Math 673/4

Multigrid Methods: A Mostly Matrix-Based Approach

Chapter 10: Convex Optimization

Abner J. Salgado and Steven M. Wise

asalgad1@utk.edu swise1@.utk.edu
University of Tennessee

F24/S25

Chapter 10, Part 1 of 2
Convex Optimization

The Goal

The purpose of this chapter is to present some background information regarding basic (linear) functional analysis, convex functions, and their minimization. As you will see, we deviate significantly from our "mostly matrix-based" approach in this chapter. Here we will set up the language and tools for the next chapter, which deals with sophisticated methods for convex optimization problems based on subspace decompositions.

Analysis on Vector Spaces

Let us begin by recalling some basic facts about linear functional analysis. For further developments we refer the reader to the book by Ciarlet (SIAM, 2013).

We recall that a Banach space is a pair $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$, where $\mathcal{V}$ is a vector space, $\|\cdot\|_{\mathcal{V}}$ is an *norm* on $\mathcal{V}$, and this space is *complete* under this norm. In addition, if this norm is induced by an *inner product* $(\cdot,\cdot)_{\mathcal{V}}$, we call this a Hilbert space.

Example (Finite Dimensional Spaces)

Let $n \in \mathbb{N}$ and consider $\mathbb{R}^n$. For $p \in [1, \infty]$ we define

$$\|\boldsymbol{v}\|_p := \begin{cases} \left(\sum_{i=1}^{n} |v_i|^p \right)^{1/p}, & p \in [1, \infty), \\ \max_{i=1}^{n} |v_i|, & p = \infty. \end{cases}$$

Then, for every $p \in [1, \infty]$, $(\mathbb{R}^n, \| \cdot \|_p)$ is a Banach space. In addition, for $p = 2$, we clearly see that this is the Euclidean norm, and so this is a Hilbert space.

It is well-known that in finite dimensions all norms are equivalent. By the previous example then, any finite dimensional vector space is complete under any norm.

The following example shows that in infinite dimensions the choice of norm matters.

Example (Continuous Functions)

Let $d \in \mathbb{N}$ and $\Omega \subset \mathbb{R}^d$ be a bounded domain. The space of functions $C(\overline{\Omega})$ can be endowed with the norm

$$\|f\|_{L^2(\Omega)} := \left(\int_\Omega |f(\boldsymbol{x})|^2 \, \mathrm{d}\boldsymbol{x} \right)^{1/2}.$$

Normed this way, however, this space is not complete. On the other hand, the norm

$$\|f\|_{L^\infty(\Omega)} := \sup_{\boldsymbol{x} \in \overline{\Omega}} |f(\boldsymbol{x})|$$

makes this space complete.

As it is customary, in what follows, if it is clear from the context we shall refer to a Banach or Hilbert space by simply mentioning the vector space, but not necessarily the norm or inner product.

We now consider the natural mappings between normed vector spaces, i.e., the continuous linear mappings.

Definition (Continuous Linear Mappings)

Let $\mathcal{X}$ and $\mathcal{Y}$ be normed spaces. The collection of all mappings with domain $\mathcal{X}$ and range $\mathcal{Y}$ that are both linear and continuous is denoted by

$$\mathcal{L}(\mathcal{X}, \mathcal{Y}).$$

A **functional** on $\mathcal{X}$ is a mapping $\mathcal{X} \to \mathbb{R}$. The **dual** space to $\mathcal{X}$ is the space

$$\mathcal{X}' := \mathcal{L}(\mathcal{X}, \mathbb{R}).$$

The following result details some properties of the continuous linear mappings.

Proposition (Properties of $\mathcal{L}(\mathcal{X}, \mathcal{Y})$)

Let $\mathcal{X}$ and $\mathcal{Y}$ be normed spaces and $T : \mathcal{X} \to \mathcal{Y}$ be linear. The following are equivalent.

1. $T \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$.
2. *T is continuous at zero, i.e., for every $\{v_n\}_{n=1}^{\infty} \subset \mathcal{X}$ such that $v_n \to 0$ in the norm of $\mathcal{X}$, we have*
$$Tv_n \to 0,$$
in the norm of $\mathcal{Y}$.
3. *The mapping T is bounded, i.e., it maps bounded sets in $\mathcal{X}$ to bounded sets in $\mathcal{Y}$, i.e., there is a constant $M > 0$ such that, for every $v \in \mathcal{X}$,*
$$\|Tv\|_{\mathcal{Y}} \leq M\|v\|_{\mathcal{X}}. \tag{1}$$

Proof.

Exercise. □

Definition (Induced Norm)

Let $\mathcal{X}$ and $\mathcal{Y}$ be normed spaces. For $T \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ the quantity

$$\|T\|_{\mathcal{L}(\mathcal{X},\mathcal{Y})} := \sup_{v \in \mathcal{X} \setminus \{0\}} \frac{\|Tv\|_{\mathcal{Y}}}{\|v\|_{\mathcal{X}}}, \tag{2}$$

is called the **induced** or **operator norm** on $\mathcal{L}(\mathcal{X}, \mathcal{Y})$.

Note that, for the special case that $T \in \mathcal{X}'$, the operator norm is simply

$$\|T\|_{\mathcal{X}'} = \sup_{v \in \mathcal{X} \setminus \{0\}} \frac{|Tv|}{\|v\|_{\mathcal{X}}}. \tag{3}$$

Proposition ($\mathcal{L}(\mathcal{X}, \mathcal{Y})$ is Banach)

Let $\mathcal{X}$ and $\mathcal{Y}$ be normed spaces. The set $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ is a vector space, and the induced norm is a bona fide norm on the space. In addition, if $\mathcal{Y}$ is Banach, the so is $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ under the induced norm. Consequently, the dual space $\mathcal{X}'$ to any normed space $\mathcal{X}$ is always complete.

Proof.

Exercise. $\square$

Example (Finite Dimensional Spaces)

Let $m, n \in \mathbb{N}$. The space of matrices $\mathbb{R}^{m\times n}$ can be identified with the space $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ under the action of matrix vector multiplication. If we endow $\mathbb{R}^m$ and $\mathbb{R}^n$ with the Euclidean norm, then $\mathbb{R}^{m\times n}$ is normed with the induced norm, which is defined as in (2).

The dual space to $\mathbb{R}^n$ can be identified with $\mathbb{R}^n$ itself. If F is a continuous linear functional on $\mathbb{R}^n$, then it can be uniquely written as

$$F(\boldsymbol{u}) = \boldsymbol{v}^\top \boldsymbol{u},$$

for some $\boldsymbol{v} \in \mathbb{R}^n$.

The following examples show that, again, in infinite dimensions things become more delicate.

Example (Not Bounded)

Consider the following normed linear spaces:

$$(\mathcal{X}, \|\cdot\|_{\mathcal{X}}) = \left(C^1([0,1]), \|\cdot\|_{L^\infty(0,1)}\right)$$

and

$$(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}}) = \left(C([0,1]), \|\cdot\|_{L^\infty(0,1)}\right).$$

The $\mathcal{X}$ space with this norm is not a Banach space, but the $\mathcal{Y}$ space is Banach. Define $T : \mathcal{X} \to \mathcal{Y}$ as

$$(Tf)(x) = f'(x). \tag{4}$$

Clearly, this is a linear mapping. However, it is not bounded.

Example (Cont.)

To see this, consider, for $N \in \mathbb{N}$, the family of functions

$$f_N(x) = \sin(2\pi N x).$$

Clearly, for all $N \in \mathbb{N}$,

$$\|f_N\|_{L^\infty(0,1)} = 1,$$

whereas

$$f_N'(x) = 2\pi N \sin(2\pi N x)$$

which implies that

$$\|f_N'\|_{L^\infty(0,1)} = 2\pi N.$$

Since N can be arbitrarily large, the mapping T is not bounded.

Example (Cont.)

If, on the other hand, we consider the same mapping but with a different domain, then this mapping may be bounded. For instance, for every $u \in C^1([0,1])$, define

$$\|u\|_{C^1([0,1])} := \|u\|_{L^\infty(0,1)} + \|u'\|_{L^\infty(0,1)},$$

and consider

$$(\mathcal{X}, \|\cdot\|_{\mathcal{X}}) = \left(C^1([0,1]), \|\cdot\|_{C^1([0,1])}\right).$$

This version of the $\mathcal{X}$ space is Banach, incidentally. Furthermore, $T : \mathcal{X} \to \mathcal{Y}$, defined as in (4), is now a bounded linear operator.

Example (Linear Functional)

Fix $g \in L^2(0,1)$. The mapping

$$L^2(0,1) \ni f \mapsto T_g(f) := (f,g)_{L^2(0,1)} := \int_0^1 f(x)g(x)\,\mathrm{d}x$$

defines a bounded linear functional on $L^2(0,1)$.

Previous examples suggest a similarity between the action of a bounded linear functional and an inner product. This similarity is the motivation for the following notation.

Definition (Duality Pairing)

Let $\mathcal{X}$ be a Banach space and $\mathcal{X}'$ be its dual. The **duality pairing** between $\mathcal{X}$ and $\mathcal{X}'$ is the mapping

$$\mathcal{X} \times \mathcal{X}' \ni (u, F) \mapsto \langle F, u \rangle_{\mathcal{X}',\mathcal{X}} \in \mathbb{R},$$

where

$$\langle F, u \rangle_{\mathcal{X}',\mathcal{X}} := F(u).$$

If the space $\mathcal{X}$ is understood from the context we shall suppress the subindices $\mathcal{X}', \mathcal{X}$ from this notation.

For Hilbert spaces this duality pairing turns out to, essentially, coincide with the inner product.

Theorem (Riesz Representation)

Let $\mathcal{V}$ be a Hilbert space. There is a linear isometry between $\mathcal{V}$ and its dual $\mathcal{V}'$. In other words, for every $F \in \mathcal{V}'$ there is a unique $\mathfrak{R}_{\mathcal{V}} F \in \mathcal{V}$ such that

$$\langle F, u \rangle_{\mathcal{V}', \mathcal{V}} = (\mathfrak{R}_{\mathcal{V}} F, u)_{\mathcal{V}}, \qquad \forall u \in \mathcal{V}.$$

Moreover,

$$\|\mathfrak{R}_{\mathcal{V}} F\|_{\mathcal{V}} = \|F\|_{\mathcal{V}'}, \qquad \forall F \in \mathcal{V}'.$$

Proof.

Exercise. □

Definition (Riesz Map)

The mapping $\mathfrak{R}_{\mathcal{V}} \in \mathcal{L}(\mathcal{V}', \mathcal{V})$ of Theorem 11 is called the **canonical Riesz map**. If it is clear from the context, the subindex $\mathcal{V}$ shall be suppressed.

Definition (Weak Convergence)

Let $\mathcal{X}$ be a Banach space and $\mathcal{X}'$ its dual. We say that the sequence $\{v_n\}_{n=1}^{\infty} \subset \mathcal{X}$ **converges weakly** to $v \in \mathcal{X}$ iff

$$\lim_{n \to \infty} \langle F, v_n - v \rangle = 0, \qquad \forall\, F \in \mathcal{X}'.$$

We shall denote this convergence via the notation $v_n \rightharpoonup v$. A set $K \subseteq \mathcal{X}$ is called **weakly closed** iff it contains all of its weak limit points. A set $O \subseteq \mathcal{X}$ is called **weakly open** iff $\mathcal{X} \setminus O$ is weakly closed.

Example (Weak Convergence)

Consider $\{v_n\}_{n=1}^{\infty} \subset L^2(-\pi, \pi)$ defined by

$$v_n(x) = \sin(nx).$$

Let us show that this sequence is not convergent in norm. If this was the case, then this sequence would be Cauchy. However, if $m, n \in \mathbb{N}$ with $m \neq n$,

$$\|v_m - v_n\|_{L^2(-\pi,\pi)}^2 = 2\pi,$$

as the reader may easily verify.

Example (Cont.)

On the other hand, we have that $v_n \rightharpoonup 0$ as we will now show. To obtain this we invoke the Riesz representation Theorem and recall that any bounded linear functional has the form

$$v \mapsto \int_{-\pi}^{\pi} f(x)v(x)\,\mathrm{d}x,$$

for some $f \in L^2(-\pi,\pi)$. We shall additionally need to use the fact that step functions are *dense* in $L^2(-\pi,\pi)$. Thus, fix $f \in L^2(-\pi,\pi)$ and $\varepsilon > 0$. There are $m \in \mathbb{N}$, $\{c_i\}_{i=0}^m \subset [-\pi,\pi]$, and $\{A_i\}_{i=1}^m \subset \mathbb{R}$ with

$$-\pi \leq c_0 < c_2 < \cdots < c_m \leq \pi,$$

such that, if we define,

$$f_\varepsilon(x) = \sum_{i=1}^{m} A_i \chi_{(c_{i-1},c_i]}(x)$$

then we have

$$\|f - f_\varepsilon\|_{L^2(-\pi,\pi)} < \frac{\varepsilon}{2\sqrt{\pi}}.$$

Example (Cont.)

Define

$$I_n = \left| \int_{-\pi}^{\pi} f(x) v_n(x) \, \mathrm{d}x \right|$$

and consider

$$\begin{aligned} I_n &\leq \|f - f_\varepsilon\|_{L^2(-\pi,\pi)} \|v_n\|_{L^2(-\pi,\pi)} + \left| \int_{-\pi}^{\pi} f_\varepsilon(x) v_n(x) \, \mathrm{d}x \right| \\ &\leq \frac{\varepsilon}{2} + \left| \sum_{i=1}^{m} A_i \int_{c_{i-1}}^{c_i} v_n(x) \mathrm{d}x \right| \\ &\leq \frac{\varepsilon}{2} + m \max_{i=1}^{m} |A_i| \max_{i=1}^{m} \left| \int_{c_{i-1}}^{c_i} v_n(x) \mathrm{d}x \right| . \end{aligned} \tag{5}$$

Example (Cont.)

As a side computation let us, for $c_L, c_R \in [-\pi, \pi]$ with $c_L < c_R$, estimate

$$\left| \int_{c_L}^{c_R} \sin(nx) \, \mathrm{d}x \right| = \frac{1}{n} \left| \cos(nc_L) - \cos(nc_R) \right| \\ \leq \frac{2}{n}.$$

This shows that, if

$$n \geq n_0 := \left\lceil \frac{4m \max_{i=1}^{m} |A_i|}{\varepsilon} \right\rceil,$$

then $I_n < \varepsilon$, and this shows the claimed weak convergence.

Nonlinear Mappings

While much more can be said about Banach spaces, the linear mappings between them, dual spaces, and their representation, we now turn our attention to nonlinear mappings and how we measure their smoothness.

Definition (Lower Semicontinuity)

Let $\mathcal{V}$ be a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$. We say that E is **lower semicontinuous (LSC)** at $v \in \mathcal{V}$ iff, whenever $v_n \to v$ in $\mathcal{V}$, we have

$$E(v) \leq \liminf_{n \to \infty} E(v_n).$$

Similarly, we say that E is **weakly lower semicontinuous (wLSC)** at $v \in \mathcal{V}$ iff, whenever $v_n \rightharpoonup v$ in $\mathcal{V}$, we have

$$E(v) \leq \liminf_{n \to \infty} E(v_n).$$

If $K \subset \mathcal{V}$ and E is (weakly) lower semicontinuous for every $v \in K$, then we say that E is (weakly) lower semicontinuous on K.

Definition (Epigraph)

Let $\mathcal{V}$ be a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$. The **epigraph** of E is the set

$$\operatorname{epi}(E) \coloneqq \{(v, \alpha) \in \mathcal{V} \times \mathbb{R} \mid E(v) \leq \alpha\} .$$

Definition (Sublevel Set)

Let $\mathcal{V}$ be a Hilbert space, $E : \mathcal{V} \to \mathbb{R}$, and $\alpha \in \mathbb{R}$. The **sublevel set** at height α of E is

$$\operatorname{level}_\alpha(E) \coloneqq \{v \in \mathcal{V} \mid E(v) \leq \alpha\} .$$

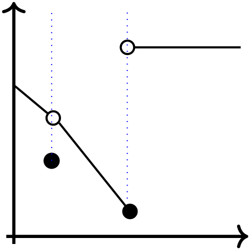

Figure: The figure on the left depicts the graph of a lower semicontinuous function, whereas the right is the graph of a function that is not lower semicontinuous. As the next proposition shows, the epigraph of the function on the left is closed. This can be seen by the dotted blue lines.

Proposition (wLSC)

Let $\mathcal{V}$ be a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$. The functional E is (weakly) LSC iff its epigraph is (weakly) closed iff for every $\alpha \in \mathbb{R}$ the set $\text{level}_\alpha E$ is (weakly) closed.

Proof.

Exercise. $\square$

We now turn our attention to differentiable mappings.

Definition (Fréchet Differentiability)

Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces and $F : \mathcal{X} \to \mathcal{Y}$. We say that F is **Fréchet differentiable** at $v \in \mathcal{X}$ iff there is a mapping $T \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ such that

$$\lim_{\|h\|_{\mathcal{X}} \to 0} \frac{\|F(v+h) - F(v) - Th\|_{\mathcal{Y}}}{\|h\|_{\mathcal{X}}} = 0.$$

The, necessarily unique, T is called the **Fréchet derivative** of F at v and it is denoted by $\mathrm{D}F(v) \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. If $K \subseteq \mathcal{X}$ and F is Fréchet differentiable at every $v \in K$, then we say that F is Fréchet differentiable on K.

Remark (Derivative of a Functional)

Let $\mathcal{V}$ be a Hilbert space, $E : \mathcal{V} \to \mathbb{R}$ be a functional, and $v \in \mathcal{V}$. The Fréchet derivative of E at v, if it exists, is an element of $\mathcal{V}'$, and so we write

$$\langle \mathrm{D}E(v), w \rangle := \mathrm{D}E(v)w, \qquad \forall\, w \in \mathcal{V}.$$

Definition

Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces and $F : \mathcal{X} \to \mathcal{Y}$. Suppose that F is Fréchet differentiable at every point of the set $K \subseteq \mathcal{X}$. The derivative defines a mapping

$$\mathcal{X} \supseteq K \ni v \mapsto \mathrm{D}F(v) \in \mathcal{L}(\mathcal{X}, \mathcal{Y}),$$

which is naturally denoted $\mathrm{D}F : K \to \mathcal{L}(\mathcal{X}, \mathcal{Y})$. If this mapping is continuous, we say that F **continuously Fréchet differentiable** on K.

Definition (Convex Sets)

Let $\mathcal{X}$ be a vector space and $K \subseteq \mathcal{X}$. We say that K is **convex** iff whenever $v, w \in K$ and $t \in [0, 1]$ we have

$$tv + (1 - t)w \in K.$$

If F is continuously Fréchet differentiable on a convex set K, we have a nice analogue of Taylor's theorem with integral remainder.

Proposition (Taylor)

Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces and $F : \mathcal{X} \to \mathcal{Y}$ be continuously Fréchet differentiable on a convex subset $K \subseteq \mathcal{X}$. Then, for every $v, w \in K$ we have

$$F(w) - F(v) = \int_0^1 \mathrm{D}F(tw + (1-t)v)(w - v)\,\mathrm{d}t.$$

Second Derivatives

In a similar way that we can study the continuity of the derivative mapping, we can study its differentiability. While this can be developed in full generality, we restrict our attention to the case of functionals, which will be more than sufficient for our purposes.

Definition (Second Derivative)

Let $\mathcal{V}$ be a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$ be a continuously Fréchet differentiable functional on the subset $K \subseteq \mathcal{V}$. If the mapping

$$\mathrm{D}E : K \to \mathcal{V}', \qquad \mathcal{V} \supseteq K \ni v \mapsto \mathrm{D}E(v) \in \mathcal{V}'$$

is Fréchet differentiable for all $v \in K \subseteq \mathcal{V}$, then its derivative

$$\mathrm{D}^2E(v) \in \mathcal{L}\left(\mathcal{V}, \mathcal{L}(\mathcal{V}, \mathbb{R})\right) = \mathcal{L}(\mathcal{V}, \mathcal{V}'),$$

is called the **second Fréchet derivative** of E at $v \in K$, and we say that E is **twice Fréchet differentiable** on K.

As before, we can explore when a functional is twice Fréchet differentiable on a set, or when it is twice continuously Fréchet differentiable at a point or on a set.

Next we observe that, since $\mathrm{D}^2 E(v) \in \mathcal{L}(\mathcal{V}, \mathcal{V}')$, for $w_1, w_2 \in \mathcal{V}$ we have

$$\mathrm{D}^2 E(v) w_1 \in \mathcal{V}', \qquad \mathrm{D}^2 E(v)(w_1, w_2) := \langle \mathrm{D}^2 E(v) w_1, w_2 \rangle.$$

In other words, $\mathrm{D}^2 E(v)$ defines a bilinear form on $\mathcal{V}$. The following result shows that this bilinear form is symmetric.

Proposition (Symmetry)

Let $\mathcal{V}$ be a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$ be twice Fréchet differentiable at $v \in \mathcal{V}$. Then,

$$\mathrm{D}^2 E(v)(w_1, w_2) = \mathrm{D}^2 E(v)(w_2, w_1), \qquad \forall\, w_1, w_2 \in \mathcal{V}.$$

We now characterize the *growth* of a functional.

Definition (Coercivity)

Let $\mathcal{V}$ be a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$. We say that E is **coercive** iff whenever $\|v_n\|_{\mathcal{V}} \to \infty$ we have

$$E(v_n) \to \infty.$$

Finally, we present a second order version of Taylor's theorem.

Proposition (Second Order Taylor)

Let $\mathcal{V}$ be a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$ be twice continuously Fréchet differentiable on a convex subset $K \subseteq \mathcal{V}$. Then, for every $v, w \in K$ we have

$$E(w) - E(v) = \langle \mathrm{D}E(v), w - v \rangle + \int_0^1 (1-t) \mathrm{D}^2 E(v + t(w-v))(w-v, w-v) \, \mathrm{d}t.$$

Moreover, for every $h \in \mathcal{V}$,

$$\langle \mathrm{D}E(w) - \mathrm{D}E(v), h \rangle = \int_0^1 \mathrm{D}^2 E(v + t(w-v))(w-v, h) \, \mathrm{d}t$$

As an application we give a sufficient condition for coercivity of a functional.

Corollary (Coercivity)

Let $\mathcal{V}$ be a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$ be twice continuously differentiable on $\mathcal{V}$. If there is a constant $\mu > 0$ such that, for every $v, w \in \mathcal{V}$, we have

$$\mathrm{D}^2 E(v)(w, w) \geq c\|w\|_{\mathcal{V}}^2,$$

then E is coercive.

Proof.

Let w be arbitrary. Owing to Proposition 1.7 we have

$$\begin{aligned} E(w) &= E(0) + \langle \mathrm{D}E(0), w \rangle + \int_0^1 (1-t)\mathrm{D}^2 E(tw)(w, w)\mathrm{d}t \\ &\geq E(0) - \|\mathrm{D}E(0)\|_{\mathcal{V}'}\|w\|_{\mathcal{V}} + \frac{\mu}{2}\|w\|_{\mathcal{V}}^2, \end{aligned}$$

where, to obtain the lower bound, we used the assumption on the second derivative. This clearly shows that, as $\|w\|_{\mathcal{V}} \to \infty$, we have $E(w) \to \infty$, and thus the functional E is coercive. $\square$

Example

Suppose that $n \in \mathbb{N}$ and $\mathsf{A} \in \mathbb{R}^{n \times n}_{\text{sym}}$. For every $\boldsymbol{x} \in \mathbb{R}^n$ define the functional

$$E(\boldsymbol{x}) := \frac{1}{2}\boldsymbol{x}^\top \mathsf{A}\boldsymbol{x} + \frac{1}{4}\sum_{i=1}^{n}(x_i)^4.$$

This functional is twice continuously Fréchet differentiable. Recall that the gradient of E and the Hessian matrix of E a point $\boldsymbol{z} \in \mathbb{R}$ are defined, respectively, as

$$\nabla E(\boldsymbol{z}) = \mathsf{A}\boldsymbol{z} + \begin{bmatrix} z_1^3 \\ z_2^3 \\ \vdots \\ z_n^3 \end{bmatrix}$$

and

$$\mathsf{H}_E(\boldsymbol{z}) = \mathsf{A} + 3\,\mathrm{diag}\left(z_1^2, z_2^2, \ldots, z_n^2\right).$$

Example (Cont.)

Then, one can prove that, for all $\boldsymbol{x} \in \mathbb{R}^n$,

$$\langle \mathrm{D}E(z), \boldsymbol{x} \rangle = \boldsymbol{x}^\top \nabla E(\boldsymbol{z})$$

and, for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,

$$\langle \mathrm{D}^2 E(z)\boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{y}^\top \mathrm{H}_E(\boldsymbol{z})\boldsymbol{x}.$$

$\mathrm{H}_E(\boldsymbol{z})$ is symmetric. Furthermore, if A is positive definite, then $\mathrm{H}_E(\boldsymbol{z})$, is also positive definite at every point $\boldsymbol{z} \in \mathbb{R}^n$. In this case, E is also strictly convex and coercive.

Convex Sets and Convex Functions

We now turn our attention more fully to the fundamental notion of convexity. For further details, we refer to Ekeland and Temam (1976). First, let us define the sum of two sets.

Definition (Minkowski sum)

Let $\mathcal{X}$ be a vector space and $A, B \subseteq \mathcal{X}$. The **(Minkowski) sum** of these sets is

$$A + B := \{a + b \in \mathcal{X} \mid a \in A, b \in B\}.$$

Similarly, for $\lambda \in \mathbb{R}$, we define

$$\lambda A := \{\lambda a \in \mathcal{X} \mid a \in A\}.$$

Proposition (Properties of Convex Sets)

Let $\mathcal{X}$ be a vector space.

1. *Let $\mathcal{A}$ be an index set and, for every $\alpha \in \mathcal{A}$, $K_\alpha \subset \mathcal{X}$ be convex. Then,*
$$\bigcap_{\alpha \in \mathcal{A}} K_\alpha$$
is convex.
2. *Let $\{K_i\}_{i=1}^\infty$ be an increasing family of convex subsets of $\mathcal{X}$, that is $K_1 \subset K_2 \subset \cdots \subset \mathcal{X}$. Then,*
$$\bigcup_{i=1}^\infty K_i$$
is convex.
3. *Let $N \in \mathbb{N}$, $\{K_i\}_{i=1}^N$ be a (finite) family of convex subsets of $\mathcal{X}$, and $\{\lambda_i\}_{i=1}^N \subset \mathbb{R}$. Then,*
$$\sum_{i=1}^N \lambda_i K_i$$
is convex.

Proposition (Cont.)

4. *Let $\mathcal{Y}$ be another vector space and $T : \mathcal{X} \to \mathcal{Y}$ be a linear, not necessarily bounded, map. The set $K \subset \mathcal{X}$ is convex iff $T(K) \subset \mathcal{Y}$ is convex.*

Proof.

We prove the first two and leave the last two for exercises.

1. Denote by W the intersection of the family of sets. Let $v, w \in W$ and $t \in [0,1]$. By definition of intersection, for every $\alpha \in \mathcal{A}$, we have $v, w \in K_\alpha$. Since K_α is convex, then $tv + (1-t)w \in K_\alpha$. Thus, $tv + (1-t)w \in W$.
2. Denote by W the union of the increasing family of sets. Let $v, w \in W$ and $t \in [0,1]$. Then, there are $N_v, N_w \in \mathbb{N}$ for which $v \in K_{N_v}$ and $w \in K_{N_w}$, respectively. Set $N = \max\{N_v, N_w\}$. Since the family of sets is increasing, we have that

$$K_{N_v} \cup K_{N_w} \subset K_N.$$

Thus, $v, w \in K_N$. Since K_N is convex $tv + (1-t)w \in K_N$, and this implies $tv + (1-t)w \in W$. □

Before we state the next result we observe that, if $\mathcal{V}$ is a Hilbert space and $v_n \to v$, then $v_n \rightharpoonup v$. Indeed, if $F \in \mathcal{V}'$, then

$$|\langle F, v_n - v \rangle| \leq \|F\|_{\mathcal{V}'} \|v_n - v\|_{\mathcal{V}} \to 0.$$

This shows that if a set is weakly closed, then it is closed. The converse, however, is in general false. The following result shows that, provided the set is convex, the converse is also true.

Lemma (Mazur)

Let $\mathcal{V}$ be a Hilbert space and $K \subseteq \mathcal{V}$. If K is convex and closed, then it is weakly closed.

We now turn our attention to convex functions. We restrict ourselves to the case we are interested here, functionals on Hilbert spaces.

Definition (Convex Functional)

Let $\mathcal{V}$ be a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$. We say that E is **convex** iff, for every $v, w \in \mathcal{V}$ and all $t \in [0, 1]$, we have

$$E(tv + (1 - t)w) \leq tE(v) + (1 - t)E(w).$$

We say that E is **strictly convex** iff, whenever $v \neq w$ and $t \in (0, 1)$, the inequality above is strict.

Example (Convex Functionals)

The following are examples of convex functionals.

- If $\mathcal{V}$ is a Hilbert space, the norm is convex.
- Let $n \in \mathbb{N}$ and $\mathsf{A} \in \mathbb{R}^{n \times n}$ be SPSD. The function
$$\Phi(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \mathsf{A}\boldsymbol{x}$$
is convex. If A is SPD then Φ is strictly convex.
- Let $n \in \mathbb{N}$ and $K \subset \mathbb{R}^n$. Its support function is
$$1_K^\star(\boldsymbol{w}) = \sup_{\boldsymbol{v} \in K} \boldsymbol{w}^\top \boldsymbol{v}.$$
This function is convex.

Proposition (Properties of Convex Functionals)

Let $\mathcal{V}$ be a Hilbert space.

1. *The functional $E : \mathcal{V} \to \mathbb{R}$ is convex iff* $\operatorname{epi}(E)$ *is convex.*
2. *If the functional $E : \mathcal{V} \to \mathbb{R}$ is convex, then for every $\alpha \in \mathbb{R}$ the set* $\operatorname{level}_\alpha(E)$ *is convex.*
3. *Let $\alpha \in \mathcal{A}$ be an index set and, for every $\alpha \in \mathcal{A}$, $E_\alpha : \mathcal{V} \to \mathbb{R}$ be a functional. Define*
$$E(v) := \sup_{\alpha \in \mathcal{A}} \{E_\alpha(v)\}, \qquad \forall\, v \in \mathcal{V}.$$
If every E_α is convex, then E is convex.
4. *If the functional $E : \mathcal{V} \to \mathbb{R}$ is convex and LSC, then it is wLSC.*
5. *Let the functional $E : \mathcal{V} \to \mathbb{R}$ be Fréchet differentiable. E is convex iff*
$$E(v) \geq E(w) + \langle \mathrm{D}E(w), v - w \rangle, \qquad \forall\, v, w \in \mathcal{V}.$$
In addition, E is strictly convex iff the inequality is strict for $v \neq w$.

Proposition (Cont.)

6. *If the functional $E : \mathcal{V} \to \mathbb{R}$ is convex and Fréchet differentiable, then its Fréchet derivative is monotone, that is,*

$$\langle \mathrm{D}E(v) - \mathrm{D}E(w), v - w \rangle \geq 0, \qquad \forall\, v, w, \in \mathcal{V}.$$

In addition, if E is strictly convex, then the derivative is strictly monotone, that is, the inequality is strict for $v \neq w$.

7. *Let the functional $E : \mathcal{V} \to \mathbb{R}$ be twice continuously Fréchet differentiable. If the second Fréchet derivative is positive semidefinite, i.e., for every $v, w \in \mathcal{V}$ we have*

$$\mathrm{D}^2 E(v)(w, w) \geq 0,$$

then E is convex. In addition, if the second derivative is positive definite, meaning the inequality above is strict for $w \neq 0$, then E is strictly convex.

We comment that the converse of property 2 is false.

We now introduce a notion that makes strict convexity more quantitative.

Definition (Strong Convexity)

Let $\mathcal{V}$ be a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$ be Fréchet differentiable. We say that E is **strongly convex** with constant $\mu > 0$ iff

$$\mu \|w - v\|_{\mathcal{V}}^2 \leq \langle \mathrm{D}E(w) - \mathrm{D}E(v), w - v \rangle, \tag{6}$$

for all $v, w \in \mathcal{V}$.

Remark (Terminology)

Some references, like Ciarlet (1989), use the term elliptic for strong convexity.

Theorem (Strong Convexity)

Let $\mathcal{V}$ be a Hilbert space and E be a strongly convex functional with constant μ. Then, for all $v, w \in \mathcal{V}$,

$$E(w) - E(v) \geq \langle \mathrm{D}E(v), w - v \rangle + \frac{\mu}{2} \|w - v\|_{\mathcal{V}}^2 . \tag{7}$$

Consequently E is strictly convex and coercive.

Proof.

Using the version of Taylor's theorem with integral remainder,

$$
\begin{aligned}
E(w) - E(v) &= \int_0^1 \langle \mathrm{D}E(v + t(w-v)), w - v \rangle \,\mathrm{d}t \\
&= \langle \mathrm{D}E(v), w - v \rangle \\
&\quad + \int_0^1 \frac{1}{t} \langle \mathrm{D}E(v + t(w-v)) - \mathrm{D}E(v), t(w-v) \rangle \,\mathrm{d}t \\
&\geq \langle \mathrm{D}E(v), w - v \rangle + \int_0^1 \frac{1}{t} \mu \left\| t(w-v) \right\|_{\mathcal{V}}^2 \,\mathrm{d}t \\
&= \langle \mathrm{D}E(v), w - v \rangle + \frac{\mu}{2} \left\| w - v \right\|_{\mathcal{V}}^2 ,
\end{aligned}
$$

which proves (7).

Next we see that, whenever $v \neq w$, (7) implies

$$
E(w) > E(v) + \langle \mathrm{D}E(v), w - v \rangle,
$$

which by a previous proposition shows that E is strictly convex.

Proof (Cont.)

Finally, setting $v = 0$ in (7) and using Young's inequality we get

$$\begin{aligned} E(w) &\geq E(0) + \langle \mathrm{D}E(0), w \rangle + \frac{\mu}{2}\|w\|_{\mathcal{V}}^2 \\ &\geq E(0) - \|\mathrm{D}E(0)\|_{\mathcal{V}'}\|w\|_{\mathcal{V}} + \frac{\mu}{2}\|w\|_{\mathcal{V}}^2 \\ &\geq E(0) - \frac{1}{\mu}\|\mathrm{D}E(0)\|_{\mathcal{V}'}^2 - \frac{\mu}{4}\|w\|_{\mathcal{V}}^2 + \frac{\mu}{2}\|w\|_{\mathcal{V}}^2 \\ &\geq C_1 + C_2\|w\|_{\mathcal{V}}^2 \end{aligned}$$

with

$$C_1 = E(0) - \frac{1}{\mu}\|\mathrm{D}E(0)\|_{\mathcal{V}'}^2, \qquad C_2 = \frac{\mu}{4}.$$

Since $C_2 > 0$, this clearly implies the coercivity of E. $\square$

The last notion we introduce is that of Lipschitz smoothness.

Definition (Lipschitz Smoothness)

Let $\mathcal{V}$ be a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$ be Fréchet differentiable. We say that E is **locally Lipschitz smooth** on $\mathcal{V}$ iff, for every bounded set $B \subset \mathcal{V}$, there is a constant $L = L_B$ such that

$$\|\mathrm{D}E(v) - \mathrm{D}E(w)\|_{\mathcal{V}'} \leq L\|v - w\|_{\mathcal{V}}, \qquad \forall\, v, w \in B.$$

Some useful properties that combine strong convexity and local Lipschitz smoothness are now presented.

Lemma (Two-Sided Bounds)

Let $\mathcal{V}$ be a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$. If E is strongly convex and locally Lipschitz smooth, then we have, for every $v, w \in \mathcal{V}$,

$$\mu \|w - v\|_{\mathcal{V}}^2 \leq \langle \mathrm{D}E(w) - \mathrm{D}E(v), w - v \rangle.$$

Let now $\alpha \in \mathbb{R}$. Then, for every $\alpha \in \mathbb{R}$ and all $v, w \in \mathrm{level}_\alpha(E)$,

$$\mu \|w - v\|_{\mathcal{V}}^2 \leq \langle \mathrm{D}E(w) - \mathrm{D}E(v), w - v \rangle \leq L \|w - v\|_{\mathcal{V}}^2 ,$$

where $L = L(E, \alpha)$ is the local Lipschitz constant of $\mathrm{D}E$ on the set $\mathrm{level}_\alpha(E)$.

Proof.

The lower bound is just strong convexity. To get the upper bound, observe that the (local) Lipschitz smoothness of E implies that, for every $w, v \in \mathrm{level}_\alpha(E)$ and all $z \in \mathcal{V}$,

$$\begin{aligned}|\langle \mathrm{D}E(w) - \mathrm{D}E(v), z\rangle| &\leq \|\mathrm{D}E(w) - \mathrm{D}E(v)\|_{\mathcal{V}'} \|z\|_{\mathcal{V}} \\ &\leq L \|w - v\|_{\mathcal{V}} \|z\|_{\mathcal{V}} .\end{aligned}$$

Setting $z = w - v$ gives the desired estimate. □