Introduction to Convex Interpolation

Pranav Reddy

University of California, San Diego

October 4, 2023

Overview

Introduction
Preliminaries
Defining the PEP

Convex Interpolation
Motivation
Convex Conjugation

Preliminaries

If we have a convex function f and we know that f is (sub)differentiable, how can we solve the problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

Preliminaries

If we have a convex function f and we know that f is (sub)differentiable, how can we solve the problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

Some options:

1. Nelder-Mead, golden section or other algorithms not requiring differentiability
2. First-order algorithms (GD, AGD, SGD, etc)
3. Second-order algorithms (Newton)

Preliminaries

If we have a convex function f and we know that f is (sub)differentiable, how can we solve the problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

Some options:

1. Nelder-Mead, golden section or other algorithms not requiring differentiability
2. First-order algorithms (GD, AGD, SGD, etc)
3. Second-order algorithms (Newton)

We focus on option 2, and ask how to analyze different variants?

Measuring Performance

How do we define a performance measure? Some options:

- Gradient norm: $\|\nabla f(x_n)\|^2$
- Objective gap: $f(x_n) - f(x_\star)$
- Distance from minimizer: $\|x_n - x_\star\|^2$

Measuring Performance

How do we define a performance measure? Some options:

- Gradient norm: $\|\nabla f(x_n)\|^2$
- Objective gap: $f(x_n) - f(x_\star)$
- Distance from minimizer: $\|x_n - x_\star\|^2$

Classic results for first-order methods typically involve showing that one of these measures decay as a function of

- Number of iterations: N
- Starting conditions of algorithm (e.g. $\|x_0 - x_\star\|^2$)
- Properties of the function class (e.g. L-smoothness, μ-strong convexity, etc.)

Measuring Performance

We can generalize this by looking at the worst-case performance of a given first-order method, $\mathcal{M}$. That is, given

- Number of iterations: N
- Initial conditions: $\mathcal{C}$
- Class of functions: $\mathcal{F}$
- Performance measure: $\mathcal{E}$

Measuring Performance

We can generalize this by looking at the worst-case performance of a given first-order method, $\mathcal{M}$. That is, given

- Number of iterations: N
- Initial conditions: $\mathcal{C}$
- Class of functions: $\mathcal{F}$
- Performance measure: $\mathcal{E}$

What is the worst possible value of $\mathcal{E}$? Equivalently, what is the worst possible performance of $\mathcal{M}$ over $\mathcal{F}$, given initial conditions $\mathcal{C}$ and N iterations, as measured using $\mathcal{E}$?

Performance Estimation Problem

We can turn this into an optimization problem, called the **Performance Estimation Problem** by Drori and Teboulle [1]:

$$\max_{f \in \mathcal{F}} \ \mathcal{E}\left(\{x_i, f_i, g_i\}_{i=1,\dots,N,\star}\right)$$

$$\text{such that } f \in \mathcal{F}$$

$$x_\star \text{ is optimal}$$

$$\{x_i, f_i, g_i\}_{i=1,\dots,N,\star} \text{ are generated by } \mathcal{M}$$

Performance Estimation Problem

We can turn this into an optimization problem, called the **Performance Estimation Problem** by Drori and Teboulle [1]:

$$\max_{f \in \mathcal{F}} \; \mathcal{E}\left(\{x_i, f_i, g_i\}_{i=1,\ldots,N,\star}\right)$$

$$\text{such that } f \in \mathcal{F}$$

$$x_\star \text{ is optimal}$$

$$\{x_i, f_i, g_i\}_{i=1,\ldots,N,\star} \text{ are generated by } \mathcal{M}$$

The problem is that this is an infinite-dimensional problem due to the decision variable being $f \in \mathcal{F}$.

Performance Estimation Problem

We can turn this into an optimization problem, called the **Performance Estimation Problem** by Drori and Teboulle [1]:

$$\max_{f \in \mathcal{F}} \; \mathcal{E}\left(\{x_i, f_i, g_i\}_{i=1,\ldots,N,\star}\right)$$

$$\text{such that } f \in \mathcal{F}$$

$$x_\star \text{ is optimal}$$

$$\{x_i, f_i, g_i\}_{i=1,\ldots,N,\star} \text{ are generated by } \mathcal{M}$$

The problem is that this is an infinite-dimensional problem due to the decision variable being $f \in \mathcal{F}$.
Is there a finite representation of a given function $f \in \mathcal{F}$?

Interpolation

This is known as an interpolation problem: Given a set of points $(x_i, f_i)_{i \in I}$, does there exist a function f such that $f(x_i) = f_i$ for all i?

Interpolation

This is known as an interpolation problem: Given a set of points $(x_i, f_i)_{i \in I}$, does there exist a function f such that $f(x_i) = f_i$ for all i?

- In general we typically also want f to satisfy some conditions, like smoothness, convexity, etc.

Interpolation

This is known as an interpolation problem: Given a set of points $(x_i, f_i)_{i \in I}$, does there exist a function f such that $f(x_i) = f_i$ for all i?

- In general we typically also want f to satisfy some conditions, like smoothness, convexity, etc.

For polynomials, we know that given $x_0, \ldots, x_n$ and $y_0, \ldots, y_n$, there is a unique polynomial of degree n such that $p(x_i) = y_i$. Similar interpolation results exist for:

- C^k Splines
- Trigonometric polynomials
- Rational functions
- Wavelets

Interpolation

This is known as an interpolation problem: Given a set of points $(x_i, f_i)_{i\in I}$, does there exist a function f such that $f(x_i) = f_i$ for all i?

- In general we typically also want f to satisfy some conditions, like smoothness, convexity, etc.

For polynomials, we know that given $x_0, \ldots, x_n$ and $y_0, \ldots, y_n$, there is a unique polynomial of degree n such that $p(x_i) = y_i$. Similar interpolation results exist for:

- C^k Splines
- Trigonometric polynomials
- Rational functions
- Wavelets

If such a result exists for convex functions, we could transform the PEP to a finite dimensional problem.

Nonsmooth Convex Interpolation

We will build the tools needed to prove such a result.
The result in the case of general convex functions is well-known [2], if we give ourselves additional first-order information.

Theorem

Given a set of points $(x_i, f_i, g_i)_{i \in I}$, there exists a convex function f such that $f(x_i) = f_i$ and $g_i \in \partial f(x_i)$ if and only if

$$f_i \geq f_j + \langle g_i, x_i - x_j \rangle$$

for all $i, j \in I$.

Nonsmooth Convex Interpolation

Proof.

Consider the function $f(x) = \max_{i \in I}\{f_i + \langle g_i, x - x_i \rangle\}$. Then,

$$\begin{aligned} f(x_j) &= \max_{i \in I}\{f_i + \langle g_i, x_j - x_i \rangle\} \\ &\geq f_j + \langle g_j, x_j - x_j \rangle \\ &= f_j. \end{aligned}$$

Additionally, by hypothesis,

$$f_j \geq f_i + \langle g_i, x_j - x_i \rangle \geq f(x_j).$$

This also implies that g_j is a subgradient at each x_j, since

$$f(x) = \max_{i \in I}\{f_i + \langle g_i, x - x_i \rangle\} \geq f(x_i) + \langle g_i, x - x_i \rangle.$$

$\square$

Smooth Convex Interpolation

Th result from before came from a simple discretization of one of the definitions of convexity. Can we do the same for smooth or strongly convex functions?

Simple Discretization is not Sufficient

We have the following two equations characterizing L-smooth convex functions

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \quad \forall x, y \in \mathbb{R}^d,$$

$$\|\nabla f(x) - \nabla f(x)\| \leq L\|x - y\|, \quad \forall x, y \in \mathbb{R}^d.$$

Simple Discretization is not Sufficient

We have the following two equations characterizing L-smooth convex functions

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \quad \forall x, y \in \mathbb{R}^d,$$

$$\|\nabla f(x) - \nabla f(x)\| \leq L\|x - y\|, \quad \forall x, y \in \mathbb{R}^d.$$

However, the discretization

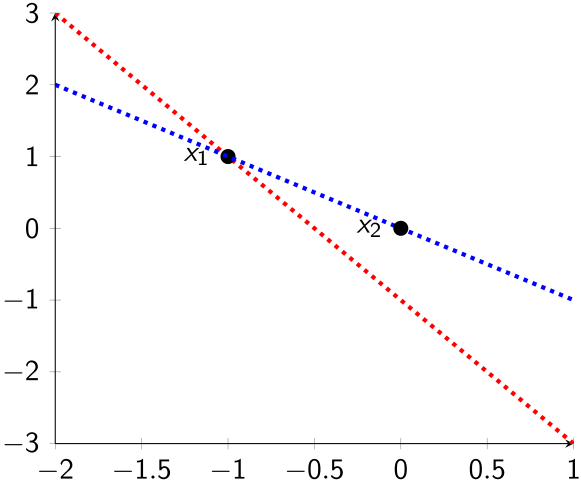
$$f(x_j) \geq f(x_i) + \langle \nabla f(x_i), x_j - x_i \rangle, \quad \forall i, j \in I$$

$$\|\nabla f(x_i) - \nabla f(x_j)\| \leq L\|x_i - x_j\|, \quad \forall i, j \in I$$

is not sufficient to guarantee interpolability.

Counterexample to Discretization

Consider $(x_1, f_1, g_1) = (-1, 1, -2)$ and $(x_2, f_2, g_2) = (0, 0, -1)$.



Counterexample to Discretization

Proof.

Since f is convex and L-smooth, we know that its derivative is nondecreasing and satisfies the intermediate value theorem.

We can integrate like so:

$$
\begin{aligned}
f(-1) &= f(0) + \int_0^{-1} f'(x)dx \\
&= \int_0^{-1} f'(x)dx \\
&> \int_0^{-1} (-1)dx \\
&> 1
\end{aligned}
$$

Counterexample to Discretization

Proof.

Since f is convex and L-smooth, we know that its derivative is nondecreasing and satisfies the intermediate value theorem. We can integrate like so:

$$
\begin{aligned}
f(-1) &= f(0) + \int_0^{-1} f'(x)dx \\
&= \int_0^{-1} f'(x)dx \\
&> \int_0^{-1} (-1)dx \\
&> 1
\end{aligned}
$$

Essentially, the curvature required by the interpolation conditions would force the function to lie strictly above its linear underapproximation, but this produces a singularity at $x = -1$. Therefore, such a function cannot exist. □

Smooth Strongly Convex Interpolation

We would like to find a characterization of L-smooth functions that can be discretized in a satisfactory way. The tool we can use to investigate this is **convex conjugation**.

Smooth Strongly Convex Interpolation

We would like to find a characterization of L-smooth functions that can be discretized in a satisfactory way. The tool we can use to investigate this is **convex conjugation**.

Definition

Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ be a function. The **Legendre-Fenchel conjugate** of f is defined as

$$f^\star(x^\star) := \sup_{x \in \mathbb{R}} \{\langle x^\star, x \rangle - f(x)\}.$$

An interpretation of this is that conjugation represents the largest global linear underestimators of f.

Convex Conjugation Example

Consider

$$f(x) = e^x$$

Then,

$$f^\star(x^\star) = \sup_x \{\langle x^\star, x\rangle - f(x)\} = \sup_x \{x^\star x - e^x\}$$

Convex Conjugation Example

Consider

$$f(x) = e^x$$

Then,

$$f^\star(x^\star) = \sup_x \{\langle x^\star, x\rangle - f(x)\} = \sup_x \{x^\star x - e^x\}$$

Convex Conjugation Example

Suppose $x^\star < 0$. Then, the quantity $x^\star x - e^x$ is unbounded, so the conjugate cannot exist.

Convex Conjugation Example

Suppose $x^\star < 0$. Then, the quantity $x^\star x - e^x$ is unbounded, so the conjugate cannot exist.

Suppose $x^\star = 0$. Then, $x^\star x - e^x = -e^x$, which is bounded above by 0, so $f^\star(0) = 0$.

Convex Conjugation Example

Suppose $x^\star < 0$. Then, the quantity $x^\star x - e^x$ is unbounded, so the conjugate cannot exist.

Suppose $x^\star = 0$. Then, $x^\star x - e^x = -e^x$, which is bounded above by 0, so $f^\star(0) = 0$.

Suppose $x^\star > 0$. Then, $x^\star x - e^x$ is differentiable in x and goes to $-\infty$ in both directions, so it must have a maximum somewhere.

Convex Conjugation Example

Suppose $x^\star < 0$. Then, the quantity $x^\star x - e^x$ is unbounded, so the conjugate cannot exist.

Suppose $x^\star = 0$. Then, $x^\star x - e^x = -e^x$, which is bounded above by 0, so $f^\star(0) = 0$.

Suppose $x^\star > 0$. Then, $x^\star x - e^x$ is differentiable in x and goes to $-\infty$ in both directions, so it must have a maximum somewhere.

If we take the derivative with respect to x, we get

$$\frac{\partial}{\partial x}(x^\star x - e^x) = x^\star - e^x$$

Convex Conjugation Example

Suppose $x^\star < 0$. Then, the quantity $x^\star x - e^x$ is unbounded, so the conjugate cannot exist.

Suppose $x^\star = 0$. Then, $x^\star x - e^x = -e^x$, which is bounded above by 0, so $f^\star(0) = 0$.

Suppose $x^\star > 0$. Then, $x^\star x - e^x$ is differentiable in x and goes to $-\infty$ in both directions, so it must have a maximum somewhere. If we take the derivative with respect to x, we get

$$\frac{\partial}{\partial x}(x^\star x - e^x) = x^\star - e^x$$

Solving this for the critical point gives

$$x = \log(x^\star),$$

so therefore

$$f^\star(x^\star) = x^\star \log(x^\star) - x^\star$$

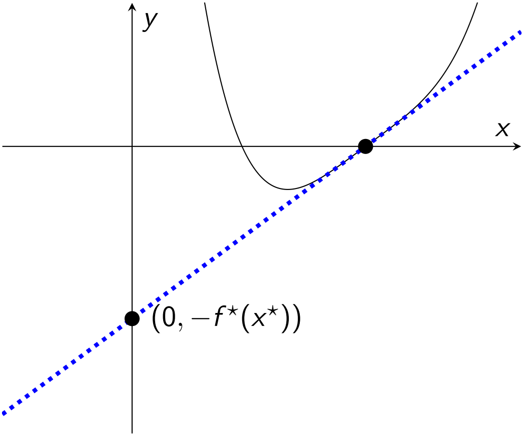
Interpretation

Suppose we say $f(x)$ is the cost to produce quantity x of a product, and suppose $x^\star$ is the price per unit. Then, the optimal profit we can get is

$$\sup_x \{\langle x, x^\star \rangle - f(x)\} = f^\star(x^\star).$$

If f is convex and differentiable, then the optimal point is given by $x^\star - f'(x) = 0$. The tangent from that point will intersect the vertical axis at $-(\langle x, x^\star \rangle - f(x))$.

Interpretation



Properties of Convex Conjugates

1. $f^\star$ is convex and closed, even when f is not.

Properties of Convex Conjugates

1. $f^\star$ is convex and closed, even when f is not.
2. Fenchel's Inequality: $f(x) + f^\star(x^\star) \geq \langle x, x^\star \rangle$, and equality holds if and only if $x^\star \in \partial f(x)$.

Properties of Convex Conjugates

1. $f^\star$ is convex and closed, even when f is not.
2. Fenchel's Inequality: $f(x) + f^\star(x^\star) \geq \langle x, x^\star \rangle$, and equality holds if and only if $x^\star \in \partial f(x)$.
3. $f^{\star\star} \leq f$, and if f is a proper closed convex function, then $f^{\star\star} = f$.

Properties of Convex Conjugates

1. $f^\star$ is convex and closed, even when f is not.
2. Fenchel's Inequality: $f(x) + f^\star(x^\star) \geq \langle x, x^\star \rangle$, and equality holds if and only if $x^\star \in \partial f(x)$.
3. $f^{\star\star} \leq f$, and if f is a proper closed convex function, then $f^{\star\star} = f$.
4. For a proper closed convex function f, $x^\star \in \partial f(x)$ if and only if $x \in \partial f^\star(x^\star)$.

Properties of Convex Conjugates

1. $f^\star$ is convex and closed, even when f is not.
2. Fenchel's Inequality: $f(x) + f^\star(x^\star) \geq \langle x, x^\star \rangle$, and equality holds if and only if $x^\star \in \partial f(x)$.
3. $f^{\star\star} \leq f$, and if f is a proper closed convex function, then $f^{\star\star} = f$.
4. For a proper closed convex function f, $x^\star \in \partial f(x)$ if and only if $x \in \partial f^\star(x^\star)$.

Property 4 is interesting: it says that convex conjugation interchanges coordinates and subgradients, which matches our economic analogy from before.

Properties of Convex Conjugates

Proof.

We will use an equivalent definition of convexity [3]. Define the epigraph of f as epi $f := \{(x, y) : f(x) \leq y\}$. Then, f is a convex function if and only if epi f is a convex set.

Properties of Convex Conjugates

Proof.

We will use an equivalent definition of convexity [3]. Define the epigraph of f as epi $f := \{(x, y) : f(x) \leq y\}$. Then, f is a convex function if and only if epi f is a convex set.

The proof of 1 follows from realizing that the mapping $x^\star \mapsto \langle x, x^\star \rangle - f(x)$ is affine, and therefore convex, so its epigraph is convex, and additionally it is also closed.

Properties of Convex Conjugates

Proof.

We will use an equivalent definition of convexity [3]. Define the epigraph of f as epi $f := \{(x, y) : f(x) \leq y\}$. Then, f is a convex function if and only if epi f is a convex set.

The proof of 1 follows from realizing that the mapping $x^\star \mapsto \langle x, x^\star \rangle - f(x)$ is affine, and therefore convex, so its epigraph is convex, and additionally it is also closed.

Taking the intersection of all epigraphs over all values of x means we take the intersection of convex and closed sets, which is also convex and closed, and this is precisely the epigraph of $f^\star$. □

Properties of Convex Conjugates

Proof.

The proof of 2 follows from the definition:

$$
\begin{aligned}
f^\star(x^\star) &= \sup_x \{\langle x^\star, x\rangle - f(x)\} \\
&\geq \langle x^\star, x\rangle - f(x) \\
&\iff \\
f(x) + f^\star(x^\star) &\geq \langle x^\star, x\rangle
\end{aligned}
$$

□

Properties of Convex Conjugates

For the case of equality, we have that

$$x^\star \in \partial f(x) \iff f(z) \geq f(x) + \langle x^\star, z - x \rangle \qquad \forall z$$

Properties of Convex Conjugates

For the case of equality, we have that

$$\begin{aligned} x^\star \in \partial f(x) &\iff f(z) \geq f(x) + \langle x^\star, z - x \rangle \qquad \forall z \\ &\iff f(z) - \langle x^\star, z \rangle \geq f(x) - \langle x^\star, x \rangle \end{aligned}$$

Properties of Convex Conjugates

For the case of equality, we have that

$$\begin{aligned} x^\star \in \partial f(x) &\iff f(z) \geq f(x) + \langle x^\star, z - x\rangle \qquad \forall z \\ &\iff f(z) - \langle x^\star, z\rangle \geq f(x) - \langle x^\star, x\rangle \\ &\iff \langle x^\star, z\rangle - f(z) \leq \langle x^\star, x\rangle - f(x) \end{aligned}$$

Properties of Convex Conjugates

For the case of equality, we have that

$$
\begin{aligned}
x^\star \in \partial f(x) &\iff f(z) \geq f(x) + \langle x^\star, z - x\rangle \qquad \forall z \\
&\iff f(z) - \langle x^\star, z\rangle \geq f(x) - \langle x^\star, x\rangle \\
&\iff \langle x^\star, z\rangle - f(z) \leq \langle x^\star, x\rangle - f(x) \\
&\iff \sup_{z \in \mathrm{dom}(f)} \{\langle x^\star, z\rangle - f(z)\} \leq \langle x^\star, x\rangle - f(x)
\end{aligned}
$$

Properties of Convex Conjugates

For the case of equality, we have that

$$\begin{aligned}
x^\star \in \partial f(x) &\iff f(z) \geq f(x) + \langle x^\star, z - x\rangle \qquad \forall z \\
&\iff f(z) - \langle x^\star, z\rangle \geq f(x) - \langle x^\star, x\rangle \\
&\iff \langle x^\star, z\rangle - f(z) \leq \langle x^\star, x\rangle - f(x) \\
&\iff \sup_{z \in \mathrm{dom}(f)} \{\langle x^\star, z\rangle - f(z)\} \leq \langle x^\star, x\rangle - f(x) \\
&\iff \langle x^\star, x\rangle - f(x) = f^\star(x^\star)
\end{aligned}$$

Properties of Convex Conjugates

For the case of equality, we have that

$$
\begin{aligned}
x^\star \in \partial f(x) &\iff f(z) \geq f(x) + \langle x^\star, z - x\rangle \qquad \forall z \\
&\iff f(z) - \langle x^\star, z\rangle \geq f(x) - \langle x^\star, x\rangle \\
&\iff \langle x^\star, z\rangle - f(z) \leq \langle x^\star, x\rangle - f(x) \\
&\iff \sup_{z \in \mathrm{dom}(f)} \{\langle x^\star, z\rangle - f(z)\} \leq \langle x^\star, x\rangle - f(x) \\
&\iff \langle x^\star, x\rangle - f(x) = f^\star(x^\star) \\
&\iff \langle x^\star, x\rangle = f(x) + f^\star(x^\star).
\end{aligned}
$$

Properties of Convex Conjugates

For the proof of 3, we will first show that $f^{\star\star}(x) \leq f(x)$.

Properties of Convex Conjugates

For the proof of 3, we will first show that $f^{\star\star}(x) \leq f(x)$.

$$
\begin{aligned}
f^{**}(x) &= \sup_z \{x^\top z - f^*(z)\} \\
&= \sup_z \{x^\top z - \sup_y \{z^\top y - f(y)\}\}
\end{aligned}
$$

Properties of Convex Conjugates

For the proof of 3, we will first show that $f^{\star\star}(x) \leq f(x)$.

$$\begin{aligned} f^{**}(x) &= \sup_z \{x^\top z - f^*(z)\} \\ &= \sup_z \{x^\top z - \sup_y \{z^\top y - f(y)\}\} \\ &= \sup_z \{x^\top z + \inf_y \{f(y) - z^\top y\}\} \end{aligned}$$

Properties of Convex Conjugates

For the proof of 3, we will first show that $f^{\star\star}(x) \leq f(x)$.

$$
\begin{aligned}
f^{**}(x) &= \sup_z \{x^\top z - f^*(z)\} \\
&= \sup_z \{x^\top z - \sup_y \{z^\top y - f(y)\}\} \\
&= \sup_z \{x^\top z + \inf_y \{f(y) - z^\top y\}\} \\
&= \sup_z \inf_y \{x^\top z + f(y) - z^\top y\}
\end{aligned}
$$

Properties of Convex Conjugates

For the proof of 3, we will first show that $f^{\star\star}(x) \leq f(x)$.

$$
\begin{aligned}
f^{**}(x) &= \sup_z \{x^\top z - f^*(z)\} \\
&= \sup_z \{x^\top z - \sup_y \{z^\top y - f(y)\}\} \\
&= \sup_z \{x^\top z + \inf_y \{f(y) - z^\top y\}\} \\
&= \sup_z \inf_y \{x^\top z + f(y) - z^\top y\} \\
&\leq \inf_y \sup_z \{z^\top (x - y) + f(y)\}
\end{aligned}
$$

Properties of Convex Conjugates

For the proof of 3, we will first show that $f^{\star\star}(x) \leq f(x)$.

$$
\begin{aligned}
f^{**}(x) &= \sup_z \{x^\top z - f^*(z)\} \\
&= \sup_z \{x^\top z - \sup_y \{z^\top y - f(y)\}\} \\
&= \sup_z \{x^\top z + \inf_y \{f(y) - z^\top y\}\} \\
&= \sup_z \inf_y \{x^\top z + f(y) - z^\top y\} \\
&\leq \inf_y \sup_z \{z^\top (x - y) + f(y)\} \\
&\leq f(x)
\end{aligned}
$$

Properties of Convex Conjugates

Now suppose f is proper, closed, and convex. Then, its epigraph is nonempty, closed, and convex.

Properties of Convex Conjugates

Now suppose f is proper, closed, and convex. Then, its epigraph is nonempty, closed, and convex.

We know from the previous results that $f^{\star\star}$ is also closed and convex, and since $f^{\star\star} \leq f$, it must also be proper. Therefore, its epigraph is also nonempty, closed, and convex.

Properties of Convex Conjugates

We want to show that $f^{\star\star} \geq f$ as well. Suppose that $f^{\star\star}(x) < f(x)$ at some point x.

Properties of Convex Conjugates

We want to show that $f^{\star\star} \geq f$ as well. Suppose that $f^{\star\star}(x) < f(x)$ at some point x.
Then, by the hyperplane separation theorem, we can strictly separate epi f from $(x, f^{\star\star}(x))$, and this hyperplane cannot be vertical. This gives, for some $\varepsilon > 0$ and vector y,

$$f(z) - \varepsilon \geq \langle y, z - x \rangle + f^{\star\star}(x) \quad \forall\ z$$

However, this contradicts Fenchel's inequality, so therefore $f^{\star\star} = f$.
This also proves that for proper closed convex functions, conjugation interchanges coordinates and gradients, which we can prove by taking the biconjugate.

Properties of Convex Conjugates

We want to show that $f^{\star\star} \geq f$ as well. Suppose that $f^{\star\star}(x) < f(x)$ at some point x.
Then, by the hyperplane separation theorem, we can strictly separate epi f from $(x, f^{\star\star}(x))$, and this hyperplane cannot be vertical. This gives, for some $\varepsilon > 0$ and vector y,

$$
\begin{aligned}
f(z) - \varepsilon &\geq \langle y, z - x\rangle + f^{\star\star}(x) \quad \forall\, z \\
\langle y, x\rangle - \varepsilon &\geq \langle y, z\rangle - f(z) + f^{\star\star}(x) \\
\langle y, x\rangle - \varepsilon &\geq f^{\star}(y) + f^{\star\star}(x)
\end{aligned}
$$

However, this contradicts Fenchel's inequality, so therefore $f^{\star\star} = f$.
This also proves that for proper closed convex functions, conjugation interchanges coordinates and gradients, which we can prove by taking the biconjugate.

Properties of Convex Conjugates

We want to show that $f^{\star\star} \geq f$ as well. Suppose that $f^{\star\star}(x) < f(x)$ at some point x.
Then, by the hyperplane separation theorem, we can strictly separate epi f from $(x, f^{\star\star}(x))$, and this hyperplane cannot be vertical. This gives, for some $\varepsilon > 0$ and vector y,

$$\begin{aligned} f(z) - \varepsilon &\geq \langle y, z - x \rangle + f^{\star\star}(x) \quad \forall\ z \\ \langle y, x \rangle - \varepsilon &\geq \langle y, z \rangle - f(z) + f^{\star\star}(x) \\ \langle y, x \rangle - \varepsilon &\geq f^{\star}(y) + f^{\star\star}(x) \\ \langle y, x \rangle &> f^{\star}(y) + f^{\star\star}(x) \end{aligned}$$

However, this contradicts Fenchel's inequality, so therefore $f^{\star\star} = f$.
This also proves that for proper closed convex functions, conjugation interchanges coordinates and gradients, which we can prove by taking the biconjugate.

Conjugation for Interpolation

We can now prove results that we will need for convex interpolation.

Theorem

Let f be a proper closed convex function. Then, f is L-smooth ($f \in \mathcal{F}_{0,L}$) if and only if $f^\star$ is $\frac{1}{L}$-strongly convex ($f \in \mathcal{F}_{\frac{1}{L},\infty}$).

Conjugation for Interpolation

Theorem

Let f be a proper closed convex function. Then, f is L-smooth ($f \in \mathcal{F}_{0,L}$) if and only if $f^\star$ is $\frac{1}{L}$-strongly convex ($f \in \mathcal{F}_{\frac{1}{L},\infty}$).

Proof.

We use the following characterizations of L-smooth and μ-strongly convex functions.

$$f \in \mathcal{F}_{0,L} \iff$$

$$\frac{1}{L}\|\nabla f(y) - \nabla f(x)\|^2 \leq \langle \nabla f(y) - \nabla f(x), y - x \rangle \leq L\|x - y\|^2$$

$$f \in \mathcal{F}_{\mu,\infty} \iff$$

$$\mu\|x - y\|^2 \leq \langle \nabla f(y) - \nabla f(x), y - x \rangle \leq \frac{1}{\mu}\|\nabla f(y) - \nabla f(x)\|^2$$

The result then follows from realizing that $\nabla f^*(\nabla f(x)) = x$. □

Minimal Curvature Subtraction

We need one last lemma to transform μ-strongly convex functions to general convex functions.

Lemma

Consider $f \in \mathcal{F}_{\mu,L}$ *with* $0 \leq \mu < L \leq \infty$. *Define* $\phi(x) := f(x) - \frac{\mu}{2}\|x - x_\star\|^1$. *Then,* $\phi \in \mathcal{F}_{0,L-\mu}$ *if and only if* $f \in \mathcal{F}_{\mu,L}$.

The mapping of $f \mapsto f - \frac{\mu}{2}\| \cdot \|^2$ is known as **minimal curvature subtraction**.

Minimal Curvature Subtraction

Proof.

Suppose $f \in \mathcal{F}_{\mu,L}$. Then,

$$\begin{aligned}\langle \nabla \phi(x) - \nabla \phi(y), x - y \rangle &= \langle \nabla f(x) - \nabla f(y) - \mu(x - y), x - y \rangle \\ &= \langle \nabla f(x) - \nabla f(y), x - y \rangle - \mu \|x - y\|^2\end{aligned}$$

We use the inequality $\langle \nabla f(x) - \nabla f(y), x - y \rangle \leq L\|x - y\|^2$ if and only if $f \in \mathcal{F}_{0,L}$ (assuming f is C^1). A proof of this can be found in Nesterov's lectures [4]. □

Minimal Curvature Subtraction

Proof.

Suppose $f \in \mathcal{F}_{\mu,L}$. Then,

$$\begin{aligned}\langle \nabla\phi(x) - \nabla\phi(y), x - y\rangle &= \langle \nabla f(x) - \nabla f(y) - \mu(x-y), x - y\rangle \\ &= \langle \nabla f(x) - \nabla f(y), x - y\rangle - \mu\|x-y\|^2 \\ &\leq L\|x-y\|^2 - \mu\|x-y\|^2 \\ &\leq (L-\mu)\|x-y\|^2.\end{aligned}$$

We use the inequality $\langle \nabla f(x) - \nabla f(y), x - y\rangle \leq L\|x-y\|^2$ if and only if $f \in \mathcal{F}_{0,L}$ (assuming f is C^1). A proof of this can be found in Nesterov's lectures [4]. □

Minimal Curvature Subtraction

Proof.

Suppose $\phi \in \mathcal{F}_{0,L-\mu}$. Then, $f(x) = \phi(x) + \frac{\mu}{2}\|x - x_\star\|^2$. By convexity and $(L-\mu)$-smoothness:

$$\begin{aligned}
\langle \nabla f(x) - \nabla f(y), x - y \rangle &= \langle \nabla\phi(x) - \nabla\phi(y) + \mu(x-y), x - y\rangle \\
&= \langle \nabla\phi(x) - \nabla\phi(y), x - y\rangle + \mu\|x - y\|^2
\end{aligned}$$

This proves L-smoothness.

Minimal Curvature Subtraction

Proof.

Suppose $\phi \in \mathcal{F}_{0,L-\mu}$. Then, $f(x) = \phi(x) + \frac{\mu}{2}\|x - x_\star\|^2$. By convexity and $(L-\mu)$-smoothness:

$$\begin{aligned}
\langle \nabla f(x) - \nabla f(y), x - y \rangle &= \langle \nabla \phi(x) - \nabla \phi(y) + \mu(x-y), x - y \rangle \\
&= \langle \nabla \phi(x) - \nabla \phi(y), x - y \rangle + \mu\|x - y\|^2 \\
&\leq (L-\mu)\|x-y\|^2 + \mu\|x-y\|^2 \\
&= L\|x-y\|^2.
\end{aligned}$$

This proves L-smoothness. For strong convexity, we have that

Minimal Curvature Subtraction

Proof.

Suppose $\phi \in \mathcal{F}_{0,L-\mu}$. Then, $f(x) = \phi(x) + \frac{\mu}{2}\|x - x_\star\|^2$. By convexity and $(L-\mu)$-smoothness:

$$\begin{aligned}
\langle \nabla f(x) - \nabla f(y), x - y \rangle &= \langle \nabla \phi(x) - \nabla \phi(y) + \mu(x-y), x - y \rangle \\
&= \langle \nabla \phi(x) - \nabla \phi(y), x - y \rangle + \mu\|x - y\|^2 \\
&\leq (L-\mu)\|x-y\|^2 + \mu\|x-y\|^2 \\
&= L\|x-y\|^2.
\end{aligned}$$

This proves L-smoothness. For strong convexity, we have that

$$\begin{aligned}
0 &\leq \langle \nabla \phi(x) - \nabla \phi(y), x - y \rangle \\
&= \langle \nabla f(x) - \nabla f(y), x - y \rangle - \mu\|x-y\|^2
\end{aligned}$$

Minimal Curvature Subtraction

Proof.

Suppose $\phi \in \mathcal{F}_{0,L-\mu}$. Then, $f(x) = \phi(x) + \frac{\mu}{2}\|x - x_\star\|^2$. By convexity and $(L-\mu)$-smoothness:

$$\begin{aligned}
\langle \nabla f(x) - \nabla f(y), x - y \rangle &= \langle \nabla \phi(x) - \nabla \phi(y) + \mu(x - y), x - y \rangle \\
&= \langle \nabla \phi(x) - \nabla \phi(y), x - y \rangle + \mu\|x - y\|^2 \\
&\leq (L - \mu)\|x - y\|^2 + \mu\|x - y\|^2 \\
&= L\|x - y\|^2.
\end{aligned}$$

This proves L-smoothness. For strong convexity, we have that

$$\begin{aligned}
0 &\leq \langle \nabla \phi(x) - \nabla \phi(y), x - y \rangle \\
&= \langle \nabla f(x) - \nabla f(y), x - y \rangle - \mu\|x - y\|^2 \\
\mu\|x - y\|^2 &\leq \langle \nabla f(x) - \nabla f(y), x - y \rangle.
\end{aligned}$$

This inequality is an equivalent definition of strong convexity for C^1 functions (Nesterov [4]).

Convex Interpolation Theorem

Now, we can state our result:

Theorem

A set $\{(x_i, f_i, g_i)\}_{i \in I}$ is $\mathcal{F}_{\mu,L}$-interpolable if and only if

$$f_i - f_j - \langle g_j, x_i - x_j \rangle \geq \frac{1}{2\left(1 - \frac{\mu}{L}\right)} \left(\frac{1}{L} \|g_i - g_j\|^2 + \mu \|x_i - x_j\|^2 - 2\frac{\mu}{L} \langle g_j - g_i, x_j - x_i \rangle \right)$$

for all $i, j \in I$.

Convex Interpolation Theorem

Proof.

The full proof is very long and doesn't add anything new. We focus on the main ideas.

1. Show that $\{(x_i, f_i, g_i)\}_{i \in I}$ is $\mathcal{F}_{\mu,L}$-interpolable if and only if $\{(g_i, f_i, x_i)\}_{i \in I}$ is $\mathcal{F}_{\frac{1}{L},\frac{1}{\mu}}$-interpolable.

□

Convex Interpolation Theorem

Proof.

The full proof is very long and doesn't add anything new. We focus on the main ideas.

1. Show that $\{(x_i, f_i, g_i)\}_{i \in I}$ is $\mathcal{F}_{\mu,L}$-interpolable if and only if $\{(g_i, f_i, x_i)\}_{i \in I}$ is $\mathcal{F}_{\frac{1}{L},\frac{1}{\mu}}$-interpolable.
2. Then, starting with a μ-strongly convex, L-smooth function f, we use minimal curvature subtraction to obtain an $(L-\mu)$-smooth function $\tilde{f}$.

□

Convex Interpolation Theorem

Proof.

The full proof is very long and doesn't add anything new. We focus on the main ideas.

1. Show that $\{(x_i, f_i, g_i)\}_{i \in I}$ is $\mathcal{F}_{\mu,L}$-interpolable if and only if $\{(g_i, f_i, x_i)\}_{i \in I}$ is $\mathcal{F}_{\frac{1}{L},\frac{1}{\mu}}$-interpolable.
2. Then, starting with a μ-strongly convex, L-smooth function f, we use minimal curvature subtraction to obtain an $(L-\mu)$-smooth function $\tilde{f}$.
3. Then, we take $\tilde{f}$ and map it to its conjugate $\tilde{f}^\star$ to obtain a $\frac{1}{L-\mu}$-strongly convex function.

□

Convex Interpolation Theorem

Proof.

The full proof is very long and doesn't add anything new. We focus on the main ideas.

1. Show that $\{(x_i, f_i, g_i)\}_{i \in I}$ is $\mathcal{F}_{\mu,L}$-interpolable if and only if $\{(g_i, f_i, x_i)\}_{i \in I}$ is $\mathcal{F}_{\frac{1}{L},\frac{1}{\mu}}$-interpolable.
2. Then, starting with a μ-strongly convex, L-smooth function f, we use minimal curvature subtraction to obtain an $(L-\mu)$-smooth function $\tilde{f}$.
3. Then, we take $\tilde{f}$ and map it to its conjugate $\tilde{f}^\star$ to obtain a $\frac{1}{L-\mu}$-strongly convex function.
4. We again use minimal curvature subtraction to obtain a convex $h^\star$, which we can use our nonsmooth convex interpolation result for.

□

Remarks

If we take $\mu = 0$, then we get the following:

$$f_i - f_j - \langle g_j, x_i - x_j \rangle \geq \frac{1}{2L} \|g_i - g_j\|^2.$$

Equivalently,

$$f_i \geq f_j + \langle g_j, x_i - x_j \rangle + \frac{1}{2L} \|g_i - g_j\|^2.$$

Remarks

If we take $\mu = 0$, then we get the following:

$$f_i - f_j - \langle g_j, x_i - x_j \rangle \geq \frac{1}{2L} \|g_i - g_j\|^2.$$

Equivalently,

$$f_i \geq f_j + \langle g_j, x_i - x_j \rangle + \frac{1}{2L} \|g_i - g_j\|^2.$$

Surprisingly, this is a discretization of the following characterization of L-smooth convex functions:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L} \|\nabla f(x) - \nabla f(y)\|^2.$$

Remarks

Similarly, if we take $L = \infty$, then we get the following:

$$f_i - f_j - \langle g_j, x_i - x_j \rangle \geq \frac{\mu}{2} \|x_i - x_j\|^2.$$

Equivalently,

$$f_i \geq f_j + \langle g_j, x_i - x_j \rangle + \frac{\mu}{2} \|x_i - x_j\|^2.$$

Remarks

Similarly, if we take $L = \infty$, then we get the following:

$$f_i - f_j - \langle g_j, x_i - x_j \rangle \geq \frac{\mu}{2} \|x_i - x_j\|^2.$$

Equivalently,

$$f_i \geq f_j + \langle g_j, x_i - x_j \rangle + \frac{\mu}{2} \|x_i - x_j\|^2.$$

This is also a discretization of the following characterization of μ-strongly convex functions:

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

Remarks

Similarly, if we take $L = \infty$, then we get the following:

$$f_i - f_j - \langle g_j, x_i - x_j \rangle \geq \frac{\mu}{2} \|x_i - x_j\|^2.$$

Equivalently,

$$f_i \geq f_j + \langle g_j, x_i - x_j \rangle + \frac{\mu}{2} \|x_i - x_j\|^2.$$

This is also a discretization of the following characterization of μ-strongly convex functions:

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

Why are some characterizations appropriate for discretization but others are not? In the continuous case they are equivalent, but in the process of discretization information is somehow lost.

Next Time

- Apply convex interpolation to discretize PEP
- Can we find interpolation results for broader classes of functions?
- How can we discretize constrained optimization?
- Can the PEP produce better asmyptotics, not just differences in constants?

References I

[1] Yoel Drori and Marc Teboulle. Performance of first-order methods for smooth convex minimization: a novel approach, 2012.

[2] Adrien B. Taylor, Julien M. Hendrickx, and François Glineur. Smooth strongly convex interpolation and exact worst-case performance of first-order methods, 2016.

[3] Ralph Tyrell Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.

[4] Yurii Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Springer Publishing Company, Incorporated, 1 edition, 2014.

[5] Benjamin Grimmer. Provably faster gradient descent via long steps, 2023.

References II

[6] Shuvomoy Das Gupta, Bart P. G. Van Parys, and Ernest K. Ryu. Branch-and-bound performance estimation programming: A unified methodology for constructing optimal optimization methods, 2023.

[7] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.