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 2 of 2
Convex Optimization

Minimization of Convex Functionals

Minimization

We now bring all the previously introduced notions together and study optimization problems. In other words, if $\mathcal{V}$ is a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$ we wish to find $u \in \mathcal{V}$ such that

$$u \in \operatorname*{argmin}_{v \in \mathcal{V}} E(v).$$

If this problem has a solution, we call such a $u \in \mathcal{V}$ a *minimizer* of E. Our goal here is to provide sufficient conditions for the existence and uniqueness of a minimizer, as well as its characterization. Again, we only present a basic introduction.

The following two results are special cases of what is commonly referred to as the *direct method of calculus of variations*. They guarantee the existence of minimizers.

Theorem (Existence I)

Let $\mathcal{V}$ be a Hilbert space, $K \subset \mathcal{V}$ be bounded and weakly closed, and $E : \mathcal{V} \to \mathbb{R}$ be wLSC. Then, the set

$$\operatorname*{argmin}_{v \in K} E(v)$$

is not empty.

Theorem (Existence II)

Let $\mathcal{V}$ be a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$ be convex, LSC, and coercive. Then, the set

$$\operatorname*{argmin}_{v \in \mathcal{V}} E(v)$$

is not empty.

Proof.

Define

$$\mathcal{B}_0 := \mathrm{level}_{E(0)}(E).$$

Owing to previous propositions, this set is convex and also it is closed. Thus, by Mazur's Lemma, it is weakly closed. Moreover, owing to the coercivity of E, this set is bounded. Thus, we may invoke the last theorem to guarantee the existence of an element of

$$\operatorname*{argmin}_{v \in \mathcal{B}_0} E(v).$$

Proof (Cont.)

On the other hand, if $w \notin \mathcal{B}_0$ we have

$$E(w) \geq E(0) \geq \inf_{v \in \mathcal{B}_0} E(v).$$

Thus,

$$\inf_{v \in \mathcal{B}_0} E(v) \leq \inf_{v \in \mathcal{V}} E(v),$$

and the minimizer over $\mathcal{B}_0$ minimizes E globally, i.e., over $\mathcal{V}$. $\square$

We now turn our attention to the question of uniqueness and characterization of minimizers.

Corollary (Existence and Uniqueness)

Let $\mathcal{V}$ be a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$ be convex, LSC, and coercive. Then, the set

$$\operatorname*{argmin}_{v \in \mathcal{V}} E(v)$$

is not empty. In addition, if E is strictly convex, this set is a singleton, i.e., there is a unique $u \in \mathcal{V}$ such that

$$E(u) = \min_{v \in \mathcal{V}} E(v).$$

Proof.

Since E is convex and LSC its epigraph is convex and closed. This implies, by Mazur's Lemma, that its epigraph is weakly closed. We, again, conclude that E is wLSC. Since E is coercive, the existence of a minimizer is guaranteed.

We now turn our attention to uniqueness. Assuming that $u_1, u_2 \in \mathcal{V}$, with $u_1 \neq u_2$, are minimizers we realize that, by strict convexity,

$$\begin{aligned} E\left(\frac{u_1 + u_2}{2}\right) &< \frac{1}{2}\left(E(u_1) + E(u_2)\right) \\ &= \min_{v \in \mathcal{V}} E(v), \end{aligned}$$

thus contradicting that u_1 and u_2 are minimizers. This shows uniqueness. $\square$

Now that we can guarantee the existence and uniqueness of a minimizer, we aim to characterize it.

Theorem (Euler-Lagrange Equation)

Let $\mathcal{V}$ be a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$ be Fréchet differentiable and strongly convex. Then the functional E has a unique minimizer $u \in \mathcal{V}$. Moreover, this minimizer is uniquely characterized by the so-called Euler-Lagrange equation

$$\langle \mathrm{D}E(u), v \rangle = 0, \qquad \forall\, v \in \mathcal{V}. \tag{1}$$

Proof.

Obviously, E being Fréchet differentiable implies that E is continuous which, along with convexity, implies that it is wLSC. Next, we invoke a theorem to see that E is coercive. This guarantees the existence and uniqueness of a minimizer.

Proof (Cont.)

We now prove that the Euler-Lagrange equation characterizes the minimizer. First, if $u \in \mathcal{V}$ solves the Euler-Lagrange equation then, by strong convexity,

$$\begin{aligned} E(v) &\geq E(u) + \langle \mathrm{D}E(u), v - u \rangle + \frac{\mu}{2}\|v - u\|_{\mathcal{V}}^2 \\ &> E(u) \end{aligned}$$

whenever $v \neq u$. This shows that u is the unique minimizer of E. On the other hand, if $u \in \mathcal{V}$ is the (unique) minimizer of E, then for $t \in \mathbb{R}_\star$ and $v \in \mathcal{V} \setminus \{0\}$

$$E(u + tv) - E(u) > 0.$$

Proof (Cont.)

Since E is Fréchet differentiable,

$$0 \leq \lim_{t \downarrow 0} \frac{1}{t} \left(E(u + tv) - E(u) \right) = \langle \mathrm{D}E(u), v \rangle$$

and

$$0 \geq \lim_{t \uparrow 0} \frac{1}{t} \left(E(u + tv) - E(u) \right) = \langle \mathrm{D}E(u), v \rangle,$$

so that $\langle \mathrm{D}E(u), v \rangle = 0$, for all $v \in \mathcal{V}$. This shows that the Euler-Lagrange equation holds. $\square$

Having guaranteed the existence and uniqueness of a minimizer, as well as its characterization via the Euler-Lagrange equations, we provide some additional facts about this minimizer.

The first result provides a relation between the energy and the norm centered at the minimizer. The following estimates can be easily proved using Taylor's theorem with integral remainder, as above. See reference in the notes for further details.

Lemma (Quadratic Trap)

Let $\mathcal{V}$ be a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$ be strongly convex and locally Lipschitz smooth. Then, for all $v, w \in \mathcal{V}$

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

In addition, for every $\alpha \in \mathbb{R}$, we have that if $v, w \in \mathrm{level}_\alpha(E)$,

$$\begin{aligned} \frac{\mu}{2} \|w - v\|_{\mathcal{V}}^2 + \langle \mathrm{D}E(v), w - v \rangle &\leq E(w) - E(v) \\ &\leq \langle \mathrm{D}E(v), w - v \rangle + \frac{L}{2} \|w - v\|_{\mathcal{V}}^2 , \end{aligned} \tag{2}$$

where $L = L(\alpha)$ is the local Lipschitz constant of $\mathrm{D}E$. Finally, if $u \in \mathcal{V}$ is the minimizer of E, then for every $\alpha \in \mathbb{R}$, and all $w \in \mathrm{level}_\alpha(E)$,

$$\frac{\mu}{2} \|w - u\|_{\mathcal{V}}^2 \leq E(w) - E(u) \leq \frac{L}{2} \|w - u\|_{\mathcal{V}}^2 . \tag{3}$$

Again, the lower bound holds for all $w \in \mathcal{V}$.

Proof.

The lower bound for (2) has already been established in a previous proposition. For the upper bound, we again apply Taylor's theorem with integral remainder and use the fact that $\mathrm{level}_\alpha(E)$ is convex. The details are left as an exercise. $\square$

Example

The function $E : \mathbb{R} \to \mathbb{R}$, defined by

$$E(v) = \frac{1}{4} v^4, \quad \forall\, v \in \mathbb{R},$$

is strictly convex and coercive, but not strongly convex. It is possible to find quadratic functions that bound E from above and below on any bounded interval of $\mathbb{R}$, but not as tightly as suggested in (3).

On the other hand, suppose that $\lambda \in \mathbb{R}$ and the function $E : \mathbb{R} \to \mathbb{R}$ is defined by

$$E(v) = \frac{1}{4} v^4 + \frac{\lambda}{2} v^2, \quad \forall\, v \in \mathbb{R}.$$

Then, for any $\lambda > 0$, E is both strictly convex and strongly convex. In fact, we can prove that $\mu = \lambda$ can be taken as the "best" strong convexity constant. It is not difficult to show that the global minimizer of E is $u = 0$, and, for all $w \in (-r, r)$,

$$\frac{\lambda}{2}(w - 0)^2 \leq E(w) - E(0) \leq \frac{\frac{r^2}{2} + \lambda}{2}(w - 0)^2.$$

We can further characterize the minimizer as follows.

Lemma (Norm of Derivative)

In the setting of the last lemma, we have

$$0 \leq E(v) - E(u) \leq \frac{1}{2\mu} \|\mathrm{D}E(v)\|_{\mathcal{V}'}^2, \qquad \forall\, v \in \mathcal{V}. \tag{4}$$

Proof.

Fix the point $v \in \mathcal{V}$. Now, for any $w \in \mathcal{V}$, using the lower bound of (2), we have

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

Define then

$$g(w) := E(v) + \langle \mathrm{D}E(v), w - v \rangle + \frac{\mu}{2} \|w - v\|_{\mathcal{V}}^2$$

and observe that the minimizer of g on $\mathcal{V}$ is

$$w_\star := v - \frac{1}{\mu} \mathfrak{R} \mathrm{D}E(v).$$

Proof (Cont.)

Therefore,

$$\begin{aligned} E(w) &\geq g(w) \\ &\geq g(w_\star) \\ &= E(v) - \frac{1}{2\mu}\|\mathfrak{R}\mathrm{D}E(v)\|_{\mathcal{V}}^2 \\ &= E(v) - \frac{1}{2\mu}\|\mathrm{D}E(v)\|_{\mathcal{V}'}^2. \end{aligned}$$

Then (4) is obtained by letting $w = u$ in the above inequality. $\square$

Lemma (Convexity of Sections)

Let $\mathcal{V}$ be a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$ be strongly convex and locally Lipschitz smooth. Fix $\alpha \in \mathbb{R}$ and set $\mathcal{B} = \mathrm{level}_\alpha(E)$. Let $\xi \in \mathcal{B}$ be arbitrary, and $\mathcal{W} \subseteq \mathcal{V}$ be a subspace. Define the section

$$J_\xi(w) := E(\xi + w), \quad \forall\, w \in \mathcal{W}.$$

Then, $J_\xi : \mathcal{W} \to \mathbb{R}$ is differentiable, strongly convex, and locally Lipschitz smoother. Furthermore, there exists a unique element $\eta \in \mathcal{W}$ such that $\xi + \eta \in \mathcal{B}$, η is the unique global minimizer of J_ξ, and

$$\langle \mathrm{D}E(\xi + \eta), w \rangle = \langle \mathrm{D}J_\xi(\eta), w \rangle = 0, \quad \forall\, w \in \mathcal{W}.$$

As a consequence, for all $w \in \mathcal{W}$ with $w + \xi \in \mathcal{B}$,

$$\frac{\mu}{2} \|w - \eta\|_{\mathcal{V}}^2 \le J(w) - J(\eta) = E(\xi + w) - E(\xi + \eta) \le \frac{L}{2} \|w - \eta\|_{\mathcal{V}}^2 .$$

The lower bound holds for any $w \in \mathcal{W}$, without restriction.

Proof.

Exercise. □

Remark (Condition Number)

For a strongly convex and a locally Lipschitz smooth functional $E : \mathcal{V} \to \mathbb{R}$, the ratio $\frac{L}{\mu}$ is sometimes referred to as the condition number of the derivative $\mathrm{D}E$; see Nesterov (2013). The rate of convergence of iterative methods for solving

$$u = \operatorname*{argmin}_{v \in \mathcal{V}} E(v)$$

usually depends on this condition number.

The Gradient Descent Method

Newton's Method

As a final consideration, we present two methods to approximate minimizers. Thus, we let $\mathcal{V}$ be a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$ be a strongly convex and locally Lipschitz smooth functional. We seek then for its (unique) minimizer $u \in \mathcal{V}$.

Since we assume that E is strongly convex, then (1) holds. Thus, provided E is twice Fréchet differentiable, we could propose a version of **Newton's method** to solve (1) and thus find the minimizer. Given $u^0 \in \mathcal{V}$, this method finds $\{u^k\}_{k=0}^{\infty} \subset \mathcal{V}$ as follows. For $k \geq 0$ find $\eta^k \in \mathcal{V}$ that solves the linear problem

$$\mathrm{D}^2 E(u^k)\eta^k = -\mathrm{D}E(u^k).$$

Notice that this equation is understood in $\mathcal{V}'$, so we can write equivalently,

$$\mathrm{D}^2 E(u^k)(\eta^k, v) = -\langle \mathrm{D}E(u^k), v\rangle, \qquad \forall\, v \in \mathcal{V}.$$

If this solution exists, then

$$u^{k+1} := u^k + \eta^k.$$

Newton's Method

Many questions, like existence and uniqueness of η^k, immediately arise. Even if this is the case, the convergence of such approach must be justified. We refer the reader to Salgado and Wise (2023) for a complete treatment of Newton's method.

We note, however, that in practice the functionals we wish to minimize are often not twice Fréchet differentiable; and for those that are, it may be difficult to compute their second derivative. Thus, we abandon this approach. Instead, we will focus on *first order methods*, that is, ones that only use first derivatives.

Definition (Gradient Descent)

Let $\mathcal{V}$ be a Hilbert space and $E : \mathcal{V} \to \mathbb{R}$ be strongly convex and locally Lipschitz smooth. The **gradient descent** method is a scheme to approximate $u \in \mathcal{V}$, the unique minimizer of E as follows. Fix a step size $\tau > 0$. Starting from arbitrary $u^0 \in \mathcal{V}$ we compute $\{u^k\}_{k=0}^{\infty} \subset \mathcal{V}$ recursively via

$$u^{k+1} = u^k - \tau \mathfrak{R} \mathrm{D} E(u^k), \tag{5}$$

for $k \geq 0$.

Let us show that, if the step size τ is sufficiently small, the method converges at least linearly.

Theorem (Convergence)

Let $\mathcal{V}$ be a Hilbert space, $E : \mathcal{V} \to \mathbb{R}$ be strongly convex and locally Lipschitz smooth, and $u \in \mathcal{V}$ be the minimizer of E. Let $u^0 \in \mathcal{V}$ be arbitrary. Define

$$\mathcal{B} = \left\{ v \in \mathcal{V} \,\middle|\, \|v - u\|_{\mathcal{V}} \le \|u^0 - u\|_{\mathcal{V}} \right\},$$

and let $L = L_{\mathcal{B}}$ be the local Lipschitz smoothness constant of E on $\mathcal{B}$. Assume that the step size satisfies

$$\tau \in \left(0, \frac{2\mu}{L^2}\right).$$

Then, the sequence $\{u^k\}_{k=0}^{\infty}$ generated by the gradient descent method of satisfies

$$\{u^k\}_{k=0}^{\infty} \subset \mathcal{B}.$$

Moreover,

$$\|u - u^{k+1}\|_{\mathcal{V}} \le \rho \|u - u^k\|_{\mathcal{V}}, \qquad \rho := \sqrt{1 - 2\mu\tau + \tau^2 L^2} \in (0, 1).$$

Consequently, the sequence $\{u^k\}_{k=0}^{\infty}$ converges at least linearly to u.

Proof.

To shorten the presentation, we introduce the error $e^k := u - u^k$. Using the Euler equation (1), we may rewrite (5) as

$$e^{k+1} = e^k - \tau\left(\mathfrak{R}\mathrm{D}E(u) - \mathfrak{R}\mathrm{D}E(u^k)\right).$$

Using the definition of the Riesz map, $\mathfrak{R}$, it follows that

$$\begin{aligned}\|e^{k+1}\|_{\mathcal{V}}^2 &= \left\|e^k - \tau\left(\mathfrak{R}\mathrm{D}E(u) - \mathfrak{R}\mathrm{D}E(u^k)\right)\right\|_{\mathcal{V}}^2 \\ &= \|e^k\|_{\mathcal{V}}^2 - 2\tau\langle \mathrm{D}E(u) - \mathrm{D}E(u^k), e^k\rangle + \tau^2\|\mathrm{D}E(u) - \mathrm{D}E(u^k)\|_{\mathcal{V}'}^2.\end{aligned}$$

By strong convexity

$$\begin{aligned}2\tau\langle \mathrm{D}E(u) - \mathrm{D}E(u^k), e^k\rangle &= 2\tau\langle \mathrm{D}E(u) - \mathrm{D}E(u^k), u - u^k\rangle \\ &\geq 2\mu\tau\|e^k\|_{\mathcal{V}}^2,\end{aligned}$$

so that

$$\|e^{k+1}\|_{\mathcal{V}}^2 \leq (1 - 2\mu\tau)\|e^k\|_{\mathcal{V}}^2 + \tau^2\|\mathrm{D}E(u) - \mathrm{D}E(u^k)\|_{\mathcal{V}'}^2. \tag{6}$$

Proof (Cont.)

Observe now that $u, u^0 \in \mathcal{B}$. We use this to inductively show that the whole sequence belongs to $\mathcal{B}$. Assume that $u^k \in \mathcal{B}$. By the local Lipschitz smoothness assumption

$$\tau^2 \|\mathrm{D}E(u) - \mathrm{D}E(u^k)\|_{\mathcal{V}'}^2 \leq \tau^2 L^2 \|e^k\|_{\mathcal{V}}^2.$$

Using this fact in (6), we conclude

$$\|e^{k+1}\|_{\mathcal{V}}^2 \leq (1 - 2\mu\tau + \tau^2 L^2)\|e^k\|_{\mathcal{V}}^2.$$

Define

$$\phi(\tau) := 1 - 2\mu\tau + \tau^2 L^2$$

and observe that, for $\tau \in (0, \frac{2\mu}{L^2})$,

$$0 < \phi(\tau) < 1.$$

Proof (Cont.)

Thus, for a step size in the stated range

$$\begin{aligned}\|e^{k+1}\|_{\mathcal{V}}^2 &\le \rho^2 \|e^k\|_{\mathcal{V}}^2 \\ &\le \|e^k\|_{\mathcal{V}}^2 \\ &\le \|u^0 - u\|_{\mathcal{V}}^2\end{aligned}$$

and we discover that $u^{k+1} \in \mathcal{B}$.

The previous inequality, i.e.,

$$\|e^{k+1}\|_{\mathcal{V}}^2 \le \rho^2 \|e^k\|_{\mathcal{V}}^2,$$

also shows the claimed convergence with linear rate. $\square$

Corollary (Optimal Rate)

In the setting of Theorem 10 the optimal rate of convergence is obtained for

$$\tau_{\mathrm{opt}} = \frac{\mu}{L^2}.$$

In fact,

$$\|u - u^{k+1}\|_{\mathcal{V}} \le \rho_{\mathrm{opt}} \|u - u^k\|_{\mathcal{V}}, \qquad \rho_{\mathrm{opt}} := \sqrt{1 - \frac{\mu^2}{L^2}}.$$

Proof.

It suffices to optimize the function ϕ defined in the course of the proof of the last theorem. The details are left as an exercise. □

More efficient methods will be presented in the next chapter.