Maximality and Monotone Operator Theory

Ernest K. Ryu

Large-Scale Convex Optimization:
Algorithms and Analyses via Monotone Operators

Stanford University
EE392F Spring 2023

Monotone operator theory

Convex optimization theory, the main subject of study in this class, focuses on the derivation and analysis of convex optimization algorithms.

Monotone operator theory views monotone operators as interesting objects in their own right and focuses on understanding them better.

One goal of this section is to provide theoretical completeness and prove several results that were simply asserted in other sections. Another goal is to provide a gentle exposure to the field of monotone operator theory.

Monotone operator theory takes place in infinite-dimensional Banach or Hilbert spaces, where a new set of interesting challenges arise. We limit our attention to finite-dimensional Euclidean spaces.

Outline

Operator extensions

$\bar{\mathbb{A}} \colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is an extension of $\mathbb{A} \colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ if $\operatorname{Gra} \bar{\mathbb{A}} \supseteq \operatorname{Gra} \mathbb{A}$.
$\bar{\mathbb{A}}$ is a proper extension of $\mathbb{A}$ if the containment $\operatorname{Gra} \bar{\mathbb{A}} \supset \operatorname{Gra} \mathbb{A}$ is strict.
Recall, a monotone operator is maximal if it has no proper monotone extension.

As discussed in §2, if $\mathbb{A} \colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone, then $\operatorname{dom} \mathbb{J}_{\mathbb{A}} = \mathbb{R}^n$.

Theorem 7.

If $f : \mathbb{R}^n \to \mathbb{R}^n \cup \{\infty\}$ is CCP, then ∂f is maximal monotone.

Proof. We know ∂f is monotone. Assume for contradiction that there is $(\tilde{x}, \tilde{g}) \notin \partial f$ such that $\{(\tilde{x}, \tilde{g})\} \cup \partial f$ is monotone. Define $(x, g) \in \partial f$ with

$$x = \operatorname*{argmin}_z \left\{ f(z) + \frac{1}{2}\|z - (\tilde{x} + \tilde{g})\|^2 \right\}, \qquad 0 = x - \tilde{x} + g - \tilde{g}.$$

Since $(\tilde{x}, \tilde{g}) \notin \partial f$, either $x \neq \tilde{x}$ or $g \neq \tilde{g}$ (or both). Using $x - \tilde{x} = -g + \tilde{g}$, we have

$$\langle g - \tilde{g}, x - \tilde{x} \rangle = -\|x - \tilde{x}\|_2^2 = -\|g - \tilde{g}\|_2^2 < 0,$$

which contradicts the assumption that $\{(\tilde{x}, \tilde{g})\} \cup \partial f$ is monotone. $\square$

Maximality of subdifferential

Key idea of proof: Given $v \in \mathbb{R}^n$,

$$v \mapsto (\underbrace{\mathrm{Prox}_f(v)}_{=x}, \underbrace{v - \mathrm{Prox}_f(v)}_{=g}) \in \partial f$$

provides a unique decomposition $v = x + g$ such that $(x, g) \in \partial f$.

Outline

Maximality of subdifferential

Fitzpatrick function

Maximality and extension theorems

Fitzpatrick function

For $\mathbb{A}\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, Fitzpatrick function $\mathbf{F}_{\mathbb{A}}\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is

$$\mathbf{F}_{\mathbb{A}}(x,u) = \langle x,u\rangle - \inf_{(y,v)\in\mathbb{A}} \langle x-y, u-v\rangle = \sup_{(y,v)\in\mathbb{A}} \left\{\langle y,u\rangle + \langle x,v\rangle - \langle y,v\rangle\right\}.$$

Lemma 3.

Assume $\mathbb{A}\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone. Then

- $\mathbf{F}_{\mathbb{A}}$ *is CCP,*
- $\mathbf{F}_{\mathbb{A}}(x,u) \geq \langle x,u\rangle$ *for all* $x,u \in \mathbb{R}^n$, *and*
- $\mathbf{F}_{\mathbb{A}}(x,u) = \langle x,u\rangle$ *if and only if* $(x,u) \in \mathbb{A}$.

Fitzpatrick function

We say $\mathbf{F}_{\mathbb{A}}$ is a representative function of $\mathbb{A}$, since $\mathbf{F}_{\mathbb{A}}$ is a convex extension of $\langle x, u \rangle$ from $\operatorname{Gra} \mathbb{A}$ to $\mathbb{R}^n \times \mathbb{R}^n$ that furthermore satisfies $\mathbf{F}_{\mathbb{A}}(x, u) \geq \langle x, u \rangle$.

Common technique in monotone operator theory: analyze a representative function to conclude results about the original operator. Analyzing $\mathbf{F}_{\mathbb{A}}$, a CCP function, with results from convex analysis is easier than directly considering $\mathbb{A}$.

Fitzpatrick function

Proof. If $(x, u) \in \mathbb{A}$, then $\langle x - y, u - v \rangle \geq 0$ for all $(y, v) \in \mathbb{A}$ by monotonicity, and the infimum

$$\inf_{(y,v)\in\mathbb{A}} \langle x - y, u - v \rangle = 0$$

is attained at (x, u). So $\mathbf{F}_{\mathbb{A}}(x, u) = \langle x, u \rangle$.

Assume $(x, u) \notin \mathbb{A}$. Then by maximality there exists a $(y, v) \in \mathbb{A}$ such that $\langle x - y, u - v \rangle < 0$. Therefore

$$\inf_{(y,v)\in\mathbb{A}} \langle x - y, u - v \rangle < 0$$

and $\mathbf{F}_{\mathbb{A}}(x, u) > \langle x, u \rangle$.

Fitzpatrick function

Define

$$f_{y,v}(x,u) = \langle y,u\rangle + \langle x,v\rangle - \langle y,v\rangle,$$

which is a closed convex function for all $(y,v) \in \mathbb{A}$. Then

$$\operatorname{epi} \mathbf{F}_{\mathbb{A}} = \bigcap_{(y,v)\in\mathbb{A}} \operatorname{epi} f_{y,v}$$

is a closed convex set as it is an intersection of closed convex sets.

Since $\mathbf{F}_{\mathbb{A}}(x,u) \geq f_{y,v}(x,u) > -\infty$ for any $(y,v) \in \mathbb{A}$, we have $\mathbf{F}_{\mathbb{A}} > -\infty$ always. On the other hand,

$$\mathbf{F}_{\mathbb{A}}(x,u) = \langle x,u\rangle < \infty$$

for any $(x,u) \in \mathbb{A}$. So $\mathbf{F}_{\mathbb{A}}$ is proper. $\square$

Minty surjectivity theorem

The Minty surjectivity theorem is foundational to operator splitting methods as it ensures that methods using resolvents are well defined.

We say $\mathbb{I} + \mathbb{A}$ is surjective if $\operatorname{range}(\mathbb{I} + \mathbb{A}) = \mathbb{R}^n$. If $\mathbb{I} + \mathbb{A}$ is surjective, then $\operatorname{dom} \mathbb{J}_{\mathbb{A}} = \mathbb{R}^n$.

Theorem 8.

If $\mathbb{A} \colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *is maximal monotone, then* $\operatorname{range}(\mathbb{I} + \mathbb{A}) = \mathbb{R}^n$.

Minty surjectivity theorem

Proof. Want to show $u \in \operatorname{range}(\mathbb{I} + \mathbb{A})$ for any $u \in \mathbb{R}^n$ and maximal monotone $\mathbb{A}$. We first establish $0 \in \operatorname{range}(\mathbb{I} + \mathbb{A})$ for any maximal monotone $\mathbb{A}$. Then the maximal monotone operator $\mathbb{B}(x) = \mathbb{A}(x) - u$ satisfies $0 \in \operatorname{range}(\mathbb{I} + \mathbb{B})$, which implies $u \in \operatorname{range}(\mathbb{I} + \mathbb{A})$ for any $u \in \mathbb{R}^d$.

Minty surjectivity theorem

We now show $0 \in \operatorname{range}(\mathbb{I} + \mathbb{A})$. Define $(y, v) \in \mathbb{R}^n \times \mathbb{R}^n$ with

$$(y, v) = \operatorname*{argmin}_{(x,u) \in \mathbb{R}^n \times \mathbb{R}^n} \left\{ \mathbf{F}_{\mathbb{A}}(x, u) + \frac{1}{2}\|x\|^2 + \frac{1}{2}\|u\|^2 \right\} = \operatorname{Prox}_{\mathbf{F}_{\mathbb{A}}}(0, 0).$$

This implies

$$\begin{bmatrix} -y \\ -v \end{bmatrix} \in \partial \mathbf{F}_{\mathbb{A}}(y, v).$$

Since $\mathbf{F}_{\mathbb{A}}$ is convex, the subgradient inequality tells us

$$\left\langle \begin{bmatrix} -y \\ -v \end{bmatrix}, \begin{bmatrix} x \\ u \end{bmatrix} - \begin{bmatrix} y \\ v \end{bmatrix} \right\rangle \leq \mathbf{F}_{\mathbb{A}}(x, u) - \mathbf{F}_{\mathbb{A}}(y, v) \qquad \forall\, (x, u) \in \mathbb{R}^n \times \mathbb{R}^n.$$

By Lemma 3,

$$\mathbf{F}_{\mathbb{A}}(x, u) - \mathbf{F}_{\mathbb{A}}(y, v) \leq \langle x, u \rangle - \langle y, v \rangle \qquad \forall\, (x, u) \in \mathbb{A}.$$

Minty surjectivity theorem

Combining the two inequalities and reorganize to get

$$\|y+v\|^2 \leq \langle x+v, u+y \rangle \qquad \forall\, (x,u) \in \mathbb{A}. \tag{1}$$

Since $0 \leq \|y+v\|^2$ and since $\mathbb{A}$ is maximal monotone, this implies $(-v,-y) \in \mathbb{A}$. By letting $(x,u) = (-v,-y)$ in (1), we get $v = -y$. Thus $(y,-y) \in \mathbb{A}$ and we have

$$0 \in (\mathbb{I} + \mathbb{A})(y).$$

□

Converse of Minty

The converse of Theorem 8 is true. As a consequence, we can show a monotone operator $\mathbb{A}\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal if $\operatorname{dom} \mathbb{J}_{\mathbb{A}} = \mathbb{R}^n$.

Theorem 9.

If $\mathbb{A}\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is monotone and $\operatorname{range}(J + \mathbb{A}) = \mathbb{R}^n$ for a symmetric positive definite $J \in \mathbb{R}^{n \times n}$, then $\mathbb{A}$ is maximal monotone.

Converse of Minty

Proof. First consider the case $J = \mathbb{I}$. Assume $\{(x, u)\} \cup \mathbb{A}$ is monotone, i.e.,

$$0 \leq \langle x - z, u - w \rangle \qquad \forall (z, w) \in \mathbb{A}.$$

To establish maximality, enough to show $(x, u) \in \mathbb{A}$. Since $\operatorname{range}(\mathbb{I} + \mathbb{A}) = \mathbb{R}^n$, there is a y such that $x + u \in (\mathbb{I} + \mathbb{A})y$. Let

$$v = x + u - y \in \mathbb{A}y.$$

Then

$$0 \leq \langle x - y, u - v \rangle = -\|x - y\|^2 = -\|u - v\|^2.$$

So $x = y$ and $u = v$, which implies $(x, u) \in \mathbb{A}$.

When $J \neq \mathbb{I}$. Then $J^{-1/2}\mathbb{A}J^{-1/2}$ is monotone and, becuase $J + \mathbb{A}$ is surjective,

$$\operatorname{range}(\mathbb{I} + J^{-1/2}\mathbb{A}J^{-1/2}) = \mathbb{R}^n.$$

This implies $J^{-1/2}\mathbb{A}J^{-1/2}$ is maximal and so is $\mathbb{A}$. $\square$

Outline

Maximality of subdifferential

Fitzpatrick function

Maximality and extension theorems

Maximality and extensions

Let P be a property of an operator such as monotonicity, θ-averagedness, or L-Lipschitz continuity. We say $\mathbb{A}\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is "maximal P" if there is no proper extension $\bar{\mathbb{A}}$ with property P. We now characterize maximal extensions of certain operator classes.

Whether a given operator can be extended while preserving certain properties is a classical question in analysis. (E.g., Hahn–Banach and Kirszbraun–Valentine theorems.)

Maximal monotone extension

Theorem 13.

A monotone operator has a maximal monotone extension.

Proof. Let $\mathbb{A}\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be monotone and let

$$\mathcal{P} = \{\mathbb{B}\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n \,|\, \mathbb{B} \text{ is monotone and } \operatorname{Gra}\mathbb{A} \subseteq \operatorname{Gra}\mathbb{B}\},$$

which is non-empty. We impose the partial order on $\mathcal{P}$ with $\mathbb{B}_1 \preceq \mathbb{B}_2$ if and only if $\operatorname{Gra}\mathbb{B}_1 \subseteq \operatorname{Gra}\mathbb{B}_2$ for all $\mathbb{B}_1, \mathbb{B}_2 \in P$. Every chain $\mathcal{C}$ in $\mathcal{P}$ has the upper bound $\bar{\mathbb{B}} \in \mathcal{P}$ given by

$$\operatorname{Gra}\bar{\mathbb{B}} = \bigcup_{\mathbb{B}\in\mathcal{C}} \operatorname{Gra}\mathbb{B}.$$

By Zorn's lemma, there is a maximal element $\bar{\mathbb{A}}$ in $\mathcal{P}$. This element $\bar{\mathbb{A}}$ extends $\mathbb{A}$ by definition of $\mathcal{P}$ and cannot be properly extended as it is maximal in $\mathcal{P}$. □

Maximal μ-strongly monotone extension

Theorem 14.

For $\mu > 0$, a μ-strongly monotone operator has a maximal μ-strongly monotone extension. Furthermore, if $\mathbb{A}\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is μ-strongly monotone, then $\mathbb{A}$ is maximal μ-strongly monotone if and only if $\mathrm{range}\,(\mathbb{A}) = \mathbb{R}^n$.

Proof. Since μ-strong monotonicity of $\mathbb{A}$ is defined as

$$\langle \mathbb{A}x - \mathbb{A}y, x - y\rangle \geq \mu\|x - y\|^2 \qquad \forall\, x, y \in \mathbb{R}^n,$$

$\mathbb{A}$ is μ-strongly monotone if and only if $\mathbb{B} = \mathbb{A} - \mu\mathbb{I}$ is monotone.

Extending $\mathbb{A}$ and $\mathbb{B}$ are equivalent in the following sense. If $\bar{\mathbb{A}}$ is a μ-strongly monotone extension of $\mathbb{A}$, then $\bar{\mathbb{A}} - \mu\mathbb{I}$ is a monotone extension of $\mathbb{B}$. If $\bar{\mathbb{B}}$ is a monotone extension of $\mathbb{B}$, then $\bar{\mathbb{B}} + \mu\mathbb{I}$ is a μ-strongly monotone extension of $\mathbb{A}$. By Theorem 13, $\mathbb{B}$ has a maximal monotone extension $\bar{\mathbb{B}}$, and $\mathbb{A}$ has a maximal μ-strongly monotone extension $\bar{\mathbb{B}} + \mu\mathbb{I}$.

Moreover, $\mathbb{A}$ is maximal μ-strongly monotone if and only if $\mathbb{B}$ is maximal monotone. By Theorems 8 and 9, $\mathbb{B}$ is maximal monotone if and only if $\mathrm{range}\,(\mathbb{A}) = \mathrm{range}\,(\mathbb{B} + \mu\mathbb{I}) = \mathbb{R}^n$. Finally, chaining the equivalences provides the second stated result. $\square$

Maximal β-cocoercive extension

Theorem 15.

Let $\beta > 0$. A β-cocoercive operator has a maximal β-cocoercive extension. Furthermore, if $\mathbb{A}\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is β-cocoercive, then $\mathbb{A}$ is maximal β-cocoercive if and only if $\operatorname{dom}\mathbb{A} = \mathbb{R}^n$.

Proof. Note $\mathbb{A}$ is β-cocoercive if and only if $\mathbb{A}^{-1}$ is β-strongly monotone.

Extending $\mathbb{A}$ and $\mathbb{A}^{-1}$ are equivalent in the following sense. If $\bar{\mathbb{A}}$ is a β-cocoercive extension of $\mathbb{A}$, then $\bar{\mathbb{A}}^{-1}$ is a β-strongly monotone extension of $\mathbb{A}^{-1}$. If $\overline{\mathbb{A}^{-1}}$ is a β-strongly monotone extension of $\mathbb{A}^{-1}$, then $(\overline{\mathbb{A}^{-1}})^{-1}$ is a β-cocoercive extension of $\mathbb{A}$. By Theorem 14, $\mathbb{A}^{-1}$ has a maximal β-strongly monotone extension $\overline{\mathbb{A}^{-1}}$, and $\mathbb{A}$ has a maximal β-cocoercive extension $(\overline{\mathbb{A}^{-1}})^{-1}$.

Maximal β-cocoercive extension

Moreover, $\mathbb{A}$ is maximal β-cocoercive if and only if $\mathbb{A}^{-1}$ is maximal β-strongly monotone. By Theorem 14, $\mathbb{A}^{-1}$ is maximal β-strongly monotone if and only if $\operatorname{range}(\mathbb{A}^{-1}) = \mathbb{R}^n$, which holds if and only if $\operatorname{dom}(\mathbb{A}) = \mathbb{R}^n$. Finally, chaining the equivalences provides the second stated result. $\square$

Remember that a β-cocoercive operators must be single-valued. By Theorem 15, [$\mathbb{A}\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal β-cocoercive] is equivalent to [$\mathbb{A}\colon \mathbb{R}^n \to \mathbb{R}^n$ is β-cocoercive] since $\mathbb{A}\colon \mathbb{R}^n \to \mathbb{R}^n$ implies $\operatorname{dom}\mathbb{A} = \mathbb{R}^n$. For the sake of conciseness, we usually avoid the former expression.

Maximal L-Lipschitz extension

Theorem 16.
Let $L > 0$. An L-Lipschitz operator has a maximal L-Lipschitz extension. Furthermore, if $\mathbb{A}\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is L-Lipschitz, then $\mathbb{A}$ is maximal L-Lipschitz if and only if $\operatorname{dom} \mathbb{A} = \mathbb{R}^n$.

This result is known as the Kirszbraun–Valentine theorem. Proof follows from similar reasoning.