lambda_m b_m = 1. \end{cases}$$

Suppose we have a set of linear inequalities $\sum_{j=1}^n a_{ij}x_j \geqslant b_i$, $i = 1, \ldots, m$, where $a_{ij}, b_i \in \mathbb{R}$ for $i \in [m]$ and $j \in [n]$. The Fourier-Motzkin elimination method is a method to eliminate variables x_k for some $k \in [n]$ from the system of inequalities, via the following steps:

(1) For each $j \in [m]$, do the following:
- If $a_{jk} > 0$, then multiply the jth inequality by $1/a_{jk}$.
- If $a_{jk} < 0$, then multiply the jth inequality by $-1/a_{jk}$.

(2) Form a new system of inequalities as follows:
- Copy down all the inequalities in which the coefficient of x_k is 0.
- For each inequality in which x_k has positive coefficient and for each inequality in which x_k has negative coefficient, form a new inequality by adding the two inequalities together.

A few remarks. The first step is to ensure that all the nonzero coefficients of x_k are ± 1, and the new system formed in the second step will not contain the variable x_k. Furthermore, if $x_1^*, \ldots, x_n^*$ is a solution to the original system, then $x_1^*, \ldots, x_{k-1}^*, x_{k+1}^*, \ldots, x_n^*$ is a solution to the new system. And if $x_1^*, \ldots, x_{k-1}^*, x_{k+1}^*, \ldots, x_n^*$ is a solution to the new system, then there exists x_k^* such that $x_1^*, \ldots, x_n^*$ is a solution to the original system, as long as the original system is consistent. Hence the original system has a solution if and only if the new solution does. So by applying the Fourier-Motzkin elimination method repeatedly, we obtain a system with at most one variable that has a solution if and only if the original system does. Since solving systems of linear inequalities with at most one variable is easy, we can conclude whether or not the original system has a solution.

Consider $S = \{x \in \mathbb{R}^n : Ax = b, x \succeq 0\} \neq \varnothing$, where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. By the Fourier-Motzkin elimination, S admits a vertex, *i.e.*, a point $v \in S$ that is not of the form

$$(1-t)x + ty, \quad t \in (0,1), \quad x \neq y, \quad x, y \neq S.$$

The following proposition follows from our discussion above:

Proposition 2.9. *Let $P \subset E$. The following statements are equivalent:*

(i) $x \in \operatorname{co} P$.

(ii) *There exists $m \in \mathbb{N}^*$, along with $x_1, \dots, x_m \in P$, and $t \in \Delta_m$ such that $x = \sum_{i=1}^m t_i x_i$.*

(iii) *There exists $m \in \mathbb{N}^*$, along with $x_1, \dots, x_m \in P$, and $t \in \mathbb{R}^m$ with nonnegative t_i such that*
$$x = \begin{pmatrix} x_1 & \cdots & x_m \end{pmatrix} \begin{pmatrix} t_1 \\ \vdots \\ t_m \end{pmatrix} \quad \text{and } \mathbb{1}^\top t = 1.$$

(iv) *There exists $m \in \mathbb{N}^*$, along with $x_1, \dots, x_m \in P$, and $t \in \mathbb{R}^m$ with nonnegative t_i such that*
$$\underbrace{\begin{pmatrix} x_1 & \cdots & x_m \\ 1 & \cdots & 1 \end{pmatrix}}_{A} t = \underbrace{\begin{pmatrix} x \\ 1 \end{pmatrix}}_{b}.$$

Recall that t is a vertex if and only if there exists a basis $\mathscr{B}$ of the range of $\begin{pmatrix} x_1 & \cdots & x_m \\ 1 & \cdots & 1 \end{pmatrix}$ such that $t_i = 0$ if $i \in \mathscr{B}$, where the number of elements in $\mathscr{B} \leqslant \min\{n+1, m\} \leqslant n+1$.

§2.3.2. Consequence of Carathéodory's theorem

The following corollary is a direct consequence of Carathéodory's theorem:

Corollary 2.10. *The convex hull of a compact subset of a finite dimensional complex vector space is compact.*

Proof. We will basically think of $\mathbb{C}^n$ as $\mathbb{R}^{2n}$, and apply Carathéodory's theorem. Let $P \subset \mathbb{C}^n$ be compact. Given positive $A \in \mathbb{C}^n$ with $\|A\| \leqslant 1$, we can write by the spectral theorem,

$$A = \sum_{i=1}^n \alpha_j P_j,$$

where $1 \geqslant \alpha_1 \geqslant \alpha_2 \geqslant \ldots \geqslant \alpha_r \geqslant 0$ (with $r \leqslant n$), and $P_1, \dots, P_n$ are orthogonal projections. We may then write, setting $\alpha_{r+1} = a_0 = 0$,

$$\begin{aligned}A &= \alpha_r \sum_{j=1}^{r} P_j + \sum_{j=1}^{r-1} (\alpha_j - \alpha_r) P_j \\ &= \alpha_r \sum_{j=1}^{r} P_j + (\alpha_{r-1} - \alpha_r) \sum_{j=1}^{r-1} P_j + \sum_{j=1}^{r-2} (\alpha_j - \alpha_{r-1}) P_j \\ &= \sum_{k=0}^{r-1} (\alpha_{r-k} - \alpha_{r-k+1}) \sum_{j=1}^{r-k} P_j.\end{aligned}$$

Since $\sum_{k=0}^{r-1} (\alpha_{r-k} - \alpha_{r-k+1}) = \alpha_r$, we have that

$$A = (1 - \alpha_r) \cdot 0 + \left(\sum_{k=0}^{r-1} (\alpha_{r-k} - \alpha_{r-k+1}) \right) \cdot Q_k,$$

where $Q_k = \sum_{j=1}^{r-k} P_j$ is a projection since it is a sum of pairwise orthogonal projections. Therefore the above expression for A expresses A as a convex combination of projections. By Carathéodory's theorem, we can write A as a convex combination of at most $n+1$ projections. Since the set of projections is compact, the convex hull of the set of projections is compact. Since the set of positive A with $\|A\| \leqslant 1$ is compact, the convex hull of the set of positive A with $\|A\| \leqslant 1$ is compact. Since the set of all $A \in \mathbb{C}^n$ with $\|A\| \leqslant 1$ is compact, the convex hull of the set of all $A \in \mathbb{C}^n$ with $\|A\| \leqslant 1$ is compact. Since the set of all $A \in \mathbb{C}^n$ is compact, the convex hull of the set of all $A \in \mathbb{C}^n$ is compact. □

§2.3.3. An equivalent result to Carathéodory's: Helly's theorem

We next prove a theorem due to Radon and cap today's lecture off with Helly's theorem. First, a helpful lemma is as follows:

Lemma 2.11. *Given a set $V = \{v_1, \ldots, v_m\}$ of vectors in $\mathbb{R}^n$, if $|V| \geqslant n+2$, then the vectors in V are affine dependent.*

Proof. Consider the set $V' = \{v_2 - v_1, \ldots, v_m - v_1\}$. Since $|V'| \geqslant n+1$, those vectors are linearly dependent, and it follows that the vectors in V are affine dependent. □

Theorem 2.2 (Radon's theorem, [Rad21])**.** *Define the set $\Omega \coloneqq \{\omega_1, \ldots, \omega_m\}$, with $|\Omega| \geqslant n+2$, where $\omega_i \in \mathbb{R}^n$. Then there exist two nonempty disjoint subsets $\Omega_1 \subset \Omega$ and $\Omega_2 \subset \Omega$ such that $\Omega_1 \cup \Omega_2 = \Omega$ and $\operatorname{co} \Omega_1 \cap \operatorname{co} \Omega_2 \neq \varnothing$. In other words, there exist $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ such that $\sum_{i \in \Omega_1} \alpha_i = 1$, $\sum_{i \in \Omega_2} \alpha_i = 1$, and $\sum_{i \in \Omega_1} \alpha_i \omega_i = \sum_{i \in \Omega_2} \alpha_i \omega_i$.*

Proof. Since we have that $|\Omega| = m \geqslant n+2$, the vectors in Ω are affine dependent by the result above. Hence there exist reals $\lambda_1, \ldots, \lambda_m$, some of which are positive, so that

$$\sum_{i=1}^{m} \lambda_i \omega_i = 0, \quad \sum_{i=1}^{m} \lambda_i = 0.$$

We may further define the index sets $I_1 \coloneqq \{1, \ldots, m : \lambda_i \geqslant 0\}$ and $I_2 \coloneqq \{1, \ldots, m : \lambda_i < 0\}$. Furthermore, $I_1, I_2 \neq \varnothing$ and $\sum_{i \in I_1} \lambda_i = -\sum_{i \in I_2} \lambda_i$. Define $\lambda \coloneqq \sum_{i \in I_1} \lambda_i$, so that we have that

$$\sum_{i \in I_1} \lambda_i \omega_i = -\sum_{i \in I_2} \lambda_i \omega_i \implies \sum_{i \in I_1} \frac{\lambda_i}{\lambda} \omega_i = -\sum_{i \in I_2} \frac{\lambda_i}{\lambda} \omega_i.$$

Now define sets $\Omega_1 \coloneqq \{\omega_i : i \in I_1\}$ and $\Omega_2 \coloneqq \{\omega_i : i \in I_2\}$. Both of these sets are nonempty and disjoint subsets of Ω with $\Omega_1 \cup \Omega_2 = \Omega$, and therefore $\operatorname{co} \Omega_1 \cap \operatorname{co} \Omega_2 \neq \varnothing$. Consequently we have that

$$\sum_{i \in I_1} \frac{\lambda_i}{\lambda} \omega_i = -\sum_{i \in I_2} \frac{\lambda_i}{\lambda} \omega_i \in \operatorname{co} \Omega_1 \cap \operatorname{co} \Omega_2.$$

The proof is done. □

We can now prove Helly's theorem, which is a generalisation of Radon's theorem:

Theorem 2.3 (Helly's theorem, [Hel23]). *Consider the space $\mathbb{O} \coloneqq \{\Omega_1, \ldots, \Omega_m\}$, a collection of convex sets in $\mathbb{R}^n$, with $\mathbb{O} \geqslant n+1$. If the intersection of any $n+1$ of these sets is nonempty, then all the sets in the collection $\mathbb{O}$ have a nonempty intersection; more formally, $\bigcap_{i=1}^{m} \Omega_m \neq \varnothing$.*

Proof. We shall prove the theorem by induction on $|\mathbb{O}|$. For $m = n+1$, the result of the theorem is trivial. For the induction step, suppose that the theorem holds for $|\mathbb{O}| = m \geqslant n+1$. Consider a collection of convex sets in $\mathbb{R}^n$, $\{\Omega_1, \ldots, \Omega_m, \Omega_{m+1}\}$, with the property that any $n = 1$ of the sets have a nonempty intersection. For $i = 1, \ldots, m+1$, define the following:

$$\Theta_i = \bigcap_{j=1,\ j \neq i}^{m+1} \Omega_j,$$

which has the property that $\Theta_i \subset \Omega_j$ whenever $j \neq i$, and $i, j \in \{1, \ldots, m+1\}$. By hypothesis, $\Theta_i \neq \varnothing$, so there exists an element $\omega_i \in \Theta_i$ for each $i \in [m+1]$. Define the set $W \coloneqq \{\omega_1, \ldots, \omega_{m+1}\}$. Applying Theorem 2.2 to W, we obtain two nonempty disjoint subsets $W_1 = \{\omega_i : i \in I_1\} \subset W$ and $W_2 = \{\omega_i : i \in I_1\} \subset W$ such that $\operatorname{co} W_1 \cap \operatorname{co} W_2 \neq \varnothing$. We may select an element $w \in \operatorname{co} W_1 \cap \operatorname{co} W_2$, and show that $w \in \bigcap_{i=1}^{m+1} \Omega_i \neq \varnothing$. In order to do this, we first note that $i \neq j$ for $i \in I_1$ and $j \in I_2$, which implies that $\theta_i \in \Omega_j$ when $i \in I_1$ and $j \in I_2$. For a particular $i \in \{1, \ldots, m+1\}$ and $i \in I_1$, $\omega_i \in \Theta_i \subset \Omega_j$ for any $j \in I_2$. Because Ω_i is convex, it follows that $\omega \in \operatorname{co} W_2 = \operatorname{co} \{\omega_i : i \in I_2\} \subset \Omega_i$. Thus we have that $\omega \in \Omega_i$ for any $i \in I_1$. By a similar argument, we can show that $\omega \in \Omega_i$ for any $i \in I_2$. Therefore we have that $\omega \in \bigcap_{i=1}^{m+1} \Omega_i \neq \varnothing$. The proof is done. □

§3. Lecture 03—02nd February, 2024

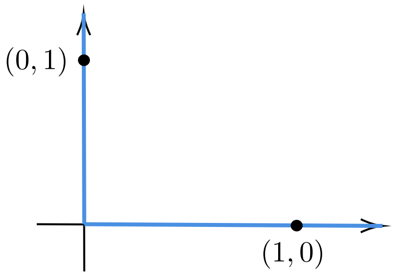
§3.1. Conic hull

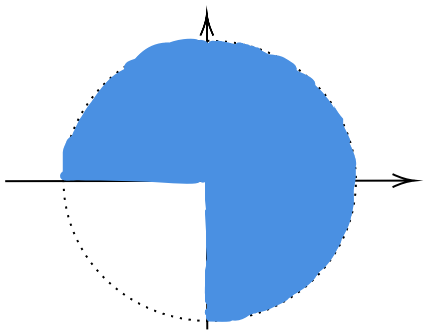
Let E be a vector space. Let $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geqslant 0\}$, and let $\mathbb{R}_{++} = \{x \in \mathbb{R} : x > 0\}$. We present the following definition:

Definition 3.1 (Cone). *A subset K of E is a* cone *if $tx \in K$ for all $x \in K$ and $t \in \mathbb{R}_+$ with $t \geqslant 0$. The set K is a* pointed cone *if $K \cap -K = \{0\}$, that is, $0 \in K$.*

> **Example 3.2.** The set $S^n_{++} = \{X \in \mathbb{R}^{n \times n} : X \text{ is symmetric and positive definite}\}$ is the cone of positive definite symmetric matrices, but it is not a pointed cone.

Note that a cone is not always convex:





Thus it is convenient to define a notion of a hull that is for convex conic sets in particular.

Definition 3.3 (Conic hull). *Let P be a subset of E. The conic hull of P is the smallest covex pointed cone that contains P, which has the form*

$$\operatorname{cone} P := \bigcap_{\substack{K \text{ is a convex} \\ \text{pointed cone of } E \\ \text{containing } P}} K.$$

Just as before, we can define the conic combination of a vector subspace:

Definition 3.4 (Conic combination). *A conic combination of a set P is a vector $x \in E$ of the form $x = \sum_{i=1}^m t_i x_i$ where $x_i \in P$, $m \in \mathbb{N}^*$, and $t_i \geqslant 0$ for all i.*

Proposition 3.5. *1. A set is a convex pointed cone if and only if it contains all conic combinations of its elements.*

2. If $P \subseteq E$, then,

$$\operatorname{cone} P = \left\{ \sum_{i=1}^m t_i x_i : m \in \mathbb{N}^*, t \geqslant 0, x_i \in P \right\}.$$

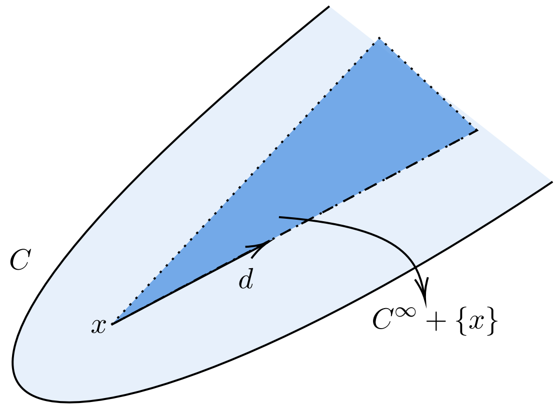
§3.2. Recession cone

The notion of a recession cone is very important, and will become especially useful when we discuss the separation of convex sets. The recession cone of a set is the set of all vectors that can be added to the set without changing its shape. More formally:

Definition 3.6 (Recession cone)**.** *Let C be a nonempty closed convex subset of E. The* recession cone *of C is defined by*

$$C^\infty = \{d \in E : C + \mathbb{R}_+ d \subseteq C\} .$$

One can think of the recession cone C^∞ as being a visualisation of the convex set C when we "zoom out" into an area with sufficiently significant perspective:



Note also—prove this for yourself—that $C^\infty = \{d \in E : C + \{d\} \subseteq C\}$. Now given $x \in C$, consider

$$\begin{aligned} C^\infty(x) &= \{d \in E : x + \mathbb{R}_+ d \subseteq C\} \\ &= \bigcap_{t>0} \frac{C \setminus \{x\}}{t}. \end{aligned}$$

We claim that this object is exactly the same as the previous one, *i.e.* vantage points do not matter:

Proposition 3.7. *Let C be a nonempty closed convex subset of E. Then $\mathbb{C}^\infty$ is a closed convex pointed cone, and*

(i) $C^\infty = \{d \in E :$ there exists $\{x_k\}_{k\geqslant 1} \subset C$ and $t_k \to \infty$ such that $x_k/t_k \to d\}$,

(ii) $C^\infty = C^\infty(x)$ for all $x \in C$.

Proof. Note that for any $t > 0$, the set $(C \setminus \{x\})/t$ is closed and convex, thus so is $C^\infty(x)$, and so is the recession cone $C^\infty = \bigcap_{x\in C} C^\infty(x)$. On the other hand, C^∞ is a pointed cone because if, for some receding direction $d \in C^\infty$ and $t > 0$, then $td \in C^\infty$. Now we prove the properties listed.

Let $x \in C$. To prove (i) and (ii), it suffices to show that $C^\infty(x) = K$, where

$$K := \left\{ d \in E : \text{ there exists } \{x_k\}_{k\geqslant 1} \subset C \text{ and } t_k \to \infty \text{ such that } \frac{x_k}{t_k} \to d \right\}.$$

First we show that $C^\infty(x) \subseteq K$. Let $d \in C^\infty(x)$. Take some sequence of scalars $\{t_k\}_{k\geqslant 1}$ with $t_k \to \infty$, and define $x_k = x + t_k d$. Then we have $x_k \in C$, and in particular, $x_k/t_k \to d$. Thus $d \in K$. Now we show that $K \subseteq C^\infty(x)$. Let the sequence $\{x_k\}_{k\geqslant 1} \subseteq C$ and $\{t_k\}_{k\geqslant 1} \to \infty$ such that $x_k/t_k \to d$. Let $t \geqslant 0$. As soon as $t_k \geqslant t$, $t/t_k \in [0,1]$, and we have the convex combination

$$x + \frac{t}{t_k}(x_k - x) \in C \xrightarrow[k\to\infty]{} x + td \in C,$$

since the set is closed (every limit point resides within the set itself). Thus $d \in C^\infty(x)$. We have now shown that $C^\infty(x) = K$ for every x, and so $C^\infty = \bigcup_{x\in C} C^\infty(x) = K$ as well. □

Another helpful property of the recession cone is that it provides us us with a nice characterisation of boundedness for convex sets:

Corollary 3.8. *Let C be closed convex nonempty. Then C is bounded if and only if $C^\infty = \{0\}$.*

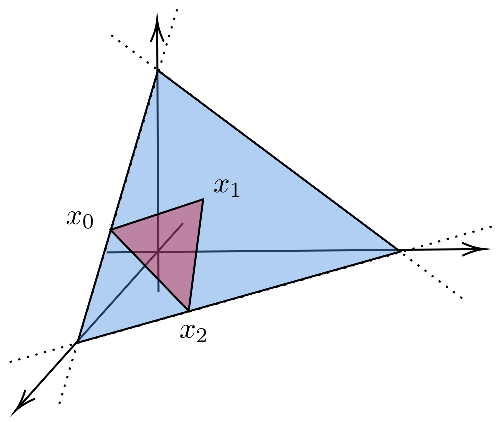
Proof. By Proposition 3.7(i), if C is bounded, then $C^\infty = \{0\}$, because the ratio x_k/t_k converging to 0 is the only feasible possibility. Conversely, if C is not bounded, then there exists some sequence $\{x_k\}_{k\geqslant 1} \subseteq C$ such that $t_k := \|x_k\| \to \infty$ and

$$\frac{x_k}{t_k} = \frac{x_k}{\|x_k\|} \to d \neq 0.$$

Thus $d \in C^\infty$, and we are done. □

§3.3. Topological aspects of convex sets: the relative interior

Consider the simplex Δ_3 residing in $\mathbb{R}^3$:



Although there is no way for this flat set cannot contain a ball for any point residing on it, it still makes sense for us to consider the set of points that lie "inside" the set, that is, within its "interior" relative to some other entity. This is the idea behind the relative interior of a set.

Definition 3.9 (Relative interior)**.** *The* relative interior *of a subset P of a vector E is*

$$\operatorname{ri} P := \{x \in P : \text{ there exists } r > 0 \text{ such that } B(x,r) \cap \operatorname{aff} P \subseteq P\}.$$

We say that C is *relatively open* if $\operatorname{ri} C = C$, and we define the relative boundary of C as the set difference $\overline{C} \setminus (\operatorname{ri} C)$. The relative interior of a set is always open, and for an n-dimensional convex set, $\operatorname{aff} C = \mathbb{R}^n$ by definition, so that $\operatorname{ri} C$ is exactly the interior of C in the usual sense.

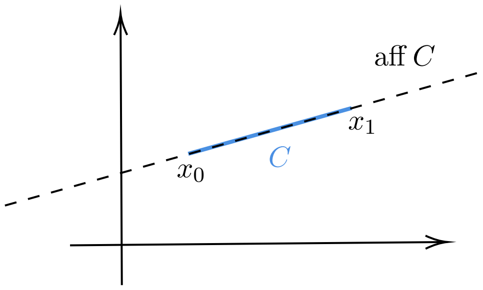
Example 3.10.
- An affine subspace (that does not equal the Euclidean space E) is not open but it is relatively open.
- The relative interior of the sphere is empty.

Now if we have convex sets C_1 and C_2 with $C_1 \subseteq C_2$, it is not necessarily true that $\operatorname{ri} C_1 \subseteq \operatorname{ri} C_2$! Consider for a counterexample the sets $C_1 = \{0\} \subseteq C_2 = [0, \infty)$. Then $\operatorname{ri} C_1 = \{0\}$, but $\operatorname{ri} C_2 = (0, \infty)$, so that $\operatorname{ri} C_1 \nsubseteq \operatorname{ri} C_2$. However, it should be obvious that if $P_1 \subseteq P_2$ and $\operatorname{aff} P_1 = \operatorname{aff} P_2$, then $\operatorname{ri} P_1 \subseteq \operatorname{ri} P_2$.

We now make a more explicit connection between convexity and the relative interior.

Proposition 3.11. *Let C be a nonempty convex set. Then its relative interior* $\operatorname{ri} C$ *is nonempty, and* $\operatorname{aff}(\operatorname{ri} C) = \operatorname{aff} C$.

Proof. Let m be the dimension $\dim(\operatorname{aff} C)$ of affine hull of C. Thus there exists $m+1$ points, say $x_0, \ldots, x_m \in C$ that are affinely independent, *i.e.* the embedding vectors $\begin{pmatrix} x_0 \\ 1 \end{pmatrix}, \ldots, \begin{pmatrix} x_m \\ 1 \end{pmatrix}$ are linearly independent. For example, the following diagram elucidates the case when $m = 1$:



If $m = 0$, then C is a singleton $\{x\}$, in which case the result is obvious, since $B(x, r) \cap \operatorname{aff} C = B(x, r) \cap \{x\} = \{x\} \subseteq C$. Now suppose that $m \geqslant 1$. It is easy to show that for any $x \in \operatorname{aff} C$, there exists a unique set of scalars $\alpha = (\alpha_1, \ldots, \alpha_m) \in \mathbb{R}^m$ such that x can be written in the form

$$x = x_0 + \sum_{i=1}^{m} \alpha_i (x_i - x_0).$$

The mapping we consider is the mapping

$$\begin{aligned} \varphi : \mathbb{R}^m &\to \operatorname{aff} C \\ \alpha &\mapsto x_0 + \sum_{i=1}^{m} \alpha_i (x_i - x_0) = \left(1 - \sum_{i=1}^{m} \alpha_i\right) x_0 + \sum_{i=1}^{m} \alpha_i x_i, \end{aligned}$$

which indeed defines a homeomorphism (*i.e.* a continuous bijection with continuous inverse) between $\mathbb{R}^m$ and $\operatorname{aff} C$.

Consider now the set of coefficients corresponding to strict convex combinations of the points $x_0, \ldots, x_m$, that is,

$$\Omega := \left\{ \alpha \in \mathbb{R}^m : 1 - \sum_{i=1}^{m} \alpha_i > 0, \alpha_i > 0, i = 1, \ldots, m \right\}.$$

Notice that Ω is an open set in $\mathbb{R}^m$, and thus $\varphi(\Omega)$ is an open set in $\operatorname{aff} C$, which concretely means that for all points x in the image $\varphi(\Omega)$, we may find some scalar radius $r > 0$ such that $B(x, r) \cap \operatorname{aff} C \subseteq \varphi(\Omega)$.[1] Hence $\operatorname{aff} \varphi(\Omega) = \operatorname{aff} C$, because the affine hull of a unit ball in a linear vector space is the entire space itself, and $\varphi(\Omega)$ is relatively open, *i.e.* $\operatorname{ri} \varphi(\Omega) = \varphi(\Omega)$. Since $\varphi(\Omega) \subseteq C$, we get that

$$\begin{aligned} \varnothing \neq \varphi(\Omega) &= \operatorname{ri} \varphi(\Omega) \subseteq \operatorname{ri} C \subseteq C \\ \operatorname{aff} \varphi(\Omega) &= \operatorname{aff}(\operatorname{ri} \varphi(\Omega)) \subseteq \operatorname{aff}(\operatorname{ri} C) \subseteq \operatorname{aff} C, \end{aligned}$$

and so $\operatorname{aff} \operatorname{ri} C = \operatorname{aff} C$. □

We will now present a lemma. Its first part is fairly intuitive: given a point in the relative interior and one on the boundary, then the points in between should be in the relative interior. The second is a little bit less intuitive but can be understood better from a geometric standpoint: for some point x in the relative interior, we can find another point $x_0 \in C$ such that when we push x a bit in the direction of x_0, we remain inside C.

Lemma 3.12. *Let C be convex nonempty. Then if $x \in \operatorname{ri} C$ and $y \in \overline{C}$, then the half-open segment $[x, y)$ is contained in $\operatorname{ri} C$. Moreover, given some $x \in E$, we have that $x \in \operatorname{ri} C$ if and only if for all $x_0 \in C$, there exists $t > 1$ such that $(1-t)x_0 + tx \in C$.*

Proof. Consider the first part for distinct x and y (because otherwise $[x, y) = \varnothing \subseteq \operatorname{ri} C$). Let $t \in [0, 1)$, and let B be the unit ball. We'll show that $z_t := (1-t)x + ty \in \operatorname{ri} C$, that is, by definition, $(z_t + \varepsilon B) \cap \operatorname{aff} C \subseteq C$ for $\varepsilon > 0$ small enough. Note that since $y \in \overline{C}$ for all $\varepsilon > 0$, we have that $y \in C + \varepsilon B$. Then, consider

$$\begin{aligned} (z_t + \varepsilon B) \cap \operatorname{aff} C &= ((1-t)x + ty + \varepsilon B) \cap \operatorname{aff} C \\ &\subseteq ((1-t)x + t(C + \varepsilon B) + \varepsilon B) \cap \operatorname{aff} C \\ &= \left(tC + (1-t) \cdot \left(x + \frac{1+t}{1-t} \varepsilon B \right) \right) \cap \operatorname{aff} C \\ &= t \underbrace{C \cap \operatorname{aff} C}_{= C \text{ by def.}} + \underbrace{(1-t) \left(x + \frac{1+t}{1-t} \varepsilon B \right) \cap \operatorname{aff} C}_{\subseteq C \text{ since } x \in \operatorname{ri} C} \\ &\subseteq tC + (1-t)C = C, \end{aligned}$$

since C is convex. Thus $z_t \in \operatorname{ri} C$, and so $[x, y) \subseteq \operatorname{ri} C$.

Now to the second part. For the forward direction, let $x \in \operatorname{ri} C$. Then there exists $\varepsilon > 0$ such that $B(x, \varepsilon) \cap \operatorname{aff} C \subseteq C$. Then for all $y \in \operatorname{aff} C$, there exists $t > 1$ but close to 1 such that

[1] This is entirely natural because in some sense the relative interior of C is the description of an open set in the affine hull of C that is "contained" in C.

$(1-t)y + tx \in B(x, \varepsilon)$, which is itself an affine combination since $y \in \operatorname{aff} C$ and $x \in C$. Thus $(1-t)y + tx \in B(x, \varepsilon) \cap \operatorname{aff} C \subseteq C$. For the reverse direction, since $C \neq \varnothing$, it follows that $\operatorname{ri} C \neq \varnothing$ by Proposition 3.11. Take $y \in \operatorname{ri} C$. Then either $y = x$, in which case $x \in \operatorname{ri} C$, or $y \neq x$, in which case it follows that by assumption, there exists $t > 1$ such that $z = (1-t)y + tx \in C$. Then $x \in [y, z) \subseteq \operatorname{ri} C$, since we know that $y \in \operatorname{ri} C$ and $z \in C$. We can then conclude that $x \in \operatorname{ri} C$. $\square$

A quick application of the above is the following (left as exercises): $\overline{\operatorname{ri} C} = \overline{C}$, and $\operatorname{ri} \overline{C} = \operatorname{ri} C$ for nonempty $C \subseteq E$.

§4. Lecture 04—09th February, 2024

§4.1. Operations on convex sets: the projection

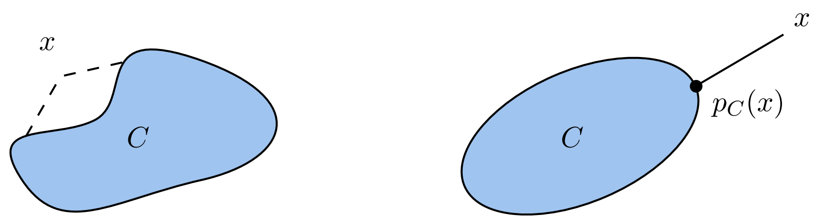
This should be a familiar notion or us. The projection of $x \in E$ on $C \subseteq E$ is related to the following optimisation problem:

$$\inf_{y \in C} \|x - y\|.$$

Speaking more precisely, the projection $p_C(x)$ of x on C is the point in C that is closest to x:

Definition 4.1 (Projection). *Let C be a nonempty closed convex subset of E. The* projection *of $x \in E$ on C is the point*

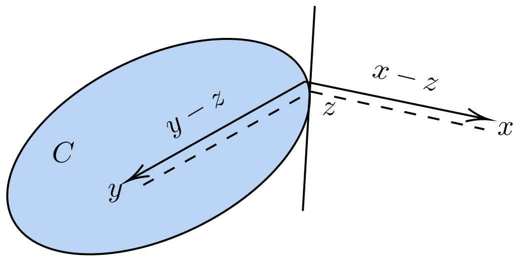
$$p_C(x) := \arg\min_{y \in C} \|x - y\|.$$



Note that althoygh projections onto convex sets are not necessarily linear, there are still some nice properties that they satisfy. We will discuss these now. Let H denote a Hilbert space over $\mathbb{R}$ and $C \subseteq H$ be a nonempty closed convex set.

Proposition 4.2 (Uniqueness of the projection). *There exists a unique $z \in C$ such that $\|x - z\| = \inf_{y \in C} \|x - y\|$, and we call z the projection of x onto C.*

Proposition 4.3 (Characterisation of the projection). *z is a projection $p_C(x)$ if and only if for all $y \in C$, we have $z \in C$ and $\langle x - z, z - y \rangle \leqslant 0$.*

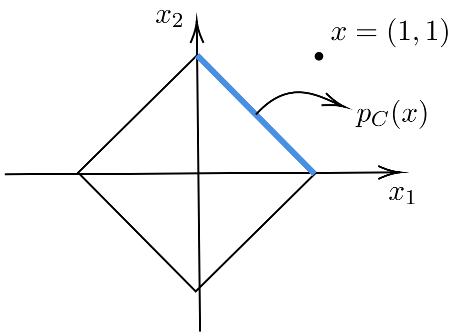


Intuitively, the inner product is never positive because $x-z$ and $y-z$ are in opposite directions relative to the hyperplane.

Proposition 4.4. *The projection is* 1*-Lipschitz continuous, i.e. for all* $x_1, x_2 \in H$*, we have*

$$\|p_C(x_1) - p_C(x_2)\| \leqslant \|x_1 - x_2\|.$$

Note that in Proposition 4.2, we cannot take any arbitrary norm. Consider the set $C \coloneqq \{x \in \mathbb{R}^2 : \|x\|_1 \leqslant 1\}$:



We now define some spaces we will often work with.

Definition 4.5 (Hilbert and Banach spaces)**.** *A space H is a* Hilbert space *if it is a vector space equipped with an inner product $\langle \cdot, \cdot \rangle$ that is complete (i.e. all Cauchy sequences converge in H) with respect to the norm $\|\cdot\| \coloneqq \sqrt{\langle \cdot, \cdot \rangle}$. A space X is a* Banach space *if it is a vector space equipped with a norm $\|\cdot\|$ that is complete with respect to the induced norm $\|\cdot\|$.*

Recall that the sequence $(x_n)_{n\in\mathbb{N}}$ is Cauchy if for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $k, p \geqslant N$, we have $\|x_k - x_p\| < \varepsilon$. A sequence is convergent if there exists $x \in H$ such that for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n \geqslant N$, we have $\|x_k - x\| < \varepsilon$. A space is complete if every Cauchy sequence converges in the space.

Example 4.6.
- $\mathbb{R}^n$ is a Hilbert space with the usual inner product and norm (and so is any finite-dimensional vector space). Indeed, let $(x_k)_{k\in\mathbb{N}}$ be a Cauchy sequence in $\mathbb{R}^n$. It is bounded and it belongs to a ball which is closed and bounded. By the Bolzano-Weierstrauss theorem, the ball is compact, and so $(x_k)_{k\in\mathbb{N}}$ has a convergent subsequence in the ball. Since the ball is closed, the limit of the subsequence is in the

ball, and so the sequence converges in $\mathbb{R}^n$.

- $L^2(\Omega)$ is a Hilbert space with the inner product $\langle f, g\rangle := \int_\Omega f(x)g(x)dx$.
- $L^p(\Omega)$ is a Banach space with the norm $\|f\|_p := \left(\int_\Omega |f(x)|^p dx\right)^{1/p}$.

We now prove Propositions 4.2-4.4 above.

Proof of Proposition 4.2. First we prove existence. By the definition of inf, there exists a minimising sequence $(y_n)_{n\in\mathbb{N}}$ in C such that $\|x-y_n\|$ converges to $d(x,C) := \inf_{y\in C}\|x-y\|$. We will show that the sequence $(y_n)_{n\in\mathbb{N}}$ is Cauchy. We apply the parallelogram law of vectors, which states that for all $a, b \in H$, we have

$$\|a+b\|^2 + \|a-b\|^2 = 2\|a\|^2 + 2\|b\|^2,$$

to the vectors $a = x - y_k$ and $b = x - y_p$ for some $k, p \in \mathbb{N}$:

$$2\|x-y_k\|^2 + 2\|x-y_p\|^2 = \|x-y_k+x-y_p\|^2 + \|x-y_k-x+y_p\|^2,$$

which implies that

$$\underbrace{4\left\|x - \underbrace{\frac{y_k+y_p}{2}}_{\substack{\in C \text{ since} \\ \text{convex } C}}\right\|^2}_{\geqslant d(x,C)^2} + \|y_k-y_p\|^2 = 2\|x-y_k\|^2 + 2\|x-y_p\|^2,$$

which means that

$$\|y_k-y_p\|^2 \leqslant 2\|x-y_k\|^2 + 2\|x-y_p\|^2 - 4d(x,C)^2.$$

Now both the terms $\|x-y_k\|^2$ and $\|x-y_p\|^2$ converge to $d(x,C)^2$ themselves, since the sequence $(y_n)_{n\in\mathbb{N}}$ is a minimising sequence. Therefore,

$$\begin{aligned}\lim_{k,p\to\infty}\|y_k-y_p\|^2 &\leqslant 2\lim_{k,p\to\infty}\|x-y_k\|^2 + 2\lim_{k,p\to\infty}\|x-y_p\|^2 - 4d(x,C)^2 \\ &= 2d(x,C)^2 + 2d(x,C)^2 - 4d(x,C)^2 = 0,\end{aligned}$$

Since H is a Hilbert space, there must exist some $z \in H$ such that for $(y_n)_{n\in\mathbb{N}}$, we have $\|y_n - z\| \to 0$, *i.e.* $y_n \to z \in H$. Since C is closed, we have that the limit point $z \in C$.

Now we prove uniqueness. Assume that there exist two distinct points $z, z' \in C$ such that $z \neq z'$ and

$$\|x-z\| = \inf_{y\in C}\|x-y\| = \inf_{y\in C}\|x-y\| = \|x-z'\| = d(x,C).$$

Then define a sequence

$$y_n = \begin{cases} z & \text{if } n \text{ is even,} \\ z' & \text{if } n \text{ is odd.}\end{cases}$$

Now $(y_n)_{n\in\mathbb{N}}$ is such that $y_n \in C$ and $\|x-y_n\| = d(x,C)$ trivially for all $n \in \mathbb{N}$. By the argument above, $(y_n)_{n\in\mathbb{N}}$ admits a limit in C, and by the uniqueness of the limit, we must have that $z = z'$, which is a contradiction. Therefore, the projection is unique. $\square$

A few points of note in the proof. The parallelogram law is a fundamental property of inner product spaces, and it is used because it allows us to relate the distance between two points to the distance between their average and the distance between the points themselves—which is engrained into the very motivation of the projection.. The fact that the sequence $(y_n)_{n\in\mathbb{N}}$ is Cauchy is a key step in the proof, and it is a consequence of the parallelogram law.

Proof of Proposition 4.3. We prove the forward direction first. Take $y \in C$. Given $0 < t < 1$, $zt = (1-t)z + ty \in C$ since C is convex. We now use the fact that that z is a projection, *i.e.* $\|x - z\| = \inf_{y\in C} \|x - y\| \leqslant \|x - zt\|$, which implies

$$\begin{aligned}
0 &\leqslant \|x - z\|^2 - \|x - zt\|^2 \\
&= \langle x - z, x - z\rangle - \langle x - zt, x - zt\rangle \\
&= \langle x - z, x - z\rangle - \langle x - z + ty - ty, x - z + ty - ty\rangle \\
&= \langle x - z, x - z\rangle - \langle x - z, x - z\rangle - 2t\langle x - z, y - z\rangle + t^2\langle y - z, y - z\rangle \\
&= 2t\langle x - z, y - z\rangle + t^2\|y - z\|^2,
\end{aligned}$$

and so $2t\langle x - z, y - z\rangle + t^2\|y - z\|^2 \geqslant 0$. Therefore for all $t \in (0, 1)$,

$$\langle x - z, y - z\rangle \leqslant -\frac{t}{2}\|y - z\|^2 \xrightarrow{t\to 0} 0,$$

and thus we conclude that $\langle x - z, y - z\rangle \leqslant 0$.

For the converse, we want to show that $C \ni \|x - z\| \leqslant \|x - y\|$ for all $y \in C$. In particular, take $y \in C$. Then

$$\begin{aligned}
\|x - y\|^2 &= \|x - z + z - y\|^2 \\
&= \|x - z\|^2 + \underbrace{\|z - y\|^2}_{\geqslant 0} - 2\underbrace{\langle x - z, z - y\rangle}_{\geqslant 0} \\
&\geqslant \|x - z\|^2,
\end{aligned}$$

and so $\|x - z\| \leqslant \|x - y\|$ for all $y \in C$. Therefore z is a projection. □

The following is a useful result across convex analysis and optimisation, and we will return to it more than once in this course.

Corollary 4.7. *The closed convex hull of a set $P \subseteq H$ is the intersection of all closed halfspaces containing P.*

Proof. We wish to show that

$$\overline{\mathrm{co}}\, P = \bigcap_{\substack{P\subseteq H^-(\xi,\alpha)\\ \xi\neq 0,\alpha\in\mathbb{R}}} H^-(\xi, \alpha),$$

where $H^-(\xi, \alpha)$ is the closed halfspace defined by $H^-(\xi, \alpha) \coloneqq \{x \in H : \langle \xi, x\rangle \leqslant \alpha\}$. Call the set $W = \bigcap_{\substack{P\subseteq H^-(\xi,\alpha)\\ \xi\neq 0,\alpha\in\mathbb{R}}} H^-(\xi, \alpha)$. We will show that W is a closed convex set containing P, and so $\overline{\mathrm{co}}\, P \subseteq W$. To do this, note that $W \subseteq H^-(\xi, \alpha)$ for all $\xi \neq 0$ and $\alpha \in \mathbb{R}$, and that $H^-(\xi, \alpha)$ is closed and convex, and therefore $\overline{\mathrm{co}}\, P \subseteq H^-(\xi, \alpha)$, as the closed convex hull is the smallest closed convex set containing $P \subseteq H^-(\xi, \alpha)$. Therefore $\overline{\mathrm{co}}\, P \subseteq W$.

We now show that $W \subseteq \overline{\mathrm{co}}\, P$. Let $x \in W$ and let $z = p_{\overline{\mathrm{co}}\, P}(x)$. Assume that $z \neq x$. Then for all $\xi \neq 0$, $\alpha \in \mathbb{R}$ and $y \in \overline{\mathrm{co}}\, P$, we have that by the characterisation of the projection (Proposition 4.3), $\langle y - z, x - z\rangle \leqslant 0$, and so

$$\langle y - z, x - z\rangle \leqslant 0 \implies \langle y, x - z\rangle \leqslant \langle z, x - z\rangle.$$

Now call $\tilde{\xi} := \langle y, x - z\rangle$ and $\tilde{\alpha} := \langle z, x - z\rangle$. Then it is obvious that $\overline{\mathrm{co}}\, P \subseteq H^-(\tilde{\xi}, \tilde{\alpha})$, and so $x \in H^-(\tilde{\xi}, \tilde{\alpha})$, where $H^-(\tilde{\xi}, \tilde{\alpha}) := \{x \in H : \langle \tilde{\xi}, x\rangle \leqslant \tilde{\alpha}\}$. It follows then that

$$\|x - z\|^2 = \langle x - z, x - z\rangle \leqslant 0 \implies x = z,$$

which is a contradiction. Therefore $z = x$, and so $x \in \overline{\mathrm{co}}\, P$, and so $W \subseteq \overline{\mathrm{co}}\, P$. □

Finally, we prove Proposition 4.4.

Proof of Proposition 4.4. Take $x_1, x_2 \in H$, and apply the characterisation of the projection to $p_C(x_1)$ to get, for all $y \in C$,

$$\begin{cases} p_C(x_1) \in C, \\ \langle x_1 - p_C(x_1), y - p_C(x_1)\rangle \leqslant 0. \end{cases}$$

Similarly, for $p_C(x_2)$, we have for all $y \in C$,

$$\begin{cases} p_C(x_2) \in C, \\ \langle x_2 - p_C(x_2), y - p_C(x_2)\rangle \leqslant 0. \end{cases}$$

Then by taking $y = p_C(x_2)$ in the first inequality and $y = p_C(x_1)$ in the second inequality, we get

$$\begin{aligned} \langle x_1 - p_C(x_1), p_C(x_2) - p_C(x_1)\rangle \leqslant 0, \\ \langle x_2 - p_C(x_2), p_C(x_1) - p_C(x_2)\rangle \leqslant 0, \end{aligned}$$

and adding these two inequalities gives

$$\langle x_1 - p_C(x_1), p_C(x_2) - p_C(x_1)\rangle + \langle x_2 - p_C(x_2), p_C(x_1) - p_C(x_2)\rangle \leqslant 0.$$

Therefore,

$$\langle -p_C(x_1) + p_C(x_2), p_C(x_2) - p_C(x_1)\rangle \leqslant \langle x_1 - x_2, p_C(x_1) - p_C(x_2)\rangle,$$

and so

$$\begin{aligned} \langle p_C(x_1) - p_C(x_2), p_C(x_2) - p_C(x_1)\rangle &= -\|p_C(x_1) - p_C(x_2)\|^2 \\ &\leqslant \langle x_1 - x_2, p_C(x_1) - p_C(x_2)\rangle \\ &\leqslant \|x_1 - x_2\| \|p_C(x_1) - p_C(x_2)\|, \end{aligned}$$

by the Cauchy-Schwarz inequality. So either $\|p_C(x_1) - p_C(x_2)\| = 0 \leqslant \|x_1 - x_2\|$ trivially, or $\|p_C(x_1) - p_C(x_2)\| \leqslant \|x_1 - x_2\|$. □

The following definitions allow us to specialise the above results to finite-dimensional spaces.

Definition 4.8 (Coercive and strongly convex functions)**.** *A function $f : H \to \mathbb{R}$ is* coercive *if for all $x \in H$, we have $\|x\| \to \infty \implies f(x) \to \infty$. A function $f : H \to \mathbb{R}$ is* strongly convex *with parameter $\alpha > 0$ if for all $x, y \in H$, we have*

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x \rangle + \frac{\alpha}{2}\|y - x\|^2.$$

Definition 4.9 (Lower semicontinuity)**.** *A function $f : H \to \mathbb{R}$ is* lower semicontinuous *at $\overline{x} \in \operatorname{dom} f$ if for all $x \in H$, we have $\liminf_{k\to\infty} f(x_k) \geqslant f(\overline{x})$, for any sequence $x_k \to x$.*

Fact 4.10. *Let $f : H \to \mathbb{R}$ be an extended real-valued function. Then f lower semicontinuous on H implies that the epigraph $\operatorname{epi} f$ of f is a closed set in $H \times \mathbb{R}$ with the product topology.*

Proof. Take a sequence $(x_k, t_k)_{k\in\mathbb{N}}$ in $\operatorname{epi} f$ such that $(x_k, t_k) \to (x, t) \in H \times \mathbb{R}$. Then $x_k \to x$ and $t_k \to t$, and so by the lower semicontinuity of f, we have $f(x) \leqslant \liminf_{k\to\infty} f(x_k) \leqslant \liminf_{k\to\infty} t_k = t$, and so $(x, t) \in \operatorname{epi} f$. Therefore $\operatorname{epi} f$ is closed. □

Proposition 4.11. *If X is a nonempty compact set and $f : X \to \overline{\mathbb{R}}$ is lower semicontinuous (i.e. $\operatorname{epi} f$ is closed on $E \times \mathbb{R}$), then $\inf_{x\in X} f(x)$ has at least one minimiser.*

Proof. First we prove that $\inf_{x\in X} f(x) \neq -\infty$. Suppose by way of contradiction that this is untrue. Then the family of closed sets $\mathcal{F} := \{f^{-1}(-\alpha, \alpha] : \alpha \in \mathbb{R}\}$ has the property of finite intersection, that is,

$$\bigcap_{i=1}^{n} f^{-1}(-\infty, \alpha] = f^{-1}\left(-\infty, \min_{1\leqslant i\leqslant n} \alpha_i\right] \neq \varnothing.$$

Because X is compact, we have that $\bigcap_{\alpha\in\mathbb{R}} f^{-1}(-\infty, \alpha] \neq \varnothing$. If it were empty, then take an open covering of X using the complement of this infinite intersection, extract a finite subcovering, and again take the minimum of the α_i to get a contradiction. Therefore $\inf_{x\in X} f(x) \neq -\infty$.

Now we show that $\inf_{x\in X} f(x)$ is indeed attained. Again, suppose not. Then the union defined by the set $\bigcup_{j=1}^{\infty} f^{-1}\left(\inf_{x\in X} f(x) + \frac{1}{2^j}, \infty\right)$ is an open covering of X which admits no finite subcovering, which is absurd, since X is compact—contradiction. Therefore $\inf_{x\in X} f(x)$ is attained. □

Corollary 4.12. *Let X be a closed nonempty subset and let $f\colon X \to \overline{\mathbb{R}}$ be lower semicontinuous and coercive. Then $\inf_{x\in X} f(x)$ is attained.*

Proof. Exercise. (See Appendix A for a proof.) □

§4.2. Normal cone and supporting hyperplane

Let E be a Euclidean space.

Definition 4.13 (Normal cone)**.** *Given a nonempty closed subset C of E, the* normal cone *to C at a point $x \in C$ is*

$$N_x C = \{\nu \in E : \langle \nu, y - x \rangle \leqslant 0 \text{ for all } y \in C\}.$$

Indeed we will show that the assumptions on C are redundant for the convexity of N_xC.

Proposition 4.14. *The normal cone N_xC is a nonempty closed convex cone in E for all $x \in C$.*

Proof. First we show that N_xC is nonempty. Take $x \in C$. Then $0 \in N_xC$ since $\langle 0, y-x\rangle = 0 \leqslant 0$ for all $y \in C$. Now we show that N_xC is closed. Take a sequence $(\nu_n)_{n\in\mathbb{N}}$ in N_xC such that $\nu_n \to \nu \in E$. Then for all $y \in C$, we have

$$\langle \nu, y-x\rangle = \lim_{n\to\infty} \langle \nu_n, y-x\rangle \leqslant 0,$$

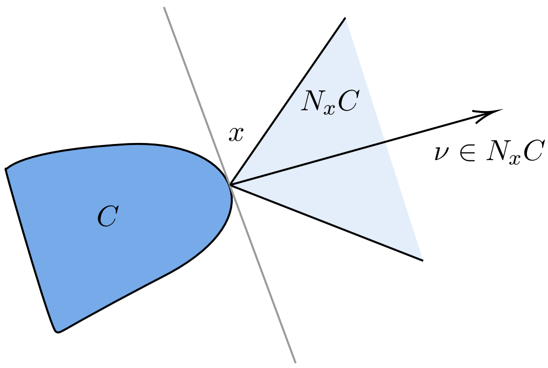
and so $\nu \in N_xC$. Therefore N_xC is closed.

Finally we show that N_xC is convex. Take $\nu_1, \nu_2 \in N_xC$ and $\lambda \in [0,1]$. Then for all $y \in C$, we have

$$\langle \lambda\nu_1 + (1-\lambda)\nu_2, y-x\rangle = \lambda\langle \nu_1, y-x\rangle + (1-\lambda)\langle \nu_2, y-x\rangle \leqslant 0,$$

and so $\lambda\nu_1 + (1-\lambda)\nu_2 \in N_xC$. Therefore N_xC is convex. ☐

The following picture makes that connection more explicit:



Proposition 4.15. *Let C be a nonempty closed convex subset of E. Let E_0 be the linear subspace parallel to* $\operatorname{aff} C$ *and let $x \in \partial_{\mathrm{rel}} C = \overline{C} \setminus \operatorname{ri} C$, where* $\operatorname{ri} C$ *is the relative interior of C. Then $N_xC \cap E_0$ contains a nonzero element.*

Proof. Without loss of generality, take $\operatorname{aff} C = E$. If $x \in \partial C$, then $C \neq E$ and there exists a sequence $(x_k)_{k\in\mathbb{N}}$ in $E \setminus C$ which converges to x. Let $y_k = p_C(x_k)$. Since $y_k \neq x_k$, we can define the unit vector

$$\nu_k := \frac{y_k - x_k}{\|y_k - x_k\|} \in N_{x_k}C.$$

Then by the characterisation of the projection (Proposition 4.3), we have for all $y \in C$,

$$\langle \nu_k, y - x_k\rangle = \frac{\langle y_k - x_k, y - x_k\rangle}{\|y_k - x_k\|} \leqslant 0 \implies \langle \nu_k, y - y_k\rangle \leqslant 0.$$

Note now that $\|y_k - x\| = \|p_C(x_k) - p_C(x)\|$, since $x \in \partial C$, and $\partial C = \overline{C} \setminus C^\circ = C \setminus C^\circ$, and so $x \in \overline{C} \setminus C^\circ$. Thus we have, since $x_k \to x$,

$$\lim_{k\to\infty} \|y_k - x\| = \lim_{k\to\infty} \|p_C(x_k) - p_C(x)\| = 0. \tag{2}$$

Since E is finite dimensional, the unit sphere is compact and so there exists a subsequence $(\nu_{k_j})_{j\in\mathbb{N}}$ of $(\nu_k)_{k\in\mathbb{N}}$ such that $\nu_{k_j} \to \nu \in S(0,1)$, *i.e.* $\|\nu\| = 1$. Taking the limit in equation (2) gives that for all $y \in C$, we have $\langle \nu, y - x\rangle \leqslant 0$, and so $\nu \in N_x C \setminus \{0\}$. Therefore $N_x C \cap E_0$ contains a nonzero element. $\square$

Now we define the supporting hyperplane to a nonempty closed convex set C at a point $x \in \partial C$.

Definition 4.16 (Supporting hyperplane)**.** *A* supporting hyperplane *of a convex set C is a hyperplane containing C in one of its closed halfspaces and containing a point on the boundary of C. It is thus a set which has the form*

$$H(\xi, \alpha) = \{x \in E : \langle \xi, x\rangle = \alpha\},$$

where $\xi \in E \setminus \{0\}$, with $0 \neq \alpha \in \mathbb{R}$ and $C \subseteq H^-(\xi, \alpha)$ satisfies $\partial C \cap H^-(\xi, \alpha) \neq \varnothing$.

Note that a supporting hyperplane of C at $x \in \overline{C}$ is a supporting hyperplane of C which contains x.

§5. Lecture 05—16th February, 2024

We will start by making a few more comments about the supporting hyperplane. Note that the supporting hyperplane is not unique. For example, in the case of a polyhedron, there are infinitely many supporting hyperplanes. However, the normal vector to the supporting hyperplane is unique.

The following property follows from many of the results we have already seen so far. Let $(E, \langle\cdot,\cdot\rangle)$ be a Euclidean space with inner product $\langle\cdot,\cdot\rangle$.

Proposition 5.1 (Supporting hyperplane)**.** *Let C be a nonempty convex subset of E, let E_0 be the linear subspace parallel to the affine hull $\operatorname{aff} C$, and let $x \in \partial_{\mathrm{rel}} C$, the relative boundary of C. Then there exists a supporting hyperplane $H(\xi, \alpha)$ of C at the point x such that $\xi \in E_0$.*

Proof. Since $\operatorname{aff} \overline{C} = \operatorname{aff} C$ and $\operatorname{ri} \overline{C} = \operatorname{ri} C$, it follows that E_0 is also the linear subspace parallel to the nonempty closed convex set $\overline{C}$, and we have that $x \in \partial_{\mathrm{rel}} \overline{C} = \partial_{\mathrm{rel}} C$. Therefore by Proposition 4.15, there exists $\xi \in \left(N_x \overline{C} \cap E_0\right) \setminus \{0\}$. Then $\langle \xi, y - x\rangle \leqslant 0$ for all $y \in \overline{C}$; in particular, $C \subseteq H^- H(\xi, \alpha)$, where $\alpha = \langle \xi, x\rangle$. In addition, $x \in H(\xi, \alpha) \cap \partial_{\mathrm{rel}} C \subseteq H(\xi, \alpha) \cap \partial C$, so that $H(\xi, \alpha)$ is a supporting hyperplane of C at x with $\xi \in E_0$. $\square$

§5.1. Separation of convex sets

Definition 5.2. *A hyperplane $H(\xi, \alpha) \coloneqq \{x \in E : \langle \xi, x\rangle = \alpha\}$ where $\xi \neq 0$* separates *two convex sets C_1 and C_2 if for all $x_1 \in C_1$ and for all $x_2 \in C_2$, we have $\langle \xi, x_1\rangle \leqslant \alpha \leqslant \langle \xi, x_2\rangle$.*

This is equivalent to

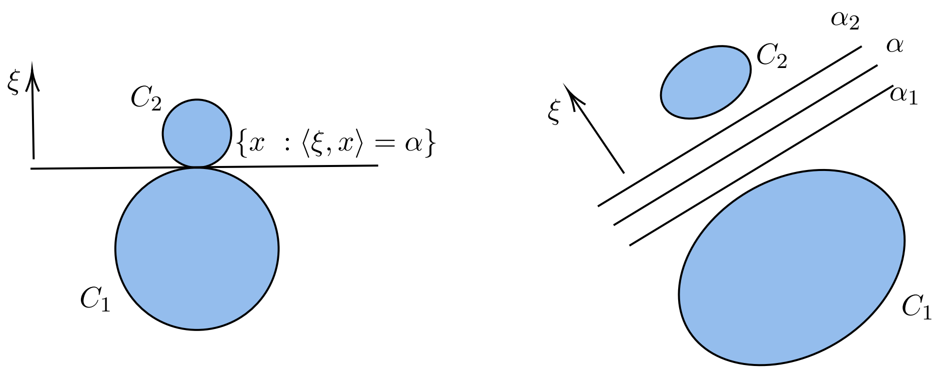
$$\sup_{x_1 \in C_1} \langle \xi, x_1\rangle \leqslant \inf_{x_2 \in C_2} \langle \xi, x_2\rangle.$$

The idea here is very intuitive: the hyperplane $H(\xi, \alpha)$ separates the two sets C_1 and C_2 if it is possible to find a hyperplane such that C_1 is on one side of the hyperplane and C_2 is on the other side. The following definition is a slightly stronger notion.

Definition 5.3. *A hyperplane $H(\xi, \alpha)$* strictly separates *two convex sets C_1 and C_2 if there exist distinct scalars α_1 and α_2 such that $\alpha_1 < \alpha_2$ and for all $x_1 \in C_1$ and for all $x_2 \in C_2$, we have $\langle \xi, x_1 \rangle < \alpha_1 < \alpha_2 < \langle \xi, x_2 \rangle$.*

Similarly, this is equivalent to

$$\sup_{x_1 \in C_1} \langle \xi, x_1 \rangle < \inf_{x_2 \in C_2} \langle \xi, x_2 \rangle.$$



Theorem 5.1. *Let C_1 and C_2 be nonempty convex sets in E. Then the following are equivalent:*

(i) One can strictly separate C_1 and C_2.

(ii) $0 \notin \overline{C_1 - C_2}$.

(iii) $\inf\{\|x_1 - x_2\| : x_1 \in C_1, x_2 \in C_2\} > 0$.

Proof. First we show that (i) implies (ii). Suppose that $H(\xi, \alpha)$ strictly separates C_1 and C_2. Then there exist distinct positive scalars α_1 and α_2 such that $\alpha_1 < \alpha_2$ and for all $x_1 \in C_1$ and for all $x_2 \in C_2$, we have $\langle \xi, x_1 \rangle < \alpha_1 < \alpha_2 < \langle \xi, x_2 \rangle$. Then $\langle \xi, x_2 - x_1 \rangle \geqslant \alpha_2 - \alpha_1 > 0$ for all $x_1 \in C_1$ and for all $x_2 \in C_2$. Thus

$$\begin{aligned} 0 &< \inf_{x_1 \in C_1, x_2 \in C_2} \langle \xi, x_2 - x_1 \rangle \\ &= \inf_{x_1 \in C_1, x_2 \in C_2} \langle -\xi, x_1 - x_2 \rangle \\ &= \inf_{x \in C_1 - C_2} \langle -\xi, x \rangle \\ &= \inf_{x \in \overline{C_1 - C_2}} \langle -\xi, x \rangle, \end{aligned}$$

thus $0 \notin \overline{C_1 - C_2}$; if $0 \in \overline{C_1 - C_2}$, then we would have $\inf_{x \in \overline{C_1 - C_2}} \langle -\xi, x \rangle = 0$, which is a contradiction.

(In the final step of the proof above we are using the fact that if $f : E \to \mathbb{R}$ is continuous and $X \in E$, then $\inf_X f = \inf_{\overline{X}} f$. We provide a proof below. Clearly $\inf_X f \geqslant \inf_{\overline{X}} f$, since the infimum over a larger set is never larger. Conversely, for a sequence $\{x_k\} \subseteq \overline{X}$ such that $f(x_k) \to \inf_{\overline{X}} f$, and thus we can find a subsequence $\{x_{k_j}\} \subseteq X$ such that $x_{k_j} \to x_k$ as $j \to \infty$. Now for an extraction φ, take another subsequence $x_{k\varphi(k)} \in X$ such that $|f(x_{k\varphi(k)}) - f(x_k)| \leqslant \frac{1}{k+1}$. Then $f(x_{k\varphi(k)}) = f(x_{k\varphi(k)}) - f(x_k) + f(x_k) \to \inf_X f$ as $k \to \infty$ since $f(x_{k\varphi(k)}) - f(x_k) \to 0$ and $f(x_k) \to \inf_{\overline{X}} f$ as $k \to \infty$. Thus $\inf_X f = \inf_{\overline{X}} f$.)

Now we show that (ii) implies (iii). The infimum can be written as

$$\begin{aligned}\inf\{\|x_1 - x_2\| : x_1 \in C_1, x_2 \in C_2\} &= \inf\{\|x\| : x \in C_1 - C_2\} \\ &= \min\{\|x\| : x \in \overline{C_1 - C_2}\} > 0,\end{aligned}$$

because the norm $\|\cdot\|$ is coercive and from Corollary 4.12. This happens if and only if $0 \notin \overline{C_1 - C_2}$.

Finally, we show that (iii) implies (i). Since the nonempty closed convex set $C \coloneqq C_1 - C_2$ does not contain 0, the projection ξ of 0 on C is nonzero, and by the characterisation of the projection, for all $x \in C$, we have that $\langle 0 - \xi, x - \xi\rangle$ is nonpositive. Then for all $x_1 \in C_1$ and for all $x_2 \in C_2$,

$$\langle -\xi, x_1 - x_2 - \xi\rangle \leqslant 0 \iff \langle -\xi, x_1\rangle + \underbrace{\|\xi\|^2}_{>0} \leqslant \langle -\xi, x_2\rangle,$$

and thus $\langle \xi, x_1\rangle < \langle \xi, x_2\rangle$ for all $x_1 \in C_1$ and for all $x_2 \in C_2$. Therefore $H(\xi, \alpha)$ separates C_1 and C_2. □

The following corollary follows immediately from the theorem.

Corollary 5.4 (Criteria for strict separation)**.** *One can strictly separate two nonempty closed disjoint convex sets C_1 and C_2 if:*

(1) $C_1 - C_2$ is closed; or,

(2) The intersection of the recession cones $C_1^\infty \cap C_2^\infty = \{0\}$; or,

(3) C_1 or C_2 is compact.

Proof. We start by establishing (1). By the implication (ii) $\Longrightarrow$ (i) in the above theorem and the fact that $0 \notin C_1 - C_2 = \overline{C_1 - C_2}$, we have that $C_1 - C_2$ is closed if C_1 and C_2 can be separated since $C_1 \cap C_2 = \emptyset$.

Now we prove (2); by (1) above, it suffices to show that $C_1 - C_2$ is closed. Take sequences $\{x_k^{(1)}\} \in C_1$ and $\{x_k^{(2)}\} \in C_2$ such that $x_k^{(1)} - x_k^{(2)} \to x \in E$. We will show that $x \in C_1 - C_2$. Notice first that the sequences $\{x_k^{(1)}\}$ and $\{x_k^{(2)}\}$ are bounded; otherwise, if say, $\{x_k^{(1)}\}$ is not bounded, then with $t_k \coloneqq \|x_k^{(1)}\|$, we can extract a converging subsequence $\{x_{\varphi(k)}^{(1)}\} \setminus t_{\varphi(k)}$ which converges to some $d \in C_1^\infty \setminus \{0\}$ and $t_{\varphi(k)} \to \infty$, by Proposition 3.7. We then have that

$$x_{\varphi(k)}^{(1)} - x_{\varphi(k)}^{(2)} \to x \iff \frac{x_{\varphi(k)}^{(1)} - x_{\varphi(k)}^{(2)}}{t_{\varphi(k)}} \to 0 \iff \frac{x_{\varphi(k)}^{(1)}}{t_{\varphi(k)}} - \frac{x_{\varphi(k)}^{(2)}}{t_{\varphi(k)}} \to 0,$$

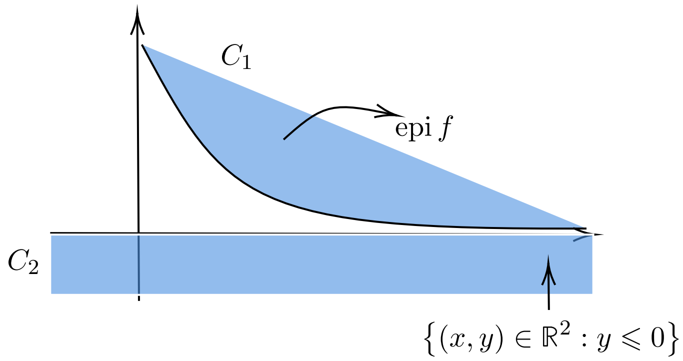
since x is fixed and $t_{\varphi(k)} \to \infty$; but then $x_{\varphi(k)}^{(1)}/t_{\varphi(k)} \to d \in C_1^\infty$ and $x_{\varphi(k)}^{(2)}/t_{\varphi(k)} \to d \in C_2^\infty$, and thus $d \in C_1^\infty \cap C_2^\infty \neq \{0\}$, which is a contradiction. Therefore, the sequences $\{x_k^{(1)}\}$ and $\{x_k^{(2)}\}$ are bounded, and thus we can extract converging subsequences $\{x_{k_j}^{(1)}\} \to x^{(1)} \in C_1$ and

$\{x^{(2)}_{k_j}\} \to x^{(2)} \in C_2$ such that $x^{(1)}_{\varphi(k)} - x^{(2)}_{\varphi(k)} = x$. Now $x_1 \in C_1$ and $x_2 \in C_2$ are limits of sequences in C_1 and C_2, and thus $x_1 \in C_1$ and $x_2 \in C_2$, and thus $x = x_1 - x_2 \in C_1 - C_2$, and thus $C_1 - C_2$ is closed.

Finally, we prove (3). If C_1 is compact, then $C_1^\infty = \{0\}$, and thus $C_1^\infty \cap C_2^\infty = \{0\}$, and thus we can simply apply (2), and similarly if C_2 is compact. $\square$

To crystallise our understanding of the above corollary, we will provide an example for which we cannot be able to perform strict separation for two nonempty closed convex disjoint unbounded sets. Consider the following function $f : \mathbb{R} \to \overline{\mathbb{R}}$ defined by

$$x \mapsto \begin{cases} 1/x & \text{if } x > 0, \\ \infty & \text{otherwise.} \end{cases}$$



Take $\mathbb{R}_+ = [0, \infty)$. Then the recession cones are $C_1 = \mathbb{R}_+ \times \mathbb{R}_+$ and $C_2 = \mathbb{R} \times \mathbb{R}_-$. Then the sufficient condition fails, because $C_1^\infty \cap C_2^\infty = \mathbb{R}_+ \times \{0\} \neq \{0\}$, and neither of the sets is compact. Thus we cannot strictly separate the two sets, which is nice because if it were, then we would have a contradiction with the fact that the recession cones intersect at a nontrivial set. So closed convex sets need not be strictly separable, and we need to be careful when dealing with unbounded sets.

We now consider the question of separability as opposed to strict separability, starting from the following elegant result:

Theorem 5.2. *One can separate two nonempty disjoint convex sets. That is, if C_1 and C_2 are nonempty disjoint convex sets, then there exists a hyperplane $H(\xi, \alpha)$ such that $C_1 \subseteq H^- H(\xi, \alpha)$ and $C_2 \subseteq H^+ H(\xi, \alpha)$.*

Proof. Let $C := C_1 - C_2$. Then C is a nonempty closed convex set, and $0 \notin C$. We will show that there exists $\xi \in E \setminus \{0\}$ such that for all $(x_1, x_2) \in C_1 \times C_2$, we have $\langle \xi, x_1 - x_2 \rangle \leqslant 0$, *i.e.* $\langle x_2 - x_1, \xi \rangle \geqslant 0$.

If $0 \notin \overline{C}$, then we can take the projection $\xi = p_{\overline{C}}(0) \neq 0$, and then we have that for all $x_1 \in C_1$ and for all $x_2 \in C_2$, it holds that $\langle 0 - \xi, (x_2 - x_1) - \xi \rangle \leqslant 0$, which then implies that $0 < \|\xi\|^2 \leqslant \langle \xi, x_2 - x_1 \rangle$, and thus $\langle \xi, x_1 \rangle \leqslant \langle \xi, x_2 \rangle$.

Otherwise if $0 \in \overline{C}$, then $0 \in \overline{C} \setminus C \subseteq \overline{C} \setminus \operatorname{ri} C = \partial_{\text{rel}} C$. Thus we can take ξ to be the opposite of a normal vector to C at 0 which is nonzero, the existence of which is guaranteed by Proposition 5.1, and which satisfies $\langle x_2 - x_1, -\xi \rangle \leqslant 0$ for all $x_1 \in C_1$ and for all $x_2 \in C_2$. $\square$

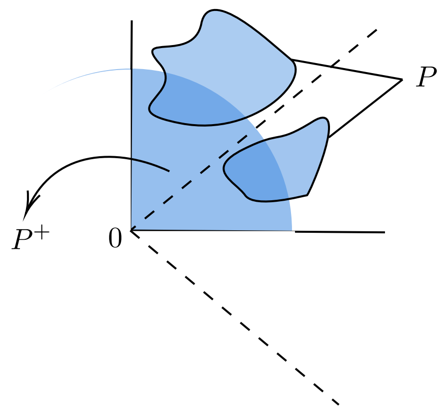
§5.2. Dual cone and Farkas' lemma

Recall that a cone is the set K such that if for all $x \in K$ and for all $\lambda \geqslant 0$, we have $\lambda x \in K$. The dual cone is the set of non-negative dot products of some $y \in \mathbb{R}^n$ with all $x \in K$; more formally:

Definition 5.5. *The* dual cone *of a subset P of E is the set*

$$P^+ := \{x \in E : \langle x, y\rangle \geqslant 0, \text{ for all } y \in P\}.$$

The dual cone is always a closed convex cone, and if P is a closed convex cone, then $P^{++} = \overline{P}$.



Observe that

$$P^+ = \bigcap_{y \in P} \{x \in E : \langle x, y\rangle \geqslant 0\},$$

and thus P^+ is the intersection of half-spaces, and thus is a closed convex cone.

The notion of the dual cone generalises that of the orthogonal subspace, since if P is a linear subspace, then $P^+ = P^\perp$, where $P^\perp$ is the orthogonal subspace of P. Recall that if $A : E \to F$ is a linear operator between finite-dimensional Euclidean vector spaces E and F, then the adjoint operator A^* must satisfy

$$\begin{aligned} \mathsf{range}(A) &= \mathsf{kern}(A^*)^\perp, \\ \overline{A(E)} = A(E) &= (A^*)^{-1}\left(E^\perp\right)^\perp \qquad\qquad \text{for } K := E, \text{ by Farkas'}. \end{aligned}$$

Farkas' lemma is a generalisation of all of this.

Lemma 5.6 (Farkas' Lemma, [Far02])**.** *Let E and F be Euclidean spaces and $A : E \to F$ a linear operator. Furthermore, let K be a nonempty convex cone of E. Then,*

$$\overline{A(K)} = \left(A^{*-1}\left(K^+\right)\right)^+.$$

Proof. We prove both inclusions separately. First, we show that $\overline{A(K)} \subseteq \left(A^{*-1}\left(K^+\right)\right)^+$. Since $\left(A^{*-1}\left(K^+\right)\right)^+$ is closed, it suffices to show that $A(K) \subseteq \left(A^{*-1}\left(K^+\right)\right)^+$. Let $y \in A(K)$, so that

$y = Ax$ for some $x \in K$. We will show that for all $x_0 \in A^{*-1}(K^+)$, we have that $\langle y, x_0 \rangle \geqslant 0$. To do so, we compute the inner product:

$$\langle y, x_0 \rangle = \langle Ax, x_0 \rangle = \langle x, A^* x_0 \rangle \geqslant 0,$$

since $x \in K$ and $x_0 \in A^{*-1}(K^+)$ with $A^* x_0 \in K^+$. Thus $y \in \left(A^{*-1}(K^+)\right)^+$, and thus $\overline{A(K)} \subseteq \left(A^{*-1}(K^+)\right)^+$.

We now turn to the second inclusion; we will show that $\left(A^{*-1}(K^+)\right)^+ \subseteq \overline{A(K)}$. Let us reason by contradiction and assume that there exists some vector $d \in F$ such that $d \in \left(A^{*-1}(K^+)\right)^+$ but $d \notin \overline{A(K)}$, that is, $d \in \left(A^{*-1}(K^+)\right)^+ \setminus \overline{A(K)}$. Then we can strictly separate $\{d\}$ and $\overline{A(K)}$ because one of these sets is compact. Thus there exists $y \in F$ such that there exists $\alpha \in \mathbb{R}$ such that for all $x \in K$, we have $\langle y, d \rangle < \alpha \leqslant \langle y, Ax \rangle$. Since this is a cone, we can take $x \to 0$, and then we immediately notice that $\langle y, d \rangle < \alpha \leqslant 0$. Recalling that K is nonempty, let $x_0 \in K$; since K is a cone, for all $t > 0$,

$$\begin{aligned} \alpha \leqslant \langle y, A(tx_0) \rangle &\iff \alpha \leqslant t \langle y, Ax_0 \rangle \\ &\iff \frac{\alpha}{t} \leqslant \langle y, Ax_0 \rangle \\ &\to 0 \leqslant \langle y, d \rangle && \text{as } t \to \infty, \\ &\iff 0 \leqslant \langle A^* y, x_0 \rangle . \end{aligned}$$

As x_0 was arbitrary, we have that $\langle A^* y, x_0 \rangle \geqslant 0$ for all $x_0 \in K$. But then by the definition of the dual cone, $A^* y \in K^+$ and $y \in A^{*-1}(K^+)$, and thus $d \in \left(A^{*-1}(K^+)\right)^+$, which implies that $\langle y, d \rangle \geqslant 0$. Yet we have that $\langle y, d \rangle < 0$—contradiction; it must be that $d \in \overline{A(K)}$, and thus $\left(A^{*-1}(K^+)\right)^+ \subseteq \overline{A(K)}$. $\square$

This result is very important for optimisation; it is at the foundation of optimality conditions for constrained optimisation (KKT conditions), and it is key in duality for linear programming.

§5.3. Convex functions

Let $f \in \operatorname{conv} E$ be a proper convex function.

Note that

$$\operatorname{ri}(\operatorname{epi} f) = \{(x, \alpha) \in E \times \mathbb{R} : x \in \operatorname{ri}(\operatorname{dom} f) \text{ such that } f(x) < \alpha\},$$

due to Lemma 3.12. (The key implication is that if for all $x_0 \in C$, there exists $t > 1$ such that $(1-t)x_0 + tx \in C$, then $x \in \operatorname{ri} C$.)

We now present some examples of convex functions.

1. **Indicator.** Given a set E, the indicator of $P \subseteq E$ is the function $\delta_p : E \to \overline{\mathbb{R}}$ defined by

$$\delta_P(x) := \begin{cases} 0 & \text{if } x \in P, \\ \infty & \text{otherwise.} \end{cases}$$

Proposition 5.7. *If $P \subseteq E$ and P is nonempty, then $\delta_P \in \operatorname{conv} E$ (resp. $\delta_P \in \overline{\operatorname{conv}}\, E$) if and only if P is convex (resp. closed and convex).*

Proof. Exercise. (See Appendix A for a proof.) □

2. **Affine functions.** A function $a : E \to \mathbb{R}$ is affine if for all $x, y \in E$ and for all $t \in \mathbb{R}$, $a\left((1-t)x + ty\right) = (1-t)a(x) + ta(y)$. Equivalently, $E \ni x \mapsto a(x) - a(0)$ is linear. An affine function has the form
$$a : x \in E \mapsto a(x) = \langle x^*, x\rangle - \alpha,$$
where $x^* \in E$ and $\alpha \in \mathbb{R}$. The function a is convex if and only if $x^* \in E$ and $\alpha \in \mathbb{R}$.

§6. Lecture 06—23rd February, 2024

§6.1. More on convex functions

We continue our discussion of affine functions from last time.

2. We begin our discussion today with the following crucial result that says that we can always find for any convex function, an affine function that is a global (except for exactly one) underestimator or a lower bound for the convex function.

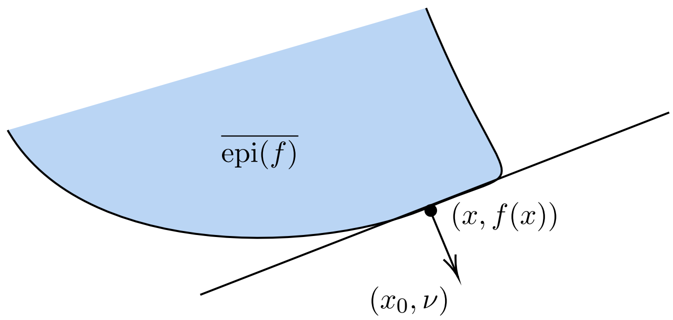
   Proposition 6.1. *If $f \in \operatorname{conv}(E)$ and $x \in \operatorname{ri}(\operatorname{dom} f)$, then there exists a point x^* in the linear subspace parallel to $\operatorname{aff}(\operatorname{dom} f)$ such that for all $y \in E$, we have*
$$f(y) \geqslant f(x) + \langle x^*, y - x\rangle.$$

   Proof. First note that $(x, f(x)) \in \partial_{\text{rel}}(\overline{\operatorname{epi} f})$. Indeed, $(x, f(x)) \in \partial_{\text{rel}}(\operatorname{epi} f)$ because $(x, f(x)) \in \operatorname{epi} f \subseteq \overline{\operatorname{epi} f}$ and $(x, \alpha) \in \operatorname{aff}(\operatorname{epi} f) \setminus \operatorname{epi} f$ for all $\alpha < f(x)$, so that $(x, f(x)) \notin \operatorname{ri}(\operatorname{epi} f)$ by the definition of the interior. As seen previously, $\operatorname{ri}(\operatorname{epi} f) = \operatorname{ri}(\overline{\operatorname{epi} f})$, so $(x, f(x)) \in \partial_{\text{rel}}(\overline{\operatorname{epi} f})$.

   Now by Proposition 4.15, there exists a nonzero normal vector $(x_0^*, \nu) \in N_{(x, f(x))}\overline{\operatorname{epi} f} \setminus \{(0,0)\}$ belonging to the parallel linear subspace to the $\operatorname{aff}(\operatorname{epi} f) = \operatorname{aff}(\operatorname{dom} f) \times \mathbb{R}$, that is, $E_0 \times \mathbb{R}$, where E_0 is the parallel linear subspace to $\operatorname{aff} \operatorname{dom} f$. We have then that for all $(y, \alpha) \in \overline{\operatorname{epi} f}$,
$$\langle (y, \alpha) - (x, f(x)), (x_0^*, \nu)\rangle \leqslant 0 \implies \langle y - x, x_0^*\rangle + (\alpha - f(x))\nu \leqslant 0. \tag{3}$$

   We now make the following remarks:

   a) For all $\alpha \geqslant f(x)$, it holds that $(x, \alpha) \in \operatorname{epi} f$, and so we can apply (3) to get
$$\underbrace{\langle x - x, x_0^*\rangle}_{=0} + (\alpha - f(x))\nu \leqslant 0.$$
   In particular, with $\alpha \coloneqq f(x) + 1$, then $(f(x) + 1 - f(x))\nu \leqslant 0$, so that $\nu \leqslant 0$.



b) Furthermore, $\nu = 0$ is impossible, since then we could take $y = x + tx_0^* \in \operatorname{dom} f$ for all $t > 0$ small (because $x \in \operatorname{ri}(\operatorname{dom} f)$ and $x_0^* \in E_0$), and then (3) would imply that

$$\langle x + tx_0^* - x, x_0^* \rangle + (\alpha - f(x))\nu \leqslant 0 \implies t\|x_0^*\|^2 \leqslant 0,$$

and so the norm of x_0^* would have to be zero, so that $(x_0^*, \nu) = (0, 0)$, which is impossible since it is a nonzero normal vector.

Therefore we can just take $x^* = -x_0^*/\nu$ and $\alpha = f(y)$ so that for all $y \in E$, (3) gives

$$\begin{aligned} \langle y - x, x_0^* \rangle + (f(y) - f(x))\nu \leqslant 0 \\ f(y) &\geqslant f(x) - \langle x_0^*/\nu, y - x \rangle \\ &= f(x) + \langle x^*, y - x \rangle, \end{aligned}$$

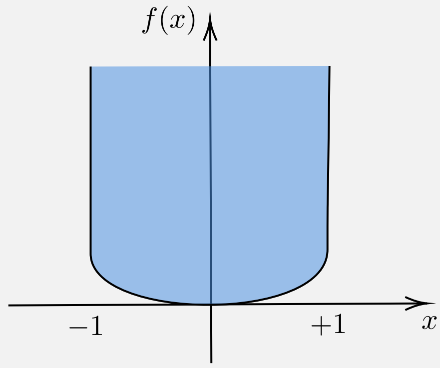
and the proof is done. □

The following example should be useful to keep in mind when thinking about the above result.

Example 6.2. Take the function

$$f(x) = \begin{cases} 1 - \sqrt{1 - x^2} & \text{if } x \in [-1, 1], \\ \infty & \text{otherwise.} \end{cases}$$

Note now that $1 \notin \operatorname{ri}(\operatorname{dom} f)$, so we cannot apply the above result to f at $x = 1$.
In particular, there is no affine lower bound that is exact at $x = 1$ (such that for affine function a, $a(1) = f(1)$).



3. **Sublinear functions.** Intuitively, a sublinear function is a function that is "less than linear" in some sense. We will make this precise in the following definition.

Definition 6.3 (Sublinear function)**.** *A function* $\sigma : E \to \mathbb{R} \cup \{\infty\}$ *is called* sublinear *if it is subadditive and positively homogeneous, that is, for all* $x_1, x_2 \in \operatorname{dom} \sigma$ *and for all* $t_1, t_2 \geqslant 0$*, we have*

$$\sigma(t_1 x_1 + t_2 x_2) \leqslant t_1 \sigma(x_1) + t_2 \sigma(x_2).$$

The following proposition gives a useful characterization of sublinear functions.

Proposition 6.4. *Suppose a function* $\sigma : E \to \mathbb{R} \cup \{\infty\}$ *has a nonempty domain. Then the following are equivalent:*

(i) σ *is sublinear.*

(ii) σ *is convex and positively homogeneous (i.e. for all* $x \in \operatorname{dom} \sigma$ *and for all* $t \geqslant 0$*, we have* $\sigma(tx) = t\sigma(x)$*).*

(iii) The epigraph $\operatorname{epi} \sigma$ *is a convex cone of* $E \times \mathbb{R}$*.*

Proof. Homework exercise; the proof is obvious. (See Appendix A for a proof.) □

Here's an example of a sublinear function.

> **Example 6.5.** If $f \in \operatorname{conv} E$ and $x \in \operatorname{dom} f$, then we can define the directional derivative
>
> $$f'(x, d) := \lim_{t \searrow 0} \frac{f(x + td) - f(x)}{t} \in \overline{\mathbb{R}},$$
>
> It is easy to check that $f'(x, \cdot)$ is a sublinear function.

4. **Support functions.** The support function of a set is a sublinear function that measures the distance of a point from the set. We will make this precise in the following definition.

Definition 6.6 (Support function)**.** *Let* $P \subseteq E$ *be a nonempty set. The* support function *of* P *is the function* $\sigma_P : E \to \mathbb{R} \cup \{\infty\}$ *defined by*

$$\sigma_P(x) := \sup_{x \in P} \langle d, x \rangle .$$

Here's another useful proposition:

Proposition 6.7. *Let* P, P_1, P_2 *be nonempty subsets of* E*. Then the following hold:*

(i) The support σ_P *is sublinear and closed (i.e.* $\operatorname{dom} \sigma_P = E$*).*

(ii) One support dominates over another if and only if its closed convex hull also dominates over the other, i.e. $\sigma_{P_1} \leqslant \sigma_{P_2}$*, if and only if* $\overline{\operatorname{co}}\, P_1 \subseteq \overline{\operatorname{co}}\, P_2$*.*

(iii) The supports are equal if and only if the closed convex hulls are the same, i.e. $\sigma_{P_1} = \sigma_{P_2}$*, if and only if* $\overline{\operatorname{co}}\, P_1 = \overline{\operatorname{co}}\, P_2$*.*

Proof. Homework exercise; as a hint, we use strict separation of convex sets. (See Appendix A for a proof.) □

§6.2. Subdifferentiability

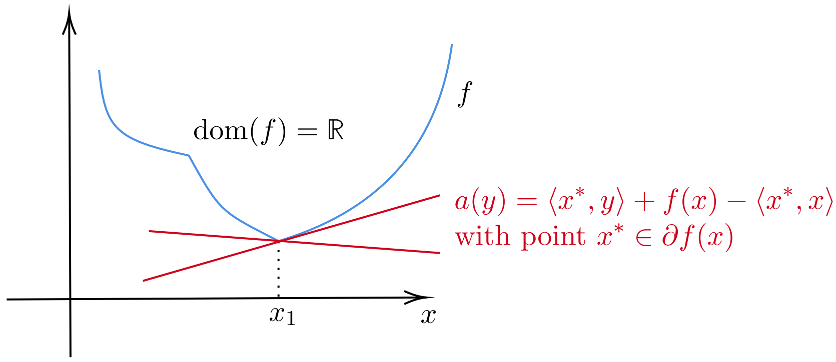
From Proposition 6.1, it is quite natural to define the subdifferential of a convex function; intuitively, the subdifferential of a convex function is a set of all the slopes of affine functions that are global underestimators of the convex function. More precisely:

Definition 6.8 (Subdifferential)**.** *Let $f \in \operatorname{conv} E$. The* subdifferential *of f is the set-valued mapping $\partial f\colon E \rightrightarrows E$ from E to the subsets of E defined by*

$$\partial f(x) \coloneqq \begin{cases} \{x^* \in E : \forall\, y \in E,\ f(y) \geqslant f(x) + \langle x^*, y - x\rangle\} & \text{if } x \in \operatorname{dom} f, \\ \emptyset & \text{otherwise.} \end{cases}$$

The subgradient *of f at x is any element of $\partial f(x)$.*

The subdifferential of f at $x \in \operatorname{dom} f$ is the set of slopes of all affine minorants of f exactly at x:



We will give a series of useful rules from calculus that follow naturally from the definition of the subdifferential. All of their proofs are left as exercises. (See Appendix A for the proofs.)

Proposition 6.9. *Let $a \geqslant 0$ and $f \in \operatorname{conv} E$ with $x \in E$. Then $\partial(\alpha f)(x) = \alpha \cdot \partial f(x)$.*

Proposition 6.10. *Let $f_1, \ldots, f_p \in \operatorname{conv} E$ and $x \in E$. Then*

$$\partial\left(\sum_{i=1}^{p} f_i\right)(x) \supseteq \sum_{i=1}^{p} \partial f_i(x),$$

with equality if

$$\bigcap_{1 \leqslant i \leqslant p} \operatorname{ri}(\operatorname{dom} f_i) \neq \emptyset.$$

Proposition 6.11. *Let $a\colon E \to F$ be an affine function from Euclidean spaces E and F, i.e. $a(x) = Ax + b$ for some linear function $A\colon E \to F$ and some vector $b \in F$. If $g \in \operatorname{conv}(F)$, then*

$$\partial(g \circ a)(x) \supseteq A^*\left(\partial g(a(x))\right),$$

where A^* is the adjoint of A, with equality if $\mathsf{range}(A) \cap \operatorname{ri}(\operatorname{dom} g) \neq \emptyset$.

Proposition 6.12. *Let $F\colon E \to (\mathbb{R} \cup \{+\infty\})^m$ be componentwise convex, i.e. for all $x \in E$, the function $F(x) = (F_1(x), \dots, F_m(x))$ is convex, and let $x \in E$. Furthermore, let $g\colon \mathbb{R}^m \to \overline{\mathbb{R}}$ be convex and increasing in each component, i.e. for all $y, z \in \mathbb{R}^m$ with $y \leqslant z$, we have $g(y) \leqslant g(z)$. Then the composition $g \circ F$ defined by*

$$(g \circ F)(x) := \begin{cases} g(F(x)) & \text{if } F(x) \in \mathbb{R}^m, \\ +\infty & \text{otherwise,} \end{cases}$$

is convex, and

$$\partial(g \circ F)(x) = \left\{ \sum_{i=1}^{m} \gamma_i x_i^* \text{ such that } \gamma \in \partial g(F(x)),\ x_i^* \in \partial F_i(x),\ 1 \leqslant i \leqslant m \right\},$$

if $F(x) \in \mathbb{R}^m$ is finite.

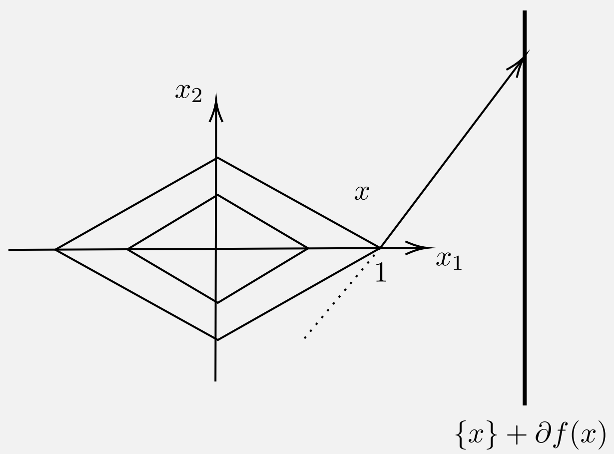
Proposition 6.13. *If $f \in \operatorname{conv}(E)$ and $x \in \operatorname{ri}(\operatorname{dom} f)$, then $\partial f(x)$ is nonempty and compact.*

Using these rules, we can now compute the subdifferential of some simple functions.

Example 6.14. Consider the function $f(x_1, x_2) = |x_1| + 2|x_2|$, and let $x = (1, 0)$. Then the subdifferential is

$$\partial f(x) = \{(1, 0)\} + \{0\} \times [-2, 2].$$

Indeed consider the level sets of this function:



The bolded line is the compact set that is the translated subdifferential at $(1, 0)$; all the arrows from the point x to the line are the slopes of the affine functions that are global underestimators of f at x; they are the subdifferential of f at x.
Indeed, the opposite direction of a subdifferential is an ascent direction of the function, even if we take an arbitrarily small displacement from the point x in that direction. This is a very strange—but we can convince ourselves of this algebraically or by the picture—we can't count on the convex values decreasing in the direction of the subdifferential. The idea from this example that we will continue to use in convex optimisation however is that we can count on progressively approaching a minimiser of a convex function by moving in the direction of the subdifferential.

§6.3. A subgradient method for minimising convex functions

We will now design an algorithm for minimising convex functions. Let $f\colon E \to \mathbb{R}$ be convex and assume that it admits at least one minimiser, that is, $\emptyset \neq \arg\min f \ni x^*$. Consider a sequence $(x_k)_k \in \mathbb{N}$ that satisfies, for all $k \in \mathbb{N}$,

$$x_{k+1} \in x_k - \alpha_k \partial f(x_k),$$

where the $\alpha_k \geqslant 0$ are the step sizes. Now we will try to design some nice step sizes α_k that guarantee that the sequence $(x_k)_k$ converges to a minimiser of f. From the above example, it is hopeless to examine the function values $f(x_{k+1})$ and $f(x_k)$ and try to control their difference; instead, we will try to control the difference of the x_i from a minimiser of f. By Proposition 6.1, we know that there always exists a subgradient if the domain is the whole space E, so there exists some $s_k \in \partial f(x_k)$, so that

$$\begin{aligned}
\|x_{k+1} - x^*\|^2 &= \|x_k - \alpha_k s_k - x^*\|^2 \\
&= \|x_k - x^*\|^2 - 2\alpha_k \langle s_k, x_k - x^* \rangle + \alpha_k^2 \|s_k\|^2 \\
&\leqslant \|x_k - x^*\|^2 - 2\alpha_k \left(f(x_k) - f(x^*) \right) + \alpha_k^2 \|s_k\|^2,
\end{aligned}$$

where we have used the fact that $s_k \in \partial f(x_k)$, so that $f(x^*) \geqslant f(x_k) + \langle s_k, x^* - x_k \rangle$. Now we want to achieve a contraction property, $\|x_{k+1} - x^*\|^2 \leqslant \|x_k - x^*\|^2$. The term $2\alpha_k \left(f(x_k) - f(x^*) \right)$ is negative, since $f(x_k) \geqslant f(x^*)$ and $\alpha_k \geqslant 0$, so we only need to control the term $\alpha_k^2 \|s_k\|^2$. Following the method due to Polyak for the step size, let the function $h(\alpha_k) \coloneqq -2\alpha_k \left(f(x_k) - f(x^*) \right) + \alpha^2 \|s_k\|^2$. When some α_k minimises h, the derivative of h at α_k is zero, so that

$$\frac{d}{d\alpha_k} h(\alpha_k) = -2 \left(f(x_k) - f(x^*) \right) + 2\alpha_k \|s_k\|^2 = 0 \implies \alpha_k = \frac{f(x_k) - f(x^*)}{\|s_k\|^2}.$$

when h is minimised, and this is well defined since $\|s_k\|^2 > 0$ by the definition of the norm and the fact that $s_k \neq 0$ by the definition of the subdifferential. With this step size, we have that

$$\begin{aligned}
\|x_{k+1} - x^*\|^2 &\leqslant \|x_k - x^*\|^2 - 2\alpha_k \left(f(x_k) - f(x^*) \right) + \alpha_k^2 \|s_k\|^2 \\
&= \|x_k - x^*\|^2 - \frac{\left(f(x_k) - f(x^*) \right)^2}{\|s_k\|^2},
\end{aligned}$$

which is the contraction property we were looking for since the term $\left(f(x_k) - f(x^*) \right)^2 / \|s_k\|^2$ is nonnegative. Thus we have that

$$\frac{\left(f(x_k) - f(x^*) \right)^2}{\|s_k\|^2} \leqslant \|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2,$$

which builds up a telescoping sum, so that provided $s_i \leqslant L$ for all $i \in \mathbb{N}$ for some $L > 0$, we have

$$\begin{aligned}
\frac{(k+1) \left(\min\limits_{i=0,\ldots,k} f(x_i) - f(x^*) \right)^2}{L^2} &\leqslant \sum_{i=0}^{k} \frac{\left(\min\limits_{j=0,\ldots,i} f(x_j) - f(x^*) \right)^2}{L^2} \\
&\leqslant \sum_{i=0}^{k} \frac{\left(\min\limits_{j=0,\ldots,i} f(x_j) - f(x^*) \right)^2}{\|s_i\|^2}
\end{aligned}$$

$$
\begin{aligned}
&\leqslant \sum_{i=0}^{k} \frac{(f(x_i) - f(x^*))^2}{\|s_i\|^2} \\
&\leqslant \|x_0 - x^*\|^2,
\end{aligned}
$$

where L is a global lower bound for the subdifferential. Hence we get that

$$\min_{i=0,\dots,k} f(x_i) - f(x^*) \leqslant \frac{L\|x_0 - x^*\|}{\sqrt{k+1}},$$

and so we have a result on the rate of convergence of the sequence $(x_k)_k$ to a minimiser of f. Are we okay with this result? Well, we have a rate of convergence, but we have no idea what the optimiser x^* is, so we have no idea what the optimal value $f(x^*)$ and consequently rate of convergence is in general. However, for some problems we do know the optimal value. For example, suppose we are trying to solve some linear system for which we want to minimise the ℓ_1 residue, that is, we want to solve the problem

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_1,$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. This is a nonsmooth convex problem, and we do know that the optimal value is zero, even if we don't know the optimal x^*. In some other cases, we tweak the step size by obtaining a (possibly learned-from-data) bound on f and then tweaking the step size to be proportional to the bound. This is the idea behind the *backtracking line search* method and others, which we will not discuss.

Instead we try a different approach for the step size. Rewrite the inequality from the contraction step as

$$2\alpha_k\left(f(x_k) - f(x^*)\right) \leqslant \|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 + \alpha_k^2\|s_k\|^2,$$

so that by a telescoping sum, we get

$$\sum_{i=0}^{k} 2\alpha_i\left(f(x_i) - f(x^*)\right) \leqslant \|x_0 - x^*\|^2 + \sum_{i=0}^{k} \alpha_i^2\|s_i\|^2.$$

Since

$$\sum_{i=0}^{k} 2\alpha_i\left(f(x_i) - f(x^*)\right) \geqslant \left(\min_{i=0,\dots,k} f(x_i) - f(x^*)\right) \sum_{i=0}^{k} 2\underbrace{\alpha_i}_{\geqslant 0},$$

we have that

$$\min_{i=0,\dots,k} f(x_i) - f(x^*) \leqslant \frac{\|x_0 - x^*\|^2 + \sum_{i=0}^{k} \alpha_i^2\|s_i\|^2}{2\sum_{i=0}^{k}\alpha_i} \leqslant \frac{\|x_0 - x^*\|^2 + L^2\sum_{i=0}^{k}\alpha_i^2}{2\sum_{i=0}^{k}\alpha_i} \to 0,$$

as $k \to \infty$, provided that $\sum_{i=0}^{k} \alpha_i \to \infty$ and $\sum_{i=0}^{k} \alpha_i^2 < \infty$, e.g. for $\alpha_k = 1/(k+1)$, which is the harmonic series starting from 0.

If we take constant step size $\alpha_k = \alpha$ for all $k \in \mathbb{N}$, then we have that

$$\min_{i=0,\dots,k} f(x_i) - f(x^*) \leqslant \frac{\|x_0 - x^*\|^2 + (k+1)\alpha^2L^2}{2(k+1)\alpha} \to \frac{\alpha L^2}{2},$$

as $k \to \infty$, so that if our convergence criterion $(\|x_0 - x^*\|^2 + (k+1)\alpha^2L^2)/(2(k+1)\alpha) < \varepsilon$, for fixed $\varepsilon > 0$, then the convergence rate is $O(1/\varepsilon)$.

Summarising all the steps above, we have the following algorithm for minimising a convex function.

> *Algorithm*: Subgradient method
>
> - Choose a starting point $x_0 \in \operatorname{dom} f$ and a sequence of step sizes $(\alpha_k)_k \in \mathbb{N}$ such that $\alpha_k \geqslant 0$ for all $k \in \mathbb{N}$ and $\sum_{k=0}^{\infty} \alpha_k = \infty$ and $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$.
> - For each $k \in \mathbb{N}$, do the following until convergence:
>   - Choose $s_k \in \partial f(x_k)$.
>   - Set $x_{k+1} \in x_k - \alpha_k s_k$.
> - Return the minimiser $f(x_k)$.

We will need more tools to analyse convex optimisation algorithms, so we will now return to the arena of convex functions and their properties.

§6.4. Regularity of convex functions

We will now discuss the regularity of convex functions, that is, the properties of convex functions that allow us to control their behaviour. We will start with the following definition.

Definition 6.15 (Lipschitz continuity). *A function $f : E \to \mathbb{R}$ is called* Lipschitz continuous *with constant $L > 0$ if for all $x, y \in E$, we have*

$$|f(x) - f(y)| \leqslant L\|x - y\|.$$

A function is called locally Lipschitz continuous *if it is Lipschitz continuous on every compact subset of its domain.*

The following fact should be familiar from calculus.

Proposition 6.16. *Let $f : E \to \mathbb{R}$ be proper convex. Then the following are equivalent if E is open:*

(i) f is Lipschitz continuous.

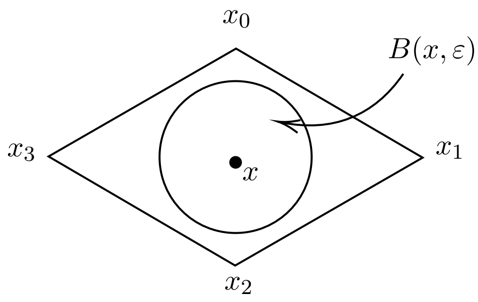
(ii) ∂f is uniformly bounded, i.e. there exists $L > 0$ such that for all $x \in E$, we have $\|x^*\| \leqslant L$ for all $x^* \in \partial f(x)$.

Proof. Homework exercise. (See Appendix A for a proof.) □

In the analysis from the previous section, we assumed that the subdifferential of the function was uniformly bounded by some $L > 0$, which is really the same as assuming that the subdifferential is globally Lipschitz continuous. We will now discuss the following result about proper convex functions.

Theorem 6.1. *If $f \in \operatorname{conv} E$ is proper, then f is locally Lipschitz on the relative interior of its domain,* $\operatorname{ri}(\operatorname{dom} f)$.

Proof. Let $x \in \operatorname{ri}(\operatorname{dom} f) \neq \emptyset$. By the equivalence of norms, there exists some radius $r > 0$ such that $\{x' : \|x - x'\|_1 \leqslant r\} \cap \operatorname{aff}\operatorname{dom} f \subseteq \operatorname{dom} f$, by the definition of the relative interior using the ℓ_1 ball instead of the Euclidean ball. In fact, there exist points $x_0, \ldots, x_m \in \operatorname{dom} f$ such that $x \in \operatorname{ri}(\operatorname{co}\{x_0, \ldots, x_m\})$ where m is the dimension of the affine hull of the domain of f, *i.e.* $m = \dim(\operatorname{aff}\operatorname{dom} f)$. Hence there exists $\varepsilon > 0$ such that $B(x, \varepsilon) \cap \operatorname{aff}\operatorname{dom} f \subseteq \operatorname{co}\{x_0, \ldots, x_m\}$. We will show that for a point x¡ there is a neighbourhood of x on which f is Lipschitz continuous:



Define the infimum

$$I \coloneqq \inf_{B(x,\varepsilon) \cap \operatorname{aff}\operatorname{dom} f} f > -\infty,$$

since f admits an affine lower bound on the compact set $B(x, \varepsilon) \cap \operatorname{aff}\operatorname{dom} f$ (for example, at x by Proposition 6.1). Similarly we define the supremum

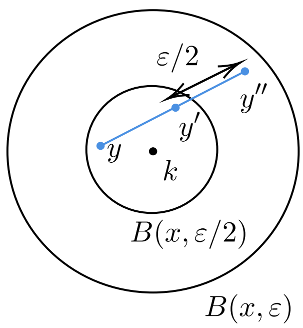
$$S \coloneqq \sup_{B(x,\varepsilon) \cap \operatorname{aff}\operatorname{dom} f} f < \infty,$$

since for all $x' \in \operatorname{co}\{x_0, \ldots, x_m\}$, we have by Jensen's inequality that

$$f(x') = f\left(\sum_{i=0}^{m} t_i x_i\right) \leqslant \sum_{i=0}^{m} t_i f(x_i) \leqslant \max\{f(x_0), \ldots, f(x_m)\},$$

where $t \in \Delta_m$ is in the (probability) simplex.

Now let $\Delta \coloneqq S - I$. We will show that f is Lipschitz continuous on $B(x, \varepsilon)$ with constant $2\Delta/\varepsilon$:



To establish connections between the values, we look at some point y'' which is controlled by S and I and guarantee by convexity that that point doesn't vary too much.

Consider $y, y' \in B(x, \varepsilon/2) \cap \operatorname{aff}(\operatorname{dom} f)$. Let y'' be such that, as the figure above suggests,

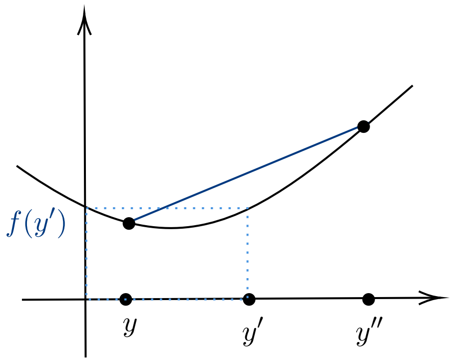
$$y'' = y' + \frac{\varepsilon}{2} \cdot \frac{y - y'}{\|y - y'\|}$$

$$= \left(1 + \frac{\varepsilon}{2\|y - y'\|}\right) y' - \frac{\varepsilon}{2\|y - y'\|} y$$
$$= \left(\frac{2\|y - y'\| + \varepsilon}{2\|y - y'\|}\right) y' - \frac{\varepsilon}{2\|y - y'\|} y.$$

Moving things around, we get that

$$y' = \frac{\varepsilon}{2\|y' - y\| + \varepsilon} \cdot y + \frac{2\|y' - y\|}{2\|y' - y\| + \varepsilon} \cdot y''.$$

Indeed, we can upper bound $f(y')$ using the fact that $f(y'')$ is bounded:



Applying the convexity of f to the point y',

$$\begin{aligned} f(y') &\leqslant \frac{\varepsilon}{2\|y' - y\| + \varepsilon} \cdot f(y) + \frac{2\|y' - y\|}{2\|y' - y\| + \varepsilon} \cdot f(y'') \\ &\leqslant f(y) + \frac{2\|y' - y\|}{2\|y' - y\| + \varepsilon} \cdot (f(y'') - f(y)) \\ &\leqslant f(y) + \frac{2\|y' - y\|}{2\|y' - y\| + \varepsilon} \cdot \Delta \\ &\leqslant f(y) + \frac{2\|y' - y\|}{\varepsilon} \cdot \Delta. \end{aligned}$$

Therefore, for all $y, y' \in B(x, \varepsilon/2)$, we have that

$$f(y) - f(y') \leqslant \frac{2\Delta}{\varepsilon} \|y - y'\|$$

and by a name change,

$$f(y') - f(y) \leqslant \frac{2\Delta}{\varepsilon} \|y' - y\|,$$

and so

$$|f(y) - f(y')| \leqslant \frac{2\Delta}{\varepsilon} \|y - y'\|.$$

Hence f is Lipschitz continuous on $B(x, \varepsilon/2)$ with constant $2\Delta/\varepsilon$, and so f is locally Lipschitz continuous on $\mathrm{ri}(\mathrm{dom}\, f)$. $\square$

§6.5. Operations on convex functions

We will now discuss some operations on convex functions that preserve convexity.

Proposition 6.17. *Let $\{f_i\}_{i\in I}$ be a family of functions $E \to \overline{\mathbb{R}}$, where $E \subseteq \mathbb{R}^n$ is a convex set. Then*

$$\operatorname{epi}\left(\sup_{i\in I} f_i\right) = \bigcap_{i\in I} \operatorname{epi}(f_i).$$

Thus $\sup_{i\in I} f_i$ is convex (resp. closed) if f_i is convex (resp. closed) for all $i \in I$.

Proof. Exercise. (See Appendix A for a proof.) □

Definition 6.18 (Marginal function)**.** *Let E and F be vector spaces and $\varphi\colon E \times F \to \overline{\mathbb{R}}$ be a coupling function. The* marginal function *$f\colon E \to \overline{\mathbb{R}}$ is defined by*

$$f(x) := \inf_{y\in F} \varphi(x, y).$$

The marginal function is a generalisation of the projection of a set onto a subspace, and it is a useful tool in optimisation, particularly in the area of proximal methods.

Proposition 6.19. *The marginal function $f\colon E \to \overline{\mathbb{R}}$ is convex if the coupling function $\varphi\colon E\times F \to \overline{\mathbb{R}}$ is convex.*

Proof. We can write:

$$\begin{aligned}
\operatorname{epi}_s f &= \{(x, \alpha) \in E \times \mathbb{R} : f(x) < \alpha\} \\
&= \left\{(x, \alpha) \in E \times \mathbb{R} : \inf_{y\in F} \varphi(x, y) < \alpha\right\} \\
&= \{(x, \alpha) \in E \times \mathbb{R} : \exists\, y \in F \text{ such that } \varphi(x, y) < \alpha\}\,.
\end{aligned}$$

Consider now the set

$$\operatorname{epi}_s \varphi = \{(x, y, \alpha) \in E \times F \times \mathbb{R} : \varphi(x, y) < \alpha\},$$

and note that

$$\operatorname{epi}_s f = \pi_{E\times\mathbb{R}}(\operatorname{epi}_s \varphi),$$

where $\pi_{E\times\mathbb{R}}$ defined the linear map that is the canonical projection onto the first and third coordinates:

$$\begin{cases} \pi_{E\times\mathbb{R}}\colon E \times F \times \mathbb{R} \to E \times \mathbb{R}, \\ (x, y, \alpha) \mapsto (x, \alpha). \end{cases}$$

Since the projection of a convex set is convex, we have that $\operatorname{epi}_s f$ is convex, and so f is convex. □

Proposition 6.20. *The subdifferential of the marginal function $f\colon E \to \overline{\mathbb{R}}$ is given by*

$$\partial f(x) = \{x^* \in E : (x^*, 0) \in \partial\varphi(x, y_x)\},$$

if φ is convex and the infimum is reached at some y_x for all $x \in \operatorname{dom} E$.

Proof. By definition,

$$\begin{aligned}
x^* \in \partial f(x) &\iff f(x') \geqslant f(x) + \langle x^*, x' - x\rangle && \text{for all } x' \in E \\
&\iff \varphi(x', y') \geqslant \varphi(x, y_x) + \langle x^*, x' - x\rangle && \text{for all } x' \in E, y' \in F \\
&\iff \varphi(x', y') \geqslant \varphi(x, y_x) + \langle (x^*, 0), (x', y') - (x, y_x)\rangle && \text{for all } x' \in E, y' \in F \\
&\iff (x^*, 0) \in \partial\varphi(x, y_x),
\end{aligned}$$

where the last equivalence follows from the definition of the subdifferential. □

This idea of going from functions to sets and back to functions is a common theme in convex analysis, and it will frequently appear as we continue our study of convex optimisation. In fact the marginal is the most general of all of the operations we have discussed so far, and it is a very useful tool in proximal algorithms. We will now discuss the infimal convolution of two functions.

The infimal convolution Take functions $f, g \colon E \to \mathbb{R} \cup \{\infty\}$. The *infimal convolution* of f and g is the function $f \,\Box\, g \colon E \to \overline{\mathbb{R}}$ defined for all $x \in E$ by

$$(f \,\Box\, g)(x) \coloneqq \inf_{y \in E}\{f(y) + g(x - y)\} = \inf\{f(y) + g(z) : y + z = x\}.$$

This is a bit of an odd notion; the following proposition helps us understand what the inf convolution corresponds to: the inf convolution is somewhat tantamount to adding two epigraphs together.

Proposition 6.21. *If $f, g \colon E \to \mathbb{R} \cup \{\infty\}$ are proper convex functions, then the infimal convolution $f \,\Box\, g$ has strict epigraph $\operatorname{epi}_s(f \,\Box\, g) = \operatorname{epi}_s f + \operatorname{epi}_s g$.*

In particular, if f and g are convex, then $f \,\Box\, g$ is convex.

Proof. Exercise. (See Appendix A for a proof.) □

> **Example 6.22.** Consider the distance to a convex set $C \subseteq E$:
>
> $$d_C(x) \coloneqq \inf_{y \in C} \|x - y\| = \inf_{y \in E}\{\delta_C(y) + \|x - y\|\} = (\delta_C \,\Box\, \|\cdot\|)(x),$$
>
> where δ_C is the indicator function of C. Thus this is convex, since the infimal convolution of a proper convex function and a norm is convex.

§7. Lecture 07—01st March, 2024

§7.1. Closure of a convex function

Definition 7.1 (Closure)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. The* closure *of f is defined as*

$$\overline{f} = \begin{cases} f(x) & \text{if } x \in \operatorname{dom} f \\ +\infty & \text{otherwise} \end{cases}$$

where dom f *is the domain of* f.

The following proposition suggests more of the specific flavour of the closure of a convex function that we are interested in.

Proposition 7.2. *Let $f\colon E \to \overline{\mathbb{R}}$ be a convex function. Then $\overline{\operatorname{epi} f}$ is the epigraph of a function $\overline{f}\colon E \to \overline{\mathbb{R}}$.*

Proof. We give a sketch of a proof. If $(x,\alpha) \in \overline{\operatorname{epi} f}$ and $\beta > \alpha$, then $(x,\beta) \in \overline{\operatorname{epi} f}$—this is because if $(x,\beta) \notin \overline{\operatorname{epi} f}$, then there exists a sequence $(x_k,\alpha_k) \in \operatorname{epi} f$ such that $(x_k,\alpha_k) \to (x,\beta)$ and $\alpha_k \to \beta$. Let $(x,\alpha) \in \overline{\operatorname{epi} f}$. Then there exists a sequence $(x_k,\alpha_k) \in \operatorname{epi} f$ such that $(x_k,\alpha_k) \to \overline{(x,\alpha)}$ and $\alpha_k \to \alpha$. For k large enough, $\alpha_k \leqslant \beta$, and therefore $(x_k,\beta) \in \overline{\operatorname{epi} f}$ converges to $(x,\beta) \in \overline{\operatorname{epi} f}$.

If $(x,\alpha_k) \in \overline{\operatorname{epi} f}$ and $\alpha_k \searrow \alpha$, then $(x,\alpha) \in \overline{\operatorname{epi} f}$. This holds because $\overline{\operatorname{epi} f}$ is closed and $\alpha_k \searrow \alpha$ implies that $(x,\alpha) \in \overline{\operatorname{epi} f}$. $\square$

Proposition 7.3. *Let $f \in \operatorname{conv} E$. Then the unique function $\overline{f}\colon E \to \overline{\mathbb{R}}$ is the supremum of all the affine lower bounds of f. That is,*

$$\overline{f}(x) = \sup\{g(x) : g \text{ is affine and } g \leqslant f\}$$

Proof. Let $\{a_i\}_{i\in I}$ be the set of all affine lower bounds of f. We will show that $\overline{f}(x) = \sup_{i\in I} a_i$, that is,

$$\operatorname{epi} \overline{f} = \bigcap_{i\in I} \operatorname{epi} a_i.$$

Since $\operatorname{epi} f$ is convex, by Corollary 4.7, the epigraph of the closure of f, namely $\overline{\operatorname{epi} f} = \operatorname{epi} \overline{f}$ is the intersection of all closed half-spaces of $E \times \mathbb{R}$ containing $\operatorname{epi} f$. Since $\operatorname{epi} a_i$ is precisely a closed half-space of $E \times \mathbb{R}$ containing $\operatorname{epi} f$, we have that $\operatorname{epi} \overline{f} \subseteq \bigcap_{i\in I} \operatorname{epi} a_i$, and consequently $\overline{\operatorname{epi} f} \subseteq \bigcap_{i\in I} \operatorname{epi} a_i$.

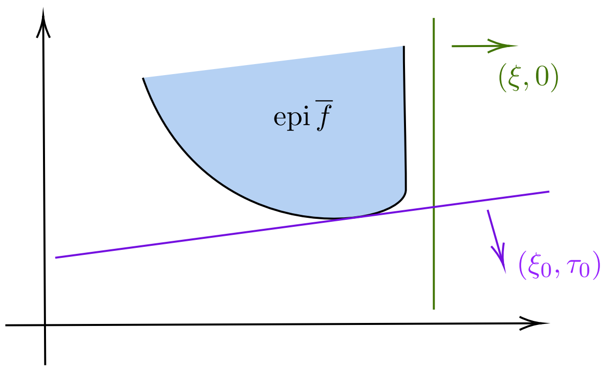
So we only need to prove the more difficult inclusion,

$$\bigcap_{i\in I} \operatorname{epi} a_i \subseteq \operatorname{epi} \overline{f}.$$

A half-space of $E \times \mathbb{R}$ is the epigraph of an affine function if and only if it is the epigraph of a lower semi-continuous function. In particular it is defined by $(\xi,\tau,t) \in \mathbb{E} \times \mathbb{R} \times \mathbb{R}$, denoted

$$H^-((\xi,\tau),t) = \{(x,\alpha) : \langle \xi,\tau\rangle + \tau\alpha \leqslant t\},$$

where ξ denotes the normal vector to the hyperplane, τ is a point on the hyperplane, and t is the distance from the origin to the hyperplane.



Now we can't find an affine lower bound for $H^-((\xi,\tau),t)$, so we have to come up with better ideas to complete the proof of this inclusion. Since the halfspace $H^-((\xi,\tau),t)$ contains $\operatorname{epi} f$, we have that for all $(x,\alpha) \in \operatorname{epi} f$, the inequality $\langle \xi, x\rangle + \tau\alpha \leqslant t$ holds. Then by the same reasoning as we previously used, it must be that $\tau \leqslant 0$, since otherwise $\alpha \to +\infty$ would contradict the inequality. Now if $\tau < 0$, then the halfspace $H^-((\xi,\tau),t)$ is in fact the epigraph of the affine lower bound of f defined by

$$x \mapsto \left\langle -\frac{\xi}{t}, x \right\rangle + \frac{t}{\tau},$$

by taking $\alpha = f(x)$ in the inequality $\langle \xi, x\rangle + \tau\alpha \leqslant t$.

Now what about halfspaces of the form $H^-(\xi, 0, t)$? Since $f \in \operatorname{conv} E$, it admits at least one affine lower bound whose epigraph is a closed half-space $H^-(\xi_0, \tau_0, t_0)$ with $\tau_0 < 0$. So if indeed we can show[2] that

$$H^-(\xi,0,t) \cap H^-(\xi_0,\tau_0,t_0) = \bigcap_{\rho \geqslant 0} H^-(\xi_0 + \rho\xi, \underbrace{\tau_0}_{<0}, t_0 + \rho t),$$

then we are done, since we would have expressed the intersection of all closed half-spaces containing $\operatorname{epi} f$ as the intersection of $\operatorname{epi} a_i$ for for all $i \in I' \subseteq I$, that is,

$$\overline{\operatorname{epi} f} = \bigcap_{i \in I'} \operatorname{epi} a_i \supseteq \bigcap_{i \in I} \operatorname{epi} a_i,$$

as desired. Indeed, $\langle \xi, x\rangle \leqslant t$ and $\langle \xi_0, x\rangle + \tau_0 f(x) \leqslant t$ hold if and only if $\langle \xi_0 + \rho\xi, x\rangle + \tau_0 f(x) \leqslant t_0 + \rho t$ for all $\rho \geqslant 0$. □

§7.2. Conjugation and conjugates

The basic idea behind duality in convex analysis is to view a (closed) convex set C as an intersection of half spaces. Applying this idea to the epigraph of a convex function f suggests that we should view f as a supremum of affine functions.

An affine minorant of f is a function $x \mapsto \langle m, x\rangle - b$ such that

$$f(x) \geqslant \langle m, x\rangle - b \quad \text{for all } x.$$

The vector m is called the "slope" of the affine minorant. Typically f has many affine minorants with a given slope m, corresponding to different values of the scalar b. We only care about the *best* affine minorant with slope m—in other words, we only care about the *best* scalar b. So our question is as follows: For a given m, which value of b is the "best"? Which value of b makes the inequality above as tight as possible?

Notice that

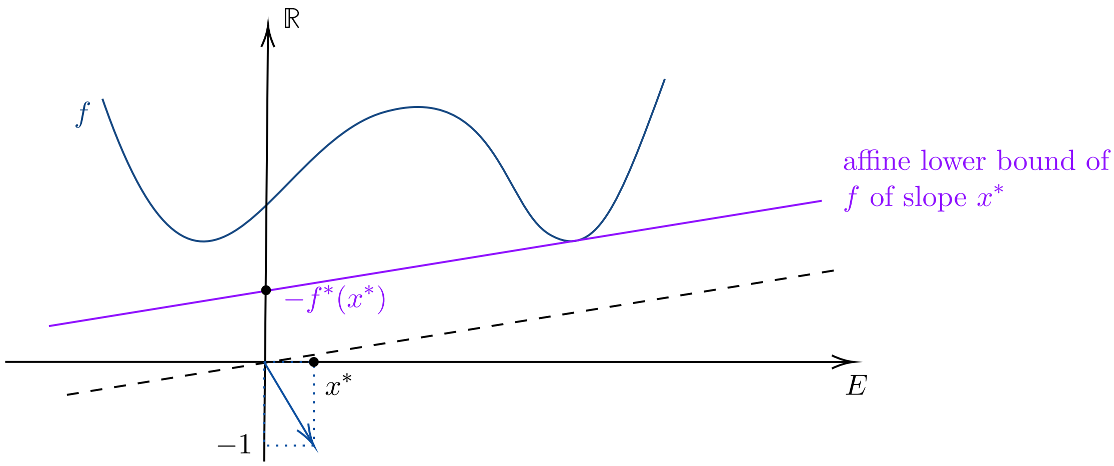
$$\begin{aligned} f(x) \geqslant \langle m, x\rangle - b \quad \text{for all } x &\iff b \geqslant \langle m, x\rangle - f(x) \quad \text{for all } x \\ &\iff b \geqslant \sup_x \langle m, x\rangle - f(x) = f^*(m). \end{aligned}$$

This shows that the best choice of b is $f^*(m)$. We have just discovered the convex conjugate f^*. The whole point of f^* is that it tells us how to view f as a supremum of affine functions. You give f^* a slope m, and it gives you the best choice of b.

[2]This was left as an exercise in lecture.

Definition 7.4 (Fenchel conjugate)**.** *Let* $f: E \to \overline{\mathbb{R}}$ *be a function. The* Fenchel conjugate *of* f *is the function* $f^*: E \to \overline{\mathbb{R}}$ *defined by*

$$\begin{aligned} f^*: E &\to \overline{\mathbb{R}} \\ x^* &\mapsto \sup_{x \in E}\{\langle x^*, x\rangle - f(x)\}. \end{aligned}$$



> **Example 7.5.** Let $f: \mathbb{R} \to \mathbb{R}$ be the function $f(x) = |x|$. Then
>
> $$f^*(x^*) = \sup_{x \in \mathbb{R}}\{x^* x - |x|\}.$$
>
> We can compute this supremum by considering the cases $x \geqslant 0$ and $x < 0$. If $x \geqslant 0$, then $x^*x - |x| = x^*x - x = x(x^* - 1)$. If $x < 0$, then $x^*x - |x| = x^*x + x = x(x^* + 1)$. So
>
> $$f^*(x^*) = \begin{cases} +\infty & \text{if } x^* > 1 \\ 0 & \text{if } x^* = 1 \\ 0 & \text{if } x^* = -1 \\ -\infty & \text{if } x^* < -1 \end{cases}$$

> **Example 7.6.** Let $S \subseteq E$. Then the indicator function of S is the function $\delta_S: E \to \overline{\mathbb{R}}$ defined by
>
> $$\delta_S(x) = \begin{cases} 0 & \text{if } x \in S \\ +\infty & \text{otherwise} \end{cases}$$
>
> Then
>
> $$\delta_S^*(x^*) = \sup_{x \in E}\{\langle x^*, x\rangle - \delta_S(x)\} = \sup_{x \in S}\langle x^*, x\rangle = \sigma_S(x^*).$$
>
> So the Fenchel conjugate of the indicator function of a set is itself the support function of the set.

The above example begs the question: what is the Fenchel conjugate of the support function? This is not a trivial question; we need to establish some more properties of the Fenchel dual before we can answer it.

Proposition 7.7. *The conjugate f^* of a function $f\colon E \to \overline{\mathbb{R}}$ is convex and closed (i.e. its epigraph is closed). In addition, the following properties hold:*

1. *f is not identically $+\infty$ if and only if the conjugate $f^* > -\infty$ if and only if the conjugate f^* is not identically $-\infty$.*
2. *f has an affine lower bound if and only if f^* is not identically $+\infty$.*
3. *f is proper and has an affine lower bound if and only if f^* is proper if and only if $f^* \in \overline{\operatorname{conv}}\, E$.*

Proof. This proposition is easy to prove; we leave it as an exercise. (See Appendix A for a proof.) □

Proposition 7.8. *The biconjugate $f^{**} = (f^*)^*$ of a function $f\colon E \to \overline{\mathbb{R}}$ is convex and closed, and f is proper and has an affine lower bound if and only if $f^{**} \in \operatorname{conv} E$.*

Proof. If f is proper and has an affine lower bound, then f^* is proper (by part 3 of the previous proposition). But f^* has an affine lower bound, since, for all $x_0 \in \operatorname{dom} f$, we have that

$$f^*(x^*) \geqslant \langle x^*, x_0\rangle - f(x_0) \text{ for all } x^* \in E.$$

Thus $f^* \in \operatorname{conv} E$, and so $f^{**} = (f^*)^* \in \operatorname{conv} E$ by part 3 of the previous proposition. Conversely, if $f^{**} \in \operatorname{conv} E$, then $f^* \in \operatorname{conv} E$, and so f^* is proper (again by part 3 of the previous proposition). But f^* has an affine lower bound, and so f is proper and has an affine lower bound (by part 3 of the previous proposition). □

Proposition 7.9. *Let $f\colon E \to \overline{\mathbb{R}}$ be proper and admit an affine lower bound. Then:*

1. *The biconjugate f^{**} is the supremum of all the affine lower bounds of f.*
2. *$f \in \operatorname{conv} E$ if and only if $\overline{f} = f^{**}$.*
3. *$f \in \overline{\operatorname{conv}}\, E$ if and only if $f = f^{**}$.*
4. *The Legendre-Fenchel transform $f \mapsto f^*$ is a bijection on $\overline{\operatorname{conv}}\, E$.*

Proof.

1. Let $x \in E$. The value at x of the supremum of all affine lower bounds of f is equal to

$$\begin{aligned}\sup_{\substack{x^*\in E,\alpha\in\mathbb{R}\\ \langle x^*,y\rangle-\alpha\leqslant f(y) \text{ for all } y\in E}} \langle x^*, x\rangle - \alpha &= \sup_{x^*\in E} \sup_{\substack{\alpha\in\mathbb{R}\\ \langle x^*,y\rangle-\alpha\leqslant f(y) \text{ for all } y\in E}} \langle x^*, x\rangle - \alpha \\ &= \sup_{x^*\in E} \langle x^*, x\rangle - f^*(x^*) \\ &= f^{**}(x),\end{aligned}$$

   since $f^*(x^*) = \sup_{y\in E}\{\langle x^*, y\rangle - f(y)\}$.

2. In this case the proof is straightforward. If $f \in \operatorname{conv} E$, then $\overline{f}$ is the supremum of all affine lower bounds of f by Proposition 7.3. This is to say, f^{**} according to the previous point. Conversely, if $\overline{f} = f^{**}$, then $f \in \operatorname{conv} E$ by Proposition 7.3.

3. If $f \in \overline{\operatorname{conv}}\, E$, then we have that

$$\operatorname{epi} f = \overline{\operatorname{epi}}\, f \qquad \text{since } \operatorname{epi} f \text{ is closed}$$

$$
\begin{aligned}
&= \operatorname{epi} \overline{f} && \text{by the definition of the closure} \\
&= \operatorname{epi} f^{**} && \text{by the previous point.}
\end{aligned}
$$

Thus $f = f^{**}$. Conversely, Proposition 7.8 tells us that if $f = f^{**}$, then $f \in \overline{\operatorname{conv}}\, E$.

4. Let $\mathscr{C}$ be the Legendre-Fenchel transform. $\mathscr{C}: f \mapsto f^*$ is defined on $\overline{\operatorname{conv}}\, E$ and takes values in $\overline{\operatorname{conv}}\, E$ by Proposition 7.7. Now $\mathscr{C}$ is injective, since if $f_1, f_2 \in \overline{\operatorname{conv}}\, E$ and $f_1^* = f_2^*$, then $f_1 = f_1^{**} = f_2^{**} = f_2$. $\mathscr{C}$ is also surjective, since if $g \in \overline{\operatorname{conv}}\, E$, then $\mathscr{C}(g^*) = g^{**} = g$ by the third point above, and thus g is the image of g^* under $\mathscr{C}$. Thus $\mathscr{C}$ is a bijection on $\overline{\operatorname{conv}}\, E$. $\square$

Example 7.10. Let us now return to the question of the Fenchel conjugate of the support function. Let $S \subseteq E$ be a nonempty convex set. Then the support function of S is the function $\sigma_S\colon E \to \overline{\mathbb{R}}$ defined by

$$\sigma_S(x) = \sup_{y \in S} \langle x, y \rangle.$$

Then,

$$(\sigma_S)^* \overset{\text{(i)}}{=} (\sigma_{\overline{\operatorname{co}}\, S})^* \overset{\text{(ii)}}{=} (\delta^*_{\overline{\operatorname{co}}\, S})^* \overset{\text{(iii)}}{=} \delta^{**}_{\overline{\operatorname{co}}\, S} \overset{\text{(iv)}}{=} \delta_{\overline{\operatorname{co}}\, S},$$

where (i) follows from Proposition 6.7, (ii) follows from Example 7.6, (iii) follows from Proposition 7.9 and the fact that $\delta_{\overline{\operatorname{co}}\, S} \in \overline{\operatorname{conv}}\, E$, and (iv) follows from Proposition 7.9.

We now introduce the notion of the dual norm.

Definition 7.11. *Let $(E, \|\cdot\|)$ be a normed vector space equipped with a norm product $\langle\cdot,\cdot\rangle$. (Here we may have $\|\cdot\| \neq \sqrt{\langle\cdot,\cdot\rangle}$.) The* dual norm *of $\|\cdot\|$ is the function $\|\cdot\|_*\colon E \to \mathbb{R}$ such that for all $x^* \in E$,*

$$\|x^*\|_* = \sup_{\|x\| \leqslant 1} \langle x^*, x\rangle = \sigma_{\mathbb{B}(0,1)}(x^*) = \sigma_{\mathbb{S}(0,1)}(x^*) = \sup_{\|x\|=1} \langle x^*, x\rangle.$$

We end today with an example.

Example 7.12. What is $(\|\cdot\|)^*$? Is it $\|\cdot\|_*$? The answer is no. We have that

$$
\begin{aligned}
(\|\cdot\|)^*(x^*) &= \sup_{x \in E} \langle x^*, x\rangle - \|x\| \\
&= \sup_{\alpha \geqslant 0} \sup_{\|x\|=\alpha} \langle x^*, x\rangle - \alpha \\
&= \max\left\{0, \sup_{\alpha>0} \alpha \left(\sup_{\|x\|=\alpha} \left\langle x^*, \frac{x}{\|x\|} \right\rangle - 1\right)\right\} \\
&= \max\left\{0, \sup_{\alpha>0} \alpha \left(\sup_{\|x\|=1} \langle x^*, x\rangle - 1\right)\right\} \\
&= \max\left\{0, \sup_{\alpha>0} \alpha\|x^*\|_* - \alpha\right\} \\
&= \max\left\{0, \sup_{\alpha>0} \alpha(\|x^*\|_* - 1)\right\} \\
&= \sup_{\alpha \geqslant 0} \alpha(\|x^*\|_* - 1)
\end{aligned}
$$

$$
\begin{aligned}
&= \begin{cases} 0 & \text{if } \|x^*\|_* \leqslant 1 \\ +\infty & \text{if } \|x^*\|_* > 1 \end{cases} \\
&= \delta_{\mathbb{B}^*(0,1)}(x^*),
\end{aligned}
$$

where $\mathbb{B}^*(0,1)$ is the closed unit ball of the dual space E^* with respect to the dual norm $\|\cdot\|_*$; indeed $(\|\cdot\|)^* = \delta_{\mathbb{B}^*(0,1)}$.
As a consequence,

$$
(\|\cdot\|) = ((\|\cdot\|)^*)^* = \left(\delta_{\mathbb{B}^*(0,1)}\right)^* = \sigma_{\mathbb{B}^*(0,1)} = \|\cdot\|_{**},
$$

since $\delta_{\mathbb{B}^*(0,1)} \in \overline{\operatorname{conv}}\, E^*$ by Proposition 7.9.

§8. Lecture 08—22nd March, 2024

§8.1. More properties of the Fenchel conjugate

We begin with this proposition that illustrates how the Fenchel conjugate links together disparate notions that we have already seen regarding operaitons on functions.

Let E be a real vector space.

Proposition 8.1. *If the functions $f, g\colon E \to \overline{\mathbb{R}}$, then the Fenchel conjugate of the infimal convolution of f and g is the sum of the Fenchel conjugates of f and g. That is, $(f \,\square\, g)^* = f^* + g^*$.*

Proof. The proof is fairly direct:

$$
\begin{aligned}
(f \,\square\, g)^*(x^*) &= \sup_{x\in E} \langle x^*, x\rangle - (f \,\square\, g)(x) \\
&= \sup_{x\in E} \left(\langle x^*, x\rangle - \inf_{y\in E} f(y) + g(x-y) \right) \\
&= \sup_{x\in E} \left(\langle x^*, x\rangle + \sup_{y\in E} -f(y) - g(x-y) \right) \\
&= \sup_{x\in E} \left(\sup_{y\in E} \langle x^*, x\rangle - f(y) - g(x-y) \right) \\
&= \sup_{y\in E} \left(\sup_{x\in E} \langle x^*, x\rangle - f(y) - g(x-y) \right) \\
&= \sup_{y\in E} \left(-f(y) + \sup_{x\in E} \langle x^*, x\rangle - g(x-y) \right) \\
&= \sup_{y\in E} \left(-f(y) + \langle x^*, y\rangle + \sup_{x\in E} \langle x^*, x-y\rangle - g(x-y) \right) \\
&= \sup_{y\in E} \left(-f(y) + \langle x^*, y\rangle + \underbrace{\sup_{z\in E} \langle x^*, z\rangle - g(z)}_{g^*(x^*)} \right) \\
&= \sup_{y\in E} \left(-f(y) + \langle x^*, y\rangle + g^*(x^*) \right) \\
&= f^*(x^*) + g^*(x^*),
\end{aligned}
$$

since the supremum of a sum is the sum of the suprema. □

There is a trip back (*i.e.* in the opposite direction), but we will not prove this:

Proposition 8.2. *If $f, g\colon E \to \overline{\mathbb{R}}$ are proper, convex, and lower semicontinuous, then the Fenchel conjugate of the sum of f and g is the infimal convolution of the Fenchel conjugates of f and g. That is, $(f+g)^* = f^* \,\square\, g^*$.*

Proof. The proof is left as an exercise. (See Appendix A for a proof.) □

Proposition 8.3. *Let $f \in \operatorname{conv} E$, let $x \in \operatorname{dom} f$, and let $x^* \in E$. Then the following are equivalent:*

(i) For all $d \in E$, the directional derivative of f in the d-direction satisfies $f'(x, d) \geqslant \langle x^*, d \rangle$.

(ii) The point $x^* \in \partial f(x)$.

(iii) The point $x \in \arg\min_{y \in E} f(y) - \langle x^*, y \rangle = \arg\max_{y \in E} \langle x^*, y \rangle - f(y)$.

(iv) (Fenchel-Young inequality.) The sum $f(x) + f^*(x^*) \leqslant \langle x^*, x \rangle$.

(v) The sum $f(x) + f^*(x^*) = \langle x, x^* \rangle$. (This relationship, particularly it's equivalence with (ii) *is important and is known as the* Fenchel-Young duality.*)*

Proof. We show all six implications.

(i) ⇒ (ii) We have that

$$\begin{aligned}
\langle x^*, y - x \rangle \leqslant f'(x, y - x) &= \lim_{t \searrow 0} \frac{f(x + t(y - x)) - f(x)}{t} \\
&= \lim_{t \searrow 0} \frac{f((1 - t)x + ty) - f(x)}{t} \\
&\leqslant \lim_{t \searrow 0} \frac{(1 - t)f(x) + tf(y) - f(x)}{t} \\
&= \lim_{t \searrow 0} \frac{tf(y) - tf(x)}{t} \\
&= f(y) - f(x),
\end{aligned}$$

which implies that $f(y) \geqslant f(x) + \langle x^*, y - x \rangle$, and so $x^* \in \partial f(x)$.

(ii) ⇒ (iii) For all $y \in E$, we observe that

$$f(y) \geqslant f(x) + \langle x^*, y - x \rangle \implies f(y) - \langle x^*, y \rangle \geqslant f(x) - \langle x^*, x \rangle,$$

and so x is a minimiser of $f(y) - \langle x^*, y \rangle$.

(iii) ⇒ (iv) For all $y \in E$, we have that

$$\begin{aligned}
\langle x^*, x \rangle - f(x) &\geqslant \langle x^*, y \rangle - f(y) \\
&\geqslant \sup_{y \in E} \langle x^*, y \rangle - f(y) \\
&= f^*(x^*),
\end{aligned}$$

and so $f(x) + f^*(x^*) \leqslant \langle x^*, x \rangle$. For the argument maximum, we have that

$$\begin{aligned}
\langle x^*, x \rangle - f(x) &\geqslant \langle x^*, y \rangle - f(y) \\
&\geqslant \sup_{y \in E} \langle x^*, y \rangle - f(y) \\
&= f^*(x^*),
\end{aligned}$$

and so $f(x) + f^*(x^*) \geqslant \langle x^*, x \rangle$.

(iv) ⇒ (v) This is immediate from the definition of the Fenchel conjugate. In particular, we have that

$$f(x) + f^*(x^*) \leqslant \langle x^*, x \rangle \implies f(x) + f^*(x^*) = \langle x^*, x \rangle,$$

since the Fenchel conjugate is $f^*(x^*) = \sup_{x \in E} \langle x^*, x \rangle - f(x)$.

(v) $\Rightarrow$ (i) We have that for all $y \in E$,

$$\langle x^*, x \rangle - f(x) = f^*(x^*) \geqslant \langle x^*, y \rangle - f(y).$$

Taking $y = x + td$ for some $d \in E$ and $t \in \mathbb{R}_+$, we get

$$\begin{aligned} f(x+td) - f(x) \geqslant \langle x^*, x+td-x \rangle = t\,\langle x^*, d \rangle &\iff \frac{f(x+td)-f(x)}{t} \geqslant \langle x^*, d \rangle \\ &\iff f'(x,d) \geqslant \langle x^*, d \rangle\,, \end{aligned}$$

by taking the limit as $t \searrow 0$.

Thus, all five statements are equivalent. □

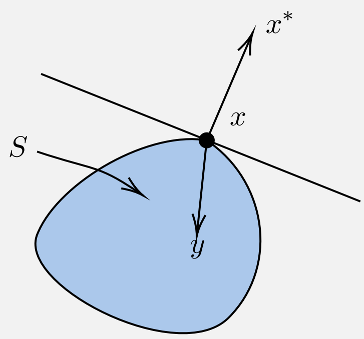
Example 8.4. We will now use these equivalences, particularly the Fenchel-Young duality, to solve the problem of determining $\partial\|\cdot\|(0)$, the subdifferentiability of the norm at the origin. We have that

$$\begin{aligned} x^* \in \partial\|\cdot\|(0) &\iff \|\cdot\|(0) + \|\cdot\|^*(x^*) = \langle 0, x^* \rangle \\ &\iff 0 + \delta_{B^*(0,1)}(x^*) = 0 \\ &\iff x^* \in B^*(0,1), \end{aligned}$$

since $\partial\|\cdot\|(0) = B^*(0,1)$ where $B^*(0,1) = \{x^* \in E \mid \|x^*\| \leqslant 1\}$ is the dual unit ball.

Example 8.5. Suppose now that we wanted to find $\partial\delta_S(x)$ for $x \in S \subseteq E$. Recalling that $\partial\delta_S(x) = N_x S$, we have

$$\begin{aligned} x^* \in \partial\delta_S(x) &\iff \delta_S(x) + \delta_S^*(x^*) = \langle x, x^* \rangle \\ &\iff 0 + \delta_{S^*}(x^*) = \langle x, x^* \rangle \\ &\iff \langle x^*, y \rangle \leqslant \langle x, x^* \rangle \quad \forall y \in S^* \\ &\iff \langle x^*, y - x \rangle \leqslant 0 \quad \forall y \in S^* \\ &\iff x^* \in N_x S. \end{aligned}$$



Corollary 8.6. *1. If $f \in \operatorname{conv} E$, then $x^* \in \partial f(x)$ implies that $x \in \partial f^*(x^*)$.*

2. If $f \in \overline{\operatorname{conv}}\, E$, then $x^* \in \partial f(x)$ if and only if $x \in \partial f^*(x^*)$; in other words, $\boxed{(\partial f)^{-1} = \partial f^*}$.

Proof. 1. If $x^* \in \partial f(x)$, we have from the Fenchel-Young duality that $f(x) + f^*(x^*) = \langle x, x^* \rangle$, and so $f(x) \leqslant \langle x, x^* \rangle - f^*(x^*)$. This implies that $x \in \partial f^*(x^*)$.

2. We have that

$$\begin{aligned} x^* \in \partial f(x) &\iff f(x) + f^*(x^*) = \langle x, x^* \rangle \\ &\iff f^*(x^*) + f^{**}(x) = \langle x, x^* \rangle \\ &\iff x \in \partial f^*(x^*), \end{aligned}$$

where the last equivalence follows from the first part of the corollary. □

We also have the following two propositions:

Proposition 8.7. *Let $f_i \colon E \to \mathbb{R} \cup \{\infty\}$ be convex, I compact, and take every $x \in \operatorname{dom} f$. With $i \mapsto f_i(x)$ upper semicontinuous, where the function $f(x) = \max_{i \in I} f_i(x)$, then for all $x \in (\operatorname{dom} f)^\circ$, we have that*

$$\partial f(x) = \operatorname{co}\left(\bigcup_{i \in I^\circ(x)} \partial f_i(x) \right),$$

where $I^\circ(x) = \{i \in I : f_i(x) = f(x)\}$. In particular, if I is finite, then

$$\partial f(x) = \operatorname{co}\left(\partial f_1(x) \cup \partial f_2(x) \cup \cdots \cup \partial f_n(x)\right).$$

Proof. The proof is left as an exercise. (See Appendix A for a proof.) □

> **Example 8.8.** Let $f(x) = \max\{f_1(x), f_2(x), f_3(x)\}$, then with $f_1(x) = f_3(x) = f(x)$ with $f_2(x) < f(x)$, we have that $\partial f(x) = \operatorname{co}\left(\partial f_1(x) \cup \partial f_3(x)\right)$.

The following proposition gives us a direct relationship between the directional derivative and the subdifferential of a function.

Proposition 8.9 (Maximum formula)**.** *If $f \in \operatorname{conv} E$ and $x \in \operatorname{ri} \operatorname{dom} f$, then for all $d \in E$, we have that*

$$f'(x, d) = \sup_{x^* \in \partial f(x)} \langle x^*, d \rangle .$$

The supremum is attained if $f'(x, d) < \infty$.

Proof. The proof is left as an exercise. (See Appendix A for a proof.) □

§8.2. Optimality conditions: Fritz-John and Karush-Kuhn-Tucker

We begin with a fairly famous result in convex analysis.

Proposition 8.10 (Fermat's rule)**.** *Let $f \in \operatorname{conv} E$ be an extended real-valued function. Then $x^* \in \arg\min_{x \in E} f(x) \iff 0 \in \partial f(x^*)$.*

Proof. We have that

$$\begin{aligned} x^* \in \arg\min_{x\in E} f(x) &\iff f(x^*) + \langle 0, x - x^* \rangle \leqslant f(x) \quad \forall x \in E \\ &\iff 0 \leqslant f(x) - f(x^*) \quad \forall x \in E \\ &\iff 0 \in \partial f(x^*), \end{aligned}$$

and we're done. □

Proposition 8.11. *Let $f \in \operatorname{conv} E$ and $C \subseteq E$ be a convex set such that $\operatorname{ri}\operatorname{dom} f \cap \operatorname{ri} C \neq \emptyset$. Then $x^* \in \arg\min_{x\in C} f(x)$ if and only if, there exists $g \in \partial f(x^*)$, such that $-g \in N_{x^*}C$.*

Proof. Note that

$$\inf_{x\in C} f(x) = \inf_{x\in E} f(x) + \delta_C(x),$$

and so we can use the previous result to write the optimality condition as

$$\begin{aligned} x^* \in \arg\min_{x\in C} f(x) &\iff 0 \in \partial\left(f + \delta_C\right)(x^*) \\ &\iff 0 \in \partial f(x^*) + \partial\delta_C(x^*) \\ &\iff 0 \in \partial f(x^*) - N_{x^*}C, \end{aligned}$$

and so $x^* \in \arg\min_{x\in C} f(x)$ if and only if there exists $g \in \partial f(x^*)$ such that $-g \in N_{x^*}C$. □

This proposition can be thought of as a generalisation of the Fermat's rule to the case where the minimum is taken over a convex set.

We now move on to more general optimality conditions.

Theorem 8.1 (Fritz John necessary optimality condition)**.** *Let $f, g_1, \ldots, g_m \colon E \to \mathbb{R}$ be convex. Then*

$$x^* \in \arg\min\{f(x) \colon g_i(x) \leqslant 0, i = 1, \ldots, m\}$$

implies that there exists $(\lambda_0, \ldots, \lambda_m) \in \mathbb{R}_+^{m+1} \setminus \{0\}$ such that

$$\begin{cases} 0 \in \lambda_0 \partial f(x^*) + \sum_{i=1}^m \lambda_i \partial g_i(x^*), \\ \lambda_i g_i(x^*) = 0 \quad \forall\, i = 1, \ldots, m. \end{cases}$$

Proof. The proof follows quite readily from what we've seen so far. Rewriting the problem as an unconstrained optimisation problem,

$$x^* \in \arg\min_{x\in E} \max\{\underbrace{f(x) - f^\star}_{:= g_0(x)}, g_1(x), \ldots, g_m(x)\},$$

where $f^\star := \inf\{f(x) : g_i(x) \leqslant 0, i = 1, \ldots, m\}$. So the existence of this minimiser implies, by Fermat's rule, that

$$0 \in \partial\left(\max\{g_0, g_1, \ldots, g_m\}\right)(x^*) \iff 0 \in \operatorname{co}\left(\bigcup_{i\in I^\circ(x^*)} \partial g_i(x^*)\right) \qquad \text{by Proposition 8.7}$$

$$\iff 0 \in \lambda_0 \partial f(x^*) + \sum_{i=1}^{m} \lambda_i \partial g_i(x^*),$$

for some $\lambda_i \in \mathbb{R}_+$, $\lambda_0 \in \mathbb{R}_+$ with $\lambda_0 + \sum_{i=1}^m \lambda_i = 1$. Thus the Fritz John necessary optimality condition holds (each $\lambda_i \geqslant 0$). $\square$

Theorem 8.2 (Karush-Kuhn-Tucker necessary optimality condition)**.** *Let $f, g_1, \ldots, g_m \colon E \to \mathbb{R}$ be convex and satisfy Slater's condition: for all $\overline{x} \in E$, $g_i(\overline{x}) < 0$ for all $i = 1, \ldots, m$. Then*

$$x^* \in \arg\min\{f(x) : g_i(x) \leqslant 0, i = 1, \ldots, m\}$$

implies that there exists $(1, \lambda_1, \ldots, \lambda_m) \in \mathbb{R}_+^{m+1} \setminus \{0\}$ such that

$$\begin{cases} 0 \in \lambda_0 \partial f(x^*) + \sum_{i=1}^m \lambda_i \partial g_i(x^*), \\ \lambda_i g_i(x^*) = 0 \quad \forall\, i = 1, \ldots, m \end{cases}$$

where $\lambda_0 = 1$, are sufficient (of course, necessary) optimality conditions.

Proof. We reason by contradiction and set $\lambda_0 = 0$. Then with $\xi_i \in \partial g_i(x^*)$, we have that

$$\begin{array}{rc} & \sum_{i=1}^{m} \lambda_i \xi_i = 0 \\ + & \Big(g_i(x^*) + \langle \xi_i, \overline{x} - x^* \rangle \leqslant g_i(\overline{x}) \Big) \cdot \lambda_i \\ \hline & 0 = \underbrace{\sum_{i=1}^{m} \lambda_i g_i(x^*)}_{=0} + \Big\langle \underbrace{\sum_{i=1}^{m} \lambda_i \xi_i}_{=0}, \overline{x} - x^* \Big\rangle \leqslant \sum_{i=1}^{m} \underbrace{\lambda_i}_{\geqslant 0} \underbrace{g_i(\overline{x})}_{<0} < 0, \end{array}$$

which is a contradiction since $(\lambda_1, \ldots, \lambda_m) \neq (0, \ldots, 0)$ and $g_i(\overline{x}) < 0$.

Now we focus on sufficiency. Define

$$h(x) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x),$$

where $\lambda_i \geqslant 0$ and $\lambda_0 = 1$. Then we have that

$$\partial h(x^*) = \partial f(x^*) + \sum_{i=1}^{m} \lambda_i \partial g_i(x^*) \ni 0,$$

and so $x^* \in \arg\min_{x \in E} h(x)$ exactly when

$$f(x^*) + \underbrace{\sum_{i=1}^{m} \lambda_i g_i(x^*)}_{=0} \leqslant f(x) + \underbrace{\sum_{i=1}^{m} \underbrace{\lambda_i}_{\geqslant 0} \underbrace{g_i(x)}_{\leqslant 0}}_{\leqslant 0} \leqslant f(x^*),$$

for all $x \in E$. Thus x^* is a minimiser of $f(x)$ subject to the feasibility constraint $g_i(x) \leqslant 0$. $\square$

So here's the overview summary: we saw that the Fritz John necessary optimality condition is a generalisation of the Fermat's rule to the case of optimisation over a convex set, and the Karush-Kuhn-Tucker necessary optimality condition is a generalisation of the Fritz John condition to the case where Slater's condition holds.

§8.3. Duality I: minimax duality

Let X be a set and let $f\colon X \to \overline{\mathbb{R}}$ be a function. Consider the problem Π defined as

$$\Pi = \inf_{x\in X} f(x).$$

Let $\mathsf{val}(\Pi)$ denote the optimal value and $\mathsf{sol}(\Pi)$ denote the set of optimal solutions. In order to introduce a dual problem, we seek to write f as

$$f(x) = \sup_{y\in Y} \varphi(x,y),$$

where $\varphi\colon X \times Y \to \overline{\mathbb{R}}$ is a coupling function and Y is a set. With this, we may write the problem also as

$$\Pi = \inf_{x\in X} \sup_{y\in Y} \varphi(x,y),$$

and its dual as

$$D_\Pi = \sup_{y\in Y} \inf_{x\in X} \varphi(x,y) = -\inf_{y\in Y} \delta(y),$$

where $\delta(y)$ is the dual function defined as

$$\begin{aligned} \delta\colon Y &\to \overline{\mathbb{R}}, \\ y &\mapsto -\inf_{x\in X} \varphi(x,y). \end{aligned}$$

We are often interested in $\mathsf{val}(D_\Pi)$ and $\mathsf{sol}(D_\Pi)$, the optimal value and set of optimal solutions of the dual problem. We have the following proposition:

Proposition 8.12 (Weak duality)**.** *We have that*

$$\sup_{y\in Y} \inf_{x\in X} \varphi(x,y) \leqslant \inf_{x\in X} \sup_{y\in Y} \varphi(x,y).$$

Proof. By definition, for all $(x',y') \in X \times Y$, we have that

$$\inf_{x\in X} \varphi(x,y') \leqslant \varphi(x',y').$$

Thus, we get

$$\sup_{y\in Y} \inf_{x\in X} \varphi(x,y) \leqslant \sup_{y\in Y} \varphi(x',y) \leqslant \inf_{x\in X} \sup_{y\in Y} \varphi(x,y),$$

and so the weak duality holds. □

§8.4. The proximal operator

The proximal operator comes out naturally from much of our prior discussions.

Let $(E, \langle \cdot, \cdot \rangle)$ be a Euclidean space with the induced norm $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$.

Definition 8.13 (Proximal mapping and Moreau envelope)**.** *Given $f \in \overline{\operatorname{conv}}\, E$, consider $p_f \colon E \rightrightarrows E$ and $\widetilde{f} \colon \mathbb{E} \to \mathbb{R}$ defined for all $x \in E$ by*

$$p_f(x) = \arg\min_{y \in E} f(y) + \frac{1}{2}\|y - x\|^2, \tag{$\clubsuit$}$$

$$\widetilde{f}(x) = \inf_{y \in E} f(y) + \frac{1}{2}\|y - x\|^2. \tag{$\dagger$}$$

The mapping ($\clubsuit$) is called the proximal mapping *of f, and ($\dagger$) is called the* Moreau envelope *of f.*

There is a natural connection between the Moreau envelope and a few of the previous notions we have discussed:

1. *Marginal.* $\widetilde{f} = \inf_{y \in E} \varphi(\cdot, y)$, where $\varphi(\cdot, y)$ is a coupling function.
2. *Inf-convolution.* $\widetilde{f} = (f \,\square\, \frac{1}{2}\| \cdot \|^2)(x) = \left(f^* \,\square\, \frac{1}{2}\| \cdot \|^2\right)^* (x)$.
3. *Conjugation.* $\widetilde{f} = \frac{1}{2}\| \cdot \|^2 - \left(f + \frac{1}{2}\| \cdot \|^2\right)^*$.

Example 8.14. We have that for the indicator function δ_S, the proximal mapping is the projection onto the closed convex set S; that is, $p_{\delta_S}(x) = p_S(x)$. To see this, note that:

$$\begin{aligned} p_{\delta_S}(x) &= \arg\min_{y \in E} \delta_S(y) + \frac{1}{2}\|y - x\|^2 \\ &= \arg\min_{y \in S} \frac{1}{2}\|y - x\|^2 \\ &= p_S(x), \end{aligned}$$

which is singleton-valued if S is nonempty closed convex, *i.e.* $\delta_S \in \overline{\operatorname{conv}}\, E$.

We can think of the proximal mapping as a projection of a point onto a function, unlike the regular projection which projects a point onto a set. For the case that the function is the indicator function of a set, the proximal mapping reduces to the regular projection onto the set, as a sort of "equality criterion".

Definition 8.15 (Strictly convex function)**.** *A function $f \colon E \to \mathbb{R} \cup \{\infty\}$ is said to be* strictly convex *if for all $x \neq y \in \operatorname{dom} f$ and for all $t \in (0, 1)$, we have that*

$$f(tx + (1 - t)y) < tf(x) + (1 - t)f(y).$$

The definition above helps us outline some properties of the proximal operator via the following proposition:

Proposition 8.16. *1. The proximal mapping p_f is singleton-valued if f is strictly convex.*

2. *Let $x, y \in E$. Then $y = p_f(x)$ if and only if $x - y \in \partial f(y)$ if and only if $y = (I + \partial f)^{-1}(x)$, for strictly convex f.*[3]

3. *If f is strictly convex, then $x \in \arg\min_E f$ if and only if $x = p_f(x)$.*

4. *The sum $p_f + p_{f^*} = I$, where I is the identity operator.*

Proof. 1. The mapping

$$E \ni y \mapsto f(y) + \frac{1}{2}\|y - x\|^2$$

is strictly convex and thus it admits at most one minimiser. Also, the mapping is coercive since

$$\lim_{\|y\| \to \infty} f(y) + \frac{1}{2}\|y - x\|^2 = \infty,$$

and so it admits at least one minimiser. Thus, the mapping is singleton-valued.

2. By Fermat's rule,

$$\begin{aligned} y = p_f(x) &\iff 0 \in \partial \left(f + \frac{1}{2}\| \cdot - x\|^2 \right)(y) \\ &\iff 0 \in \partial f(y) + y - x \\ &\iff y + \partial f(y) \ni x \\ &\iff (I + \partial f)(y) \ni x \\ &\iff y = (I + \partial f)^{-1}(x). \end{aligned}$$

3. Just as we might expect, we have

$$\begin{aligned} x \in \arg\min_E f &\iff 0 \in \partial f(x) \\ &\iff x - x \in \partial f(x) \\ &\iff x = p_f(x). \end{aligned}$$

4. Part 2 above and Corollary 8.6 lead us to reason that

$$\begin{aligned} y = p_f(x) &\iff x - y \in \partial f(y) \iff x - (x - y) \in \partial f^*(x - y) \\ &\iff x - y \in p_{f^*}(x), \end{aligned}$$

and so $p_f + p_{f^*} = I$ since p_f and p_{f^*} are inverses of each other. ☐

[3] The notation $(I + \partial f)^{-1}$ is called the *resolvent* of ∂f.

§9. Lecture 09—29th March, 2024

§9.1. Duality II: duality by perturbation

This is the most powerful viewpoint on duality, and it will heavily rely on properties that we have seen in the previous lectures[4]. Consider the primal problem

$$\inf_{x\in X} f(x), \tag{P}$$

and consider a perturbation function

$$\begin{aligned}\Phi\colon X\times\mathcal{P} &\to \overline{\mathbb{R}},\\ (x,p)&\mapsto \Phi(x,p),\end{aligned}$$

where $\mathcal{P}$ is some set of perturbations (a subset of Euclidean space E), such that $\Phi(x,0)=f(x)$ for all $x\in X$. Thus we can naturally define some primal problem indexed by some perturbation $p\in\mathcal{P}$:

$$\inf_{x\in X}\phi(x,p), \tag{P_p}$$

Consider the value function $v\colon\mathcal{P}\to\overline{\mathbb{R}}$ defined naturally as

$$\begin{aligned}v(p)&\coloneqq \inf_{x\in X}\Phi(x,p),\\ v(0)&=\inf_{x\in X}\Phi(x,0)=\inf_{x\in X} f(x)=\mathsf{val}(P).\end{aligned}$$

Assume that $\operatorname{dom} v\neq\emptyset$ and that v admits an affine lower bound. Then $v^*\in\overline{\operatorname{conv}}\,\mathcal{P}$. We can then define the dual form of the primal as

$$-\inf_{p^*\in\mathcal{P}} v^*(p^*), \tag{D}$$

This function is convex, and so it is reasonable to try to minimise it. We have the following two properties of this dual problem:

$$\mathsf{val}(D)=v^{**}(0),\qquad \mathsf{sol}(D)=\partial v^{**}(0).$$

Both of these are easy to check; indeed,

$$\mathsf{val}(D)=-\inf_{p^*\in\mathcal{P}} v^*(p^*)=\sup_{p^*\in\mathcal{P}}\langle 0,p^*\rangle - v^*(p^*)=v^{**}(0),$$

and

$$\begin{aligned}p\in\mathsf{sol}(D)&\iff p\in\arg\min_{p^*\in\mathcal{P}} v^*(p^*)-\langle 0,p^*\rangle\\ &\iff 0\in\partial v^*(p)\\ &\iff p\in\partial v^{**}(0),\end{aligned}$$

since $v^*\in\overline{\operatorname{conv}}\,\mathcal{P}$.

[4]Modulo some new perspectives.

Let us examine the dual function:

$$\begin{aligned}
v^*(p^*) &= \sup_{p\in\mathcal{P}} \langle p, p^*\rangle - v(p) \\
&= \sup_{p\in\mathcal{P}} \langle p, p^*\rangle - \inf_{x\in X} \Phi(x,p) \\
&= \sup_{p\in\mathcal{P}} \sup_{x\in X} \langle p, p^*\rangle - \Phi(x,p) \\
&= \sup_{(x,p)\in X\times\mathcal{P}} \langle (0,p^*), (x,p)\rangle - \Phi(x,p) \\
&= \Phi^*(0,p^*),
\end{aligned}$$

where Φ^* is the Fenchel conjugate of Φ. So we can rewrite the dual problem as

$$-\inf_{p^*\in\mathcal{P}} \Phi^*(0,p^*). \tag{D}$$

Now attribute a Lagrangian to the perturbation function thus:

$$\begin{aligned}
\ell_\Phi \colon X \times \mathcal{P} &\to \overline{\mathbb{R}}, \\
(x,p^*) &\mapsto -\sup_{p\in\mathcal{P}} \langle p^*, p\rangle - \Phi(x,p).
\end{aligned}$$

Based on the above,

$$\begin{aligned}
v^*(p^*) &= -\inf_{x\in X} \left\{ -\sup_{p\in\mathcal{P}} \langle p^*, p\rangle - \Phi(x,p) \right\} \\
&= -\inf_{x\in X} \ell_\Phi(x,p^*).
\end{aligned}$$

Now we present an analogous result as we had in the previous lecture in this setting with the Lagrangian:

Proposition 9.1 (Weak duality)**.** *We have that* $\mathsf{val}(P) \geqslant \mathsf{val}(D)$.

Proof. Proceeding straightforwardly,

$$\begin{aligned}
\mathsf{val}(D) &= -\inf_{p^*\in\mathcal{P}} v^*(p^*) = \sup_{p^*\in\mathcal{P}} -v^*(p^*) = \sup_{p^*\in\mathcal{P}} \inf_{x\in X} \ell_\Phi(x,p^*) \\
&\leqslant \inf_{x\in X} \sup_{p^*\in\mathcal{P}} \ell_\Phi(x,p^*) = \inf_{x\in X} \left\{ \sup_{p^*\in\mathcal{P}} \inf_{p\in\mathcal{P}} \Phi(x,p) - \langle p^*, p\rangle \right\} \\
&\leqslant \inf_{x\in X} \inf_{p\in\mathcal{P}} \sup_{p^*\in\mathcal{P}} \Phi(x,p) - \langle p^*, p\rangle = \inf_{x\in X} \Phi(x,0) \\
&= \inf_{x\in X} f(x) \\
&= \mathsf{val}(P),
\end{aligned}$$

as desired. $\square$

The idea here is that we are building the coupling function from perturbation, and all of the ideas follow right through. We can also derive a number of strong duality results:

Proposition 9.2 (Strong duality).

1. *If the value function $v \in \operatorname{conv} P$ and $0 \in \operatorname{dom} v$, we have:*
   a) *v is lower semicontinuous at 0 if and only if $\mathsf{val}(P) = \mathsf{val}(D)$.*
   b) *The subdifferential $\partial v(0) \neq \emptyset$ if and only if $\mathsf{val}(P) = \mathsf{val}(D)$ and it is true that whenever $\mathsf{sol}(D) \neq \emptyset$, we have $\mathsf{sol}(D) = \partial v(0)$.*
   c) *If $0 \in \operatorname{ri} \operatorname{dom} v$, then it holds that $\mathsf{val}(P) = \mathsf{val}(D)$ and $\mathsf{sol}(D) = \partial v(0) \neq \emptyset$.*
   d) *If $0 \in (\operatorname{dom} v)^\circ$, then $\mathsf{val}(P) = \mathsf{val}(D)$ and $\mathsf{sol}(D) = \partial v(0)$ is a compact nonempty set and v is Lipschitz continuous around 0.*
2. *If $v \in \overline{\operatorname{conv}} P$, then $\mathsf{val}(P) = \mathsf{val}(D) \in \overline{\mathbb{R}}$.*

Proof. The proof is left as an exercise. (See Appendix A for a proof.) □

§9.2. Duality III: Fenchel duality

This perspective on duality is particularly useful for composite optimisation problems. Consider the primal problem

$$\inf_{x \in E} f(x) + g(Ax), \tag{P}$$

where $f : E \to \overline{\mathbb{R}}$, $g : F \to \overline{\mathbb{R}}$, and $A : E \to F$ is a linear operator. We can define the value function as

$$\begin{aligned} v_F : F &\to \overline{\mathbb{R}}, \\ p &\mapsto \inf_{x \in E} f(x) + g(Ax + p), \end{aligned}$$

where $v_F(0) = \mathsf{val}(P)$.

We can determine a dual as follows:

$$\begin{aligned} v_F^*(\lambda) &= \sup_{p \in F} \langle \lambda, p \rangle - v_F(p) \\ &= \sup_{x \in E} \sup_{p \in F} \langle \lambda, p \rangle - f(x) - g(Ax + p) \\ &= \sup_{x \in E} -f(x) - \langle \lambda, Ax \rangle + \sup_{p \in F} \langle \lambda, Ax + p \rangle - g(Ax + p) \\ &= \sup_{x \in E} -f(x) - \langle \lambda, Ax \rangle - g^*(\lambda), \end{aligned}$$

where g^* is the Fenchel conjugate of g. So

$$\begin{aligned} v_F^*(\lambda) &= \sup_{x \in E} \langle -A^*\lambda, x \rangle - f(x) - g^*(\lambda) \\ &= f^*(-A^*\lambda) - g^*(\lambda), \end{aligned}$$

where A^* is the adjoint of A. Therefore we can define the dual problem as

$$\sup_{\lambda \in F} -f^*(-A^*\lambda) - g^*(\lambda). \tag{D_F}$$

Theorem 9.1. *For Euclidean spaces E, F, assume that $f \in \operatorname{conv} E$, $g \in \operatorname{conv} F$, $\mathsf{val}(P_F) \in \mathbb{R}$, and*

$$0 \in (\operatorname{dom} g - A(\operatorname{dom} f))^\circ ,$$

where $A\colon E \to F$ is a linear operator. Then we have that $\mathsf{val}(P_F) = \mathsf{val}(D_F)$ and $\mathsf{sol}(D_F)$ is a compact nonempty set.

Proof. The proof is left as an exercise. (See Appendix A for a proof.) □

§9.3. Duality IV: Lagrangian duality

Following the same discussion from above, take the problem $(P) \equiv (P_{X,EI})$ as

$$\inf_{x \in X} f(x) \text{ s.t. } \begin{cases} c_E(x) = 0, \\ c_I(x) \leqslant 0, \end{cases} \tag{$P_{X,EI}$}$$

where there are m_E equality constraints and m_I inequality constraints with $m = m_E + m_I$. We can define the duality via perturbation; we have the value function

$$v_0(p) = \inf_{\substack{x \in X \\ c_E(x)+p_E=0 \\ c_I(x)+p_I \leqslant 0}} f(x) = \inf_{x \in X} \Phi_0(x,p),$$

where the perturbation function is

$$\begin{aligned} \Phi_0\colon X \times \mathbb{R}^m &\to \overline{\mathbb{R}}, \\ (x,p) &\mapsto \begin{cases} f(x) & \text{if } c_E(x) + p_E = 0, c_I(x) + p_I \leqslant 0, \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

Consider the Lagrangian ℓ_0 associated to Φ_0:

$$\begin{aligned} \ell_0(x,\lambda) &= - \sup_{p=(p_E,p_I) \in \mathbb{R}^m} \langle \lambda, p \rangle - \Phi_0(x,p) \\ &= \sup_{\substack{p=(p_E,p_I) \\ p_E = -c_E(x) \\ p_I \leqslant -c_I(x)}} \langle \lambda, p \rangle - f(x) \\ &= \begin{cases} f(x) + \sum_{i \in E \cup I} \lambda_i c_i(x) & \text{if } \lambda_I \geqslant 0, \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

The following proposition characterises Lagrangian strong duality:

Proposition 9.3 (Lagrange strong duality)**.** *Suppose $v\colon \mathcal{P} \to \overline{\mathbb{R}}$ is convex with $v(0) = \mathsf{val}(P_{X,EI})$. Then we have that $\mathsf{val}(P_{X,EI}) = \mathsf{val}(D_{X,EI})$ if and only if v is lower semicontinuous at 0.*

Proof. The proof is left as an exercise. (See Appendix A for a proof.) □

Now consider the cases where we have functions $f, g_1, \dots, g_m$ all convex. Then v is convex, and in particular we have that v is lower semicontinuous at 0 if and only if there exists a Slater point, *i.e.* a point $x_s \in X$ such that $g_i(x_s) < 0$ for all i. For this cases we may strengthen the result with existence:

Proposition 9.4 (Lagrange strong duality with Slater point)**.** *Suppose $f, g_1, \ldots, g_m$ are convex functions with some point $x_s \in X$ such that $g_i(x_s) < 0$ for all i, where*

$$v(p) = \inf_{x \in X} f(x) + \sum_{i=1}^{m} p\lambda_i g_i(x).$$

Then we have that $\mathsf{val}(P_{X,EI}) = \mathsf{val}(D_{X,EI})$ and $\mathsf{sol}(D_{X,EI}) \neq \emptyset$ whenever we have finite optimal solutions for the primal problem.

Proof. The proof is left as an exercise. (See Appendix A for a proof.) $\square$

§9.4. Example: semidefinite optimisation

One of the concrete applications of the duality theory we have developed is in semidefinite optimisation. Semidefinite optimisation is a generalisation of linear optimisation where we consider the problem

$$\inf_{X \in \mathbb{S}^n} \langle C, X \rangle, \text{ s.t. } \begin{cases} A(X) = b, \\ X \succeq 0, \end{cases} \tag{$P_{\mathbb{S}^n}$}$$

where $C \in \mathbb{S}$, the mapping $A \colon \mathbb{S}^n \to \mathbb{R}^m$ is linear, and $b \in \mathbb{R}^m$ with $X \in \mathbb{S}^n$ and $\langle U, V \rangle = \mathsf{trace}\left(U^\top V\right)$. By the Riesz representation theorem, there exist $A_1, \ldots, A_m \in \mathbb{S}^n$ such that for all $X \in \mathbb{S}^n$,

$$A(X) = \begin{pmatrix} \langle A_1, X \rangle \\ \vdots \\ \langle A_m, X \rangle \end{pmatrix},$$

and we can write the adjoint $A^* \colon \mathbb{R}^m \to \mathbb{S}^n$ satisfying $\langle A(X), y \rangle = \langle X, A^* y \rangle$ for all $X \in \mathbb{S}^n$ as

$$A^* y = \sum_{i=1}^{m} y_i A_i.$$

Let us now discover the dual versions of the primal problem $(P_{\mathbb{S}^n})$ according to two duality perspectives we have discussed.

Minimax duality We can take the coupling function $\varphi(X, y) = \langle C, X \rangle + \langle y, b - A(X) \rangle$ and get that the primal problem is

$$\begin{aligned} \inf_{X \succeq 0} \langle C, X \rangle + \delta_{\{A(X)=b\}}(X) &= \inf_{X \succeq 0} \langle C, X \rangle + \sup_{y \in \mathbb{R}^m} \langle y, b - A(X) \rangle \\ &= \inf_{X \succeq 0} \sup_{y \in \mathbb{R}^m} \langle C, X \rangle + \langle y, b - A(X) \rangle, \end{aligned}$$

in which case the dual itself is

$$\begin{aligned} \sup_{y \in \mathbb{R}^m} \inf_{X \succeq 0} \langle C, X \rangle + \langle y, b - A(X) \rangle &= \sup_{y \in \mathbb{R}^m} \langle b, y \rangle + \inf_{X \succeq 0} \langle C, X \rangle - \langle A^* y, X \rangle \\ &= \sup_{y \in \mathbb{R}^m} \langle b, y \rangle - \inf_{X \succeq 0} \langle C - A^* y, X \rangle \\ &= \sup_{y \in \mathbb{R}^m} \langle b, y \rangle \text{ s.t. } A^* y \leqslant C, \end{aligned}$$

and there is a clear nicer dependence.

Duality by perturbation Here we take the value function

$$v(p) = \inf_{\substack{X \succeq 0 \\ A(X)=b+p}} \langle C, X \rangle ,$$

so that

$$\begin{aligned} v^*(y) &= \sup_{p \in \mathbb{R}^m} \left\{ \langle y, p \rangle - \inf_{X \succeq 0} \langle C, X \rangle + \delta_{\{A(X)=b+p\}}(X) \right\} \\ &= \sup_{X \succeq 0} - \langle C, X \rangle + \sup_{p \in \mathbb{R}^m} \langle y, p \rangle - \delta_{\{A(X)-b=p\}}(X) \\ &= \sup_{X \succeq 0} - \langle C, X \rangle + \delta^*_{\{A(X)-b\}}(p), \\ &= \sup_{X \succeq 0} - \langle C, X \rangle + \sup_{y \in \{A(X)-b\}} \langle y, p \rangle , \end{aligned}$$

since the Fenchel conjugate of the indicator function is the support function. Thus we get that

$$\begin{aligned} v^*(y) &= \sup_{X \succeq 0} - \langle C, X \rangle + \langle y, A(X) - b \rangle \\ &= - \langle b, y \rangle - \inf_{X \succeq 0} \langle C - A^* y, X \rangle \\ &= \begin{cases} - \langle b, y \rangle & \text{if } A^* y \leqslant C, \\ +\infty & \text{otherwise,} \end{cases} \end{aligned}$$

and we then recover the dual problem as

$$- \inf_{p^* \in \mathcal{P}} v^*(p^*) = \sup_{y \in \mathbb{R}^m} - \langle b, y \rangle \text{ s.t. } C - A^* y \succeq 0.$$

Proposition 9.5. *If* $\mathsf{sol}(P_{\mathbb{S}^n}) \neq \emptyset$ *and there exsists* $y \in \mathbb{R}^m$ *such that* $C - A^* y \succeq 0$*, then* $\mathsf{val}(P_{\mathbb{S}^n}) = \mathsf{val}(D_{\mathbb{S}^n})$*. If there exists* $X \succeq 0$ *such that* $A(X) = b$ *and* $\mathsf{sol}(D) = \emptyset$ *and* A *is surjective, then* $\mathsf{val}(P_{\mathbb{S}^n}) = \mathsf{val}(D_{\mathbb{S}^n})$*.*

Proof. The proof is left as an exercise. (See Appendix A for a proof.) □

Proposition 9.6. *Assume that there exists* $(\overline{X}, \overline{y})$ *such that* $\overline{X} \succeq 0$*,* $A(\overline{X}) = b$*, and* $C - A^* \overline{y} \succeq 0$*. Then* (X, y, S) *is a primal-dual solution if and only if*

$$\begin{cases} A^*(y) + S = C, & S \succeq 0, \\ A(X) = b, & X \succeq 0, \\ \langle X, S \rangle = 0, \end{cases}$$

where the dual problem is

$$\sup_{(y,S) \in \mathbb{R}^m \times \mathbb{S}^n} \langle b, y \rangle \text{ s.t. } A^*(y) + S = C, S \succeq 0.$$

Proof. The proof is left as an exercise. (See Appendix A for a proof.) □

This Schur complement is quite useful, so we elaborate on it here:

Lemma 9.7 (Schur complement)**.** *If A is invertible, then for any $B \in \mathbb{R}^{n\times m}$ and $C \in \mathbb{R}^{m\times m}$, we have*

$$\begin{pmatrix} A & B \\ B^\top & C \end{pmatrix} = \begin{pmatrix} I & 0 \\ -B^\top A^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & C - B^\top A^{-1} B \end{pmatrix} \begin{pmatrix} I & -A^{-1}B \\ 0 & I \end{pmatrix}.$$

In particular, if $A \succeq 0$, then

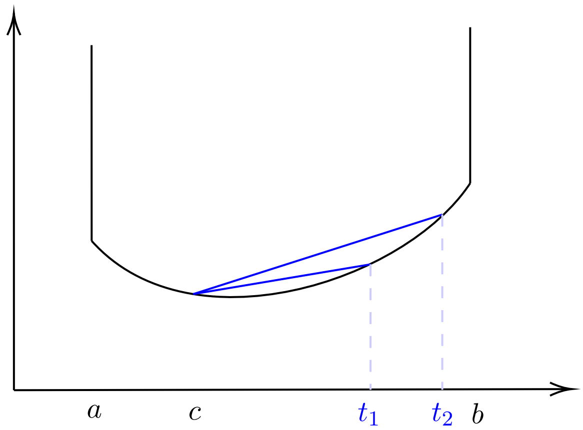
$$\begin{pmatrix} A & B \\ B^\top & C \end{pmatrix} \succeq 0 \iff C - B^\top A^{-1} B \succeq 0.$$

We call $C - B^\top A^{-1} B$ the Schur complement of A in the matrix $\begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}$.

§9.5. Strong convexity

We start with the following result, which harkens back to Theorem 6.1:

Lemma 9.8. *If $f \in \overline{\operatorname{conv}}\,\mathbb{R}$, then $f \in C^\circ(\operatorname{dom} f)$.*



Proof of Lemma 9.8. From the diagram above, we get

$$\frac{f(t_1) - f(c)}{t_1 - c} \leqslant \frac{f(t_2) - f(c)}{t_2 - c} \leqslant \frac{f(b) - f(c)}{b - c}. \tag{$\clubsuit$}$$

Indeed,

$$\lim_{t \nearrow b} \frac{f(t) - f(c)}{t - c} \leqslant \frac{f(b) - f(c)}{b - c},$$

and thus

$$\lim_{t \nearrow b} f(t) \leqslant f(c) + (b - c) \cdot \frac{f(b) - f(c)}{b - c} = f(b),$$

and $\lim_{t\nearrow b} f(t) \geqslant f(b)$, since lower semicontinuous f means that $\lim_{\varepsilon \searrow 0} \inf_{y \in B(x,\varepsilon)} f(y) \geqslant f(x)$. Thus $\lim_{t \searrow b} f(t) = f(b)$, and so f is continuous at b. It then remains to show ($\clubsuit$) for $t_1, t_2 \in \operatorname{dom} f$. Rewrite it as

$$f(t_1) - f(c) \leqslant \frac{t_1 - c}{t_2 - c} \cdot (f(t_2) - f(c)),$$

which may be equivalently written as

$$f(t_1) \leqslant \frac{t_1 - c}{t_2 - c} f(t_2) + \left(1 - \frac{t_1 - c}{t_2 - c}\right) f(c).$$

Therefore we may obtain the convex combination for t_1 as

$$t_1 = \frac{t_1 - c}{t_2 - c} t_2 + \left(1 - \frac{t_1 - c}{t_2 - c}\right) c,$$

and so f is convex. $\square$

We now introduce some strengthening of convexity:

Definition 9.9 (Strong convexity). *A function $f \colon E \to \mathbb{R} \cup \{+\infty\}$ is said to be μ-strongly convex if $\mu \geqslant 0$ and for all $x, y \in \operatorname{dom} f$ and for all $t \in [0, 1]$,*

$$f(tx + (1-t)y) \leqslant tf(x) + (1-t)f(y) - \frac{\mu}{2} t(1-t)\|x - y\|^2,$$

where the norm $\|\cdot\|$ is generic.

We have the following generalisation of the fundamental theorem of calculus from the definition of the subdifferential:

Lemma 9.10. *Let $f \in \overline{\operatorname{conv}}\, \mathbb{R}$ and let $[a, b] \subset \operatorname{dom} f$ with $a < b$, then*

$$f(b) - f(a) = \int_a^b h(t)\, \mathrm{d}t,$$

where $h \colon (a, b) \to \mathbb{R}$ satisfies $h(t) \in \partial f(t)$ for all $t \in (a, b)$.

Proof. The proof is left as an exercise. (See Appendix A for a proof.) $\square$

Now let $\operatorname{dom} \partial f \coloneqq \{x \in E : \partial f(x) \neq \emptyset\}$.

Theorem 9.2. *Let $f \in \overline{\operatorname{conv}}\, E$ and $\mu \geqslant 0$. The following are equivalent:*

(i) f is μ-strongly convex.

(ii) For all $x \in \operatorname{dom} \partial f$, for all $y \in \operatorname{dom} f$, and for all $g \in \partial f(x)$,

$$f(y) \geqslant f(x) + \langle g, y - x \rangle + \frac{\mu}{2}\|y - x\|^2.$$

(iii) For all $x, y \in \operatorname{dom} \partial f$, for all $g_x \in \partial f(x)$, and for all $g_y \in \partial f(y)$,

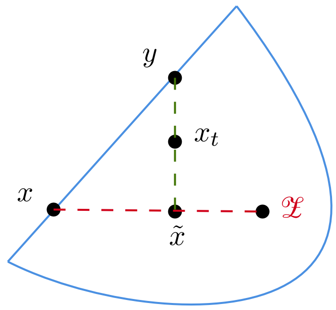
$$\langle g_x - g_y, x - y \rangle \geqslant \mu \|x - y\|^2.$$

Proof. We do each implication in turn.

(ii) $\implies$ (i) Let $x, y \in \operatorname{dom} f$ and $t \in (0,1)$; furthermore take $\mathscr{X} \in \operatorname{ri}\operatorname{dom} f$. Define, for $\alpha \in (0,1]$,

$$\begin{aligned}\tilde{x} &= (1-\alpha)x + \alpha\mathscr{X} \in \operatorname{ri}\operatorname{dom} f,\\ x_t &= (1-t)\tilde{x} + ty \in \operatorname{ri}\operatorname{dom} f,\end{aligned}$$

so that $x_t \in \operatorname{ri}\operatorname{dom} f$.



Let $g_t \in \partial f(x_t)$, so that

$$\begin{aligned}f(\tilde{x}) &\geqslant f(x_t) + \langle g_t, \tilde{x} - x_t\rangle + \frac{\mu}{2}\|\tilde{x} - x_t\|^2\\ f(\tilde{x}) &\geqslant f(x_t) + \langle g_t, t\tilde{x} - ty\rangle + \frac{\mu}{2}\|t\tilde{x} - ty\|^2\\ &\geqslant f(x_t) + t\,\langle g_t, \tilde{x} - y\rangle + \frac{\mu}{2}t^2\|\tilde{x} - y\|^2. \qquad (\ddagger)\end{aligned}$$

Similarly,

$$f(y) \geqslant f(x_t) + \langle g_t, y - x_t\rangle + \frac{\mu}{2}\|y - x_t\|^2,$$

and so we get the analogous

$$f(y) \geqslant f(x_t) + (1-t)\,\langle g_t, y - \tilde{x}\rangle + \frac{\mu}{2}(1-t)^2\|y - \tilde{x}\|^2. \qquad (\dagger)$$

Therefore with $(1-t)\cdot(\ddagger) + t\cdot(\dagger)$ we get

$$(1-t)f(\tilde{x}) + tf(y) \geqslant f(x_t) + t(1-t)\cdot 0 + \frac{\mu}{2}t(1-t)\|y - \tilde{x}\|^2.$$

With the substitution $\tilde{x} = (1-\alpha)x + \alpha\mathscr{X}$, we get

$$\begin{aligned}f\left((1-t)(1-\alpha)x + (1-t)\alpha\mathscr{X} + t\alpha y\right) \leqslant (1-t)f\left((1-\alpha)x + \alpha\mathscr{X}\right)&\\ + tf(y) - \frac{\mu}{2}t(1-t)\left\|(1-\alpha)x + \alpha\mathscr{X} - y\right\|^2&,\end{aligned}$$

and with $\alpha \searrow 0$ we get

$$f(tx + (1-t)y) \leqslant tf(x) + (1-t)f(y) - \frac{\mu}{2}t(1-t)\|x - y\|^2.$$

(iii) $\implies$ (ii) Let $x \in \operatorname{dom}\partial f$ and $g_x \in \partial f(x)$. Furthermore let $\mathscr{X} \in \operatorname{ri}(\operatorname{dom} f)$, and let $\widetilde{y} := (1-\alpha)y + \alpha\mathscr{X} \in \operatorname{ri}(\operatorname{dom} f)$ for $\alpha \in (0,1]$. Consider the mapping

$$\varphi\colon \mathbb{R} \to \overline{\mathbb{R}},$$

$$t \mapsto \begin{cases} f(x_t) & \text{where } x_t := (1-t)x + t\widetilde{y} \in \operatorname{ri}(\operatorname{dom} f) \text{ if } t \in (0,1), \\ +\infty & \text{otherwise.} \end{cases}$$

Let $t \in (0,1)$; there exists a subdifferentiable $g_t \in \partial f(x_t)$ for which $\langle g_t, \widetilde{y} - x\rangle \in \partial\varphi(t)$. Then we can write, by Lemma 9.10,

$$f(\widetilde{y}) - f(x) = \varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)\,\mathrm{d}t = \int_0^1 \langle g_t, \widetilde{y} - x\rangle\,\mathrm{d}t.$$

Since $g \in \partial f(x)$ and $g_t \in \partial f(x_t)$, we have the strong monotonicity property

$$\langle g_t - g, \widetilde{y} - x\rangle \geqslant \mu\|\widetilde{y} - x\|^2,$$

and plugging in the value for x_t makes this

$$\langle g_t, \widetilde{y} - x\rangle \geqslant \langle g_t, \widetilde{y} - x\rangle + \mu(1-t)\|\widetilde{y} - x\|^2.$$

Thus,

$$\begin{aligned} f(\widetilde{y}) - f(x) &\geqslant \int_0^1 \left(\langle g_t, \widetilde{y} - x\rangle + \mu(1-t)\|\widetilde{y} - x\|^2\right)\mathrm{d}t \\ &= \langle g, \widetilde{y} - x\rangle + \mu\|\widetilde{y} - x\|^2 \int_0^1 (1-t)\,\mathrm{d}t \\ &= \langle g, \widetilde{y} - x\rangle + \frac{\mu}{2}\|\widetilde{y} - x\|^2. \end{aligned}$$

Recalling the definition of $\widetilde{y}$,

$$\begin{gathered} f\left((1-\alpha)y + \alpha\mathscr{X}\right) \geqslant f(x) + \langle g, (1-\alpha)y + \alpha\mathscr{X} - x\rangle + \frac{\mu}{2}\|(1-\alpha)y + \alpha\mathscr{X} - x\|^2 \\ \Big\downarrow \alpha \searrow 0 \\ f(y) \geqslant f(x) + \langle g, y - x\rangle + \frac{\mu}{2}\|y - x\|^2. \end{gathered}$$

(i) $\implies$ (iii) Let $x, y \in \operatorname{dom}\partial f$, $g_x \in \partial f(x)$, and $g_y \in \partial f(y)$, and $t \in (0,1)$. Take $x_t \in \operatorname{dom} f$ to be $x_t := tx + (1-t)y$, so that we have

$$f(x_t) \leqslant (1-t)f(x) + tf(y) - \frac{\mu}{2}t(1-t)\|x - y\|^2.$$

Note now that by the definition of the subdifferential, we have

$$\langle g_x, y - x\rangle = \frac{\langle g_x, x_t - x\rangle}{t} \leqslant \frac{f(x_t) - f(x)}{t} \leqslant f(y) - f(x) - \frac{\mu}{2}(1-t)\|x - y\|^2,$$

and with $t \searrow 0$ we get

$$\langle g_x, y - x\rangle \leqslant f(y) - f(x) - \frac{\mu}{2}\|x - y\|^2,$$

and similarly

$$\langle g_y, x - y\rangle \leqslant f(x) - f(y) - \frac{\mu}{2}\|x - y\|^2.$$

Adding these two inequalities gives us that

$$\langle g_x - g_y, y - x\rangle \leqslant -\mu\|x - y\|^2 \implies \langle g_x - g_y, x - y\rangle \geqslant \mu\|x - y\|^2,$$

the desired result. $\square$

§10. Lecture 10—05th April, 2024

§10.1. More on strong convexity

The following proposition is very important for establishing the linear convergence for many convex optimisation algorithms.

Proposition 10.1 (Polyak-Lojasiewicz inequality)**.** *Take any continuously differentiable function $f\colon E \to \mathbb{R}$ that is μ-strongly convex on E with μ positive. Let its subdifferential be $g_x \in \partial f(x)$. Then, for any $x \in E$,*

$$\frac{1}{2}\|g_x\|^2 \geqslant \mu(f(x) - f(x^*)),$$

where x^* is the minimizer of f.

Proof. Recall the definition of strong convexity:

$$f(tx + (1-t)y) \leqslant tf(x) + (1-t)f(y) - \frac{\mu}{2}t(1-t)\|x-y\|^2.$$

Taking the minimisation with respect to y on both sides, we get $f(x^*) \geqslant f(x) - \frac{1}{2}\mu\|x - x^*\|^2$. Rearranging gives us the Polyak-Lojasiewicz inequality. □

We write $\overline{\operatorname{conv}}_\mu E$ to denote the set of functions that are μ-strongly closed convex on E.

Proposition 10.2 (Existence and uniqueness of solutions)**.** *If $f \in \overline{\operatorname{conv}}_\mu E$ with $\mu > 0$, then f has a unique minimiser.*

Proof. First we show existence. Let $x_0 \in \operatorname{ri}\operatorname{dom} f$ and $g \in \partial f(x_0)$. By the characterisation of strong convexity and the fact that the definition applies for a generic norm., we have that for all $x \in E$,

$$\begin{aligned} f(x) &\geqslant f(x_0) + \langle g, x - x_0\rangle + \frac{\mu}{2}\|x - x_0\|^2 \\ &\geqslant f(x_0) + \langle g, x - x_0\rangle + \frac{\mu C}{2}|x - x_0|^2, \qquad\qquad \text{where } |x| = \sqrt{\langle x, x\rangle} \\ &\geqslant f(x_0) - \frac{1}{2\mu C}|g|^2 + \frac{\mu C}{2}\left|x - \left(x_0 - \frac{1}{\mu C}g\right)\right|^2, \end{aligned}$$

where $C \geqslant 0$ by the equivalence of norms. Thus the level set $[f \leqslant f(x_0)] \subseteq B_{|\cdot|}(x_0 - \frac{1}{\mu C}g, \frac{1}{\mu C}|g|)$. This means that $[f \leqslant f(x_0)]$ is bounded and closed, so it has a minimiser.

Now we prove uniqueness. Suppose $x_1 \neq x_2$ are distinct minimisers of f, *i.e.*

$$f(x_1) = f(x_2) = \inf_{x\in E} f(x) = f^*.$$

Then the infimum is

$$\begin{aligned} \inf_{x\in E} f(x) &\leqslant f\left(\frac{x_1 + x_2}{2}\right) \\ &\leqslant \frac{1}{2}f(x_1) + \frac{1}{2}f(x_2) - \frac{\mu}{8}\|x_1 - x_2\|^2 \\ &= \inf_{x\in E} f(x) - \frac{\mu}{8}\|x_1 - x_2\|^2, \end{aligned}$$

and so $\|x_1 - x_2\| = 0 \implies x_1 = x_2$, a contradiction. Hence, the minimiser is unique. □

We now pause our discussion on nonsmooth optimisation to talk about smooth optimisation.

§10.2. Dual averaging

We start by revisiting the subgradient method and will present a way by which we can accelerate it.

Let $f \in \operatorname{conv} E$ and $\operatorname{dom} f = E$. Recall the subgradient method (Section 6.3): for some $x_0 \in E$, perform the iteration

$$\begin{aligned} x_{k+1} &= x_k - \alpha_k s_k, \\ s_k &\in \partial f(x_k), \end{aligned}$$

where $\alpha_k > 0$ is the step size. We have already seen that if $x^* \in \arg\min_E f$ and $\|s_k\| \leqslant L$ for all $k \in \mathbb{N}$, then

$$\min_{i=0,\ldots,k} f(x_i) - f(x^*) \leqslant \frac{\|x_0 - x^*\|^2 + L^2 \sum_{i=0}^k \alpha_i^2}{2\sum_{i=0}^k \alpha_i} = \begin{cases} O\left(1/\ln k\right) & \text{if } \alpha_k = 1/(k+1), \\ O\left(1/k^{1/2+\varepsilon}\right) & \text{if } \alpha_k = 1/(k+1)^{1/2+\varepsilon}, \\ O\left(k^{-1/2}\ln k\right) & \text{if } \alpha_k = 1/\sqrt{k+1}, \end{cases}$$

using the fact that $\sum_{i=1}^k \sim \ln k$ and $\sum_{i=1}^k i^{-1/2} \sim \sqrt{k}$. Recall that if we know $f(x^*)$, then Polyak's step size gives us a rate of $O(1/\sqrt{k})$. But what do we do if we don't know $f(x^*)$?

Let us examine the subgradient method. Expanding the tower of iterates, we have

$$x_{k+1} = x_k - \alpha_k s_k = x_{k-1} - \alpha_{k-1}s_{k-1} - \alpha_k s_k = \ldots = x_0 - \sum_{i=0}^k \alpha_i s_i,$$

where $s_i \in \partial f(x_i)$. One bit of criticism of the subgradient method that might explain our shortcomings (towards a fast rate) is that the new updates are getting very small weights as we go along. This is because we basically "expend" a lot of our weights early on in the algorithm, so when we come closer to the optimum, we are not making as much progress as we would like. This essentially motivates the idea of *dual averaging*; here we will try to balance the weights of the iterates by taking the average of the subgradients before we update the iterate. This is the idea behind the dual averaging method.

Algorithm: DUAL AVERAGING METHOD—SETUP

- Set the starting point $x_0 = 0$ and a sequence of step sizes $(\beta_k)_k \in \mathbb{N}$ such that β_k is well defined for all $k \in \mathbb{N}$.
- For each $k \in \mathbb{N}$, do the following until convergence:
  - Set $x_{k+1} \in x_0 - \beta_k \sum_{i=0}^k s_i$, where $s_i \in \partial f(x_i)$.
- Return the minimiser $f(x_k)$.

How then do we choose the new step size β_k? Define for all $k \in \mathbb{N}^*$ the quantity

$$E_k := \gamma_k \|x_k - x^*\|^2 + \sum_{i=0}^{k-1} f(x_i) - f(x^*),$$

where $\gamma_k > 0$ is a monotonic sequence of real numbers. We will examine the variation of this quantity over the evaluation of the iterates and see if we gain any insights. Define $\Delta_E \coloneqq E_{k+1} - E_k$. Then

$$\begin{aligned}
\Delta_E &= \gamma_{k+1}\|x_{k+1} - x^*\|^2 - \gamma_k\|x_k - x^*\|^2 + f(x_k) - f(x^*) \\
&= (\gamma_{k+1} - \gamma_k)\|x^*\|^2 + \gamma_{k+1}\|x_{k+1}\|^2 - \gamma_k\|x_k\|^2 + \langle 2\gamma_k x_k - 2\gamma_{k+1}x_{k+1}, x^*\rangle \\
&\qquad + 2\gamma_k\|x_k\|^2 - 2\gamma_{k+1}\langle x_{k+1}, x_k\rangle + f(x_k - f(x^*)) \\
&= (\gamma_{k+1} - \gamma_k)\|x^*\|^2 + \gamma_{k+1}\|x_{k+1}\|^2 + \gamma_k\|x_k\|^2 - 2\gamma_{k+1}\langle x_{k+1}, x_k\rangle \\
&\qquad + \underbrace{\langle 2\gamma_k x_k - 2\gamma_{k+1}x_{k+1}, x^* - x_k\rangle + f(x_k) - f(x^*)}_{\leqslant 0 \text{ by the definition of } \partial f}.
\end{aligned}$$

We would like for $s_k = 2\gamma_k x_k - 2\gamma_{k+1}x_{k+1} \in \partial f(x_k)$, so that we can use the intuitions we have already built up from the subgradient inequality. First we find γ_k; consider the iterates x_{k+1} and x_k:

$$\begin{aligned}
\frac{1}{\beta_k}x_{k+1} &= -\sum_{i=0}^{k} s_i \\
\frac{1}{\beta_{k-1}}x_k &= -\sum_{i=0}^{k-1} s_i.
\end{aligned}$$

Subtracting the two, we get

$$\frac{1}{\beta_k}x_{k+1} - \frac{1}{\beta_{k-1}}x_k = -s_k \implies \gamma_k = \frac{1}{2\beta_{k-1}}.$$

Returning to the expression for Δ_E, we have

$$\begin{aligned}
\Delta_E &= (\gamma_{k+1} - \gamma_k)\|x^*\|^2 + \gamma_{k+1}\|x_{k+1}\|^2 + \gamma_k\|x_k\|^2 - 2\gamma_{k+1}\langle x_{k+1}, x_k\rangle \\
&\qquad + \langle 2\gamma_k x_k - 2\gamma_{k+1}x_{k+1}, x^* - x_k\rangle + f(x_k) - f(x^*) \\
&\leqslant (\gamma_{k+1} - \gamma_k)\|x^*\|^2 + \frac{1}{\gamma_k}(\underbrace{\gamma_{k+1}\underbrace{\gamma_k}_{\leqslant \gamma_{k+1}}}_{\leqslant \gamma_{k+1}^2}\|x_{k+1}\|^2 + \gamma_k^2\|x_k\|^2 - 2\gamma_{k+1}\gamma_k\langle x_{k+1}, x_k\rangle) \\
&\leqslant (\gamma_{k+1} - \gamma_k)\|x^*\|^2 + \frac{1}{\gamma_k}\|\underbrace{\gamma_{k+1}x_{k+1} - \gamma_k x_k}_{= s_k/2}\|^2 \\
&\leqslant (\gamma_{k+1} - \gamma_k)\|x^*\|^2 + \frac{1}{4\gamma_k}\|s_k\|^2,
\end{aligned}$$

which is an interestingly simple reduction of the original problem, as it gives us a bound on how much E_k can increase per round. We can now use this bound to telescope:

$$\begin{aligned}
E_{k+1} - E_1 = \sum_{i=1}^{k} E_{i+1} - E_i &\leqslant \sum_{i=1}^{k}\left[(\gamma_{i+1} - \gamma_i)\|x^*\|^2 + \frac{1}{4\gamma_i}\|\underbrace{s_i}_{\leqslant L}\|^2\right] \\
&\leqslant \gamma_{k+1}\|x^*\|^2 + \frac{L^2}{4}\sum_{i=1}^{k}\frac{1}{\gamma_i}.
\end{aligned}$$

Notice now that by Jensen's inequality, we have

$$(k+1)\cdot\left(f\left(\frac{1}{k+1}\sum_{i=0}^{k}x_i\right)-f(x^*)\right)\leqslant(k+1)\cdot\left(\sum_{i=0}^{k}\frac{1}{k+1}f(x_i)-\sum_{i=0}^{k}\frac{1}{k+1}f(x^*)\right)$$
$$=\sum_{i=0}^{k}f(x_i)-f(x^*)\leqslant E_{k+1}.$$

Now define $\widetilde{x}_k=\frac{1}{k+1}\sum_{i=0}^{k}x_i$. Then we can write

$$\begin{aligned}f(\widetilde{x}_k)-f(x^*)&\leqslant\frac{E_{k+1}}{k+1}\\&\leqslant\frac{E_1+\gamma_{k+1}\|x^*\|^2+\frac{L^2}{4}\sum_{i=1}^{k}\frac{1}{\gamma_i}}{k+1}\\&=\frac{E_1}{k+1}+\frac{\|x^*\|^2}{2\sqrt{k+1}}+\frac{L^2}{2(k+1)}\sum_{i=1}^{k}\frac{1}{\sqrt{i}}\qquad\text{with }\gamma_k=\frac{1}{2}\sqrt{k}.\end{aligned}$$

From calculus we know that

$$\sum_{i=1}^{k}\int_{i-1}^{i}\frac{1}{\sqrt{x}}\,\mathrm{d}x=\int_0^k\frac{1}{\sqrt{x}}\,\mathrm{d}x=2\sqrt{k}\leqslant2\sqrt{k+1},$$

and so

$$\begin{aligned}f(\widetilde{x}_k)-f(x^*)&\leqslant\frac{\|x_1-x^\|+f(0)-f(x^*)}{k+1}+\frac{\|x^*\|^2}{2\sqrt{k+1}}+\frac{L^2}{2(k+1)}\\&=\frac{1}{\sqrt{k+1}}\left(\frac{1}{2}\|x^*\|^2+L^2+\frac{\|x_1-x^*\|^2+f(0)-f(x^*)}{\sqrt{k+1}}\right)\\&=O\left(\frac{1}{\sqrt{k}}\right).\end{aligned}$$

So we must choose $\beta_k=\frac{1}{\sqrt{k+1}}$ to achieve the optimal rate. This is the dual averaging method:

Algorithm: Dual averaging method

- For each $k\in\mathbb{N}$, do the following until convergence:
  - Set $x_{k+1}\in-\frac{1}{\sqrt{k+1}}\sum_{i=0}^{k}s_i$, where $s_i\in\partial f(x_i)$.
- Return the minimiser $f(x_k)$.

In fact, we cannot do better than $O(1/\sqrt{k})$—an improvement over the subgradient method of Section 6.3—for this method.

The dual averaging method was presented 15 years ago by Nesterov [Nes09] and has many applications. One particularly amazing fact about this method is that it doesn't require much; we don't need tuning of the step size as we progress towards the optimum, and yet we have a reliable stopping criterion (as is evident in the paper by Nesterov).

There are other variants of the dual averaging method such as dual averaging with a stochastic objective and distributed dual averaging, but we do not discuss these here since their extensions are rather natural in the context of the individual problems.

§10.3. Differentiability

§10.3.1. First-order differentiability

Let E and F be two normed vector spaces, $\Omega \subseteq E$ an open set, and $f\colon \Omega \to F$ a function. Pick $x \in \Omega$.

Definition 10.3 (Directional differentiability). *The* directional derivative *of f at x in the direction $h \in E$ is*

$$f'(x,h) = \lim_{t \searrow 0} \frac{f(x+th) - f(x)}{t} \in \mathbb{R},$$

Here we're requiring that the limit exists for all $h \in E$ and is finite. In order for this definition to make any sense, we're implicitly saying that there must be some $\varepsilon > 0$ such that $x + th \in \Omega$ for all $t \in (-\varepsilon, \varepsilon)$, so that $f(x+th)$ is well-defined.

Lemma 10.4. *The directional derivative is degree-one positively homogeneous, i.e. $f'(x, \lambda h) = \lambda f'(x,h)$ for all $\lambda \geqslant 0$.*

Proof. Just use the definition of the directional derivative:

$$\begin{aligned} f'(x,\lambda h) &= \lim_{t \searrow 0} \frac{f(x+t\lambda h) - f(x)}{t} = \lim_{t \searrow 0} \frac{\lambda f(x+th) - \lambda f(x)}{t} \\ &= \lambda \lim_{t \searrow 0} \frac{f(x+th) - f(x)}{t} \\ &= \lambda f'(x,h). \end{aligned}$$ □

A stronger notion of differentiability is that of Gâteaux differentiability:

Definition 10.5 (Gâteaux differentiability). *The function f is* Gâteaux differentiable *at x if the directional derivative $f'(x,h)$ exists for all $h \in E$ and the mapping $h \mapsto f'(x,h)$ is linear and continuous. We denote by $f'(x)$ this linear continuous mapping.*

The "linear and continuous" specification is important here; linear maps need not be continuous in infinite dimensions. In fact, Gâteaux differentiability doesn't imply continuity of the derivative:

Example 10.6. Consider $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} \frac{y}{x}(x^2+y^2) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

It is easy to check that the Gâteaux derivative of f at $(0,0)$ is the zero linear map:

$$\frac{f(tx,ty) - f(0,0)}{t} = \frac{\frac{ty}{tx}\left(t^2x^2 + t^2y^2\right)}{t} = \frac{ty}{x}(x^2+y^2) \to 0.$$

However, f is not continuous at $(0,0)$. For instance, consider $c(\varepsilon) = (\varepsilon^4, \varepsilon)$. Then $c(\varepsilon) \to 0$ as

$\varepsilon \to 0$, but
$$f(c(\varepsilon)) = \frac{\varepsilon}{\varepsilon^4} \cdot \left(\varepsilon^8 + \varepsilon^2\right) = \varepsilon^5 + \frac{1}{\varepsilon}.$$
Then $f(c(\varepsilon)) \to \infty$ as $\varepsilon \searrow 0$, and $f(c(\varepsilon)) \to -\infty$ as $\varepsilon \nearrow 0$. Thus f is not continuous at $(0,0)$.

An even stronger notion of differentiability is that of Fréchet differentiability:

Definition 10.7 (Fréchet differentiability)**.** *The function f is* Fréchet differentiable *at x if there exists a bounded, linear, and continuous map $L\colon E \to F$ such that*

$$f(x+h) = f(x) + Lh + o(\|h\|),$$

where $o(\|h\|)$ is a function such that $o(\|h\|)/\|h\| \to 0$ as $h \to 0$.

We denote by $f'(x)$ this linear map.

We write $f \in \mathcal{D}^1(\Omega, F)$ to denote that f is Fréchet differentiable on Ω with values in F (sometimes we write $f \in \mathcal{D}^1(\Omega)$ if $F = \mathbb{R}$). Likewise, we write $f \in \mathcal{G}^1(\Omega, F)$ if f is Gâteaux differentiable on Ω with values in F (and $f \in \mathcal{G}^1(\Omega)$ if $F = \mathbb{R}$). We say that $f \in \mathcal{C}^1(\Omega)$ if f is Fréchet differentiable on Ω and f' is continuous on Ω. Clearly, for convex f, if $f \in \mathcal{D}^1(\Omega)$, then $f \in \mathcal{G}^1(\Omega)$, and the differentials coincide. The converse is also true:

Lemma 10.8. *If $f \in \mathcal{G}^1(\Omega)$ and f is convex, then $f \in \mathcal{D}^1(\Omega)$.*

Proof. The proof is left as an exercise. (See Appendix A for a proof.) □

Example 10.9. Consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x^3/(x^2+y^2) & \text{if } (x,y) \neq (0,0), \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to show that f is Gâteaux differentiable at $(0,0)$, but not Fréchet differentiable at $(0,0)$.

Recall the composition rule for (that is, which applies to) all Fréchet differentiable functions:

Proposition 10.10 (Composition rule)**.** *Let $f\colon \Omega \to F$ and $g\colon F \to G$ be Fréchet differentiable functions. If f is Fréchet differentiable at $x \in \Omega$ and g is Fréchet differentiable at $f(x)$, then $g \circ f$ is Fréchet differentiable at x and*

$$(g \circ f)'(x) = g'(f(x)) \circ f'(x).$$

Proof. The proof is left as an exercise. (See Appendix A for a proof.) □

We can supply a definition of the gradient of a function f at a point x:

Definition 10.11 (Gradient)**.** *Let $f\colon \Omega \to \mathbb{R}$ be Fréchet differentiable at $x \in \Omega$. The* gradient *of f at x is the unique vector $\nabla f(x) \in E$ such that*

$$f'(x, h) = \langle \nabla f(x), h \rangle \quad \forall\, h \in E.$$

§10.3.2. Second-order differentiability

Definition 10.12 (Second derivative)**.** *Let $f\colon \Omega \to F$ be Fréchet differentiable at $x \in \Omega$. We say that f is* twice differentiable *at x if the linear mapping S is Fréchet differentiable at x, where S is defined as*

$$\begin{aligned} S\colon E &\to \mathcal{L}(E, F), \\ x &\mapsto f'(x) \end{aligned}$$

The second order derivative $f'' \in \mathcal{L}(E, \mathcal{L}(E, F))$.

Proposition 10.13 (Properties of the second derivative)**.** *1. The second derivative f'' is the directional derivative of $x \mapsto f'(x)h$ in the direction k:*

$$f''(x)h = \lim_{t \searrow 0} \frac{f'(x + tk)h - f'(x)h}{t}.$$

2. The function $(h, k) \mapsto f''(x)(h, k)$ (i.e. the second derivative $\partial^2 f$) is bilinear and symmetric.

Proof. The proof is left as an exercise. (See Appendix A for a proof.) □

Definition 10.14 (Hessian)**.** *Let $f\colon \Omega \to \mathbb{R}$ be Fréchet differentiable at $x \in \Omega$. The* Hessian *of f at x is the unique bounded, symmetric, bilinear form $\nabla^2 f(x)$ on E such that*

$$f''(x)(h, k) = \left\langle \nabla^2 f(x)h, k \right\rangle \quad \forall h, k \in E.$$

Proposition 10.15. *Let $f \in \operatorname{conv} E$ and $x \in (\operatorname{dom} f)^\circ$. If f is differentiable*[5] *at x, then the subdifferential $\partial f(x) = \{\nabla f(x)\}$. Conversely, if the subdifferential $\partial f(x)$ is a singleton $\{x^*\}$, then f is differentiable at x and the gradient $\nabla f(x) = x^*$.*

Proof. First we work the forward direction. Let $y \in E$. Then

$$\langle \nabla f(x), y - x \rangle = f'(x)(y - x) = f'(x, y - x) = \lim_{t \searrow 0} \frac{f(x + t(y - x)) - f(x)}{t} \leqslant f(y) - f(x),$$

so that

$$f(x) + \langle \nabla f(x), y - x \rangle \leqslant f(y) \implies \nabla f(x) \in \partial f(x)$$

by the definition of the subdifferential. Now let $x^* \in \partial f(x)$. Then for all $d \in E$, Proposition 8.3 gives us

$$f'(x, d) \geqslant \langle x^*, d \rangle \implies f'(x)d \geqslant \langle x^*, d \rangle \implies f'(x)(-d) \geqslant \langle x^*, -d \rangle \implies f'(x)(d) \leqslant \langle x^*, d \rangle.$$

[5]Here we mean Fréchet differentiable, and that will be our default meaning for "differentiable" from now on.

Thus $f'(x)d = \langle x^*, d\rangle$ for all $d \in E$, so that $x^* = \nabla f(x)$.

For the converse, suppose that $\partial f(x) = \{x^*\}$. Then for all $d \in E$, we have by the maximum formula (Proposition 8.9) that

$$f'(x,d) \geqslant \sup_{x_0^* \in \partial f(x)} \langle x_0^*, d\rangle = \langle x^*, d\rangle \,,$$

and so f is Gâteaux differentiable at x. By Lemma 10.8, f is Fréchet differentiable at x, and the gradient is $\nabla f(x) = x^*$. □

§10.4. Properties of convex functions with Lipschitz gradients

We say that $f \in \mathcal{C}^{1,1}(\Omega)$ if $f \in \mathcal{C}^1(\Omega)$ and there exists a constant $L \geqslant 0$ such that for all $x, y \in \Omega$,

$$\|f'(x) - f'(y)\|_{E,F} \leqslant L\|x - y\|_E,$$

where $L \in \mathcal{L}(E,F)$ is defined as

$$\|L\|_{E,F} \coloneqq \sup_{\|x\|_E \leqslant 1} \|Lx\|_F.$$

If E is Euclidean and $F = \mathbb{R}$, then

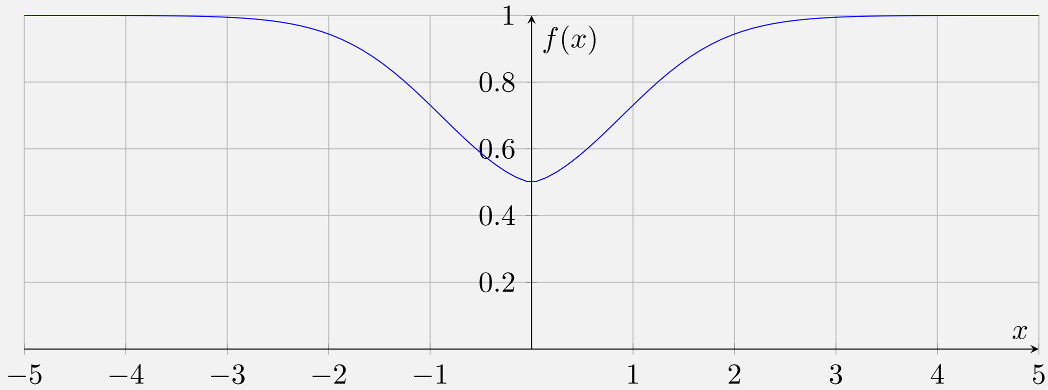
$$\begin{aligned}\|f'(x) - f'(y)\|_{E,F} &= \sup_{\|h\|_E \leqslant 1} |f'(x)h - f'(y)h| = \sup_{\|h\|_E \leqslant 1} \big|\big\langle f'(x) - f'(y), h\big\rangle\big| \\ &= \|\nabla f(x) - \nabla f(y)\|_*,\end{aligned}$$

where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$ (*i.e.* the support function of the unit ball of $\|\cdot\|$).

Example 10.16. We consider the following inclusions:

1. *Is $\mathcal{C}^{1,1}(\mathbb{R}) \subseteq \mathcal{C}^{0,1}(\mathbb{R})$ (the set of Lipschitz continuous functions)?* The answer is no; consider the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$.
2. *Is $\mathcal{C}^{0,1}(\mathbb{R}) \cap \mathcal{C}^1(\mathbb{R}) \subseteq \mathcal{C}^{1,1}(\mathbb{R})$?* The answer is no; consider the function

$$f(x) = \frac{1}{1 + e^{-|x|^{3/2}}}.$$



The moral lesson here is that there is no connection between Lipschitz continuity at different orders; the Lipschitz continuity of the objective is unrelated to the Lipschitz continuity of the nth gradient.

§11. Lecture 11—19th April, 2024

§11.1. More properties of convex functions

Before we start going into analysing algorithms, we will look at some more properties of convex functions that will be useful in this regard. Recall that we write $\mathcal{C}([a,b],\mathbb{R})$ to denote the set of continuous functions from $[a,b]$ to $\mathbb{R}$ and $\mathcal{D}([a,b],\mathbb{R})$ to denote the set of Fréchet differentiable functions from $[a,b]$ to $\mathbb{R}$.

Proposition 11.1 (Mean value theorem)**.** *Suppose $a \neq b$. If $f \in \mathcal{C}[[a,b],\mathbb{R}] \cap \mathcal{D}((a,b),\mathbb{R})$ is convex, then there exists some $x \in (a,b)$ such that $f(b) - f(a) = f'(x)(b-a)$.*

Proof sketch. Take

$$g(t) \coloneqq f(t) - \left[f(a) + \frac{f(b)-f(a)}{b-a}(t-a)\right].$$

Then $g(a) = g(b) = 0$. Since f is convex, g is non-negative. Now, apply Rolle's theorem, which says that $\exists\, x \in (a,b)$ such that $g'(x) = 0$. The result is that

$$g'(x) = f'(x) - \frac{f(b)-f(a)}{b-a} = 0 \implies f(b) - f(a) = f'(x)(b-a),$$

as required. □

The next result is a property that holds for functions with Lipschitz continuous gradients.

Proposition 11.2. *Let E, F be normed vector spaces, let $C \subset E$ be open convex, take $f \in \mathcal{D}(E,F)$, and pick any $L \geqslant 0$. Then the following are equivalent:*

(i) For all $x \in C$, $\|f'(x)\| \leqslant L$.

(ii) For all $x, y \in C$, $\|f(x) - f(y)\| \leqslant L\|x-y\|$.

Proof. First we show that the second condition implies the first. Indeed,

$$\begin{aligned}\left\|f'(x)\right\| &= \sup_{\|h\|\leqslant 1}\left\|f'(x)\cdot h\right\| = \sup_{\|h\|\leqslant 1}\left\|\lim_{t\searrow 0}\frac{f(x+th)-f(x)}{t}\right\| \\ &= \sup_{\|h\|\leqslant 1}\lim_{t\searrow 0}\left\|\frac{f(x+th)-f(x)}{t}\right\| \leqslant \sup_{\|h\|\leqslant 1}\lim_{t\searrow 0}\frac{L\|x+th-x\|}{t} \\ &= L.\end{aligned}$$

Now we show that the first condition implies the second. Fix $x, y \in C$ and take $t \in (0,1)$. Let $g(t) \coloneqq f(x+t(y-x))$. Then $g'(t) = f'(x+t(y-x))(y-x)$. By the mean value theorem,

$$\begin{aligned} f(y) - f(x) = g(1) - g(0) = g'(x) &= f'(x+t(y-x))(y-x) \\ &= f'(x+t(y-x))(y-x) - f'(x)(y-x) + f'(x)(y-x) \\ &\leqslant \left\|f'(x+t(y-x)) - f'(x)\right\|\|y-x\| + \left\|f'(x)\right\|\|y-x\| \\ &\leqslant L\|y-x\|. \end{aligned}$$

□

These lead to the following corollary.

Corollary 11.3. *Let f, C, and L be as above, and let $\|\cdot\|_*$ be the dual norm. Then $\|\nabla f(x)\|_* \leqslant L$ for all $x \in C$ if and only if $|f(y) - f(x)| \leqslant L \|y - x\|$ for all $x, y \in C$.*

Proof. Exercise. $\square$

The following result will be important for us:

Lemma 11.4. *If $f \in \mathcal{C}_L^{1,1}(\Omega)$ (i.e. f is continuously differentiable with Lipschitz continuous gradient) with $\Omega \subseteq E$ open and convex, then for all $x, y \in \Omega$,*

$$\left|f(y) - f(x) - \left\langle f'(x), y - x\right\rangle\right| \leqslant \frac{L}{2} \|y - x\|^2 .$$

Proof. We will prove this first by working in one dimension. Let $g(t) := f(x + t(y - x))$. Then $g'(t) = f\left(x + t(y - x)\right)(y - x)$. By the fundamental theorem of calculus,

$$\begin{aligned} f(y) - f(x) = g(1) - g(0) &= \int_0^1 g'(t)\,\mathrm{d}t = \int_0^1 f'(x + t(y - x))(y - x)\,\mathrm{d}t \\ &= \int_0^1 \left\langle f'(x + t(y - x)), y - x\right\rangle \,\mathrm{d}t. \end{aligned}$$

Thus we get

$$\begin{aligned} \left|f(y) - f(x) - \left\langle f'(x), y - x\right\rangle\right| &= \left|\int_0^1 \langle \nabla f(x + t(y - x)), y - x\rangle - \langle \nabla f(x), y - x\rangle \,\mathrm{d}t\right| \\ &= \left|\int_0^1 \langle \nabla f(x + t(y - x)) - \nabla f(x), y - x\rangle \,\mathrm{d}t\right| \\ &\leqslant \int_0^1 |\langle \nabla f(x + t(y - x)) - \nabla f(x), y - x\rangle| \,\mathrm{d}t \\ &\leqslant \int_0^1 \|\nabla f(x + t(y - x)) - \nabla f(x)\|_* \cdot \|y - x\| \,\mathrm{d}t \\ &\leqslant \int_0^1 L \cdot \|t(y - x)\| \cdot \|y - x\| \,\mathrm{d}t \\ &= L \cdot \|y - x\|^2 \int_0^1 t\,\mathrm{d}t \\ &= \frac{L}{2} \|y - x\|^2 . \end{aligned}$$

(Here we used the Cauchy-Schwarz inequality in the fourth line.) The general case follows by applying the one-dimensional result to each coordinate. $\square$

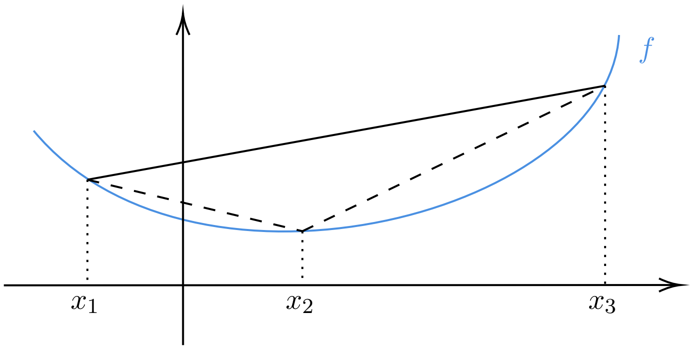
Our next result tugs more on this relationship between convexity and the derivative that we are trying to establish.

Lemma 11.5. *Let $f \in \mathcal{D}(\mathbb{R})$. Then the following are equivalent:*

(i) f *is convex.*

(ii) *For all* $t_1, t_2 \in \mathbb{R}$*, if* $t_1 \leqslant t_2$*, then* $f'(t_1) \leqslant f'(t_2)$*.*

Proof. Consider the following figure:



Note that f is convex if and only if, for all $x_1 < x_2 < x_3$, the secant line connecting $(x_1, f(x_1))$ and $(x_3, f(x_3))$ lies below the graph of f:

$$\frac{f(x_3) - f(x_1)}{x_3 - x_1} \leqslant \frac{f(x_2) - f(x_1)}{x_2 - x_1} \leqslant \frac{f(x_3) - f(x_2)}{x_3 - x_2}. \tag{$\star$}$$

Now we show that ($\star$) implies that $f'(x_1) \leqslant f'(x_2)$. Let $t \in (0, 1)$. Then we have

$$\begin{aligned}\frac{f(x_1 + t(x_2 - x_1)) - f(x_1)}{t(x_2 - x_1)} \leqslant \frac{f(x_2) - f(x_1)}{x_2 - x_1} &\implies \lim_{t \searrow 0} \frac{f(x_1 + t(x_2 - x_1)) - f(x_1)}{t(x_2 - x_1)} \leqslant \frac{f(x_2) - f(x_1)}{x_2 - x_1} \\ &\implies f'(x_1)(x_2 - x_1) \leqslant f(x_2) - f(x_1) \\ &\implies f'(x_1) \leqslant \frac{f(x_2) - f(x_1)}{x_2 - x_1} = f'(x_2).\end{aligned}$$

The converse is straightforward from the definition of the derivative and the mean value theorem: if $f'(x_1) \leqslant f'(x_2)$, then

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} = f'(x_1 + t(x_2 - x_1)) \leqslant f'(x_2)$$

for some $t \in (0, 1)$. Thus we obtain that f is convex. □

Lemma 11.6. *Let* $f \in \mathcal{D}(E)$*. Then the following are equivalent:*

(i) f *is convex.*

(ii) *For all* $x, y \in \mathbb{R}^n$*,* $f(y) \geqslant f(x) + \langle \nabla f(x), y - x \rangle$*.*

(iii) *For all* $x, y \in \mathbb{R}^n$*,* $\langle \nabla f(y) - \nabla f(x), y - x \rangle \geqslant 0$*.*

Proof. We prove each implication in turn.

(i) $\implies$ (ii) For all $t \in (0, 1)$, we have

$$f\left(x + t(y - x)\right) \leqslant t f(y) + (1 - t) f(x) \implies \frac{f(x + t(y - x)) - f(x)}{t} \leqslant f(y) - f(x)$$

$$\implies \lim_{t\searrow 0}\frac{f(x+t(y-x))-f(x)}{t}\leqslant f(y)-f(x)$$
$$\implies \langle \nabla f(x), y-x\rangle \leqslant f(y)-f(x).$$

(ii) $\implies$ (iii) Take $x, y\in\mathbb{R}^n$. Then

$$\begin{aligned}
f(y)\geqslant f(x)+\langle\nabla f(x), y-x\rangle &\implies f(x)\geqslant f(y)+\langle\nabla f(y), x-y\rangle\\
&\implies \langle\nabla f(x), y-x\rangle\geqslant\langle\nabla f(y), y-x\rangle .
\end{aligned}$$

(iii) $\implies$ (i) Take $g(t)\coloneqq f(x+t(y-x))$, so that $g'(t)=\langle\nabla f(x+t(y-x)), y-x\rangle$. Then for $t_2>t_1$, we have

$$g'(t_2)-g'(t_1)=\frac{1}{t_2-t_1}\langle\nabla f(x+t_2(y-x))-\nabla f(x+t_1(y-x)),(y-x)(t_2-t_1)\rangle\geqslant 0.$$

This implies that g is convex, and so $g(t)\leqslant tg(1)+(1-t)g(0)$, so that f is convex. □

Lemma 11.7. *Let $f\in\mathcal{D}^1(E)$. Then for all $x,y\in E$ and for all $t\in[0,1]$, the following are equivalent:*

(i) $f\in\mathcal{F}_L^{1,1}(E)\coloneqq\mathrm{conv}(E)\cap\mathcal{C}_L^{1,1}(E)$, *where* $\mathcal{C}_L^{1,1}(E)\coloneqq\{f\in\mathcal{D}^1(E):\|\nabla f(x)\|\leqslant L\}$.

(ii) $0\leqslant f(y)-f(x)-\langle\nabla f(x), y-x\rangle\leqslant\frac{L}{2}\|y-x\|^2$.

(iii) $f(x)+\langle\nabla f(x), y-x\rangle+\frac{1}{2L}\|\nabla f(y)-\nabla f(x)\|_*^2\leqslant f(y)$.

(iv) $\frac{1}{L}\|\nabla f(y)-\nabla f(x)\|_*^2\leqslant\langle\nabla f(x)-\nabla f(y), x-y\rangle$.

(v) $0\leqslant\langle\nabla f(x)-\nabla f(y), x-y\rangle\leqslant L\|x-y\|^2$.

(vi) $f(tx+(1-t)y)+\frac{(1-t)t}{2L}\|\nabla f(y)-\nabla f(x)\|_*^2\leqslant(1-t)f(x)+tf(y)$.

(vii) $0\leqslant(1-t)f(x)+tf(y)-f((1-t)x+ty)\leqslant\frac{L}{2}t(1-t)\|x-y\|^2$.

Proof. Again we prove the implications in turn.

(i) $\implies$ (ii) This follows from Lemma 11.4 and Lemma 11.6.

(ii) $\implies$ (iii) We check that

$$\begin{aligned}
f(x)-f(y)&=\inf_{z\in E}\{(f(x)-f(z))+(f(z)-f(y))\}\\
&\leqslant\inf_{z\in E}\left\{\langle\nabla f(x), x-z\rangle+\langle\nabla f(z), z-y\rangle+\frac{L}{2}\|z-y\|^2\right\}\\
&=\inf_{z\in E}\left\{\langle\nabla f(x), x-y\rangle+\langle\nabla f(y)-\nabla f(x), z-y\rangle+\frac{L}{2}\|z-y\|^2\right\}\\
&=\langle\nabla f(x), x-y\rangle+\inf_{z\in E}\left\{\langle\nabla f(y)-\nabla f(x), z-y\rangle+\frac{L}{2}\|z-y\|^2\right\}\\
&=\langle\nabla f(x), x-y\rangle-\sup_{z\in E}\left\{\langle\nabla f(y)-\nabla f(x), y-z\rangle+\frac{L}{2}\|y-z\|^2\right\}\\
&=\langle\nabla f(x), x-y\rangle-L\cdot\sup_{z\in E}\left\{\left\langle\frac{\nabla f(y)-\nabla f(x)}{L}, u\right\rangle-\frac{1}{2}\|u\|^2\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \langle \nabla f(x), x-y\rangle - L\cdot \sup_{u\in E}\left\{\left\langle \frac{\nabla f(y)-\nabla f(x)}{L}, u\right\rangle - \frac{1}{2}\|u\|^2\right\} \\
&= \langle \nabla f(x), x-y\rangle - L\cdot \frac{1}{2}\left\|\frac{\nabla f(y)-\nabla f(x)}{L}\right\|_*^2 \\
&= \langle \nabla f(x), x-y\rangle + \frac{1}{2L}\|\nabla f(y)-\nabla f(x)\|_*^2.
\end{aligned}
$$

(iii) $\implies$ (iv) We have

$$
\begin{aligned}
f(x) + \langle \nabla f(x), y-x\rangle + \frac{1}{2L}\|\nabla f(y)-\nabla f(x)\|_*^2 &\leqslant f(y) \\
f(y) + \langle \nabla f(y), x-y\rangle + \frac{1}{2L}\|\nabla f(x)-\nabla f(y)\|_*^2 &\leqslant f(x).
\end{aligned}
$$

Adding these two inequalities gives

$$
\langle \nabla f(x)-\nabla f(y), y-x\rangle + \frac{1}{2L}\left(\|\nabla f(y)-\nabla f(x)\|_*^2 + \|\nabla f(x)-\nabla f(y)\|_*^2\right) \leqslant 0,
$$

which simplifies to

$$
\frac{1}{L}\|\nabla f(y)-\nabla f(x)\|_*^2 \leqslant \langle \nabla f(x)-\nabla f(y), x-y\rangle .
$$

(iv) $\implies$ (i) We apply Lemma 11.6 and Cauchy-Schwarz to get

$$
\frac{1}{L}\|\nabla f(y)-\nabla f(x)\|_*^2 \leqslant \langle \nabla f(x)-\nabla f(y), x-y\rangle \leqslant \|\nabla f(x)-\nabla f(y)\|_*\cdot\|x-y\| \leqslant L\|x-y\|^2 .
$$

(ii) $\implies$ (v) We have both

$$
\begin{aligned}
0 \leqslant f(y) - f(x) - \langle \nabla f(x), y-x\rangle &\leqslant \frac{L}{2}\|y-x\|^2 \\
0 \leqslant f(x) - f(y) - \langle \nabla f(y), x-y\rangle &\leqslant \frac{L}{2}\|x-y\|^2 .
\end{aligned}
$$

Adding gives us

$$
0 \leqslant \langle \nabla f(x)-\nabla f(y), x-y\rangle \leqslant L\|x-y\|^2 .
$$

(v) $\implies$ (ii) Again we have

$$
\begin{aligned}
f(y) - f(x) - \langle \nabla f(x), y-x\rangle &= \int_0^1 \frac{1}{t}\langle \nabla f(x+t(y-x)) - \nabla f(x), t(y-x)\rangle \,\mathrm{d}t \\
&\leqslant \int_0^1 \frac{1}{t}\frac{L}{2}\|t(y-x)\|^2 \,\mathrm{d}t = \frac{L}{2}\|y-x\|^2\cdot\int_0^1 t\,\mathrm{d}t \\
&= \frac{L}{2}\|y-x\|^2 .
\end{aligned}
$$

(ii) $\implies$ (vii) Per usual we define $z_t \coloneqq x + t(y-x)$, so that

$$
f(x) \leqslant f(z_t) + \langle \nabla f(z_t), x-z_t\rangle + \frac{L}{2}\|x-z_t\|^2 \tag{$\dagger$}
$$

$$f(y) \leqslant f(z_t) + \langle \nabla f(z_t), y - z_t \rangle + \frac{L}{2} \|y - z_t\|^2 . \tag{‡}$$

Then $(1-t)$ times (†) plus t times (‡) gives

$$\begin{aligned}
(1-t) \cdot f(x) + t \cdot f(y) &\leqslant f(z_t) + \left\langle \nabla f(z_t), \underbrace{(1-t)(x - z_t) + t(y - z_t)}_{=0} \right\rangle + \frac{L(1-t)}{2} \|x - z_t\|^2 \\
&\qquad + \frac{Lt}{2} \|y - z_t\|^2 \\
&\leqslant f(x + t(y-x)) + \frac{(1-t)^2 tL}{2} \|x - y\|^2 + \frac{t^2(1-t)L}{2} \|x - y\|^2 .
\end{aligned}$$

This then implies that

$$0 \leqslant (1-t) f(x) + t f(y) - f\left((1-t)x + ty\right) \leqslant \frac{L}{2} t(1-t) \|x - y\|^2 .$$

(vii) $\implies$ (ii) The inequality in (vii) is equivalent to

$$f(y) - f(x) \leqslant \frac{f(x + t(y-x)) - f(x)}{t} + \frac{(1-t)L}{2} \|y - x\|^2 .$$

Taking the limit as $t \searrow 0$ gives us that

$$f(y) - f(x) \leqslant \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|y - x\|^2 .$$

(iii) $\implies$ (vi) Define $z_t \coloneqq x + t(y-x)$. Then

$$f(x) \geqslant f(z_t) + \langle \nabla f(z_t), x - z_t \rangle + \frac{1}{2L} \|\nabla f(x) - \nabla f(z_t)\|_*^2 \tag{♣}$$

$$f(y) \geqslant f(z_t) + \langle \nabla f(z_t), y - z_t \rangle + \frac{1}{2L} \|\nabla f(y) - \nabla f(z_t)\|_*^2 . \tag{♢}$$

Then $(1-t)$ times (♣) plus t times (♢) gives

$$\begin{aligned}
(1-t) f(x) + t f(y) &\geqslant f(z_t) + \left\langle \nabla f(z_t), \underbrace{(1-t)(x - z_t) + t(y - z_t)}_{=0} \right\rangle \\
&\qquad + \frac{1-t}{2L} \|\nabla f(x) - \nabla f(z_t)\|_*^2 + \frac{t}{2L} \|\nabla f(y) - \nabla f(z_t)\|_*^2 \\
&\geqslant f(x + t(y-x)) + \frac{t(1-t)}{2L} \|\nabla f(y) - \nabla f(x)\|_*^2 ,
\end{aligned}$$

where the last inequality follows from the fact that $a\|x\|^2 + (1-a)\|y\|^2 \geqslant a(1-a)\|x - y\|^2$.

(vi) $\implies$ (iii) The inequality in (vi) is equivalent to

$$f(y) - f(x) \geqslant \frac{f(x + t(y-x)) - f(x)}{t} + \frac{1-t}{2L} \|\nabla f(y) - \nabla f(x)\|_*^2 .$$

Taking the limit as $t \searrow 0$ gives us that

$$f(y) - f(x) \geqslant \langle \nabla f(x), y - x \rangle + \frac{1}{2L} \|\nabla f(y) - \nabla f(x)\|_*^2 .$$

The schematic for the implications is complete. $\square$

Theorem 11.1. *Let $L > 0$ and let $\mu > 0$. Then the following statements hold:*

1. *If $f \in \mathcal{F}_L^{1,1}(E)$, then f^* is $\frac{1}{L}$-strongly convex with respect to $\|\cdot\|_*$.*
2. *If $f \in \overline{\text{conv}}_\mu(E)$, then $f^* \in \mathcal{F}^{1,1}_{\mu^{-1}}$.*

Proof. 1. Let $y_1, y_2 \in \text{dom}(\partial f^*)$ and take $v_1 \in \partial f^*(y_1)$ and $v_2 \in \partial f^*(y_2)$. Then since $f \in \overline{\text{conv}}(E)$, we have by Corollary 8.6 that $y_1 \in \partial f(v_1)$ and $y_2 \in \partial f(v_2)$. This then implies that, by Proposition 10.15, $y_1 = \nabla f(v_1)$ and $y_2 = \nabla f(v_2)$. Then by cocoercivity (Lemma 11.7(iv)), we have that

$$\langle y_1 - y_2, v_1 - v_2 \rangle \geqslant \frac{1}{L} \|y_1 - y_2\|_*^2,$$

which, together with Theorem 9.2, implies that f^* is $\frac{1}{L}$-strongly convex with respect to the norm $\|\cdot\|_*$.

2. We have that, by Proposition 8.3 and Corollary 8.6,

$$\partial f^*(x^*) = \arg\min_{y \in E} \{f(y) - \langle x^*, y\rangle\},$$

which is a singleton, and so $f^* \in \mathcal{D}$. Now let $y_1, y_2 \in E$ and take

$$\begin{aligned} v_1 &= \nabla f^*(y_1), \qquad & v_2 &= \nabla f^*(y_2), \\ y_1 &\in \partial f(v_1), \qquad & y_2 &\in \partial f(v_2). \end{aligned}$$

Then by Theorem 9.2, we get $\langle y_1 - y_2, v_1 - v_2\rangle \geqslant \mu \|y_1 - y_2\|^2$, and this gives us that

$$\begin{aligned} \|y_1 - y_2\|_* \|v_1 - v_2\| \geqslant \langle y_1 - y_2, v_1 - v_2 \rangle \geqslant \mu \|v_1 - v_2\|^2 &\implies \|y_1 - y_2\|_* \geqslant \mu \|v_1 - v_2\| \\ &\implies \frac{\|\nabla f^*(y_1) - \nabla f^*(y_2)\|}{\|y_1 - y_2\|} \geqslant \frac{1}{\mu}, \end{aligned}$$

which implies that $f^* \in \mathcal{F}^{1,1}_{\mu^{-1}}$.

$\square$

§12. Lecture 12—26th April, 2024

§12.1. Convergence rate

Let $(V, \|\cdot\|)$ be a normed vector space and let $x = \{x_k\}_{k\in\mathbb{N}}$ denote a sequence in V. Assume $\lim_{k\to+\infty} x_k = x^* \in V$.

Definition 12.1 (Quotient factor)**.** *Let $p \in [1, +\infty)$. The* quotient factor *of the sequence x is defined as*

$$Q_p(x) = \begin{cases} 0 & \text{if } x_k = x^* \text{ for all but finitely many } k \\ \limsup_{k\to+\infty} \dfrac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^p} & \text{if } x_k \neq x^* \text{ for all but finitely many } k \\ +\infty & \text{otherwise.} \end{cases}$$

where $\limsup_{k\to+\infty} \mu_k = \lim_{k\to+\infty} \sup_{n\geqslant k} \mu_n$.

Definition 12.2 (Root factor)**.** *Let $p \in [1, +\infty)$. The* root factor *of the sequence x is defined as*

$$R_p(x) = \begin{cases} \limsup_{k\to+\infty} \|x_k - x^*\|^{1/k} & \text{if } p = 1 \\ \limsup_{k\to+\infty} \|x_k - x^*\|^{1/p^k} & \text{if } p > 1 \end{cases}$$

In the following table, we summarise the characterisations of a sequence based on its (quotient, root) factor properties:

Factor property	Sequence characterisation
$Q_1(x) = 0$	Q-superlinear
$Q_1(x) \in [0, 1)$	Q-linear
$Q_1(x) = 1$	Q-sublinear
$Q_2(x) = 0$	Q-superquadratic
$Q_2(x) \in [0, \infty)$	Q-quadratic
$Q_2(x) = +\infty$	Q-subquadratic
$R_1(x) = 0$	R-superlinear
$R_1(x) \in [0, 1)$	R-linear
$R_1(x) = 1$	R-sublinear
$R_2(x) \in [0, 1)$	R-quadratic

We will now explore the properties of these factors. Let $x = \{x_k\}_{k\in\mathbb{N}}$ be a sequence in V and let $p \in [1, +\infty)$.

Proposition 12.3. *$R_p(x)$ is independent of the choice of the norm $\|\cdot\|$, if V is finite-dimensional.*

Proof. Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be two norms on V. Since V is finite-dimensional, there exist $c_1, c_2 > 0$ such that for all $x \in V$, $c_1 \|x\| \leqslant \|x\|' \leqslant c_2 \|x\|$. Now consider a positive sequence μ_k such that $\mu_k \searrow 0$ as $k \to +\infty$. Then

$$\limsup_{k\to+\infty} \|x_k - x^*\|^{\mu_k} \leqslant \lim_{k\to\infty} \left(\frac{1}{c_1}\right)^{\mu_k} \limsup_{k\to+\infty} \|x_k - x^*\|'^{\mu_k}$$

$$\leqslant \underbrace{\lim_{k\to\infty} \left(\frac{c_2}{c_1}\right)^{\mu_k}}_{=1 \text{ since } a^{\mu_k} \to 1} \limsup_{k\to+\infty} \|x_k - x^*\|^{\mu_k},$$

and since μ_k is arbitrary, we have $R_p(x)$ is the same regardless of the choice of norm. □

Proposition 12.4. *We have that $R_1(x) \leqslant Q_1(x)$ for all sequences x.*

Proof. The case for $Q_1(x) = \infty$ is trivial. So suppose $Q_1(x) < \infty$ and let $\varepsilon_k = \|x_k - x^*\|$. Furthermore let $\varepsilon > 0$ and $\gamma = Q_1(x) + \varepsilon$. Then there must exist some $k_0 \in \mathbb{N}$ such that

$$\varepsilon_k \leqslant \gamma\varepsilon_{k-1} \leqslant \ldots \leqslant \gamma^{k-k_0}\varepsilon_{k_0},$$

for any $k \geqslant k_0$. This implies that

$$\begin{aligned} R_1(x) &= \limsup_{k\to+\infty} \varepsilon_k^{1/k} \leqslant \limsup_{k\to+\infty} \gamma^{(k-k_0)/k}\varepsilon_{k_0}^{1/k} = \limsup_{k\to+\infty} \left(\gamma^{k-k_0}\varepsilon_{k_0}\right)^{1/k} = \gamma \limsup_{k\to+\infty} \frac{\varepsilon_{k_0}}{\gamma^{k_0}} \\ &= \gamma \limsup_{k\to+\infty} \frac{\varepsilon_{k_0}}{\varepsilon_{k_0}} = \gamma = Q_1(x) + \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we have $R_1(x) \leqslant Q_1(x)$. □

Proposition 12.5. *$R_1(x) < 1$ if and only if there exists some $y \in \mathbb{R}^{\mathbb{N}}$ such that $\|x_k - x^*\| \leqslant y_k$ and $Q_1(y) < 1$.*

Proof. The reverse direction is easy: since y is an upper bound for x and by Proposition 12.4, we have $R_1(x) \leqslant R_1(y) \leqslant Q_1(y) < 1$. For the forward direction, recall that $\limsup_{k\to+\infty} \|x_k - x^*\|^{1/k} < 1$ implies that there exists some $\gamma \in (0,1)$ such that $\|x_k - x^*\| \leqslant \gamma^k$ for all $k \in \mathbb{N}$. Let $y_k = \gamma^k$ for very large k and $y_k = 1$ for small k. Then y is an upper bound for x and $Q_1(y) < 1$. □

Proposition 12.6. *Given a sequence satisfying $\mu \in \mathbb{R}^{\mathbb{N}}$, $\mu_k \geqslant 0$, and $\mu_k \searrow 0$ as $k \to +\infty$, then $\mu_k \leqslant \frac{2}{k+2}\sum_{i=\lfloor k/2\rfloor}^{k} \mu_i$ for all $k \in \mathbb{N}$.*

Proof. Clearly $\lfloor k/2\rfloor \leqslant k/2$ and so

$$\left(\frac{k}{2} + 1\right)\mu_k \leqslant \left(k - \left\lfloor\frac{k}{2}\right\rfloor + 1\right)\mu_k \leqslant \sum_{i=\lfloor k/2\rfloor}^{k} \mu_i,$$

since $\mu_k \geqslant 0$. Dividing by $k+2$ gives the desired result. □

§12.2. The gradient method

Let $(H, \langle\cdot,\cdot\rangle)$ be a Hilbert space equipped with the norm $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$. Let $f\colon H \to \mathbb{R}$ be a continuously differentiable function and let $x^* \in \operatorname{dom} f$ be a minimiser of f.

The gradient method is a first-order optimisation algorithm that iteratively updates the sequence $x = \{x_k\}_{k\in\mathbb{N}}$ as $x_{k+1} = x_k - \alpha_k \nabla f(x_k)$, where α_k is the step size at iteration k; it is presented below.

Algorithm: Gradient method

- Choose a starting point $x_0 \in H$ and a sequence $\{\alpha_k\}_{k\in\mathbb{N}}$ of step sizes.
- For $k = 0, 1, 2, \ldots$ until convergence:
  - Compute the gradient $\nabla f(x_k)$.
  - Update the sequence as $x_{k+1} = x_k - \alpha_k \nabla f(x_k)$.

The earliest use of this method dates back to the 19th century, when Cauchy (1847) proposed the method for solving non-linear equations like

$$f_1(x) = \cdots = f_m(x) = 0 \longrightarrow \inf_{x\in\mathbb{R}^n} f_1(x)^2 + \cdots + f_m(x)^2.$$

The method was later popularised by Rosenbrock (1960) and Fletcher and Reeves (1964) for unconstrained optimisation problems. We present some convergence properties of the gradient method below.

Proposition 12.7 (Sublinear convergence of the gradient method)**.** *If $f \in \mathcal{C}_L^{1,1}$, $\inf f > -\infty$, and $0 < \varepsilon_1 \leqslant \alpha_k \leqslant 2/L - \varepsilon_2$, then the gradient method gives a sequence $x = \{x_k\}_{k\in\mathbb{N}}$ such that*

$$\min_{i=0,\ldots,k} \|\nabla f(x_i)\| = o\left(\frac{1}{\sqrt{k}}\right).$$

Proof. First we achieve a following bound for successive evaluations on the sequence x:

$$\begin{aligned}
f(x_{k+1}) &\leqslant f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k\rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2 \\
&= f(x_k) - \alpha_k \|\nabla f(x_k)\|^2 + \frac{L}{2}\alpha_k^2 \|\nabla f(x_k)\|^2 \\
&= f(x_k) - \alpha_k \frac{L}{2}\left(\frac{2}{L} - \alpha_k\right)\|\nabla f(x_k)\|^2 \\
&\leqslant f(x_k) - \frac{L}{2}\varepsilon_1\varepsilon_2 \|\nabla f(x_k)\|^2 .
\end{aligned}$$

The first inequality follows from the Lipschitzness of f, as shown in Proposition 11.7(ii). The second inequality follows from the choice of α_k and the fact that $\varepsilon_1 \leqslant 2/L - \varepsilon_2$. The last inequality follows from the fact that $\alpha_k \leqslant 2/L - \varepsilon_2$.

The above bound implies that

$$\begin{aligned}
\min_{i=0,\ldots,k} \|\nabla f(x_i)\|^2 \leqslant \|\nabla f(x_k)\|^2 &\leqslant \frac{2}{L\varepsilon_1\varepsilon_2}\left(f(x_k) - f(x_{k+1})\right) \\
&\leqslant \frac{2}{k+2}\sum_{i=\lfloor k/2\rfloor}^{k} \frac{2}{L\varepsilon_1\varepsilon_2}\left(f(x_i) - f(x_{i+1})\right) \\
&\leqslant \frac{4}{L\varepsilon_1\varepsilon_2(k+2)}\left(f\left(x_{\lfloor k/2\rfloor}\right) - f(x_{k+1})\right) \\
&\leqslant \frac{4}{L\varepsilon_1\varepsilon_2(k+2)}\left(f\left(x_{\lfloor k/2\rfloor}\right) - \lim_{i\to+\infty} f(x_i)\right),
\end{aligned}$$

and taking the equare root of both sides gives

$$\min_{i=0,\dots,k} \|\nabla f(x_i)\| \leqslant \sqrt{\frac{4\left(f\left(x_{\lfloor k/2 \rfloor}\right) - \lim_{i\to+\infty} f(x_i)\right)}{L\varepsilon_1\varepsilon_2(k+2)}} \leqslant \varepsilon,$$

and after $k = o\left(1/\sqrt{\varepsilon}\right)$ iterations, we have $\|\nabla f(x_i)\| \leqslant \varepsilon$ for all $0 \leqslant i \leqslant k$. □

We now obtain such a rate result for the gradient method applied on functions that belong to a more restricted class.

Proposition 12.8. *If $f \in \mathcal{F}_L^{1,1}$, there exists $x^* \in E$ such that $f(x^*) = \inf f = 0$, and $0 < \varepsilon_1 \leqslant \alpha_k \leqslant 2/L - \varepsilon_2$, then the gradient method gives a sequence $x = \{x_k\}_{k\in\mathbb{N}}$ such that $f(x_k) = o\left(\frac{1}{k}\right)$.*

Proof. Given that f is convex, we have that by the (sub)gradient inequality that

$$\underbrace{f(x^*)}_{=0} \geqslant f(x_k) + \langle \nabla f(x_k), x^* - x_k\rangle \implies \langle \nabla f(x_k), x_k - x^*\rangle \geqslant f(x_k).$$

Since $f \in \mathcal{C}_L^{1,1}$, we have from the proof of the previous proposition that

$$f(x_{k+1}) \leqslant f(x_k) - \frac{L}{2}\alpha_k\left(\frac{2}{L} - \alpha_k\right)\|\nabla f(x_k)\|^2,$$

and so

$$\begin{aligned}
\|x_{k+1} - x^*\|^2 &= \|x_k - \alpha_k \nabla f(x_k) - x^*\|^2 \\
&= \|x_k - x^*\|^2 - 2\alpha_k \langle \nabla f(x_k), x_k - x^*\rangle + \alpha_k^2 \|\nabla f(x_k)\|^2 \\
&\leqslant \|x_k - x^*\|^2 - 2\alpha_k f(x_k) + \frac{2}{L}\cdot\frac{\alpha_k}{\frac{2}{L} - \alpha_k}(f(x_k) - f(x_{k+1})) \\
&\leqslant \|x_k - x^*\|^2 - 2\varepsilon_1 f(x_k) + \frac{2}{L}\cdot\frac{\varepsilon_2}{\varepsilon_1}(f(x_k) - f(x_{k+1})).
\end{aligned}$$

So we can write

$$f(x_k) \leqslant \frac{1}{2\varepsilon_1}\left(\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2\right) + \frac{\varepsilon_2}{L\varepsilon_1^2}\left(f(x_k) - f(x_{k+1})\right),$$

and summing over $i = 0, \dots, k$ gives

$$\sum_{i=0}^{k} f(x_i) \leqslant \frac{1}{2\varepsilon_1}\left(\|x_0 - x^*\|^2\right) + \frac{\varepsilon_2}{L\varepsilon_1^2} f(x_0) \implies \sum_{i=0}^{k} f(x_i) < \infty.$$

Therefore $f(x_k) \leqslant \frac{2}{k+2}\sum_{i=\lfloor k/2\rfloor}^{k} f(x_i) \leqslant \frac{2}{k+2}\sum_{i=\lfloor k/2\rfloor}^{\infty} f(x_i) = o\left(\frac{1}{k}\right)$. □

As it turns out, if f is μ-*strongly* convex as opposed to merely convex (and L-smooth of course), then we have the following stronger lower bound.

Proposition 12.9. *If $f \in \mathcal{S}^{1,1}_{\mu,L}$, then for all $x, y \in E$, we have*

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geqslant \frac{\mu L}{\mu + L}\|x - y\|^2 + \frac{1}{\mu + L}\|\nabla f(x) - \nabla f(y)\|^2.$$

Proof. We have $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geqslant \mu\|x-y\|^2$ and $\|\nabla f(x) - \nabla f(y)\| \leqslant L\|x-y\|$. Construct an auxiliary function $\phi = f - \frac{\mu}{2}\|\cdot\|^2$, and note that: (1) ϕ is convex (clearly), (2) $\nabla \phi = \nabla f - \mu x$ (by the chain rule), and (3) the following relation holds:

$$\begin{aligned}\langle \nabla \phi(x) - \nabla \phi(y), x - y \rangle &= \underbrace{\langle \nabla f(x) - \nabla f(y), x - y \rangle - \mu\|x - y\|^2}_{\geqslant 0} \\ &\leqslant (L - \mu)\|x - y\|^2,\end{aligned}$$

by Cauchy-Schwarz. This implies that ϕ is $(L-\mu)$-smooth, and so $\phi \in \mathcal{F}^{1,1}_{L-\mu}$. We distinguish cases as follows, noting that $\mu > L$ is not possible:

- *Case I:* $\mu = L$. In this case, we have

$$\begin{aligned}\langle \nabla f(x) - \nabla f(y), x - y \rangle &= \frac{1}{2}\langle \nabla f(x) - \nabla f(y), x - y \rangle + \frac{1}{2}\langle \nabla f(x) - \nabla f(y), x - y \rangle \\ &\geqslant \underbrace{\frac{\mu}{2}\|x - y\|^2}_{\text{by strong convexity}} + \underbrace{\frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2}_{\text{by cocoercivity (Lemma 11.7(iv))}},\end{aligned}$$

  as desired.

- *Case II:* $\mu < L$. By the cocoercivity of $\phi \in \mathcal{F}^{1,1}_{L-\mu}$, we know that

$$\begin{aligned}\langle \nabla \phi(x) - \nabla \phi(y), x - y \rangle &\geqslant \frac{1}{L-\mu}\|\nabla \phi(x) - \nabla \phi(y)\|^2 \\ \langle \nabla f(x) - \nabla f(y), x - y \rangle &\geqslant \mu\|x - y\|^2 + \frac{1}{L-\mu}\|\nabla f(x) - \nabla f(y) - \mu(x - y)\|^2 \\ \left(1 + \frac{2\mu}{L-\mu}\right)\langle \nabla f(x) - \nabla f(y), x - y \rangle &\geqslant \left(\mu + \frac{\mu^2}{L-\mu}\right)\|x - y\|^2 \frac{1}{L-\mu}\|\nabla f(x) - \nabla f(y)\|^2 \\ \frac{L+\mu}{L-\mu}\langle \nabla f(x) - \nabla f(y), x - y \rangle &\geqslant \frac{\mu L}{L-\mu}\|x - y\|^2 + \frac{1}{L-\mu}\|\nabla f(x) - \nabla f(y)\|^2.\end{aligned}$$

  Multiplying through by $\frac{L-\mu}{L+\mu}$ gives the desired result. □

Again, another result, before a break.

Proposition 12.10. *If $f \in \mathcal{S}^{1,1}_{\mu,L}$, there exists $x^* \in E$ such that $f(x^*) = \inf f > -\infty$, and $0 < \varepsilon_1 \leqslant \alpha_k \leqslant \frac{2}{\mu+L}$, then x_k converges Q-linearly under the gradient method.*

Proof. By the update rule for the gradient method,

$$\begin{aligned}\|x_{k+1} - x^*\|^2 &= \|x_k - \alpha_k \nabla f(x_k) - x^*\|^2 \\ &= \|x_k - x^*\|^2 - 2\alpha_k \langle \nabla f(x_k), x_k - x^* \rangle + \alpha_k^2 \|\nabla f(x_k)\|^2\end{aligned}$$

$$
\begin{aligned}
&\leqslant \|x_k - x^*\|^2 - 2\alpha_k \frac{\mu L}{\mu + L}\|x_k - x^*\|^2 - 2\alpha_k \frac{1}{\mu + L}\|\nabla f(x_k) - \nabla f(x^*)\|^2 + \alpha_k^2 \|\nabla f(x_k)\|^2 \\
&\leqslant \left(1 - 2\alpha_k \frac{\mu L}{\mu + L}\right)\|x_k - x^*\|^2 + \alpha_k \underbrace{\left(\alpha_k - \frac{2}{\mu + L}\right)}_{\leqslant 0}\|\nabla f(x_k)\|^2 \\
&\leqslant \left(1 - \frac{2\alpha_k \mu L}{\mu + L}\right)\|x_k - x^*\|^2.
\end{aligned}
$$

If $\alpha_k = \frac{2}{\mu+L}$, then we have that $1 - \frac{2\alpha_k \mu L}{\mu+L} = \left(\frac{\mu-L}{\mu+L}\right)^2$, and so

$$
\|x_{k+1} - x^*\|^2 \leqslant \left(\frac{K-1}{K+1}\right)^2 \|x_k - x^*\|^2,
$$

where $K = \frac{L}{\mu} \in [1, +\infty)$. Also, $f(x_k) \leqslant f(x^*) + \langle \nabla f(x^*), x_k - x^* \rangle + \frac{L}{2}\|x_k - x^*\|^2$, and since $\nabla f(x^*) = 0$, we have

$$
f(x_k) - f(x^*) \leqslant \frac{L}{2}\|x_k - x^*\|^2 \leqslant \frac{L}{2}\left(\frac{K-1}{K+1}\right)^{2k}\|x_0 - x^*\|^2,
$$

which implies that $\limsup_{k\to+\infty} \frac{\|x_{k+1}-x^*\|}{\|x_k - x^*\|} \in (0,1)$. □

Proposition 12.11. *If $f \in \mathcal{C}([a,b],E) \cap \mathcal{D}((a,b),E)$, then there exists some $x \in (a,b)$ such that $\|f(a) - f(b)\| \leqslant (b-a)\|f'(x)\|$.*

Proof. Define $g(t) = \langle f(b) - f(a), f(t) \rangle \in \mathbb{R}$, then there exists $x \in (a,b)$ such that $g(b) - g(a) = g'(x)(b-a)$, and so

$$
\begin{aligned}
\langle f(b) - f(a), f(b) \rangle - \langle f(b) - f(a), f(a) \rangle &= \left\langle f(b) - f(a), f'(x) \right\rangle (b-a) \\
&\leqslant \|f(b) - f(a)\|\|f'(x)\| \cdot (b-a),
\end{aligned}
$$

and so

$$
\|f(b) - f(a)\|^2 \leqslant \|f(b) - f(a)\|\|f'(x)\| \cdot (b-a) \iff \|f(b) - f(a)\| \leqslant (b-a)\|f'(x)\|,
$$

as desired. □

Proposition 12.12. *If $f \in \mathcal{S}^{1,1}_{\mu,L} \cap \mathcal{D}^2$, there exists some $x^* \in E$ such that $f(x^*) = \inf f > -\infty$, and $0 < \varepsilon_1 \leqslant \alpha_k \leqslant \varepsilon_2 < \frac{2}{L}$, then x_k converges Q-linearly under the gradient method.*

Proof. We have (using Proposition 12.9 and for $z_k \in (x^*, x_k)$) that

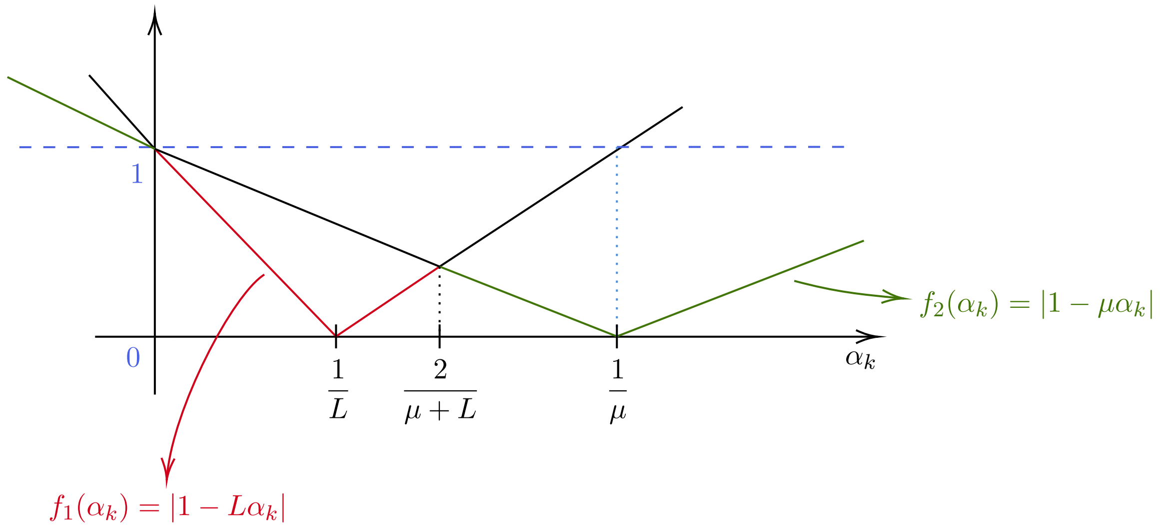
$$
\begin{aligned}
\|x_{k+1} - x^*\| &= \|x_k - \alpha_k \nabla f(x_k) - (x^* - \alpha_k \nabla f(x^*))\| \\
&= \|(I - \alpha_k \nabla f)(x_k) - (I - \alpha_k \nabla f)(x^*)\| \\
&\leqslant \left\|\left(I - \alpha_k \nabla^2 f(z_k)\right)(x_k - x^*)\right\| \\
&\leqslant \left\|I - \alpha_k \nabla^2 f(z_k)\right\| \|x_k - x^*\|.
\end{aligned}
$$

Indeed, let $g(t) = (I - \alpha_k \nabla f)(x^* + (x_k - x^*)t) \in \mathcal{C}([0,1], E) \cap \mathcal{D}((0,1), E)$, then there exists some $t \in (0,1)$ such that $\|g(1) - g(0)\| \leqslant \|g'(t)\|$, and so

$$\|(I - \alpha_k \nabla f)(x_k) - (I - \alpha_k \nabla f)(x^*)\| \leqslant \left\|I - \alpha_k \nabla^2 f(z_k)\right\| \|x_k - x^*\|,$$

where $z_k = x^* + (x_k - x^*)t \to z$ as $k \to +\infty$. We have, for $f \in \mathcal{S}_{\mu,L}^{1,1}$, that $\mu I \leqslant \nabla^2 f \leqslant LI$, and so

$$\left\|(I - \alpha_k \nabla^2 f)(z)\right\| = \max\{|1 - \alpha_k \mu|, |1 - \alpha_k L|\} < 1.$$



If $\alpha_k = 2/(\mu + L)$, then we have that $\|x_{k+1} - x^*\| \leqslant \left(1 - \frac{2\mu L}{\mu + L}\right) \|x_k - x^*\|$, as shown previously. Also,

$$\begin{aligned} f(x_k) - f(x^*) &\leqslant \left\langle \nabla \underbrace{f(x^*)}_{=0}, x_k - x^* \right\rangle + \frac{L}{2}\|x_k - x^*\|^2 \\ &\leqslant \frac{L}{2}\left(\frac{K-1}{K+1}\right)^{2k} \|x_0 - x^*\|^2, \end{aligned}$$

which implies that $\limsup_{k \to +\infty} \dfrac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \in (0,1)$. Thus x_k converges Q-linearly. $\square$

We now discuss an extension of this linear convergence rate for strongly convex functions. First we state a famous inequality due to Polyak-Łojasiewicz.

Proposition 12.13 (Polyak-Łojasiewicz inequality)**.** *For strongly convex $\mathcal{D}^1$ functions (with norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$) of the form*

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|y - x\|^2,$$

where $f \colon E \to \mathbb{R}$ is μ-strongly convex with $\mu > 0$ and $x, y \in E$, we have that

$$\frac{1}{2\mu}\|\nabla f(x)\|^2 \geqslant f(x) - f(x^*) \geqslant \frac{\mu}{2}\|x - x^*\|^2 \implies \|\nabla f(x)\|^2 \geqslant 2\mu(f(x) - f(x^*)),$$

where x^* is the minimiser of f.

Proposition 12.14. *If $f \in \mathcal{C}_L^{1,1}$ is non-constant, $f^* = \inf f > -\infty$, there exists $\mu > 0$ for all $x \in E$ such that $\|\nabla f(x)\|^2 \geqslant 2\mu(f(x) - f^*)$, and $0 < \varepsilon_1 \leqslant \alpha_k \leqslant 2/L - \varepsilon_2$, then $f(x_k)$ converges Q-linearly and x_k converges R-linearly under the gradient method.*

Proof. Once again,

$$\begin{aligned}
f(x_{k+1}) &\leqslant f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2 \\
&= f(x_k) - \alpha_k \|\nabla f(x_k)\|^2 + \frac{L}{2}\alpha_k^2 \|\nabla f(x_k)\|^2 \\
&= f(x_k) - \alpha_k \frac{L}{2}\left(\frac{2}{L} - \alpha_k\right) \|\nabla f(x_k)\|^2 \\
&\leqslant f(x_k) - \mu L \alpha_k \left(\frac{2}{L} - \alpha_k\right)(f(x_k) - f^*).
\end{aligned}$$

This implies that

$$f(x_{k+1}) - f^* \leqslant \underbrace{\left(1 - \mu\left(\alpha_k\left(\frac{2}{L} - \alpha_k\right)\right)\right)}_{\leqslant 1 - \mu L \varepsilon_1 \varepsilon_2 =: q \in [0,1)}(f(x_k) - f^*),$$

and

$$\begin{aligned}
\|x_{k+1} - x_k\| &= \alpha_k \|\nabla f(x_k)\| \\
&\leqslant \sqrt{\frac{(f(x_k) - f(x_{k+1}))\,\alpha_k}{\frac{L}{2}\cdot\left(\frac{2}{L} - \alpha_k\right)}} \\
&\leqslant \sqrt{\frac{q^k\,(f(x_0) - f^*)\,\alpha_k}{\frac{L}{2}\cdot\left(\frac{2}{L} - \alpha_k\right)}} \\
&\leqslant C \cdot \sqrt{q^k}
\end{aligned}$$

where $C = \sqrt{(f(x_0) - f^*)\cdot\left(\frac{4}{L^2\varepsilon_2} - \frac{2}{L}\right)}$. Now let $n > \ell \geqslant 0$ be integers. Then

$$\begin{aligned}
\|x_\ell - x_n\| &= \left\|\sum_{i=\ell}^{n-1}(x_{i+1} - x_i)\right\| \leqslant \sum_{i=\ell}^{n-1}\|x_{i+1} - x_i\| \\
&\leqslant C\sum_{i=\ell}^{n-1}\sqrt{q^i} = C\sqrt{q^\ell}\sum_{i=0}^{n-\ell-1} q^i \leqslant C\sqrt{q^\ell}\frac{1}{1 - \sqrt{q}}.
\end{aligned}$$

Thus $\{x_k\}_{k\in\mathbb{N}}$ is a Cauchy sequence, and so converges to some $x^* \in E$. Hence

$$\|x_k - x^*\| \leqslant \frac{C}{1 - \sqrt{q}}\sqrt{q^k} \implies \lim_{k\to+\infty}\|x_k - x^*\| = 0.$$

The proof is completed by noting that $f(x_k) - f^* \leqslant q^k(f(x_0) - f^*)$. $\square$

Two final results.

Proposition 12.15 (Lojasiewicz, 1963)**.** *Let $x^* \in \mathbb{R}^n$, and let $f\colon \mathbb{R}^n \to \mathbb{R}$ be analytic in a neighbourhood of x^* such that $f(x^*) = 0$. Then there exists $\mathscr{U} \in N(x^*)$, $c > 0$, and $\vartheta \in (0,1)$ such that for all $x \in \mathscr{U}$, we have $\|\nabla f(x)\| \geqslant c|f(x)|^\vartheta$.*

Polyak is the special case where $\vartheta = \frac{1}{2}$ and $\mathscr{U} = \mathbb{R}^n$:

$$f(x_{k+1}) \leqslant f(x_k) - \frac{L}{2}\alpha_k \left(\frac{2}{L} - \alpha_k\right) \|\nabla f(x_k)\|^2 \leqslant f(x_k) - C \cdot f(x_k)^{2\vartheta}.$$

If $\vartheta \in (0, \frac{1}{2}]$, then we get linear convergence using Polyak's proof. If $\vartheta \in (\frac{1}{2}, 1)$, then we get a sublinear rate; we present a quick proof now. Assume $f(x_k) \leqslant \alpha k^{-\beta}$ (i.e. a sublinear rate). We will prove the result by induction (noting that $x \mapsto x - Cx^{2\vartheta}$ is increasing). We have

$$\begin{aligned}
f(x_{k+1}) \leqslant \alpha k^{-\beta} - C \cdot k^{-2\vartheta\beta} &= \alpha(k+1)^\beta \left(\left(\frac{k+1}{k}\right)^\beta - C\alpha^{2\vartheta}\frac{(k+1)^\beta}{k^{2\vartheta\beta}}\right) \\
&\leqslant \alpha(k+1)^{-\beta}\left(\left(1 + \frac{1}{k}\right)^\beta - C\alpha^{2\vartheta}k^{-(2\vartheta-1)\beta}\right) \\
&\leqslant \alpha(k+1)^{-\beta}\left(1 + 2\frac{\beta}{k} - C\alpha^{2\vartheta}k^{-(2\vartheta-1)\beta}\right) \\
&= \alpha(k+1)^{-\beta}\left(1 + \frac{2\beta - C\alpha^{2\vartheta}}{k}\right) \leqslant \alpha(k+1)^{-\beta}.
\end{aligned}$$

Here we have set $\beta = (2\vartheta - 1)^{-1}$, and the result holds if $2\beta - C\alpha^{2\vartheta} \leqslant 0$, *i.e.* $2\beta \leqslant C\alpha^{2\vartheta}$.

Proposition 12.16 (Kurdyka, 1998)**.** *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be analytic and $X \subset \mathbb{R}^n$ be bounded. Let $\nu \in \mathbb{R}$. Then there exists $\rho > 0$ and a strictly increasing continuous analytic function $\psi\colon [0,\rho) \to (0,\infty) \in \mathcal{C}^1$ with $\psi(0) = 0$, such that for all $x \in X$,*

$$0 < |f(x) - \nu| < \rho \implies \|\nabla(\psi_0|f - \nu|)(x)\| \leqslant 1.$$

Actually we can even take concave ψ.

§13. Lecture 13—03rd May, 2024

§13.1. The heavy ball method

When we apply the gradient method to a convex quadratic, one observes that there are many oscillations when the quadratic is ill-conditioned. The heavy ball method is a modification of the gradient method that mitigates these oscillations. This method is implied by the following ordinary differential equation:

$$\ddot{x} + (1-\beta)\dot{x} + \nabla f(x) = 0.$$

Ignoring the first term, the term $(1-\beta)\dot{x} + \nabla f(x) = 0$ is like a gradient flow with a friction term. The heavy ball method is a discretisation of this ODE in the following way:

$$\alpha \cdot \frac{\frac{x_{k+1}-x_k}{\alpha} - \frac{x_k - x_{k-1}}{\alpha}}{\alpha} + (1-\beta)\frac{x_{k+1}-x_k}{\alpha} + \nabla f(x_k) = 0,$$

which simplifies to $x_{k+1} - x_k - (x_k - x_{k-1}) + (1-\beta)(x_{k+1} - x_k) + \alpha \nabla f(x_k) = 0$. Thus we obtain the following canonical multi-step form:

$$x_{k+1} = x_k - \alpha \nabla f(x_k) + \beta(x_k - x_{k-1}),$$

where $\beta \in [0,1)$ is the momentum parameter that keeps a memory of the displacement in the previous iteration. We then summarise the heavy ball method as follows:

Algorithm: Heavy ball method

- Choose $x_0 \in \mathbb{R}^n$, $\alpha > 0$, $\beta \in [0,1)$.
- For $k = 0, 1, 2, \ldots$ do
  - $x_{k+1} = x_k - \alpha \nabla f(x_k) + \beta(x_k - x_{k-1})$.
  - If $\|\nabla f(x_k)\| < \varepsilon$ for sufficiently small ε, stop.

There are many details we're glossing over, but we will focus on concrete results for the quadratic case (as opposed to the more general (and, of course, more difficult) case proposed by Polyak for functions $f \in \mathcal{S}^{1,1}_{\mu,L}$). We will focus on the quadratic case in this course.

Proposition 13.1. *Let $f(x) = \frac{1}{2}(x-x^*)^\top A(x-x^*)$ where $0 < \mu < L$ are the smallest and largest eigenvalues of A respectively. Take the parameters*

$$\alpha = \left(\frac{2}{\sqrt{L} + \sqrt{\mu}}\right)^2, \quad \beta = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1},$$

where $\kappa = \frac{L}{\mu} \geqslant 1$ is the condition number of A. Then

$$\|x_k - x^*\| \leqslant \sqrt{2} \cdot \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k \cdot \|x_0 - x^*\|,$$

for all k sufficiently large.

Proof. This is a multi-step method—the update depends on the two previous steps as opposed to the immediate past step, which is new for us—so we will try to reduce this to a single step system by increasing the dimension of the state. For $k \in \mathbb{N}^* = \{1, 2, 3, \dots\}$, we have

$$\begin{aligned}
\begin{pmatrix} x_{k+1} - x^* \\ x_k - x^* \end{pmatrix} &= \begin{pmatrix} x_k - x^* - \alpha\nabla f(x_k) + \beta(x_k - x_{k-1}) \\ x_{k-1} - x^* \end{pmatrix} \\
&= \begin{pmatrix} x_k - \alpha A(x_k - x^*) + \beta(x_k - x_{k-1}) - x^* \\ x_{k-1} - x^* \end{pmatrix} \\
&= \begin{pmatrix} (I - \alpha A + \beta I)(x_k - x_{k-1}) - \alpha A(x_{k-1} - x^*) - x^* \\ x_{k-1} - x^* \end{pmatrix} \\
&= \begin{pmatrix} (1+\beta)I - \alpha A & -\beta I \\ I & 0 \end{pmatrix} \begin{pmatrix} x_k - x^* \\ x_{k-1} - x^* \end{pmatrix}.
\end{aligned}$$

Taking norms then gives

$$\begin{aligned}
\left\| \begin{pmatrix} x_{k+1} - x^* \\ x_k - x^* \end{pmatrix} \right\| &= \left\| \begin{pmatrix} (1+\beta)I - \alpha A & -\beta I \\ I & 0 \end{pmatrix}^k \begin{pmatrix} x_1 - x^* \\ x_0 - x^* \end{pmatrix} \right\| \\
&\leqslant \left\| \begin{pmatrix} (1+\beta)I - \alpha A & -\beta I \\ I & 0 \end{pmatrix}^k \right\| \cdot \left\| \begin{pmatrix} x_1 - x^* \\ x_0 - x^* \end{pmatrix} \right\| \\
&\leqslant \left[\rho \begin{pmatrix} (1+\beta)I - \alpha A & -\beta I \\ I & 0 \end{pmatrix} + \varepsilon \right]^k \cdot \left\| \begin{pmatrix} x_1 - x^* \\ x_0 - x^* \end{pmatrix} \right\| \qquad (\star) \\
&= \left[\max_{i=1,\dots,n} \rho \cdot \begin{pmatrix} (1+\beta) - \alpha\lambda_i & -\beta \\ 1 & 0 \end{pmatrix} + \varepsilon \right]^k \cdot \left\| \begin{pmatrix} x_1 - x^* \\ x_0 - x^* \end{pmatrix} \right\| \\
&= (\sqrt{\beta} + \varepsilon)^k \cdot \left\| \begin{pmatrix} x_1 - x^* \\ x_0 - x^* \end{pmatrix} \right\| \\
&= (\sqrt{\beta} + \varepsilon)^k \cdot \|x_0 - x^*\|,
\end{aligned}$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A and ρ is the spectral radius of the matrix. In ($\star$), we used Gelfand's formula which is valid for every norm and which states that for any matrix $M \in \mathbb{R}^{n \times n}$,

$$\lim_{k\to\infty} \left\| M^k \right\|^{1/k} = \rho(M) \coloneqq \sup\{\lambda \in \mathbb{C} : \det(M - \lambda I) = 0\}.$$

The eigenvalues of the matrix $\begin{pmatrix} (1+\beta) - \alpha\lambda_i & -\beta \\ 1 & 0 \end{pmatrix}$ are exactly the roots of the characteristic polynomial $X^2 - (1 + \beta - \alpha\lambda_i)X + \beta = 0$, whose determinant is equal to

$$(1 + \beta - \alpha\lambda_i)^2 - 4\beta = 8 \cdot \frac{(\lambda_i - L)(\lambda_i - \mu)}{(\sqrt{L} + \sqrt{\mu})^2} \leqslant 0.$$

Thus the polynomial is of the form $(x - z)(x - \overline{z}) = 0$ where $z \in \mathbb{C}$, and the spectral radius is equal to $|z| = \sqrt{z\overline{z}} = \sqrt{\beta}$. Since $\beta \in [0, 1)$, we have $\sqrt{\beta} < 1$ and the result follows. □

§13.2. Nesterov's accelerated gradient method

We will only focus on the case for convex functions (as opposed to the typical strongly convex case) in this course. First we take a detour and revisit the gradient method from last time.

1. Consider the continuous time version of the gradient method:

$$\dot{x} + \nabla f(x) = 0,$$

where $f \in \mathcal{F}^{1,1}$ and $x^* \in \arg\min f \neq \varnothing$ and $x\colon [0,\infty) \to E$ is $\mathcal{C}^1$. We will show that we can derive a rate in continuous-time quickly using the idea of finding a Lyapunov function along which the method is monotonic. Due to convexity, we can take the energy functional to be the distance to the minimiser of f:

$$\begin{aligned}
\mathcal{E}(x) &\coloneqq t(f(x) - f(x^*)) + \frac{1}{2}\|x - x^*\|^2 \\
\dot{\mathcal{E}}(x) &= f(x) - f(x^*) + t\langle \nabla f(x), \underbrace{\dot{x}}_{=-\nabla f(x)} \rangle + \underbrace{\langle x - x^*, \underbrace{\dot{x}}_{=-\nabla f(x)} \rangle}_{\leqslant f(x^*)-f(x) \text{ since } f \text{ is convex}} \\
&\leqslant -t\left\|\nabla f(x)\right\|^2 \leqslant 0,
\end{aligned}$$

and hence $\mathcal{E}(x(t)) \leqslant \mathcal{E}(x(0))$ for all $t \geqslant 0$. (This is a Lyapunov function for the gradient method.) Thus we can write

$$t\left(f(x(t)) - f(x^*)\right) \leqslant t\left(f(x(t)) - f(x^*)\right) + \underbrace{\frac{1}{2}\left\|x(t) - x^*\right\|^2}_{\geqslant 0} \leqslant \frac{1}{2}\left\|x(0) - x^*\right\|^2,$$

and hence

$$f(x(t)) - f(x^*) \leqslant \frac{1}{2t}\left\|x(0) - x^*\right\|^2.$$

This is somewhat nice because, more or less, the entire difficulty relied on simply finding the energy functional, and life was fairly straightforward from there.

2. Inspired by this—and despite the inescapable difficulty of doing acceleration—we consider a new problem:

$$\ddot{x} + \frac{3}{t}\dot{x} + \nabla f(x) = 0.$$

Here the term $3/t$ corresponds to a friction term decreasing with time (like we have in the model of the motion of a particle in a fluid f) and the unit coefficient of the $\ddot{x}$ term corresponds to the mass. The energy functional is then

$$\begin{aligned}
\mathcal{E}(x) &= t^2 \cdot (f(x) - f(x^*)) + 2\left\| x + \frac{t}{2}\cdot\dot{x} - x^* \right\|^2 \\
\dot{\mathcal{E}}(x) &= 2t\left(f(x) - f(x^*)\right) + t^2\left\langle \nabla f(x), \dot{x}\right\rangle + 4\left\langle x + \frac{t}{2}\dot{x} - x^*, \underbrace{\dot{x} + \frac{1}{2}\dot{x} + \ddot{x}}_{=\frac{3}{2}\dot{x}+\frac{t}{2}\ddot{x}=-\frac{t}{2}\nabla f(x)} \right\rangle \\
&= 2t\left(f(x) - f(x^*)\right) + t^2\left\langle \nabla f(x), \dot{x}\right\rangle + 4\left\langle x + \frac{t}{2}\dot{x} - x^*, -\frac{t}{2}\nabla f(x)\right\rangle \\
&= 2t\left(f(x) - f(x^*)\right) + \underbrace{t^2\left\langle \nabla f(x), \dot{x}\right\rangle + 4\left\langle \frac{t}{2}\dot{x}, -\frac{t}{2}\nabla f(x)\right\rangle}_{=0} + 4\left\langle x - x^*, -\frac{t}{2}\nabla f(x)\right\rangle \\
&= 2t\left(f(x) - f(x^*) - \langle x^* - x, \nabla f(x)\rangle\right) \leqslant 0,
\end{aligned}$$

since f is convex. Thus we obtain

$$f(x(t)) - f(x^*) \leqslant \frac{2}{t^2} \left\| x(0) - x^* \right\|^2,$$

if $\dot{x}(0) = 0$.

3. Consider now the system

$$\begin{cases} x_k = y_{k-1} - s\nabla f(y_{k-1}) \\ y_k = x_k + \dfrac{k-1}{k+1} \cdot (x_k - x_{k-1}), \end{cases}$$

Here the momentum/inertia is $(k-1)/(k+1)$, which is a decreasing function in the variable k, and converges to 1 as $k \to \infty$. Let us try to motivate the connection between this system and the differential equation we saw in **2.** We have the following compact form for this system:

$$x_{k+1} = y_k - s\nabla f(y_k) = x_k + \frac{k-1}{k+2}(x_k - x_{k-1}) - s\nabla f(y_k),$$

and so we can write

$$\frac{x_{k+1} - x_k}{\sqrt{s}} = \frac{k-1}{k+2} \cdot \frac{x_k - x_{k-1}}{\sqrt{s}} - \nabla f(y_k).$$

Let $x_k \approxeq x(t) \approxeq x(k\sqrt{s})$, and let $t = \sqrt{s}k$. Then as $s \to 0$, we have

$$\begin{aligned} x(t) &\approxeq x_{t/\sqrt{s}} = x_k \\ x(t+\sqrt{s}) &\approxeq x_{(t+\sqrt{s})/\sqrt{s}} = x_{k+1}, \end{aligned}$$

so that by a Taylor approximation we have

$$\begin{aligned} \frac{x_{k+1} - x_k}{\sqrt{s}} &= \frac{x(t+\sqrt{s}) - x(t)}{\sqrt{s}} = \dot{x}(t) + \frac{1}{2}\ddot{x}(t)\sqrt{s} + o(\sqrt{s}) \\ \frac{x_k - x_{k-1}}{\sqrt{s}} &= \frac{x(t) - x(t-\sqrt{s})}{\sqrt{s}} = \dot{x}(t) - \frac{1}{2}\ddot{x}(t)\sqrt{s} + o(\sqrt{s}), \end{aligned}$$

Thus

$$\begin{aligned} \dot{x}(t) + \frac{1}{2}\ddot{x}(t)\sqrt{s} &= \frac{k-1}{k+2}\left(\dot{x}(t) - \frac{1}{2}\ddot{x}(t)\sqrt{s} + o(\sqrt{s})\right) - \sqrt{s}\nabla f(y_k) \\ &= \left(1 - \frac{3}{k+2}\right) \cdot \left(\dot{x}(t) + \frac{1}{2}\ddot{x}(t)\sqrt{s} + o(\sqrt{s})\right) - \sqrt{s}\nabla f(x(t)) + o(\sqrt{s}) \\ &\approxeq \left(1 - \frac{3\sqrt{s}}{t}\right) \cdot \left(\dot{x}(t) + \frac{1}{2}\ddot{x}(t)\sqrt{s} + o(\sqrt{s})\right) - \sqrt{s}\nabla f(x(t)) + o(\sqrt{s}), \end{aligned}$$

and identifying the terms of order $\sqrt{s}$ gives

$$\begin{aligned} \frac{1}{2}\ddot{x}(t) &= -\frac{1}{2}\ddot{x}(t) - \frac{3}{t}\dot{x}(t) - \nabla f(x(t)) \\ \ddot{x}(t) + \frac{3}{t}\dot{x}(t) + \nabla f(x(t)) = 0. \end{aligned}$$

We have managed to recover the differential equation we saw in **2.** from the discrete system we have in **3.**, hinting at a connection between the two. Let us explore that connection further.

4. Now we try to use the analysis for the ordinary differential equation in developing analyses for Nesterov's accelerated gradient method. Consider again

$$\begin{cases} x_k = y_{k-1} - s\nabla f(y_{k-1}) \\ y_k = x_k + \dfrac{k-1}{k+1} \cdot (x_k - x_{k-1}), \end{cases}$$

for all $k \in \mathbb{N}^*$ with $x_0 = y_0 \in \mathbb{R}$. Previously we saw the functional

$$\mathcal{E}(t) = t^2 \left(f(x(t)) - f(x^*)\right) + 2 \left\| x + \frac{t}{2}\dot{x} - x^* \right\|^2 .$$

Can we somehow find a discrete version of this functional? Define

$$\mathcal{E}_k = (k+1)^2 s \left(f(x_k) - f(x^*)\right) + 2 \left\| \underbrace{\frac{k+2}{2} y_k - \frac{k}{2} x_k}_{=:z_k} - x^* \right\|^2 .$$

Let $f \in \mathcal{F}_L^{1,1}$ and $s \leqslant \frac{1}{L}$. We have, for all $x, y \in E$, that

$$\begin{aligned} & f(y - s\nabla f(y)) \leqslant f(y) - \frac{s}{2} \|\nabla f(y)\|^2 && \text{since } \nabla f \text{ is } L\text{-Lipschitz} \\ & f(y - s\nabla f(y)) \leqslant f(x) + \langle \nabla f(x), y - x\rangle - \frac{s}{2} \|\nabla f(x)\|^2 && \text{by convexity of } f. \end{aligned}$$

Now let $x_k \coloneqq y_{k-1} - s\nabla f(y_{k-1})$. We have

$$\begin{array}{rl} & \left(f(y_{k-1} - s\nabla f(y_{k-1})) \leqslant f(x_{k-1}) + \langle \nabla f(y_{k-1}), y_{k-1} - x_{k-1}\rangle - \frac{s}{2} \|\nabla f(y_{k-1})\|^2 \right) \cdot \dfrac{k-1}{k+2} \\ + & \left(f(y_{k-1} - s\nabla f(y_{k-1})) \leqslant f(x^*) + \langle \nabla f(y_{k-1}), y_{k-1} - x^*\rangle - \dfrac{s}{2} \|\nabla f(y_{k-1})\|^2 \right) \cdot \dfrac{2}{k+1} \\ \hline & f(x_k) \leqslant \dfrac{k-1}{k+2} f(x_{k-1}) + \dfrac{2}{k+1} f(x^*) + \varpi, \end{array}$$

where

$$\begin{aligned} \varpi &= \frac{2}{k+1} \left\langle \nabla f(y_{k-1}), \frac{k+1}{2} \cdot y_{k-1} - \frac{k-1}{2} \cdot x_{k-1} - x^* \right\rangle - \frac{s}{2} \|\nabla f(y_{k-1})\|^2 \\ &= \frac{2}{s(k+1)^2} \left(\|z_{k-1} - x\|^2 - \|z_k - x^*\|^2 \right). \end{aligned}$$

This is because

$$\begin{aligned} \|z_k - x^*\|^2 &= \left\| z_{k-1} - \frac{s(k+1)}{2} \nabla f(y_{k-1}) - x^* \right\|^2 \\ &= \|z_{k-1} - x^*\|^2 - 2 \left\langle z_{k-1} - x^*, -\frac{s(k+1)}{2} \nabla f(y_{k-1}) \right\rangle + \frac{s^2(k+1)^2}{4} \|\nabla f(y_{k-1})\|^2 , \end{aligned}$$

and because

$$\begin{aligned}
z_k &= \frac{k+2}{2} \cdot y_k - \frac{k}{2} \cdot x_k \\
&= \frac{k+2}{2} \cdot \left(x_k + \frac{k-1}{k+2}(x_k - x_{k-1}) \right) - \frac{k}{2} \cdot (y_{k-1} - s\nabla f(y_{k-1})) \\
&= \frac{k+2}{2} x_k + \frac{k-1}{2} (x_k - x_{k-1}) - \frac{k}{2} y_{k-1} + \frac{ks}{2} \nabla f(y_{k-1}) \\
&= \frac{2k+1}{2} \cdot x_k + \frac{k-1}{2} \cdot x_{k-1} - \frac{k}{2} \cdot y_{k-1} + \frac{ks}{2} \nabla f(y_{k-1}) \\
&= \frac{2k+1}{2} \cdot (y_{k-1} - s\nabla f(y_{k-1})) + \frac{k-1}{2} \cdot x_{k-1} - \frac{k}{2} \cdot y_{k-1} + \frac{ks}{2} \nabla f(y_{k-1}) \\
&= \frac{k+1}{2} y_{k-1} - \frac{k-1}{2} x_{k-1} - \frac{s(k+1)}{2} \nabla f(y_{k-1}) \\
&= z_{k-1} - \frac{s(k+1)}{2} \nabla f(y_{k-1}).
\end{aligned}$$

Therefore we have obtained that

$$\begin{aligned}
f(x_k) - f(x^*) \leqslant \frac{k-1}{k+1} \cdot (f(x_{k-1}) - f(x^*)) + \frac{2}{k+1} \cdot (f(x^*) - f(x^*)) \\
+ \frac{2}{s(k+1)^2} \cdot \left(\|z_{k-1} - x^*\|^2 - \|z_k - x^*\|^2 \right)
\end{aligned}$$

That is,

$$\begin{aligned}
\underbrace{s(k+1)^2 \left(f(x_k) - f(x^*)\right) + 2\left\|z_k - x^*\right\|}_{=\mathcal{E}_k} &\leqslant s(k-1)(k+1) \cdot (f(x_{k-1}) - f(x^*)) + 2\left\|z_{k-1} - x^*\right\|^2 \\
&= \underbrace{sk^2 \left(f(x_{k-1}) - f(x^*)\right) - 2\left\|z_{k-1} - x^*\right\|^2}_{=\mathcal{E}_{k-1}} \\
&\qquad - s\left(f(x_{k-1}) - f(x^*)\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathcal{E}_k + s\left(f(x_{k-1}) - f(x^*)\right) &\leqslant \mathcal{E}_{k-1} \\
\mathcal{E}_k + s\sum_{i=1}^{k} \left(f(x_i) - f(x^*)\right) &\leqslant \mathcal{E}_0 = s\left(f(x_0) - f(x^*)\right) + 2\left\|z_0 - x^*\right\|^2.
\end{aligned}$$

Putting things together,

$$s(k+1)^2 \left(f(x_k) - f(x^*)\right) \leqslant \mathcal{E}_k + s\sum_{i=1}^{k-1} f(x_i) - f(x^*) \leqslant 2\left\|x_0 - x^*\right\|^2,$$

if $y_0 = x_0$. Thus,

$$f(x_k) - f(x^*) \leqslant \frac{2}{s(k+1)^2} \left\|x_0 - x^*\right\|^2.$$

Now we summarise Nesterov's accelerated gradient method. We omit some other details regarding the setup, but they all follow from our discussion above.

Algorithm: NESTEROV'S ACCELERATED GRADIENT METHOD

- Choose $y_0 \in \mathbb{R}^n$, $a_0 = 1$, $x_{-1} = y_0$, and

$$\alpha_{-1} = \frac{\|y_0 - z\|}{\|\nabla f(y_0) - \nabla f(z)\|},$$

  where $z \neq y_0$, $\nabla f(z) \neq \nabla f(y_0)$, and $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$.
- For $k = 0, 1, 2, \ldots$ do
  - Find the smallest index $i \in \mathbb{N}$ for which

$$f(y_k) - f\left(y_k - 2^{-i}\alpha_{k-1}\nabla f(y_k)\right) \geqslant 2^{-i-1}\alpha_{k-1}\left\|\nabla f(y_k)\right\|^2.$$

  - Set

$$\begin{aligned} \alpha_k &= 2^{-i}\alpha_{k-1}, \\ x_k &= y_k - \alpha_k \nabla f(y_k), \\ a_{k+1} &= \frac{1 + \sqrt{1 + 4a_k^2}}{2}, \\ y_{k+1} &= x_k + \frac{a_k - 1}{a_{k+1}}(x_k - x_{k-1}). \end{aligned}$$

  - If $\|\nabla f(y_k)\| < \varepsilon$ for sufficiently small ε, stop.

§13.3. More on the proximal operator

Now we continue our discussion from Section 8.4.

Proposition 13.2. *The proximal operator p_f is monotone and Lipschitz continuous.*

Proof. Take the respective settings

$$\begin{aligned} y_1 &= p_f(x) & y_2 &= p_f(x_2) \\ g_1 &= x_1 - y_1 \in \partial f(y_1) & g_2 &= x_2 - y_2 \in \partial f(y_2). \end{aligned}$$

Since f is convex, it is monotone; *i.e.*

$$\begin{aligned} \langle y_1 - y_2, g_1 - g_2 \rangle &\geqslant 0 \\ \langle y_1 - y_2, x_1 - y_1 - (x_2 - y_2) \rangle &\geqslant 0 \\ \langle y_1 - y_2, x_1 - x_2 \rangle &\geqslant \|y_1 - y_2\|^2 \\ \langle p_f(x_1) - p_f(x_2), x_1 - x_2 \rangle &\geqslant \|p_f(x_1) - p_f(x_2)\|^2, \end{aligned}$$

and so p_f is (strictly) monotone. To show that p_f is Lipschitz continuous, we write $x_1 - x_2 = y_1 - y_2 + g_1 - g_2$ and so

$$\|x_1 - x_2\|^2 = \|y_1 - y_2\|^2 + \|g_1 - g_2\|^2 + \underbrace{2\langle y_1 - y_2, g_1 - g_2 \rangle}_{\geqslant 0}$$

$$\geqslant \|p_f(x_1) - p_f(x_2)\|^2 + \|p_{f^*}(x_1) - p_{f^*}(x_2)\|^2 ,$$

where f^* is the convex conjugate of f. Thus p_f is Lipschitz continuous with constant 1. $\square$

Proposition 13.3. *The Moreau envelope $\widetilde{f} \in \mathcal{F}_1^{1,1}$. Also, $\nabla \widetilde{f} = p_{f^*}$.*

Proof. Recall that $\widetilde{f}(x) = \inf_{y \in E} \varphi(x, y)$, where $\varphi(x, y) = f(y) + \frac{1}{2}\|x - y\|^2$. Recall also that $\partial \widetilde{f}(x) = \{x^* : (x^*, 0) \in \partial \varphi(x, y_x)\}$, where $\widetilde{f}(x) = \varphi(x, y_x)$. Then

$$\begin{aligned}
\partial \varphi(x, y) &= \partial\left((x, y) \mapsto f(y)\right) + \partial \left(\frac{1}{2}\|\cdot - \cdot\|\right)^2 (x, y) \\
&= \begin{pmatrix} 0 \\ \partial f(y) \end{pmatrix} + \begin{pmatrix} x - y \\ y - x \end{pmatrix} \\
&= \begin{pmatrix} x - y \\ \partial f(y) + y - x \end{pmatrix}.
\end{aligned}$$

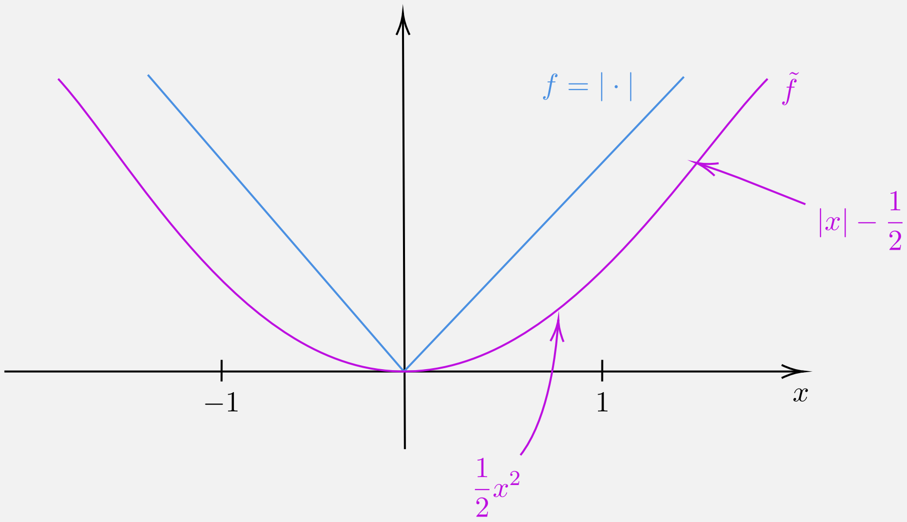
We then have that

$$\begin{aligned}
\partial \widetilde{f}(x) = \left\{ x^* : \begin{pmatrix} x^* \\ 0 \end{pmatrix} \in \begin{pmatrix} x - y_x \\ \partial f(y_x) + y_x - x \end{pmatrix} \right\} &= \left\{ x^* : \begin{array}{c} x^* = x - y \\ x - y \in \partial f(y) \end{array} \right\} \\
&= \left\{ x^* : \begin{array}{c} x^* = x - y \\ y = p_f(x) \end{array} \right\} \\
&= \{p_{f^*}(x)\}.
\end{aligned}$$

$\square$

One of the big appeals of the proximal operator is this smoothing effect it has on functions, as the following example shows.

Example 13.4. Consider the function $f = |\cdot|$.



We compute both the proximal operator and Moreau envelope for this f as

$$p_f(x) = \operatorname{sign}(x) \max\{|x| - 1, 0\},$$

$$\widetilde{f}(x) = f(p_f(x)) + \frac{1}{2}|p_f(x) - x|^2 = |p_f(x)|^2 + \frac{1}{2}|p_f(x) - x|^2 = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \leqslant 1, \\ |x| - \frac{1}{2} & \text{if } |x| > 1. \end{cases}$$

Proposition 13.5. *For all $x \in E$ and for all $\lambda > 0$, we have $p_{\lambda f}(x) + p_{\lambda^{-1} f^*}(\lambda x) = x$.*

We will not prove this, but we will furnish an example.

Example 13.6. Consider an arbitrary norm $\|\cdot\|_a$. This norm can be viewed as the support function of the dual ball, *i.e.* $\|\cdot\|_a = \sigma_{B^*(0,1)}$, and so $\|\cdot\|_a^* = \left(\sigma_{B^*(0,1)}\right)^* = I_{B(0,1)}$. Then $p_{\lambda^{-1}\|\cdot\|_a^*} = p_{\lambda^{-1} I_{B(0,1)}} = p_{B^*(0,1)}$. But of course, $p_{\lambda\|\cdot\|_a}(x) = x - \lambda p_{B^*(0,1)}(\lambda x)$.

Proposition 13.7. *For any f, $\arg\min f = \arg\min \widetilde{f}$.*

Proof. Proceeding directly gives

$$y \in \arg\min f \iff x = p_f(x) \iff \nabla \widetilde{f}(x) = p_{f^*}(x) = 0 \iff x \in \arg\min \widetilde{f},$$

since $p_f + p_{f^*} = \mathrm{id}$ and $\widetilde{f} \in \mathcal{F}_1^{1,1}$. □

§13.4. Proximal algorithms: proximal gradient descent

The proximal gradient method We will now get our first taste of proximal algorithms via the proximal gradient method. Here we are concerned with minimising a composite function $h = f + g$, where $f \in \mathcal{C}_L^{1,1}(E)$, and $g \in \operatorname{conv} E$. The idea is to first minimise f via a gradient step, and then minimise g via a proximal step:

$$x_{k+1} \in \arg\min_{x \in E} + \frac{1}{2}\|x_k - \alpha_k \nabla f(x_k) - x\|^2 + \alpha_k g(x) = p_{\alpha_k g}(x_k - \alpha_k \nabla f(x_k)).$$

Consider the typical Taylor-type bounds:

$$\begin{aligned} f(x_{k+1}) &\leqslant f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2 \\ g(x_{k+1}) &\leqslant g(x_k) + \langle \partial g(x_{k+1}), x_{k+1} - x_k \rangle . \end{aligned}$$

Adding these gives

$$h(x_{k+1}) \leqslant h(x_k) + \langle \nabla f(x_k) + \partial g(x_{k+1}), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2 .$$

By Fermat's rule, we have

$$0 \in x_{k+1} - (x_k - \alpha_k \nabla f(x_k)) + \alpha_k \partial g(x_{k+1}) \iff 0 \in \frac{x_{k+1} - x_k}{\alpha_k} + \nabla f(x_k) + \partial g(x_{k+1}).$$

Now,

$$\mathsf{d}(0, \partial h(x_{k+1}))^2 = \mathsf{d}(0, \nabla f(x_{k+1}) + \partial g(x_{k+1}))^2$$

$$
\begin{aligned}
&\leqslant \|\nabla f(x_{k+1}) - \nabla f(x_k) + \nabla f(x_k) + s\|^2 && \text{for all } s \in \partial g(x_{k+1}) \\
&\leqslant 2\|\nabla f(x_{k+1}) - \nabla f(x_k)\|^2 + 2\|\nabla f(x_k) + s\|^2 && \text{by the parallelogram law} \\
&\leqslant 2L\|x_{k+1} - x_k\|^2 + 2\mathsf{d}(0, \nabla f(x_k) + \partial g(x_{k+1}))^2.
\end{aligned}
$$

Here $\mathsf{d}(x, S) = \inf_{y \in S} \|x - y\|$ is the distance from x to the set S. Indeed,

$$
\begin{aligned}
\langle \nabla f(x_k) + \partial g(x_{k+1}), x_{k+1} - x_k \rangle &= \alpha_k \left\langle \nabla f(x_k) + \partial g(x_{k+1}), \frac{x_{k+1} - x_k}{\alpha_k} \right\rangle \\
&= \frac{\alpha_k}{2} \left\| \nabla f(x_k) + \partial g(x_{k+1}) + \frac{x_{k+1} - x_k}{\alpha_k} \right\|^2 \\
&\qquad - \frac{\alpha_k}{2} \|\nabla f(x_k) + \partial g(x_{k+1})\|^2 - \frac{1}{2\alpha_k} \|x_{k+1} - x_k\|^2 \\
&\leqslant -\frac{\mathsf{d}(0, \partial h(x_{k+1}))^2}{4\alpha_k} - \frac{L\|x_{k+1} - x_k\|^2}{4} - \frac{1}{2\alpha_k} \|x_{k+1} - x_k\|^2 .
\end{aligned}
$$

Then

$$
h(x_{k+1}) \leqslant h(x_k) - \frac{\mathsf{d}(0, \partial h(x_{k+1}))^2}{4\alpha_k} - \frac{1}{2} \cdot \left(\left(\frac{L}{2} + 1 \right) \frac{1}{\alpha_k} - L \right) \|x_{k+1} - x_k\|^2 .
$$

Then we can choose $\alpha_k := \alpha$ small enough so that

$$
\mathsf{d}(0, \partial h(x_{k+1}))^2 \leqslant 4\alpha \left(h(x_k) - h(x_{k+1}) \right) \implies \sum_{k=0}^{\infty} \mathsf{d}(0, \partial h(x_{k+1}))^2 < \infty,
$$

and then

$$
(k+1) \min_{i=1,\dots,k+1} \mathsf{d}(0, \partial h(x_i))^2 \leqslant 4\alpha h(x_0) \implies \min_{i=1,\dots,k} \mathsf{d}(0, \partial h(x_i)) \leqslant \frac{2\sqrt{\alpha h(x_0)}}{\sqrt{k}}.
$$

We can now summarise the proximal gradient method. In the algorithm, we have defined the proximal mapping as $p_{f,t}(x) = \arg\min_{y \in E} \frac{1}{2t} \|y - x\|^2 + f(y)$.

Algorithm: Proximal gradient method

- Choose a starting point $x_0 \in H$ and a sequence $\{\alpha_k\}_{k \in \mathbb{N}}$ of step sizes.
- For $k = 0, 1, 2, \dots$ until convergence:
  - Compute the generalised gradient $G_\alpha(x) = \dfrac{x - p_{f,\alpha}(x - \alpha \nabla f(x))}{\alpha}$.
  - Update the sequence as $x_k = x_{k-1} - \alpha_k \cdot G_{\alpha_k}(x_{k-1})$.
  - If $\|\nabla f(x_k)\| < \varepsilon$ for sufficiently small ε, stop.

§A. Omitted proofs

This section houses all the proofs that Prof. Josz omitted during the lectures. Most of them are my work obtained while studying for the exams—so please do not treat them as gospel; in fact treat them as the treacherous heresy of a charlatan—and I will explicitly state when they are not.

Proof of Proposition 2.7. 1. The fact that a set that contains all its convex combinations is convex is immediate from the definition of convexity. We will prove the converse by induction on the number of vectors m in the set S.

The validity of the statement is clear for $k = 1$ since the set is a singleton and the only convex combination is the vector itself. Now, suppose the statement holds for $m = k$; we will show that it holds for $m = k + 1$. Let

$$s = \lambda_1 s_1 + \cdots + \lambda_{k+1} s_{k+1}$$

be a convex combination of the elements in S. Observe that

$$s' = \frac{\lambda_2}{\sum_{i=2}^{k+1} \lambda_i} s_2 + \cdots + \frac{\lambda_{k+1}}{\sum_{i=2}^{k+1} \lambda_i} s_{k+1}$$

is a convex combination of k elements in S, and so by the induction hypothesis, $s' \in S$. But then

$$s = \lambda_1 s_1 + \sum_{i=2}^{k+1} \lambda_i \left(\frac{\lambda_i}{\sum_{i=2}^{k+1} \lambda_i} s_i \right) = \lambda_1 s_1 + \sum_{i=2}^{k+1} \lambda_i s' = \lambda_1 s_1 + (1 - \lambda_1) s',$$

and since S is convex, $s \in S$. This completes the proof.

2. Write $K := \{\sum_{i=1}^m t_i x_i : m \in \mathbb{N}^*, t \in \Delta_m, x_i \in P\}$. First we show that $K \subseteq \operatorname{co} P$. Let $s = \lambda_1 s_1 + \cdots + \lambda_m s_m \in K$ be an arbitrary element of K for some collection $\{s_1, \ldots, s_m\} \subseteq P$. Now let C be any convex set containing S. By part 1 above, C contains all the convex combinations of its elements, and so $s \in C$. Since C was arbitrary, $s \in \operatorname{co} P$, and so $K \subseteq \operatorname{co} P$.

Now we to show that $\operatorname{co} P \subseteq K$, all we need do is show that K is convex. Take $x = \sum_{i=1}^m \alpha_i x_i$ and $y = \sum_{i=1}^m \beta_i y_i$ to be two arbitrary elements of K with $\sum_{i=1}^m \alpha_i = \sum_{i=1}^m \beta_i = 1$ and $x_i, y_i \in P$. Then for any $\lambda \in [0, 1]$,

$$\lambda x + (1 - \lambda) y = \lambda \sum_{i=1}^m \alpha_i x_i + (1 - \lambda) \sum_{i=1}^m \beta_i y_i.$$

But then this is a convex combination of the elements of P, as $x_i, y_i \in P$ for all i and so

$$\lambda \sum_{i=1}^m \alpha_i + (1 - \lambda) \sum_{i=1}^m \beta_i = \sum_{i=1}^m (\lambda \alpha_i + (1 - \lambda) \beta_i) = 1,$$

and so $\lambda x + (1 - \lambda) y \in K$. This completes the proof. □

Proof of Corollary 4.12. If f is identically $+\infty$ or $-\infty$, then every point of X is a minimum point for f. If f takes the value $-\infty$, all of these points are minimisers of f. Suppose now that f is not identically $+\infty$ or $-\infty$. Let $\{x_n\}_{n \in \mathbb{N}}$ be a sequence in X such that

$$\lim_{n \to \infty} f(x_n) = \inf_{x \in X} f(x) =: \alpha.$$

Clearly such a sequence exists, as $\inf_{x\in X} f(x) \in \mathbb{R}$. Without loss of generality, take $f(x_n) < +\infty$ for all n, and let $\beta := \sup_{n\in\mathbb{N}} f(x_n) < +\infty$. Since f is coercive, the sublevel set $\{f \leqslant \beta\}$ is compact and therefore there is a subsequence $\{x_k\}_{k\in\mathbb{N}}$ of $\{x_n\}_{n\in\mathbb{N}}$ which converges to a point $x \in X$. Since f is lower semicontinuous, we obtain

$$\alpha = \inf_{y\in X} f(y) \leqslant f(x) \leqslant \liminf_{k\to\infty} f(x_k) = \alpha,$$

and so $f(x) = \alpha$. This completes the proof. □

Proof of Proposition 5.7. The proposition follows by setting $f(x) = 0$ in the definition of convexity for $x \in P$. □

Proof of Proposition 6.4. The implication (i) $\implies$ (ii) is immediate from the definition of sublinearity, and the implications (ii) $\implies$ (iii) and (iii) $\implies$ (i) are immediate once we show that σ is sublinear if and only if $\operatorname{epi}\sigma$ is a convex cone in $E \times \mathbb{R}$.

For the forward direction, suppose that σ is convex and positively homogeneous, so that $\operatorname{epi}\sigma$ is convex. So we only need to verify that if $(x,t) \in \operatorname{epi}\sigma$ then $\lambda(x,t) = (\lambda x, \lambda t) \in \operatorname{epi}\sigma$ for all $\lambda \geqslant 0$. If $\lambda = 0$, then the result follows from the asumption that $\sigma(0) = 0$ by sublinearity. Now consider $\lambda > 0$; since $(x,t) \in \operatorname{epi}\sigma$, we have that $\sigma(x) \leqslant t$, and by the positive homogeneity criterion of σ, $\sigma(\lambda x) = \lambda\sigma(x) \leqslant \lambda t$, and so $(\lambda x, \lambda t) \in \operatorname{epi}\sigma$.

Now the reverse direction: from the assumption that $\operatorname{epi}\sigma$ is a convex cone, σ is convex, and we only need to verify that σ is positively homogeneous. First we verify that for all $\lambda > 0$, $\sigma(\lambda x) \leqslant \lambda\sigma(x)$. Since $\operatorname{epi}\sigma$ is a convex cone, and $(x, \sigma(x)) \in \operatorname{epi}\sigma$, we have that $(\lambda x, \lambda\sigma(x)) \in \operatorname{epi}\sigma$, and so $\sigma(\lambda x) \leqslant \lambda\sigma(x)$. Now for any particular $\overline{\lambda} > 0$ and $\overline{x} \in E$, we we have that $\sigma\left(\overline{\lambda}\overline{x}\right) \leqslant \overline{\lambda}\sigma(\overline{x})$. But then using the observation we just proved with $\lambda = 1/\overline{\lambda}$ and $x = \overline{\lambda}\overline{x}$, we have that $\sigma(\overline{x}) \leqslant \overline{\lambda}\sigma\left(\overline{\lambda}\overline{x}\right)$, and so $\sigma(\overline{x}) \geqslant \overline{\lambda}\sigma\left(\overline{\lambda}\overline{x}\right)$. Therefore $\sigma(\overline{x}) = \overline{\lambda}\sigma\left(\overline{\lambda}\overline{x}\right)$, and so σ is positively homogeneous. □

Proof of Proposition 6.7. (i) We first show that σ_P is sublinear. To start, we first check positive homogeneity: for any $x \in E$ and $\lambda > 0$,

$$\sigma_P(\lambda x) = \sup\langle \lambda x, p\rangle : p \in P = \lambda \sup\langle x, p\rangle : p \in P = \lambda\sigma_P(x).$$

Now we just need to check subadditivity. Let $x_1, x_2 \in E$; then for any $p \in P$, we have that

$$\begin{aligned}\sigma_P(x_1 + x_2) &= \sup_{p\in P}\langle x_1 + x_2, p\rangle = \sup_{p\in P}\langle x_1, p\rangle + \langle x_2, p\rangle \\ &\leqslant \sup_{p\in P}\langle x_1, p\rangle + \sup_{p\in P}\langle x_2, p\rangle = \sigma_P(x_1) + \sigma_P(x_2),\end{aligned}$$

and since σ_P is the supremum of linear functions $\langle\cdot, p\rangle$, $\operatorname{epi}\sigma_P$ is the intersection of closed half-spaces and so it is closed, and by Proposition 6.4, σ_P is a convex cone.

(ii) First we show the forward direction. Suppose that $\sigma_{P_1} \leqslant \sigma_{P_2}$, *i.e.* for all $x \in E$, the set $\sup\langle x, p_1\rangle : p_1 \in P_1 \leqslant \sup\langle x, p_2\rangle : p_2 \in P_2$. If $p_1 \in \overline{\operatorname{co}}\, P_1$, then p_1 can be approximated by convex combinations of points in P_1. Since $\sigma_{P_1} \leqslant \sigma_{P_2}$, any convex combination of points in P_1 that approximates p_1 has its support less than or equal to the support of points in P_2, implying $p_1 \in \overline{\operatorname{co}}\, P_2$. Thus, $\overline{\operatorname{co}}\, P_1 \subseteq \overline{\operatorname{co}}\, P_2$. The reverse direction is similar.

(iii) Follows directly from the previous case. In particular $\overline{\mathrm{co}}\, P_1 = \overline{\mathrm{co}}\, P_2$ if $\overline{\mathrm{co}}\, P_1 \subseteq \overline{\mathrm{co}}\, P_2$ and $\overline{\mathrm{co}}\, P_2 \subseteq \overline{\mathrm{co}}\, P_1$. In these cases, part (ii) above indicates that $\sigma_{P_1} \leqslant \sigma_{P_2}$ and $\sigma_{P_2} \leqslant \sigma_{P_1}$, and so $\sigma_{P_1} = \sigma_{P_2}$. $\square$

Proof of Proposition 6.9. We wish to show that for $a \geqslant 0$ and $f \in \operatorname{conv} E$ with $x \in E$, the subdifferential $\partial(\alpha f)(x) = \alpha \cdot \partial f(x)$. Indeed, for any $y \in E$,

$$\begin{aligned}\alpha f(y) + \langle x - y, \alpha \cdot \partial f(x)\rangle &\leqslant \alpha f(y) + \alpha\langle x - y, \partial f(x)\rangle = \alpha f(y) + \langle \alpha x - \alpha y, \partial f(x)\rangle \\ &= \alpha f(y) + \langle x - y, \partial f(x)\rangle = f(y) + \langle x - y, \partial f(x)\rangle,\end{aligned}$$

and so $\alpha \cdot \partial f(x) \in \partial(\alpha f)(x)$. Conversely, for any $y \in E$,

$$\begin{aligned}f(y) + \langle x - y, \alpha \cdot \partial f(x)\rangle &\leqslant f(y) + \alpha\langle x - y, \partial f(x)\rangle = \alpha f(y) + \langle x - y, \partial f(x)\rangle \\ &= \alpha f(y) + \langle \alpha x - \alpha y, \partial f(x)\rangle = \alpha f(y) + \langle x - y, \alpha \cdot \partial f(x)\rangle,\end{aligned}$$

and so $\partial(\alpha f)(x) \subseteq \alpha \cdot \partial f(x)$. Thus $\partial(\alpha f)(x) = \alpha \cdot \partial f(x)$. $\square$

Proof of Proposition 6.10. We prove the fact for $p = 2$; the general case follows by induction. Let $f_1, f_2 \in \operatorname{conv} E$ and $x \in E$, and define $f = f_1 + f_2$, so that $\operatorname{dom} f = \operatorname{dom}(f_1 + f_2) = \operatorname{dom} f_1 \cap \operatorname{dom} f_2$. Now let $x \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$, and let $x^* \in \partial f_1(x) + \partial f_2(x)$. Then there exist $x_1^* \in \partial f_1(x)$ and $x_2^* \in \partial f_2(x)$ such that $x^* = x_1^* + x_2^*$. For any $y \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$, we have by the subgradient inequality that

$$\begin{aligned}f_1(y) &\geqslant f_1(x) + \langle y - x, x_1^*\rangle, \\ f_2(y) &\geqslant f_2(x) + \langle y - x, x_2^*\rangle,\end{aligned}$$

and summing gives us that

$$f_1(y) + f_2(y) \geqslant f_1(x) + f_2(x) + \langle y - x, x_1^* + x_2^*\rangle = (f_1 + f_2)(x) + \langle y - x, x^*\rangle.$$

Thus $x^* = x_1^* + x_2^* \in \partial(f_1 + f_2)(x) = \partial f(x)$, and so $\partial f_1(x) + \partial f_2(x) \subseteq \partial f(x)$. For the equality case, if $\bigcap_{i=1,2} \operatorname{ri}(\operatorname{dom} f_i) \neq \emptyset$, then since $\operatorname{dom} f = \operatorname{dom} f_1 \cap \operatorname{dom} f_2$, we have that $\operatorname{ri} \operatorname{dom} f \neq \emptyset$ and so $\partial f(x) = \partial f_1(x) + \partial f_2(x)$. $\square$

Proof of Proposition 6.11. Let $x \in \operatorname{dom}(g \circ a)$, so that $x' = Ax + b \in \operatorname{dom} a$ is such that $a(x') = a(Ax + b) < \infty$. Furthermore, let $t \in A^*(\partial a(x'))$, so that there is a $d \in F^*$ such that $t = A^*(d)$ with $d \in \partial a(x')$. With $y \in \operatorname{dom}(g \circ a)$, we have that $y' = Ay + b \in \operatorname{dom} a$ and $a(y') < \infty$ with $(g \circ a)(y) = a(y')$. Applying the subgradient inequality for a at x' with subgradient d gives

$$a(y') \geqslant a(x') + \left\langle y' - x', d\right\rangle \implies (g \circ a)(y) \geqslant (g \circ a)(x) + \langle A(y - x), d\rangle\,.$$

Noting that $\langle A(y - x), d\rangle = \langle y - x, A^*(d)\rangle$, and for any $y \in \operatorname{dom}(g \circ a)$, we have that $(g \circ a)(y) \geqslant (g \circ a)(x) + \langle y - x, A^*(d)\rangle$, and so $t = A^*(d) \in \partial(g \circ a)(x)$. Therefore $A^*(\partial a(x')) \subseteq \partial(g \circ a)(x)$. $\square$

Proof of Proposition 6.12. Convexity is straightforward: since each F_i is convex and g is both convex and componentwise nondecreasing, we have for all $x, y \in E$ and $\lambda \in [0, 1]$ that

$$g\left(\sum_{i=1}^{m} \lambda F_i(x) + (1 - \lambda)F_i(y)\right) \leqslant \sum_{i=1}^{m} g\left(\lambda F_i(x) + (1 - \lambda)F_i(y)\right)$$

$$\leqslant \sum_{i=1}^{m} \lambda g(F_i(x)) + (1-\lambda) g(F_i(y)) = \lambda g(F(x)) + (1-\lambda) g(F(y)),$$

so that $g \circ F$ is convex. The proof for the second part below is adapted from Theorem 4.3.1 of Hiriart-Urruty and Lemaréchal's *Fundamentals of Convex Analysis* (1993). Indeed we want to show that

$$\partial(g \circ F)(x) = \left\{ \sum_{i=1}^{m} \gamma_i x_i^* \text{ such that } \gamma \in \partial g(F(x)),\ x_i^* \in \partial F_i(x),\ 1 \leqslant i \leqslant m \right\},$$

for finite $F(x) = (F_1(x), \ldots, F_m(x))$ via support functions, and hence we need to establish the convexity and compactness of $S = \{\sum_{i=1}^{m} \gamma_i x_i^* : \gamma \in \partial g(F(x)),\ x_i^* \in \partial F_i(x),\ 1 \leqslant i \leqslant m\}$. Obviously S is closed and bounded because the subdifferentials (be they $\partial g(F(x))$ or $\partial F_i(x)$, are compact by Proposition 6.13), and so S is compact. It then remains to show that S is convex. Pick points in S and form their convex combination thus:

$$x^* = \alpha \sum_{i=1}^{m} \gamma_i x_i^* + (1-\alpha) \sum_{i=1}^{m} \delta_i y_i^* = \sum_{i=1}^{m} \left(\alpha \gamma_i x_i^* + (1-\alpha)\delta_i y_i^*\right),$$

where $\gamma \in \partial g(F(x))$, $x_i^* \in \partial F_i(x)$, $\delta \in \partial g(F(y))$, and $y_i^* \in \partial F_i(y)$, with $\alpha \in [0,1]$. We need to show that $x^* \in S$. Since g is convex, we have that

$$g\left(\sum_{i=1}^{m} \alpha F_i(x) + (1-\alpha) F_i(y)\right) \leqslant \sum_{i=1}^{m} \alpha g(F_i(x)) + (1-\alpha) g(F_i(y)),$$

and so $\alpha\gamma_i + (1-\alpha)\delta_i \in \partial g(F_i(x))$ for all i. Similarly, since F_i is convex, we have that

$$F_i\left(\sum_{i=1}^{m} \alpha x_i + (1-\alpha) y_i\right) \leqslant \sum_{i=1}^{m} \alpha F_i(x_i) + (1-\alpha) F_i(y_i),$$

and so $\alpha x_i^* + (1-\alpha) y_i^* \in \partial F_i(x)$ for all i. Therefore $x^* \in S$, and so S is convex.

To show that S is the subdifferential of $g \circ F$, we need to show that for all $x \in E$ and $x^* \in S$, we have

$$\langle x^*, y \rangle \leqslant (g \circ F)(y) - (g \circ F)(x),$$

for all $y \in E$. Since $x^* \in S$, we have that $x^* = \sum_{i=1}^{m} \gamma_i x_i^*$ for some $\gamma \in \partial g(F(x))$ and $x_i^* \in \partial F_i(x)$. Then for all $y \in E$, we have that

$$\begin{aligned} \langle x^*, y \rangle &= \sum_{i=1}^{m} \langle \gamma_i x_i^*, y \rangle = \sum_{i=1}^{m} \langle x_i^*, \gamma_i y \rangle \\ &\leqslant \sum_{i=1}^{m} F_i(y) - F_i(x) = g(F(y)) - g(F(x)) = (g \circ F)(y) - (g \circ F)(x), \end{aligned}$$

and so $x^* \in \partial(g \circ F)(x)$. Conversely, for all $x \in E$ and $x^* \in \partial(g \circ F)(x)$, we have

$$\langle x^*, y \rangle \leqslant (g \circ F)(y) - (g \circ F)(x),$$

for all $y \in E$. Since g is convex, we have that

$$g(F(y)) - g(F(x)) \leqslant \langle x^*, y \rangle,$$

and so $x^* \in \partial g(F(x))$. Similarly, since F_i is convex, we have that

$$F_i(y) - F_i(x) \leqslant \langle x_i^*, y \rangle ,$$

and so $x_i^* \in \partial F_i(x)$. Therefore $x^* = \sum_{i=1}^m \gamma_i x_i^*$ for some $\gamma \in \partial g(F(x))$ and $x_i^* \in \partial F_i(x)$, and so $x^* \in S$. Thus $S = \partial(g \circ F)(x)$. □

Proof of Proposition 6.13. Let $x_0 \in \operatorname{ri}\operatorname{dom} f$, and assume that $f(x_0) > -\infty$ (otherwise the proof is trivial). It suffices to only consider the case where $x_0 = 0$ and $f(x_0) = 0$ (modulo replacing f by $x \mapsto f(x + x_0) - f(x_0)$). In this situation, for every vector $\nu \in E$, $\nu \in \partial f(0)$ if and only if for all $x \in \operatorname{dom} f$, $\langle \nu, x \rangle \leqslant f(x)$.

Let $A = \operatorname{aff}\operatorname{dom} f$. This vector space contains 0 and is therefore a subspace of the Euclidean space. Furthermore, let C be the closure of $\operatorname{epi} f \cap (A \times \mathbb{R})$, which is a closed convex set in $A \times \mathbb{R}$ endowed with the product $\langle (x, t), (x', t') \rangle = \langle x, x' \rangle + tt'$. Observe that $(0, 0) = (x_0, f(x_0)) \in \partial C$, and so we can apply Proposition 5.1 in $A \times \mathbb{R}$. There is a nonzero vector $\alpha \in A \times \mathbb{R}$ such that for all $z \in C$, $\langle \alpha, z \rangle \leqslant 0$. Write $\alpha = (\nu, t) \in A \times \mathbb{R}$. For $z = (x, s) \in C$, we have that $\langle \nu, x \rangle + ts \leqslant 0$. Let $x \in \operatorname{dom} f$; in particular, $f(x) < \infty$ and for all $s \geqslant f(x)$, $(x, s) \in C$. Therefore for all $x \in \operatorname{dom} f$ and for all $s \geqslant f(x)$, $\langle \nu, x \rangle + ts \leqslant 0$, and with $s \to \infty$, we get that $t \leqslant 0$.

We now prove by contradiction that $t < 0$. Suppose not; *i.e.* $t = 0$. Then $\langle \nu, x \rangle \leqslant 0$ for all $x \in \operatorname{dom} f$, and as $0 \in \operatorname{ri}\operatorname{dom} f$, there is a set B open in A such that $0 \in B \subseteq \operatorname{dom} f$. Thus for $x \in A$, there is an $\varepsilon > 0$ such that $\varepsilon x \in B \subseteq \operatorname{dom} f$. By the inequality $\langle \nu, \varepsilon x \rangle + ts \leqslant 0$, we have $\langle \nu, \varepsilon x \rangle \leqslant 0$, and so $\langle \nu, x \rangle \leqslant 0$. Similarly, $\langle \nu, -x \rangle \leqslant 0$, and therefore $\langle \nu, x \rangle = 0$ for all $x \in A$. Since $\nu \in A$, we have $\langle \nu, \nu \rangle = 0$, and so $\nu = 0$; but $s \neq 0$ and so $t \neq 0$, a contradiction. Therefore $t < 0$, and so with $s = f(x)$, we have $\langle \nu, x \rangle + tf(x) \leqslant 0$, and so $\nu \in \partial f(x)$, so that $\partial f(x)$ is nonempty.

To show that $\partial f(x)$ is compact, we need to show that $\partial f(x)$ is bounded. Let $x \in \operatorname{dom} f$ and $\nu \in \partial f(x)$. Then for all $y \in \operatorname{dom} f$, $\langle \nu, y \rangle \leqslant f(y)$. In particular, for $y = 0$, we have $\langle \nu, 0 \rangle \leqslant f(0) = 0$, and so $\nu \in B(0, \|\nu\|)$. Therefore $\partial f(x)$ is bounded, and so compact. □

Proof of Proposition 6.16. Suppose that f is Lipschitz continuous with constant L, that E is open, and take $x \in E$. As x is an interior point of $\operatorname{dom} f$, the subdifferential $\partial f(x)$ is nonempty and thus we can take $x^* \in \partial f(x)$. Let $x^\dagger \in E$ be a vector for which $\left\|x^\dagger\right\| = 1$ and $\left\langle x^\dagger, x^* \right\rangle = \|x^*\|$. Since E is open, we can choose sufficiently small $\varepsilon > 0$ such that $x + \varepsilon x^\dagger \in E$. Then by the subgradient inequality, $f(x + \varepsilon x^\dagger) \leqslant f(x) + \left\langle x^*, \varepsilon x^\dagger \right\rangle$. Therefore,

$$\begin{aligned} \varepsilon \|x^*\| = \left\langle \varepsilon x^\dagger, x^* \right\rangle &\leqslant f(x + \varepsilon x^\dagger) - f(x) \\ &\leqslant L\|x + \varepsilon x^\dagger - x\| = L\varepsilon, \end{aligned}$$

and so $\|x^*\| \leqslant L$. For the converse, suppose that $\|x^*\| \leqslant L$ for all $x^* \in \partial f(x)$ and $x \in E$. Since f is proper and convex, with $x, y \in \operatorname{ri}\operatorname{dom} f$ and $x' \in \partial f(x)$, $y' \in \partial f(y)$, the subgradient inequality gives

$$\begin{aligned} f(y) \geqslant f(x) + \left\langle y - x, x' \right\rangle &\implies f(x) - f(y) \leqslant \left\langle x - y, x' \right\rangle , \\ f(x) \geqslant f(y) + \left\langle x - y, y' \right\rangle &\implies f(y) - f(x) \leqslant \left\langle y - x, y' \right\rangle . \end{aligned}$$

Adding these two inequalities gives

$$|f(x) - f(y)| \leqslant \left\langle x - y, x' - y' \right\rangle \leqslant \|x - y\|\|x' - y'\| \leqslant L\|x - y\|,$$

and so f is Lipschitz continuous with constant L. □

Proof of Proposition 6.17. For any $x \in E$, the section $(\{x\} \times \mathbb{R}) \cap \left(\bigcap_{i\in I} \operatorname{epi} f_i\right)$ is

$$\bigcap_{i\in I} ((\{x\} \times \mathbb{R}) \cap \operatorname{epi} f_i) = \{x\} \times \bigcap_{i\in I} [f_i(x), +\infty) = \{x\} \times \left[\sup_{i\in I} f_i(x), +\infty\right)$$
$$= (\{x\} \times \mathbb{R}) \cap \operatorname{epi} \sup_{i\in I} f_i,$$

and so $\operatorname{epi} \sup_{i\in I} f_i = \bigcap_{i\in I} \operatorname{epi} f_i$. $\square$

Proof of Proposition 6.21. Let $(x, \alpha) \in \operatorname{epi}_s (f \,\square\, g)$, so that $(f \,\square\, g)(x) < \alpha$. Then by the definition of the infimal convolution, there exists some $y \in E$ such that $f(y) + g(x - y) < \alpha$. Now choose any $\beta, \gamma \in \mathbb{R} \cup \{\infty\}$ such that $f(y) < \beta$, $g(x - y) < \gamma$, and $\beta + \gamma = \alpha$. Then $(y, \beta) \in \operatorname{epi}_s f$ and $(x - y, \gamma) \in \operatorname{epi}_s g$, and so $(x, \alpha) = (y, \beta) + (x - y, \gamma) \in \operatorname{epi}_s f + \operatorname{epi}_s g$. For the reverse inclusion, if $(x, \alpha) \in \operatorname{epi}_s f + \operatorname{epi}_s g$, then there exist $(y, \beta) \in \operatorname{epi} f$ and $(z, \gamma) \in \operatorname{epi} g$ such that $(x, \alpha) = (y, \beta) + (z, \gamma)$. Then $\alpha = \beta + \gamma \geqslant f(y) + g(z) \geqslant f(y) + g(x - y) = (f \,\square\, g)(x)$, and so $(x, \alpha) \in \operatorname{epi}_s (f \,\square\, g)$, so that $\operatorname{epi}_s (f \,\square\, g) \subseteq \operatorname{epi}_s f + \operatorname{epi}_s g$. Therefore $\operatorname{epi}_s (f \,\square\, g) = \operatorname{epi}_s f + \operatorname{epi}_s g$.

For the second statement, define $h(x, y) = f(y) + g(x - y)$. Since f, g are convex, h is convex, and we can show that for any $x \in E$, there exists $y \in E$ such that $h(x, y) < \infty$. Indeed pick any $x \in E$ as well as a $y \in \operatorname{dom} f$, so that $f(y) < \infty$. Then since $g(x - y) < \infty$ as g is real valued, it must be that $h(x, y) = f(y) + g(x - y) < \infty$, and so $h(x, y) < \infty$. Therefore h is proper, and

$$(f \,\square\, g)(x) = \inf_{y\in E} h(x, y) = \inf_{y\in E} \{f(y) + g(x - y)\}$$

is convex, interpreted as a partial minimization of a convex function with respect to y. $\square$

Proof of Proposition 7.7. It suffices to prove that for $f \colon E \to \overline{\mathbb{R}}$, the following hold:

(1) $f(x) \geqslant f^{**}(x)$ for all $x \in E$.

(2) if f is closed, proper, and convex, then $f(x) = f^{**}(x)$.

For part (i), we know that for all x and y, we have $f^*(y) \geqslant \langle x, y\rangle$. So for any $x \in E$,

$$f(x) = \sup_{y\in E} \{\langle x, y\rangle - f^*(y)\} \geqslant \langle x, y\rangle - f^*(y) \geqslant f^{**}(x).$$

Part (ii) is less straightforward; by part (i), we have that $\operatorname{epi} f \subseteq \operatorname{epi} f^{**}$, and we need to show that $\operatorname{epi} f^{**} \subseteq \operatorname{epi} f$; towards this, it suffices to show that $(x, f^{**}(x)) \in \operatorname{epi} f$ for all $x \in E$. Suppose not. Since $\operatorname{epi} f$ is closed and convex, $(x, f^{**}(x))$ can be strictly separated from $\operatorname{epi} f$, and hence $\langle y, z\rangle + bs < c < \langle y, x\rangle + b \cdot f^{**}(x)$ for some $y \in E$, $z \in \mathbb{R}$, $b \in \mathbb{R}$, and $s \in \mathbb{R}$ with $(z, s) \in \operatorname{epi} f$. Without loss of generality, assume $b \neq 0$ (otherwise, add $\varepsilon(\overline{y} - 1)$ to (y, b) where $\overline{y} \in \operatorname{dom} f^*$), and so we must have $b < 0$, since $b > 0$ leads to a contradiction for large s. Again without loss of generality, assume $b = -1$ (otherwise, divide by $-b$). Then we have $\langle y, z\rangle - s < c < \langle y, x\rangle - f^{**}(x)$. Taking the supremum over z, we get that $f^*(y) + f^{**}(x) < \langle x, y\rangle$, which contradicts $f(x) \geqslant \langle x, y\rangle - f^*(y)$, and hence $(x, f^{**}(x)) \in \operatorname{epi} f$, as desired. Therefore $\operatorname{epi} f = \operatorname{epi} f^{**}$, and so $f = f^{**}$. $\square$

Proof of Proposition 8.2. We require the following result—which we will not prove—to prove this proposition:

Theorem A.1 (Fenchel's duality theorem)**.** *Let $f, g\colon E \to \mathbb{R} \cup \{\infty\}$ be proper convex functions. If* $\operatorname{ri}\operatorname{dom} f \cap \operatorname{ri}\operatorname{dom} g \neq \emptyset$*, then*

$$\inf_{x\in E}\{f(x) + g(x)\} = \sup_{y\in E}\{-f^*(y) - g^*(y)\},$$

and the supremum of the dual problem (the right-hand side) is attained whenever the infimum of the primal problem (the left-hand side) is finite.

With this theorem we can prove the proposition. Pick any $y \in E$, and elaborate the conjugate function

$$\begin{aligned}(f+g)^*(y) &= \sup_{x\in E}\{\langle x, y\rangle - (f+g)(x)\} = \sup_{x\in E}\{\langle x, y\rangle - f(x) - g(x)\} \\ &= -\inf_{x\in E}\{f(x) + g(x) - \langle x, y\rangle\} = -\inf_{x\in E}\{f(x) + h(x)\},\end{aligned}$$

where $h(x) = g(x) - \langle x, y\rangle$. Since g is proper and convex, h is also proper and convex, and furthermore, $(\operatorname{ri}\operatorname{dom} f \cap \operatorname{ri}\operatorname{dom} h) = (\operatorname{ri}\operatorname{dom} f) \cap E = \operatorname{ri}\operatorname{dom} f \neq \emptyset$. Therefore by Fenchel's duality theorem, we have

$$\inf_{x\in E}\{f(x) + h(x)\} = \sup_{z\in E}\{-f^*(z) - h^*(-z)\}.$$

Furthermore,

$$\begin{aligned}h^*(-z) &= \sup_{x\in E}\{\langle x, -z\rangle - h(x)\} = \sup_{x\in E}\{\langle x, -z\rangle - g(x) + \langle x, y\rangle\} \\ &= \sup_{x\in E}\{\langle x, y-z\rangle - g(x)\} = g^*(y-z),\end{aligned}$$

and consequently,

$$\begin{aligned}(f+g)^*(y) &= -\inf_{x\in E}\{f(x) + h(x)\} = -\sup_{z\in E}\{-f^*(z) - h^*(-z)\} \\ &= \inf_{z\in E}\{f^*(z) + h^*(-z)\} = \inf_{z\in E}\{f^*(z) + g^*(y-z)\} \\ &= (f^* \,\square\, g^*)(y),\end{aligned}$$

and so $(f+g)^* = f^* \,\square\, g^*$. ∎

Proof of Proposition 8.7. For any nonzero direction $d \in E$, $f'(x, d) = \max_{i\in I^\circ(x)} f_i'(x, d)$. Without loss of generality, assume $I^\circ(x) = \{1, \dots, m\}$ for some $m \in \mathbb{N}$ (otherwise, reorder the indices). Then by Proposition 8.9, $f_i'(x, d) = \sup_{x^*\in\partial f_i(x)} \langle x^*, d\rangle$, and so $f'(x, d) = \max_{i\in I^\circ(x)} \sup_{x_i^*\in\partial f_i(x)} \langle x_i^*, d\rangle$. Recalling that for $a_1, \dots, a_k \in \mathbb{R}$,

$$\max\{a_1, \dots, a_k\} = \sup_{t\in\Delta_k} \sum_{i=1}^k t_i a_i,$$

we can expand $f'(x, d)$ as

$$f'(x, d) = \max_{i\in I^\circ(x)} \sup_{x_i^*\in\partial f_i(x)} \langle x_i^*, d\rangle = \sup_{t\in\Delta_k} \left\{ \sum_{i=1}^k t_i \sup_{x_i^*\in\partial f_i(x)} \langle x_i^*, d\rangle \right\}$$

$$
\begin{aligned}
&= \sup_{t\in\Delta_k}\left\{\sum_{i=1}^{k}\sup_{x_i^*\in\partial f_i(x)}\langle t_i x_i^*, d\rangle\right\} = \sup_{\substack{t\in\Delta_k\\ x_i^*\in\partial f_i(x)}}\left\{\sum_{i=1}^{k}\langle t_i x_i^*, d\rangle\right\}\\
&= \sup_{x^*\in\operatorname{co}\left(\bigcup_{i=1}^k \partial f_i(x)\right)}\langle x^*, d\rangle\\
&= \sigma_{\operatorname{co}\left(\bigcup_{i=1}^k \partial f_i(x)\right)}(d),
\end{aligned}
$$

where $\sigma_{\operatorname{co}\left(\bigcup_{i=1}^k \partial f_i(x)\right)}(d)$ is the support function of the convex hull of the subdifferentials of the functions f_i at x. Now since $x \in (\operatorname{dom} f)^\circ$, Proposition 8.9 gives us that $f'(x,d) = \sigma_{\partial f(x)}(d) = \sigma_{\operatorname{co}\left(\bigcup_{i=1}^k \partial f_i(x)\right)}(d)$. Now, it is clear from previous results we have seen that the $\partial f_i(x)$ are convex, nonempty, closed, and bounded, and so $\bigcup_{i=1}^k \partial f_i(x)$ is also convex, nonempty, closed, and bounded. Its convex hull $\operatorname{co}\left(\bigcup_{i=1}^k \partial f_i(x)\right)$ is therefore also convex, nonempty, amd compact since E is finite-dimensional, and since $\sigma_{\partial f(x)}(d) = \sigma_{\operatorname{co}\left(\bigcup_{i=1}^k \partial f_i(x)\right)}(d)$, it follows from Proposition 6.7(iii) that $\partial f(x) = \operatorname{co}\left(\bigcup_{i=1}^k \partial f_i(x)\right)$. □

Proof of Proposition 8.9. Let $x \in \operatorname{ri}\operatorname{dom} f$ and $d \in E$, and let $t > 0$. Then by the subgradient inequality, $f(x+td) - f(x) \geqslant \langle td, x^*\rangle$ for all $x^* \in \partial f(x)$. Therefore for all $x^* \in \partial f(x)$,

$$
\frac{f(x+td)-f(x)}{t} \geqslant \langle x^*, d\rangle \implies f'(x,d) = \lim_{t\searrow 0}\frac{f(x+td)-f(x)}{t} \geqslant \langle x^*, d\rangle\,.
$$

Taking the supremum over $x^* \in \partial f(x)$, we get $f'(x,d) \geqslant f^*(d)$, and so $f'(x,d) \geqslant \sup_{x^*\in\partial f(x)}\langle x^*, d\rangle$.

To prove the reverse inequality, let $h(\nu) = f'(\nu, d)$; almost by definition, h is real valued and nonnegative homogeneous, and $\operatorname{dom}\partial f = E$. In particular, h is subdifferentiable at d, and so we can pick $x^* \in \partial h(d)$. Then for all $x \in \operatorname{ri}\operatorname{dom} f$ and $t > 0$, we have $tf'(x,d) = t\cdot h(x) = h(tx)$ since h is nonnegative homogeneous. By the subdifferential inequality,

$$
t\cdot h(x) = h(tx) \geqslant h(d) + \langle x^*, tx - d\rangle = f'(x,d) + \langle x^*, tx - d\rangle\,.
$$

Rearranging, we get $t\cdot(f'(x,d) - \langle x^*, x\rangle) \geqslant \langle x^*, d\rangle$. Since this inequality holds for all $t > 0$, the term in parentheses must be nonnegative (otherwise large enough t makes the inequality invalid), and so $f'(x,d) \geqslant \langle x^*, d\rangle$. We already know that for any $y \in \operatorname{dom} f$, $f(y) \geqslant f(x) + f'(x, y-x)$, and so $f'(x, y-x) \geqslant \langle x^*, y-x\rangle$, and so $x^* \in \partial f(x)$. Taking $t = 0$ in the subgradient inequality for h, we get that $0 \geqslant f'(x,d) - \langle x^*, d\rangle$, and so $f'(x,d) \leqslant \langle x^*, d\rangle$. Consequently, $f'(x,d) \leqslant \sup_{x^*\in\partial f(x)}\langle x^*, d\rangle$, and so combining this with the previous fact, $f'(x,d) = \sup_{x^*\in\partial f(x)}\langle x^*, d\rangle$. □

Proof of Lemma 9.10. If $a = t_0 < t_1 < \cdots < t_n = b$ is a partition of $[a,b]$, then

$$
f'_-(t_{k-1}) \leqslant f'_+(t_{k-1}) \leqslant \frac{f(t_k) - f(t_{k-1})}{t_k - t_{k-1}} \leqslant f'_-(t_k) \leqslant f'_+(t_k),
$$

for all $1 \leqslant k \leqslant n$. Since

$$
f(b) - f(a) = \sum_{k=1}^{n} f(t_k) - f(t_{k-1}) = \sum_{k=1}^{n}(t_k - t_{k-1})\frac{f(t_k) - f(t_{k-1})}{t_k - t_{k-1}},
$$

we have that

$$f(b)-f(a)\leqslant\sum_{k=1}^{n}(t_k-t_{k-1})f'_+(t_{k-1})\leqslant\sum_{k=1}^{n}(t_k-t_{k-1})f'_-(t_k).$$

Taking the limit as $n\to\infty$ and the mesh of the partition goes to zero, we get that

$$f(b)-f(a)\leqslant\int_a^b f'_+(t)dt\leqslant\int_a^b f'_-(t)dt,$$

and from the other direction,

$$f(b)-f(a)\geqslant\int_a^b f'_-(t)dt\geqslant\int_a^b f'_+(t)dt.$$

As $f'_-(t)\leqslant h(t)\leqslant f'_+(t)$ for all $t\in[a,b]$, we get the equality in Lemma 9.10. □

Proof of Lemma 10.8. Suppose by contradiction that f is not Fréchet differentiable. Then there exists an $\varepsilon_0>0$ and a sequence $\{h_k\}_{k\in\mathbb{N}}$ with $h_k\to 0$ such that

$$\varepsilon_0\leqslant\frac{|f(\overline{x}+h_k)-f(\overline{x})-\langle v,h_k\rangle|}{\|h_k\|}.$$

Define $t_k=\|h_k\|$ and let $d_k=h_k/\|h_k\|$. Then we have that $t_k\to 0$ as $k\to\infty$, and also that $\|d_k\|=1$ for all $k\in\mathbb{N}$. Thus $\{d_k\}_{k\in\mathbb{N}}$ is a bounded sequence, and so it has a convergent subsequence $\{d_{k_j}\}_{j\in\mathbb{N}}$ (by the Bolzano-Weierstrass theorem). Without loss of generality, suppose that $d_k\to d$ as $k\to\infty$. Rewriting the inequality above, we get

$$\begin{aligned}\varepsilon_0&\leqslant\frac{|f(\overline{x}+h_k)-f(\overline{x})-\langle v,h_k\rangle|}{\|h_k\|}\\&=\frac{|f(\overline{x}+t_kd_k)-f(\overline{x})-\langle v,t_kd_k\rangle|}{t_k}\\&=\frac{|f(\overline{x}+t_kd_k)-f(\overline{x}+t_kd)+f(\overline{x}+t_kd)-f(\overline{x})-\langle v,t_kd_k\rangle+\langle v,t_kd\rangle-\langle v,t_kd\rangle|}{t_k}\\&\leqslant\frac{|f(\overline{x}+t_kd_k)-f(\overline{x}+t_kd)|}{t_k}+\frac{|f(\overline{x}+t_kd)-f(\overline{x})-\langle v,t_kd\rangle|}{t_k}+\frac{|\langle v,t_kd\rangle-\langle v,t_kd_k\rangle|}{t_k}.\end{aligned}$$

Now because $\overline{x}\in\int(\operatorname{dom} f)$, it must be the case that f is locally Lipschitz continuous around $\overline{x}$, and thus there exists an $\ell\geqslant 0$ and a $\delta>0$ such that $|f(x)-f(y)|\leqslant\ell\|x-y\|$ for all $x,y\in B(\overline{x},\delta)$. Applying this fact and the Cauchy-Schwarz inequality to the previous inequality, we get

$$\begin{aligned}\varepsilon_0&\leqslant\frac{\ell\,\|f(\overline{x}+t_kd_k)-f(\overline{x}+t_kd)\|}{t_k}+\frac{|f(\overline{x}+t_kd)-f(\overline{x})-\langle v,t_kd\rangle|}{t_k}+\frac{\|v\|\,\|t_kd-t_kd_k\|}{t_k}\\&=\ell\,\|d_k-d\|+\frac{|f(\overline{x}+t_kd)-f(\overline{x})-\langle v,t_kd\rangle|}{t_k}+\|v\|\,\|d-d_k\|.\end{aligned}$$

Now, using the Gâteaux differentiability of f at $\overline{x}$ on the middle term, we can see that the final term above goes to 0 as $k\to\infty$. But that contradicts the fact that it was greater than ε_0 for all $k\in\mathbb{N}$, and so f must be Fréchet differentiable at $\overline{x}$. □

Proof of Proposition 10.10. In some sufficiently small neighbourhood N_1 of a, we have the following Taylor expansion:

$$f(a+h) = f(a) + f'(a)\cdot h + o(\|h\|) = b + f'(a)\cdot h + o_f(\|h\|),$$

where $b = f(a) \in F$. Similarly, in some small neighbourhood M_2 of b, we have the Taylor expansion

$$g(b+k) = g(b) + g'(b)\cdot k + o_g(\|k\|),$$

where $g(b) \in G$. Now since

$$\left\|f'(a)\cdot h + o_f(\|h\|)\right\| \leqslant \left(\left\|f'(a)\right\| + \left\|\frac{o_f(\|h\|)}{h}\right\|\right)\|h\|,$$

one can restricth $h \in N_2 \subset N_1$ such that $k = f'(a)\cdot h + o_f(\|h\|) \in M_2$ has norm $\leqslant (\|f'(a)\|+1)\|h\|$. Should the case arise, further restrict $h \in N_3 \subset N_2$ such that $b+k \in M_2$ has norm $\leqslant (\|f'(a)\|+1)\|h\|$. Then we have that for $h \in N_3$,

$$\begin{aligned}(g\circ f)(a+h) &= g(b+k) = g(b) + g'(b)\cdot k + o_g(\|k\|)\\ &= g(b) + g'(b)\cdot(f'(a)\cdot h + o_f(\|h\|)) + o_g(\|k\|)\\ &= g(b) + g'(f(a))\cdot f'(a)\cdot h + g'(b)\cdot o_f(\|h\|) + o_g(\|k\|)\\ &= g(f(a)) + g'(f(a))\cdot f'(a)\cdot h + \varphi(h),\end{aligned}$$

where $\varphi(h) = g'(f(a))\cdot o_f(\|h\|) + o_g(\|k\|)$ is well-defined for $h \in N_3$; we need to show that when divided by $\|h\|$, the quotient tends to zero. But this follows nicely, because $\|k\| \leqslant (\|f'(a)\|+1)\|h\|$, and so $\varphi(h)$ is the sum of two $o(\|h\|)$ terms and thus is $o(\|h\|)$ itself. Therefore $(g\circ f)'(a) = g'(f(a))\cdot f'(a)$. ☐

Proof of Proposition 10.13.

1. Since f is twice differentiable at x, we have that

   $$f(x+tk) = f(x) + f'(x)tk + \frac{1}{2}f''(x)tk^2 + o(t),$$

   and so

   $$\frac{f'(x+tk)h - f'(x)h}{t} = f''(x)hk + o(1).$$

   Taking the limit as $t \searrow 0$, we get that

   $$f''(x)h = \lim_{t\searrow 0}\frac{f'(x+tk)h - f'(x)h}{t},$$

   as desired.

2. For $h, k \in E$, fix some dummy variable $a \in E$ that is the evaluation of some function in h and k. Consider the map

   $$\begin{aligned}F\colon \mathbb{R} &\to F\\ t &\mapsto f(a+th+tk) - f(a+th) - f(a+tk) + f(a),\end{aligned}$$

   and likewise the ancilliary map

   $$g\colon \mathbb{R} \to F$$

$$s \mapsto f(a + sh + tk) - f(a + sh).$$

Because $\partial^2 f(a)$ exists, there exists a neighbourhood N_1 of a such that $\partial f(x)$ exists for all $x \in N_1$. Hence f is differentiable on N_1, and it follows that there exists a neighbourhood N_2 of 0 such that g is differentiable on N_2. By the mean value theorem for vector-valued functions, we have

$$F(t) = g(t) - g(0) = \int_0^1 \partial g(\theta t)(t) \, \mathrm{d}\theta,$$

and by the chain rule, we have the derived map

$$\begin{aligned} \partial g(\theta) \colon R &\to F \\ \ell &\mapsto \partial f(a + \theta h + tk)(h\ell) - \partial f(a + \theta h)(h\ell). \end{aligned}$$

Hence

$$\begin{aligned} \partial g(\theta t)(t) &= \partial f(a + \theta th + tk)(ht) - \partial f(a + \theta th)(ht) \\ &= t\partial f(a + \theta th + tk)h - t\partial f(a + \theta th)h \\ &= t\left(\partial f(a + \theta th + tk) - \partial f(a + \theta th)\right)h. \end{aligned}$$

Because ∂f is differentiable at a, we have, for all $x \in E$,

$$\partial f(x) = \partial f(a) + \partial^2 f(a)(x - a) + \|x - a\| \cdot r(x),$$

where $r \colon E \to \mathcal{L}(E, F)$ is continuous at a and $r(a) = 0$. Therefore

$$\begin{aligned} \partial f(a + \theta th + tk) &= \partial f(a) + \partial^2 f(a)(\theta th + tk) + \|\theta th + tk\| \cdot r(a + \theta th + tk) \\ \partial f(a + \theta th) &= \partial f(a) + \partial^2 f(a)(\theta th) + \|\theta th\| \cdot r(a + \theta th), \end{aligned}$$

and consequently

$$\begin{aligned} \partial f(a + \theta th + tk) - \partial f(a + \theta th) &= \big(\partial^2 f(a)(\theta th + tk) - \partial^2 f(a)(\theta th)\big) \\ &\qquad + \|\theta th + tk\| \cdot r(a + \theta th + tk) - \|\theta th\| \cdot r(a + \theta th) \\ &= \partial^2 f(a)(tk) + \|\theta th + tk\| \cdot r(a + \theta th + tk) \\ &\qquad - \|\theta th\| \cdot r(a + \theta th) \\ &= t\partial^2 f(a)(k) + \|\theta th + tk\| \cdot r(a + \theta th + tk) \\ &\qquad - \|\theta th\| \cdot r(a + \theta th). \end{aligned}$$

Therefore

$$\begin{aligned} F(t) &= \int_0^1 t\left(t\partial^2 f(a)(k) + \|\theta th + tk\| \cdot r(a + \theta th + tk) - \|\theta th\| \cdot r(a + \theta th)\right) \mathrm{d}\theta \\ &= t^2\partial^2 f(a)(k) + t\int_0^1 \|\theta th + tk\| \cdot r(a + \theta th + tk)(h) \, \mathrm{d}\theta - t\int_0^1 \|\theta th\| \cdot r(a + \theta th)(h) \, \mathrm{d}\theta. \end{aligned}$$

Let $M = \|h\| + \|k\|$. For all $\theta \in [0, 1]$, we have

$$\begin{aligned} \|\|\theta th + tk\| \cdot r(a + \theta th + tk)(h)\| &= \|\theta th + tk\| \cdot \|r(a + \theta th + tk)(h)\| \\ &\leqslant (|\theta t| \, \|h\| + |t| \, \|k\|) \cdot \|r(a + \theta th + tk)\| \cdot \|h\| \end{aligned}$$

$$
\begin{aligned}
&\leqslant (|t|\,\|h\| + |t|\,\|k\|) \cdot \|r(a+\theta th+tk)\| \cdot \|h\| \\
&\leqslant (M|t| + M|t|) \cdot \|r(a+\theta th+tk)\| \cdot M \\
&= 2M^2|t| \cdot \|r(a+\theta th+tk)\|\,.
\end{aligned}
$$

It follows then that

$$
\begin{aligned}
\lim_{t\to 0}\left\|\frac{\|\theta th+tk\|\cdot r(a+\theta th+tk)(h)}{t}\right\| &= \lim_{t\to 0}\frac{\|\|\theta th+tk\|\cdot r(a+\theta th+tk)(h)\|}{|t|} \\
&\leqslant \lim_{t\to 0}\frac{2M^2|t|\cdot\|r(a+\theta th+tk)\|\cdot M}{|t|} \\
&= 2M^2\lim_{t\to 0}\|r(a+\theta th+tk)\| = 0,
\end{aligned}
$$

since, for all $\theta\in[0,1]$, we have $\theta t\to 0$, $a+\theta th+tk\to a$ as $t\to 0$ and since r is continuous at a and $r(a)=0$. So we have that

$$
\lim_{t\to 0}\frac{\|\theta th+tk\|\cdot r(a+\theta th+tk)(h)}{t} = 0,
$$

and it follows that for all $\delta>0$, there is some $\varepsilon>0$ such that for all $t\in(-\varepsilon,\varepsilon)$ and $\theta\in[0,1]$, we have

$$
\left\|\frac{\|\theta th+tk\|\cdot r(a+\theta th+tk)(h)}{t}\right\| < \delta \implies \int_0^1\left\|\frac{\|\theta th+tk\|\cdot r(a+\theta th+tk)(h)}{t}\right\|\mathrm{d}\theta < \delta.
$$

Hence

$$
\left\|\int_0^1\frac{\|\theta th+tk\|\cdot r(a+\theta th+tk)(h)}{t}\,\mathrm{d}\theta\right\| < \delta,
$$

and consequently

$$
\lim_{t\to 0}\int_0^1\frac{\|\theta th+tk\|\cdot r(a+\theta th+tk)(h)}{t}\,\mathrm{d}\theta = 0.
$$

With pretty much the same argument, we can show that

$$
\lim_{t\to 0}\int_0^1\frac{\|\theta th\|\cdot r(a+\theta th)(h)}{t}\,\mathrm{d}\theta = 0.
$$

As such, we have that

$$
\begin{aligned}
\lim_{t\to 0}\frac{F(t)}{t^2} &= \frac{t^2\partial^2 f(a)(k) + t\int_0^1\|\theta th+tk\|\cdot r(a+\theta th+tk)(h)\,\mathrm{d}\theta - t\int_0^1\|\theta th\|\cdot r(a+\theta th)(h)\,\mathrm{d}\theta}{t^2} \\
&= \lim_{t\to 0}\partial^2 f(a)(k)(h) + \lim_{t\to 0}\int_0^1\frac{\|\theta th+tk\|\cdot r(a+\theta th+tk)(h)}{t}\,\mathrm{d}\theta \\
&\qquad - \lim_{t\to 0}\int_0^1\frac{\|\theta th\|\cdot r(a+\theta th)(h)}{t}\,\mathrm{d}\theta \\
&= \partial^2 f(a)(k)(h).
\end{aligned}
$$

Now consider the map

$$
\overline{g}\colon \mathbb{R}\to F
$$

$$s \mapsto f(a+th+sk) - f(a+sk).$$

With a similar reasoning as above, we get that

$$\lim_{t \to 0} \frac{\overline{g}(t)}{t} = \partial^2 f(a)(h)(k),$$

and so $\partial^2 f(a)(h)(k) = \partial^2 f(a)(k)(h)$. Since $h, k \in E$ were arbitrary, we have that $\partial^2 f(a)$ is symmetric. Bilinearity follows from the linearity of the derivative. $\square$

References

[Bec17] Amir Beck. *First-order methods in optimization.* SIAM, 2017. 3

[BV04] Stephen P Boyd and Lieven Vandenberghe. *Convex optimization.* Cambridge university press, 2004. 3

[Car07] C. Caratheodory. Uber den variabilitatsbereich der koeffizienten von potenzreihen, die gegebene werte nicht annehmen. *Math. Ann.*, 64:95–115, 1907. 15

[CT05] Emmanuel J Candes and Terence Tao. Decoding by linear programming. *IEEE transactions on information theory*, 51(12):4203–4215, 2005. 8

[Far02] Julius Farkas. Theorie der einfachen ungleichungen. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 1902(124):1–27, 1902. 36

[Hel23] E. Helly. Uber mengen konvexer kbrper mit gemeinschaftlichen punkten. *Jber. Deutsch. Math. Verein.*, 32:175–176, 1923. 19

[N+18] Yurii Nesterov et al. *Lectures on convex optimization*, volume 137. Springer, 2018. 3

[Nes09] Yurii Nesterov. Primal-dual subgradient methods for convex problems. *Mathematical programming*, 120(1):221–259, 2009. 78

[Rad21] J. Radon. Mengen konvexer korper, die einen gemeinsamen punkt enthalten. *Math. Ann.*, 83:113–115, 1921. 18

[Roc15] Ralph Tyrell Rockafellar. Convex analysis. *Princeton University Press*, 2015. 3