Iterative Convex Optimization Algorithms; Part One: Using the Baillon–Haddad Theorem

Charles Byrne
(Charles_Byrne@uml.edu)
http://faculty.uml.edu/cbyrne/cbyrne.html
Department of Mathematical Sciences
University of Massachusetts Lowell
Lowell, MA 01854, USA

March 26, 2015

University of Massachusetts Lowell

Talk Available on Web Site

This slide presentation and accompanying article are available on my web site, http://faculty.uml.edu/cbyrne/cbyrne.html ; click on “Talks”.

Outline

1. fixed-point iteration and convergent operators;
2. orthogonal projection and gradient operators;
3. nonexpansive and firmly nonexpansive operators;
4. an elementary proof of the Baillon–Haddad Theorem;
5. Fenchel conjugate and Moreau envelope;
6. Bauschke-Combettes BH Theorem;
7. averaged operators and the Krasnosel'skii-Mann Theorem;
8. using the BH Theorem.

TFAE

Mathematics is sometimes said to be the science of patterns. Patterns often show up in the way theorems are formulated. One popular pattern, or template, for mathematical theorems is the "TFAE" pattern,

The following are equivalent:

The three versions of the Baillon–Haddad (BH) Theorem we present here all follow the TFAE pattern.

Convergent Operators

Here $\mathcal{H}$ is a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. We are concerned with the behavior of sequences $\{x^k\}$ defined by

$$x^k = Tx^{k-1},$$

where $T : \mathcal{H} \to \mathcal{H}$ is a continuous operator.

Definition

Let $T : \mathcal{H} \to \mathcal{H}$. A vector z is a **fixed point** of T if $Tz = z$. The set of fixed points of T is denoted $\text{Fix}(T)$. The operator T is **convergent** if, for any starting vector x^0, the sequence $\{x^k\}$ generated by $x^k = Tx^{k-1}$ converges weakly to a fixed point of T, whenever T has fixed points.

Objectives

We have two basic objectives.

- Given a particular operator, we want to decide if it is convergent.
- We want a large class of convergent operators from which to construct iterative algorithms.

Differentiation

Definition

A function $f : \mathcal{H} \to \mathbb{R}$ is Gâteaux differentiable at x if there is a vector $\nabla f(x)$ such that, for every x and d, the directional derivative of f at x, in the direction d, is given by $\langle \nabla f(x), d \rangle$. The function f is Gâteaux differentiable, or just differentiable, if f is Gâteaux differentiable at each x.

Theorem

If $T = \nabla f$ is continuous, then f is Fréchet differentiable and f is continuous. For a convex $f : \mathbb{R}^J \to \mathbb{R}$, Gâteaux differentiable and Fréchet differentiable are equivalent.

Orthogonal Projections and Gradient Operators

1. Let C be a nonempty, closed, convex subset of $\mathcal{H}$. For every x in $\mathcal{H}$ there is a unique member of C, denoted $P_C x$, that is the closest member of C to x. The operators P_C are the **orthogonal projection operators.**
2. Let $f : \mathcal{H} \to (-\infty, +\infty]$ be Gâteaux differentiable. Its **gradient operator** is $T = \nabla f$.

University of Massachusetts
UMASS Lowell

Fundamental Problems

1. **The Convex Feasibility Problem (CFP):** Let $C_i, i = 1, ..., I$, be nonempty, closed, convex subsets of $\mathcal{H}$, with nonempty intersection C. Find a vector x in C.
2. **The Constrained Minimization Problem:** Let C be a nonempty closed subset of $\mathcal{H}$ and $f : \mathcal{H} \to \mathbb{R}$. Find x in C such that $f(x) \leq f(y)$, for all y in C.
3. **Combine the Two:** Given x, find the point in $C = \cap_{i=1}^{I} C_i$ closest to x.

Solutions as Fixed Points

1. **CFP:** A vector x solves the CFP if and only if $x = P_{C_i}x$, for $i = 1, ..., I$. How about

$$x = \frac{1}{I}\sum_{i=1}^{I} P_{C_i}x?$$

2. **Constrained Minimization:** Let C be a closed, nonempty, convex subset and $f : \mathcal{H} \to \mathbb{R}$ be convex and differentiable. A vector x in C solves the constrained minimization problem if and only if

$$x = P_C(x - \gamma \nabla f(x)),$$

for any $\gamma > 0$.

Projected Gradient Descent Algorithms

For $k = 1, 2, ...$ let

$$x^k = P_C(x^{k-1} - \gamma_k \nabla f(x^{k-1})),$$

for selected $\gamma_k > 0$. Can we find some $\gamma > 0$ such that the operator $P_C(I - \gamma \nabla f)$ is convergent? Is there $\gamma > 0$ such that the sequence

$$x^k = P_C(x^{k-1} - \gamma \nabla f(x^{k-1}))$$

converges to a minimizer of f over C?

The CQ Algorithm

Let A be a real I by J matrix, $C \subseteq \mathbb{R}^J$, and $Q \subseteq \mathbb{R}^I$, both closed convex sets. The **split feasibility problem** (SFP) is to find x in C such that Ax is in Q. The function

$$f(x) = \frac{1}{2}\|P_Q Ax - Ax\|_2^2$$

is convex, differentiable and $\nabla f(x) = A^T(I - P_Q)Ax$ is L-Lipschitz for $L = \rho(A^T A)$. We want to minimize the function $f(x)$ over x in C. The CQ algorithm has the iterative step

$$x^k = P_C\Big(x^{k-1} - \gamma A^T(I - P_Q)Ax^{k-1}\Big).$$

The sequence converges to a solution whenever f has a minimum on the set C, for $0 < \gamma \leq 2/L$.

University of Massachusetts UMASS Lowell

Intensity Modulated Radiation Therapy

Yair Censor and colleagues modified the CQ algorithm to obtain efficient algorithms for designing protocols for **intensity modulated radiation therapy (IMRT)**.

- Censor, Y., Bortfeld, T., Martin, B., and Trofimov, A. "A unified approach for inversion problems in intensity-modulated radiation therapy." *Physics in Medicine and Biology* 51 (2006), 2353-2365.
- Censor, Y., Elfving, T., Kopf, N., and Bortfeld, T. (2005) "The multiple-sets split feasibility problem and its application for inverse problems." *Inverse Problems*, **21** , pp. 2071-2084.

Orthogonal Projection as Gradient Operators

The operators P_C and ∇f are fundamental. The functions

$$f(x) = \frac{1}{2}(\|x\|^2 - \|x - P_C x\|^2)$$

and

$$g(x) = \frac{1}{2}\|x - P_C x\|^2$$

are convex and differentiable. We have

$$\nabla f(x) = P_C x$$

and

$$\nabla g(x) = x - P_C x.$$

Proximity-Function Minimization

The function

$$f(x) = \frac{1}{I}\sum_{i=1}^{I}\|x - P_{C_i}x\|^2$$

has the gradient

$$\nabla f(x) = x - \frac{1}{I}\sum_{i=1}^{I} P_{C_i}x.$$

This suggests the iteration

$$x^k = \frac{1}{I}\sum_{i=1}^{I} P_{C_i}x^{k-1}.$$

Do fixed points exist? Do they solve the CFP?

Operators of Interest

Summarizing, operators of interest to us include

1. $$T = P_C(I - \gamma \nabla f),$$

2. $$T = \frac{1}{I} \sum_{i=1}^{I} P_{C_i} x,$$

   and

3. $$T = P_{C_I} P_{C_{I-1}} \cdots P_{C_2} P_{C_1}.$$

Strict Contractions

Definition

Let (X, d) be a complete metric space. An operator $T : X \to X$ is a strict contraction if there is r in $(0, 1)$ such that

$$d(Tx, Ty) \leq r\, d(x, y),$$

for all x and y in X.

Theorem

(Banach-Picard Theorem) *Let $T : X \to X$ be a strict contraction. Then T has a unique fixed point, to which the sequence $\{T^k x^0\}$ converges, for all x^0.*

Contrasts with Banach-Picard

- The operators are not strict contractions;
- There may be multiple fixed points or no fixed points;
- Convergence will occur if a fixed point exists, but not otherwise;
- The limit may depend on the starting vector x^0;
- An algorithm typically involves functions and convex sets to be determined in each application;
- Whether or not fixed points exist will depend on the choice of functions and convex sets.

Firmly Nonexpansive Operators

Definition

An operator $N : \mathcal{H} \to \mathcal{H}$ is **nonexpansive** (ne) if, for all x and y in $\mathcal{H}$, we have

$$\|Nx - Ny\| \leq \|x - y\|.$$

Definition

An operator $F : \mathcal{H} \to \mathcal{H}$ is **firmly nonexpansive** (fne) if, for all x and y in $\mathcal{H}$, we have

$$\langle Fx - Fy, x - y \rangle \geq \|Fx - Fy\|^2.$$

Averaged Operators

Lemma

An operator F is fne if and only if there is a ne operator N such that

$$F = \frac{1}{2}I + \frac{1}{2}N,$$

where I is the identity operator.

Definition

An operator $A : \mathcal{H} \to \mathcal{H}$ is α-**averaged** (α-**av**) if, for some α in the interval $(0, 1)$, and for some ne operator N, we have

$$A = (1 - \alpha)I + \alpha N.$$

If A is α-av for some α then A is **averaged** (av).

Properties of Averaged Operators

1. All fne operators are av; all av operators are ne.
2. Not all ne operators are convergent.
3. All av operators are convergent (Krasnosel'skii-Mann Theorem).
4. The class of fne operators is not closed to finite products.
5. The class of av operators is closed to finite products.
6. The P_C operators are ne gradients of convex functions and are fne;
7. When are other ne gradients of convex functions fne? ALWAYS! (The Baillon–Haddad Theorem).

The Baillon–Haddad Theorem

The simplest version of the BH Theorem (actually a corollary in Baillon and Haddad, 1977) is the following.

Theorem

Let $f : \mathcal{H} \to \mathbb{R}$ be convex and Gâteaux differentiable, and $T = \nabla f$. The following are equivalent:

1. *T is nonexpansive;*
2. *T is firmly nonexpansive.*

Bauschke and Combettes (2010) describe the BH theorem as

"*a remarkable result that has important applications in optimization*".

University of Massachusetts
UMASS Lowell

A New Baillon–Haddad Theorem (CB, 2014)

Theorem

Let $f : \mathcal{H} \to \mathbb{R}$ be convex and Gâteaux differentiable, and let $q(x) = \frac{1}{2}\|x\|^2$. The following are equivalent:

1. $g(x) = q(x) - f(x) = \frac{1}{2}\|x\|^2 - f(x)$ *is convex;*
2. $\frac{1}{2}\|x - y\|^2 \geq D_f(x, y) \doteq f(x) - f(y) - \langle \nabla f(y), x - y \rangle \geq 0$, *for all x and y;*
3. $D_f(x, y) \geq \frac{1}{2}\|\nabla f(x) - \nabla f(y)\|^2$, *for all x and y;*
4. $T = \nabla f$ *is firmly nonexpansive;*
5. $T = \nabla f$ *is nonexpansive and f is Fréchet differentiable.*

University of Massachusetts UMASS Lowell

Characterizing Convexity

Theorem

Let $g : \mathcal{H} \to \mathbb{R}$ be Gâteaux differentiable. The following are equivalent:

1. *g is convex;*
2. *for all x and z*
$$g(z) \geq g(x) + \langle \nabla g(x), z - x \rangle;$$
3. *for all x and z*
$$\langle \nabla g(z) - \nabla g(x), z - x \rangle \geq 0.$$

Bregman Distances

Let $f : \mathcal{H} \to \mathbb{R}$ be convex and Gâteaux differentiable.

Definition

The Bregman distance from x to y is

$$D_f(x,y) \doteq f(x) - f(y) - \langle \nabla f(y), x-y \rangle \geq 0.$$

If f is strictly convex, then $D_f(x,y) = 0$ if and only if $x = y$. From

$$D_f(x,y) + D_f(y,x) = \langle \nabla f(x) - \nabla f(y), x-y \rangle$$

we see how the Bregman distance has a role to play in the definition of firmly nonexpansive gradient and the BH Theorem.

1. implies 2.

Because $g(x) = \frac{1}{2}\|x\|^2 - f(x)$ is convex, we have

$$g(x) \geq g(y) + \langle \nabla g(y), x - y \rangle,$$

for all x and y, which is easily shown to be equivalent to

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

Remark: The Bregman distance D_f does not uniquely determine f; if we add an affine linear function to f the Bregman distance is unchanged. This is important in the next part of the proof.

2. implies 3.

Fix y and define $d(x)$ by

$$d(x) = D_f(x, y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle \geq 0.$$

Then $\nabla d(x) = \nabla f(x) - \nabla f(y)$ and $D_f(z, x) = D_d(z, x)$ for all z and x. Therefore, we have

$$\frac{1}{2}\|z - x\|^2 \geq D_d(z, x) = d(z) - d(x) - \langle \nabla d(x), z - x \rangle.$$

Now let $z - x = \nabla f(y) - \nabla f(x)$. Then 3. holds, since $d(z) \geq 0$ and

$$d(x) = D_f(x, y) \geq d(z) + \frac{1}{2}\|\nabla f(x) - \nabla f(y)\|^2.$$

University of Massachusetts UMASS Lowell

3. implies 4.

From 3. we have both

$$D_f(x,y) \geq \frac{1}{2}\|\nabla f(x) - \nabla f(y)\|^2,$$

and

$$D_f(y,x) \geq \frac{1}{2}\|\nabla f(x) - \nabla f(y)\|^2.$$

Adding these two inequalities gives

$$D_f(x,y) + D_f(y,x) = \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \|\nabla f(x) - \nabla f(y)\|^2.$$

Therefore, ∇f is firmly nonexpansive.

4. implies 5.

Clearly, if ∇f is firmly nonexpansive, then it is also nonexpansive, by Cauchy's Inequality. Since ∇f is then continuous, f must be Fréchet differentiable.

5. implies 1.

From $\nabla g(x) = x - \nabla f(x)$ we get

$$\langle \nabla g(x) - \nabla g(y), x - y \rangle = \|x - y\|^2 - \langle \nabla f(x) - \nabla f(y), x - y \rangle$$
$$\geq \|x - y\|(\|x - y\| - \|\nabla f(x) - \nabla f(y)\|) \geq 0,$$

since ∇f is nonexpansive. Therefore, g is convex.

The Fenchel Conjugate

Let $f : \mathcal{H} \to \mathbb{R}$ be convex, lower semi-continuous (closed), and therefore continuous. The affine function $h(x) = \langle a, x \rangle - \gamma$ satisfies $h(x) \leq f(x)$ for all x if and only if

$$\gamma \geq \langle a, x \rangle - f(x),$$

for all x. The smallest value of γ for which this is true is $f^*(a)$, the **Fenchel conjugate** of f at a, given by

$$f^*(a) = \sup_{x \in \mathcal{H}} \{\langle a, x \rangle - f(x)\}. \tag{1}$$

The conjugate of f^* is defined in the obvious way and $f^{**} = f$.

University of Massachusetts UMASS Lowell

Some Examples of the Fenchel Conjugate

- Let $f(x) = \frac{1}{p}\|x\|^p$. Then $f^*(a) = \frac{1}{q}\|a\|^q$, for $\frac{1}{p} + \frac{1}{q} = 1$.
- Let $f(x) = \frac{1}{2}\|x\|^2$. Then $f^*(a) = \frac{1}{2}\|a\|^2$.
- Let $f(x) = \iota_B(x)$, where $B \subseteq \mathcal{H}$ is the unit ball and $\iota_B(x)$ is the indicator function of B, which is 0 on B and $+\infty$ off of B. Then $f^*(a) = \|a\|_2$.

University of Massachusetts UMASS Lowell

The Fenchel Conjugate of the Gradient

Suppose that $a = \nabla f(y)$. Then the sup is attained at $x = y$. We have

$$f(x) - f(y) \geq \langle \nabla f(y), x - y \rangle = \langle a, x - y \rangle,$$

so that

$$\langle a, y \rangle - f(y) \geq \langle a, x \rangle - f(x)$$

and

$$f^*(\nabla f(y)) + f(y) = \langle \nabla f(y), y \rangle.$$

The Moreau Envelope

The **Moreau envelope** of the convex function $f : \mathcal{H} \to (-\infty, +\infty]$ is the continuous convex function

$$m_f(x) = \inf_{y \in \mathcal{H}} \{f(y) + \frac{1}{2}\|x - y\|^2\}. \qquad (2)$$

If f is closed, the infimum is uniquely attained at the point $\mathbf{prox}_f(x)$.

1. The operator prox_f is fne.
2. The Moreau envelope is Fréchet differentiable.
3. $\nabla m_f(x) = x - \text{prox}_f(x)$.
4. For $f(x) = \iota_C(x)$ we have

$$\mathrm{prox}_{\iota_C}(x) = P_C(x).$$

University of Massachusetts Lowell

Infimal Convolution

Let $f : \mathcal{H} \to \mathbb{R}$ and $g : \mathcal{H} \to \mathbb{R}$ be arbitrary. Then the **infimal convolution** of f and g, written $f \oplus g$, is

$$(f \oplus g)(x) = \inf_{y} \{ f(y) + g(x - y) \}. \tag{3}$$

For convex f and $g(x) = q(x) = \frac{1}{2}\|x\|^2$, we have $f \oplus q = m_f$.

Properties of the Infimal Convolution

Let f and g be functions from $\mathcal{H}$ to $\mathbb{R}$. Then $(f \oplus g)^* = f^* + g^*$. With f convex and $q(x) = \frac{1}{2}\|x\|^2 = q^*(x)$ in place of $g(x)$, we have

1. $(m_f)^* = (f \oplus q)^* = f^* + q$;
2. $m_f = f \oplus q = (f^* + q)^*$; and
3. $m_{f^*} = f^* \oplus q = (f + q)^*$.

The BH Theorem of Bauschke and Combettes (2010)

Theorem

The following are equivalent:

1. f *is Fréchet differentiable and the operator* $T = \nabla f$ *is nonexpansive;*
2. $g(x) = \frac{1}{2}\|x\|^2 - f(x)$ *is convex;*
3. $h(x) = f^*(x) - \frac{1}{2}\|x\|^2$ *is convex;*
4. $f = m_{h^*}$;
5. $\nabla f = \mathrm{prox}_h = I - \mathrm{prox}_{h^*}$;
6. f *is Fréchet differentiable and the operator* $T = \nabla f$ *is firmly nonexpansive.*

Helpful Identities

The following two identities are easy to prove and are quite helpful. For any operator $T : \mathcal{H} \to \mathcal{H}$ and $G = I - T$ we have

$$\|x - y\|^2 - \|Tx - Ty\|^2 = 2\langle Gx - Gy, x - y\rangle - \|Gx - Gy\|^2, \quad (4)$$

and

$$\langle Tx - Ty, x - y\rangle - \|Tx - Ty\|^2 =$$

$$\langle Gx - Gy, x - y\rangle - \|Gx - Gy\|^2. \quad (5)$$

From (1), if T is ne, then $\langle Gx - Gy, x - y\rangle \geq 0$, so G is a **maximal monotone** operator.

University of Massachusetts UMASS Lowell

Using the Complementary Operator

If T is ne, then the identity in (4) shows how to express this property of T in terms of a property of $G = I - T$. Similarly, the identity in (5) shows that T is fne if and only if $G = I - T$ is fne.

Definition

An operator $G : \mathcal{H} \to \mathcal{H}$ is ν**-inverse strongly monotone** (ν**-ism**), for some $\nu > 0$, if, for all x and y in $\mathcal{H}$,

$$\langle Gx - Gy, x - y \rangle \geq \nu \|Gx - Gy\|^2.$$

Properties of T and G

1. T is ne if and only if G is ν-ism for $\nu = \frac{1}{2}$;
2. T is α-av if and only if G is ν-ism for $\nu = \frac{1}{2\alpha}$, for some $0 < \alpha < 1$;
3. T is fne if and only if G is ν-ism for $\nu = 1$.
4. T is fne if and only if G is fne;
5. If G is ν-ism and $0 < \mu \le \nu$, then G is μ-ism.
6. If G is ν-ism and $\gamma > 0$ then γG is $\frac{\nu}{\gamma}$-ism.

The Krasnosel'skii-Mann Theorem

Theorem

Let $A : \mathcal{H} \to \mathcal{H}$ be α-averaged, for some $\alpha \in (0, 1)$. Then A is convergent.

Proof: Let $\mathcal{H}$ be $\mathbb{R}^N$ and $Az = z$. The identity in Equation (4) is the key to the proof. Using $Az = z$ and $(I - A)z = 0$ and setting $G = I - A$ we have

$$||z - x^k||^2 - ||z - x^{k+1}||^2 = 2\langle Gz - Gx^k, z - x^k\rangle - ||Gz - Gx^k||^2.$$

Since G is $\frac{1}{2\alpha}$-ism, we have

$$||z - x^k||^2 - ||z - x^{k+1}||^2 \geq (\frac{1}{\alpha} - 1)||x^k - x^{k+1}||^2.$$

The Proof, Continued

Consequently, the sequence $\{||z - x^k||\}$ is decreasing, the sequence $\{x^k\}$ is bounded, and $\{||x^k - x^{k+1}||\}$ converges to zero. Let x^* be a cluster point of $\{x^k\}$. Then we have $Tx^* = x^*$, so we may use x^* in place of the arbitrary fixed point z. It follows then that the sequence $\{||x^* - x^k||\}$ is decreasing. Since a subsequence converges to zero, the entire sequence converges to zero. ∎

Using the Baillon–Haddad (BH) Theorem

Let $f : \mathcal{H} \to \mathbb{R}$ be convex and Gâteaux differentiable, and the gradient of f L-Lipschitz continuous, that is,

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|.$$

Then

- the gradient of $g = \frac{1}{L}f$ is ne, so ∇g is fne and 1-ism;
- for any $0 < \gamma < \frac{2}{L}$ the operator $\gamma\nabla f = \gamma L\nabla g$ is $\nu = \frac{1}{\gamma L}$-ism;
- $I - \gamma\nabla f$ is averaged;
- the operator $P_C(I - \gamma\nabla f)$ is averaged. Therefore,

$$x^{k+1} = P_C(x^k - \gamma\nabla f(x^k))$$

converges weakly to a minimizer of f over C, if such minimizers exist.

Some Definitions and Propositions

Now we present some of the fundamental definitions and propositions in convex analysis. It is useful to allow the value $+\infty$, as, for example, in the definition of the indicator function of a set C; we take $\iota_C(x) = 0$ for x in C and $\iota_C(x) = +\infty$ for x not in C. Then ι_C is a convex function whenever C is a convex set, whereas the characteristic function of C, defined by $\chi_C(x) = 1$, for x in C, and $\chi_C(x) = 0$, for x not in C, is not convex.

Some Definitions

- The *domain* of a function $f : \mathcal{H} \to [-\infty, +\infty]$ is the set

$$\mathrm{dom}(f) = \{x | f(x) < +\infty\}.$$

- A function $f : \mathcal{H} \to [-\infty, +\infty]$ is *proper* if there is no x with $f(x) = -\infty$ and some x with $f(x)$ finite.
- A function $f : \mathcal{H} \to [-\infty, +\infty]$ is *lower semi-continuous* (lsc) or *closed* if

$$f(x) = \liminf_{y \to x} f(y).$$

- The *epigraph* of $f : \mathcal{H} \to [-\infty, +\infty]$ is the set

$$\mathrm{epi}(f) = \{(x, \gamma) | f(x) \leq \gamma\}.$$

University of Massachusetts UMASS Lowell

Some Propositions

Proposition

A function $f : \mathcal{H} \to [-\infty, +\infty]$ *is lsc (or closed) if and only if* $\operatorname{epi}(f)$ *is closed, and convex if and only if* $\operatorname{epi}(f)$ *is convex. If* f *is convex, then* $\operatorname{dom}(f)$ *is convex.*

Proposition

If $f : \mathcal{H} \to [-\infty, +\infty]$ *is proper and convex, and either* f *is lsc or* $\mathcal{H}$ *is finite-dimensional, then* f *is continuous in the interior of* $\operatorname{dom}(f)$*.*

Corollary

If $f : \mathcal{H} \to \mathbb{R}$ *is lsc and convex, then* f *is continuous. If* $\mathcal{H}$ *is finite-dimensional and* f *is convex, then* f *is continuous.*

The Subdifferential

We shall restrict our attention to functions $f : \mathcal{H} \to \mathbb{R}$, although most of the results we present are valid, with some restrictions, for proper functions.

Definition

The subdifferential at x of the function f is the set

$$\partial f(x) = \{u | f(y) \geq f(x) + \langle u, y - x \rangle \text{, for all } y\}.$$

Proposition

A function $f : \mathcal{H} \to \mathbb{R}$ is closed and convex if and only if $\partial f(x)$ is nonempty, for all x.

Corollary

Let $\mathcal{H}$ be finite-dimensional. A function $f : \mathcal{H} \to \mathbb{R}$ is convex if and only if $\partial f(x)$ is nonempty, for all x.

Differentiability

Proposition

Let $f : \mathcal{H} \to \mathbb{R}$ be convex and Gâteaux differentiable. Then the operator ∇f is strong-to-weak continuous. If $\mathcal{H}$ is finite-dimensional, then ∇f is continuous and f is Fréchet differentiable.

The Convex Case

Generally, a function $f : \mathcal{H} \to \mathbb{R}$ can be Gâteaux differentiable at x, but $\partial f(x)$ can be empty; we need not have $\nabla f(x)$ in $\partial f(x)$. However, we do have the following proposition.

Proposition

Let $f : \mathcal{H} \to \mathbb{R}$ be convex. Then f is Gâteaux differentiable at x if and only if $\partial f(x) = \{u\}$, in which case $u = \nabla f(x)$.

Corollary

Let $f : \mathcal{H} \to \mathbb{R}$ and $g : \mathcal{H} \to \mathbb{R}$ be closed and convex. If $f + g$ is Gâteaux differentiable at x, then so are f and g.

More Properties

Let $f : \mathcal{H} \to \mathbb{R}$ be closed and convex. The following hold:

1. $m_f = q - (f+q)^*$;
2. $m_f + m_{f^*} = q$; and
3. $\mathrm{prox}_f + \mathrm{prox}_{f^*} = I$.

Some Propositions

Let $f : \mathcal{H} \to \mathbb{R}$ be closed and convex, $q(x) = \frac{1}{2}\|x\|^2$, $g(x) = q(x) - f(x)$, and $h(x) = f^*(x) - q(x)$.

1. If g is convex, then so is h.
2. If h is convex, then $f = m_{h^*}$.

The End

THE END

References

- Baillon, J.-B., and Haddad, G. (1977) “Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones.” *Israel J. of Mathematics*, **26**, pp. 137-150.
- Bauschke, H., and Combettes, P. (2010) “The Baillon-Haddad Theorem revisited.” *J. Convex Analysis*, **17**, pp. 781–787.
- Bauschke, H., and Combettes, P. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, New York: Springer CMS Books in Mathematics, 2011.
- Byrne, C. (2014) “On a generalized Baillon–Haddad Theorem for convex functions on Hilbert space.” to appear in the *Journal of Convex Analysis*.
- Byrne, C. (2014) *Iterative Optimization in Inverse Problems*. Boca Raton, FL: CRC Press.

- Golshtein, E., and Tretyakov, N. (1996) *Modified Lagrangians and Monotone Maps in Optimization*. New York: John Wiley and Sons.
- Krasnosel'skii, M. (1955) "Two observations on the method of sequential approximations." *Uspeki Mathematicheskoi Nauki* (in Russian), **10(1)**.
- Mann, W. (1953) "Mean value methods in iteration."*Proc. Amer. Math. Soc.* **4**, pp. 506–510.
- Moreau, J.-J. (1965) "Proximité et dualité dans un espace hilbertien." *Bull. Soc. Math. France*, **93**, pp. 273–299.
- Rockafellar, R.T. and Wets, R. J-B. (2009) *Variational Analysis* (3rd printing), Berlin: Springer-Verlag.