Weak convexity and approximate subdifferentials

Claudia Sagastizábal
(co-author F. Atenas)
IMECC-UNICAMP Brazil
CeMEAI
CEPID - Center for Mathematical Sciences Applied to Industry

JPP-Fest *Challenges and advances in modern variational analysis*
March 15th, 2023

A quotation from 1997

Hunting for a Smaller Convex Subdifferential

V. F. Demyanov & V. Jeyakumar

Journal of Global Optimization **10**, 305–326 (1997)

With the discovery of the convex subdifferential and the subdifferential of a max-type function it was generally understood that in the nonsmooth case it is not sufficient to employ a singleton - the gradient - to study properties of a function.

A quotation from 1997

Hunting for a Smaller Convex Subdifferential

V. F. Demyanov & V. Jeyakumar

Journal of Global Optimization **10**, 305–326 (1997)

With the discovery of the convex subdifferential and the subdifferential of a max-type function it was generally understood that in the nonsmooth case it is not sufficient to employ a singleton - the gradient - to study properties of a function.
The introduction of the Clarke sub-differential was a great breakthrough, and a safari season started in the Wilderness of Endolandia.

A quotation from 1997

Hunting for a Smaller Convex Subdifferential

V. F. Demyanov & V. Jeyakumar

Journal of Global Optimization **10**, 305–326 (1997)

With the discovery of the convex subdifferential and the subdifferential of a max-type function it was generally understood that in the nonsmooth case it is not sufficient to employ a singleton - the gradient - to study properties of a function.
The introduction of the Clarke sub-differential was a great breakthrough, and a safari season started in the Wilderness of Endolandia.

Many different generalizations of the concept of gradient have been proposed.

A quotation from 1997

Hunting for a Smaller Convex Subdifferential

V. F. Demyanov & V. Jeyakumar

Journal of Global Optimization **10**, 305–326 (1997)

With the discovery of the convex subdifferential and the subdifferential of a max-type function it was generally understood that in the nonsmooth case it is not sufficient to employ a singleton - the gradient - to study properties of a function.
The introduction of the Clarke sub-differential was a great breakthrough, and a safari season started in the Wilderness of Endolandia.

Many different generalizations of the concept of gradient have been proposed.

The most productive hunter is Jean-Paul Penot.
He discovered and studied many convex objects, one of the most promising and popular being that of "small subdifferential" (nurtured jointly with P. Michel)

A quotation from 1997

Hunting for a Smaller Convex Subdifferential

V. F. Demyanov & V. Jeyakumar

Journal of Global Optimization **10**, 305–326 (1997)

With the discovery of the convex subdifferential and the subdifferential of a max-type function it was generally understood that in the nonsmooth case it is not sufficient to employ a singleton - the gradient - to study properties of a function.
The introduction of the Clarke sub-differential was a great breakthrough, and a safari season started in the Wilderness of Endolandia.

Many different generalizations of the concept of gradient have been proposed.

The most productive hunter is Jean-Paul Penot.
He discovered and studied many convex objects, one of the most promising and popular being that of "small subdifferential" (nurtured jointly with P. Michel)

today we join the expedition, now hunting for ε-subdifferentials

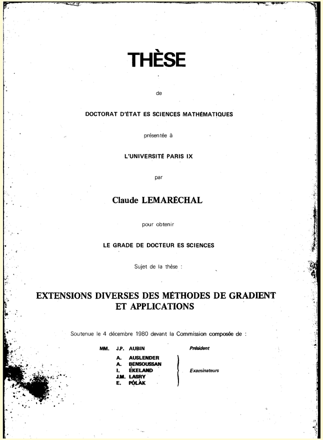
Why bother about ε-subgradients?

Two prolific hunters in the convex case

JBHU (calculus!) and CLL (algorithms)

Two prolific hunters in the convex case

JBHU (calculus!) and CLL (algorithms)Chapter 5, Dec 4th, 1980:



THÈSE

de

DOCTORAT D'ÉTAT ES SCIENCES MATHÉMATIQUES

présentée à

L'UNIVERSITÉ PARIS IX

par

Claude LEMARÉCHAL

pour obtenir

LE GRADE DE DOCTEUR ES SCIENCES

Sujet de la thèse :

EXTENSIONS DIVERSES DES MÉTHODES DE GRADIENT
ET APPLICATIONS

Soutenue le 4 décembre 1980 devant la Commission composée de :

MM.	J.P. AUBIN	*Président*
	A. AUSLENDER	*Examinateurs*
	A. BENSOUSSAN	
	I. EKELAND	
	J.M. LASRY	
	E. POLAK	

1.2. *Construction des ε-sous-différentiels*.

Etant donnée une fonction convexe, on peut en général calculer en un point x : la valeur de la fonction, et un sous-gradient ; plus rarement, on peut calculer tout $\partial f(x)$. Il est exceptionnel de pouvoir calculer directement tout $\partial_\varepsilon f(x)$ pour un $\varepsilon > 0$ (cf. [11]). La question se pose donc de savoir comment calculer des éléments de $\partial_\varepsilon f(x)$ qui ne soient pas dans $\partial f(x)$. Ce paragraphe montre qu'on peut le faire en calculant des éléments de $\partial f(y)$, pour des points y bien choisis.

Théorème 1.2.1.

Soient x et y appartenant à dom f, $g \in \partial f(y)$. Une condition nécessaire et suffisante pour que g soit également dans $\partial_\varepsilon f(x)$ est :

(11) $$f(y) \geq f(x) + (g, y-x) - \varepsilon$$

Démonstration : La condition et évidemment nécessaire : la relation de définition (7) doit au moins être satisfaite pour z=y.

<u>Remarque 1.2.2.</u> : Ce théorème très simple est fondamental, et nous en ferons un usage constant. Nous l'appellerons <u>théorème de transport des sous-gradients</u>.

Un façon simple d'interpréter la relation (11) est de considérer le nombre

$$\alpha(y,g,x) = f(x) - [f(y) + (g,x-y)]$$

Ce nombre est positif. Il représente l'erreur faite en remplaçant f(x) par la valeur en x de la linéarisation de f en y. Le théorème 1.2.1. s'écrit : soit $g \in \partial f(y)$. Alors :

$$g \in \partial_\varepsilon f(x) \quad \text{si et seulement si} \quad \alpha(y,g,x) \leq \varepsilon$$

ce qui peut s'énoncer : $g \in \partial_\varepsilon f(x)$ si f(x) est approché à ε près par la linéarisation de f en y. //

Why bother about ε-subgradients?

The inclusion $0 \in \partial f(\bar{x})$ can **fail** as optimality certificate

Why bother about ε-subgradients?

The inclusion $0 \in \partial f(\bar{x})$ can **fail** as optimality certificate

- As a set-valued mapping $\partial \;\; f(x)$ is osc:

$$\left(\quad x^k, g(x^k) \in \partial \;\; f(x^k) \right) : \begin{cases} x^k \to \bar{x} \\ g(x^k) \to \bar{g}\,. \end{cases} \Longrightarrow \bar{g} \in \partial f(\bar{x})$$

Why bother about ε-subgradients?

The inclusion $0 \in \partial f(\bar{x})$ can **fail** as optimality certificate

- As a set-valued mapping $\partial \; f(x)$ is osc:

$$\left(\quad x^k, g(x^k) \in \partial \; f(x^k) \right) : \begin{cases} \\ x^k \to \bar{x} \\ g(x^k) \to \bar{g}\,. \end{cases} \implies \bar{g} \in \partial f(\bar{x})$$

- As a set-valued mapping, $\partial \; f(x)$ is **not** isc: Given $\bar{g} \in \partial f(\bar{x})$

Why bother about ε-subgradients?

The inclusion $0 \in \partial f(\bar{x})$ can **fail** as optimality certificate

- As a set-valued mapping $\partial \; f(x)$ is osc:

$$\left(\quad x^k, g(x^k) \in \partial \; f(x^k) \right) : \begin{cases} \quad x^k \to \bar{x} \\ g(x^k) \to \bar{g}\,. \end{cases} \Longrightarrow \bar{g} \in \partial f(\bar{x})$$

- As a set-valued mapping, $\partial \; f(x)$ is **not** isc: Given $\bar{g} \in \partial f(\bar{x})$

$$\exists \left(x^k, g(x^k) \in \partial \; f(x^k) \right) : \begin{cases} \quad x^k \to \bar{x} \\ g(x^k) \to \bar{g}\,. \end{cases}$$

Why bother about ε-subgradients?

The inclusion $0 \in \partial f(\bar{x})$ can fail as optimality certificate

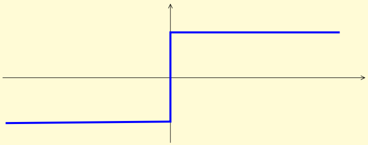
- As a set-valued mapping $\partial \ f(x)$ is osc:

$$\left(\quad x^k, g(x^k) \in \partial \ f(x^k)\right) : \begin{cases} x^k \to \bar{x} \\ g(x^k) \to \bar{g}\,. \end{cases} \Longrightarrow \bar{g} \in \partial f(\bar{x})$$

- As a set-valued mapping, $\partial \ f(x)$ is **not** isc: Given $\bar{g} \in \partial f(\bar{x})$

$$\exists \left(x^k, g(x^k) \in \partial \ f(x^k)\right) : \begin{cases} x^k \to \bar{x} \\ g(x^k) \to \bar{g}\,. \end{cases}$$

???



Why bother about ε-subgradients?

The inclusion $0 \in \partial f(\bar{x})$ can fail as optimality certificate

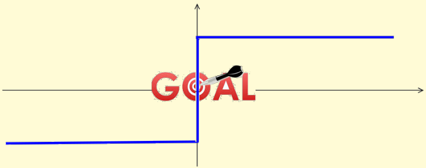
- As a set-valued mapping $\partial \; f(x)$ is osc:

$$\left(\quad x^k, g(x^k) \in \partial \; f(x^k) \right) : \begin{cases} x^k \to \bar{x} \\ g(x^k) \to \bar{g}\,. \end{cases} \Longrightarrow \bar{g} \in \partial f(\bar{x})$$

- As a set-valued mapping, $\partial \; f(x)$ is **not** isc: Given $\bar{g} \in \partial f(\bar{x})$

$$\exists \left(x^k, g(x^k) \in \partial \; f(x^k) \right) : \begin{cases} x^k \to \bar{x} \\ g(x^k) \to \bar{g}\,. \end{cases}$$

???



Why bother about ε-subgradients?

The inclusion $0 \in \partial f(\bar{x})$ can fail as optimality certificate

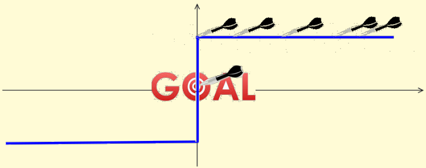
- As a set-valued mapping $\partial\ f(x)$ is osc:

$$\left(\quad x^k, g(x^k) \in \partial\ f(x^k)\right) : \begin{cases} x^k \to \bar{x} \\ g(x^k) \to \bar{g}\,. \end{cases} \Longrightarrow \bar{g} \in \partial f(\bar{x})$$

- As a set-valued mapping, $\partial\ f(x)$ is **not** isc: Given $\bar{g} \in \partial f(\bar{x})$

$$\exists\left(x^k, g(x^k) \in \partial\ f(x^k)\right) : \begin{cases} x^k \to \bar{x} \\ g(x^k) \to \bar{g}\,. \end{cases}$$

???



Why bother about ε-subgradients?

The inclusion $0 \in \partial f(\bar{x})$ can fail as optimality certificate**: use instead** $0 \in \partial_\varepsilon f(\bar{x})$

Why bother about ε-subgradients?

The inclusion $0 \in \partial f(\bar{x})$ can fail as optimality certificate**: use instead** $0 \in \partial_\varepsilon f(\bar{x})$

- As a set-valued mapping $\partial_{\varepsilon_k} f(x)$ is osc:

$$\left(\varepsilon_k, x^k, g(x^k) \in \partial_\varepsilon f(x^k)\right) : \begin{cases} \varepsilon_k \to 0 \\ x^k \to \bar{x} \\ g(x^k) \to \bar{g}\,. \end{cases} \implies \bar{g} \in \partial f(\bar{x})$$

Why bother about ε-subgradients?

The inclusion $0 \in \partial f(\bar{x})$ can fail as optimality certificate**: use instead** $0 \in \partial_\varepsilon f(\bar{x})$

- As a set-valued mapping $\partial_{\varepsilon_k} f(x)$ is osc:

$$\left(\varepsilon_k, x^k, g(x^k) \in \partial_\varepsilon f(x^k)\right) : \begin{cases} \varepsilon_k \to 0 \\ x^k \to \bar{x} \\ g(x^k) \to \bar{g}\,. \end{cases} \implies \bar{g} \in \partial f(\bar{x})$$

- As a set-valued mapping, $\partial_{\varepsilon_k} f(x)$ is also isc: Given $\bar{g} \in \partial f(\bar{x})$

$$\exists \left(x^k, g(x^k) \in \partial_{\varepsilon_k} f(x^k)\right) : \begin{cases} \varepsilon_k \to 0 \\ x^k \to \bar{x} \\ g(x^k) \to \bar{g}\,. \end{cases}$$

Why bother about ε-subgradients?

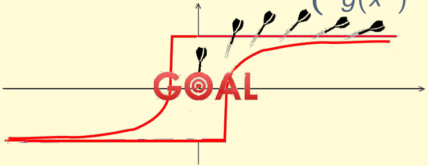
The inclusion $0 \in \partial f(\bar{x})$ can fail as optimality certificate**: use instead** $0 \in \partial_\varepsilon f(\bar{x})$

- As a set-valued mapping $\partial_{\varepsilon_k} f(x)$ is osc:

$$\left(\varepsilon_k, x^k, g(x^k) \in \partial_\varepsilon f(x^k)\right) : \begin{cases} \varepsilon_k \to 0 \\ x^k \to \bar{x} \\ g(x^k) \to \bar{g}\,. \end{cases} \implies \bar{g} \in \partial f(\bar{x})$$

- As a set-valued mapping, $\partial_{\varepsilon_k} f(x)$ is also isc: Given $\bar{g} \in \partial f(\bar{x})$

$$\exists \left(x^k, g(x^k) \in \partial_{\varepsilon_k} f(x^k)\right) : \begin{cases} \varepsilon_k \to 0 \\ x^k \to \bar{x} \\ g(x^k) \to \bar{g}\,. \end{cases}$$



Why bother about ε-subgradients?

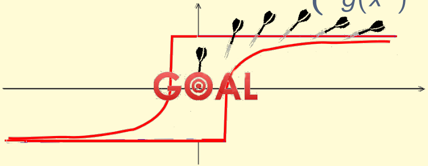
The inclusion $0 \in \partial f(\bar{x})$ can fail as optimality certificate**: use instead** $0 \in \partial_\varepsilon f(\bar{x})$

- As a set-valued mapping $\partial_{\varepsilon_k} f(x)$ is osc:

$$\left(\varepsilon_k, x^k, g(x^k) \in \partial_\varepsilon f(x^k)\right) : \begin{cases} \varepsilon_k \to 0 \\ x^k \to \bar{x} \\ g(x^k) \to \bar{g}\,. \end{cases} \implies \bar{g} \in \partial f(\bar{x})$$

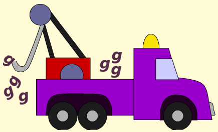
- As a set-valued mapping, $\partial_{\varepsilon_k} f(x)$ is also isc: Given $\bar{g} \in \partial f(\bar{x})$

$$\exists \left(x^k, g(x^k) \in \partial_{\varepsilon_k} f(x^k)\right) : \begin{cases} \varepsilon_k \to 0 \\ x^k \to \bar{x} \\ g(x^k) \to \bar{g}\,. \end{cases}$$



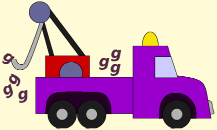
In an algorithmic scheme, ε-subgradients are built using subgradient information

The transportation formula for convex f: from $\partial f(x_0)$ to $\partial_\varepsilon f(u_0)$



Or how to express subgradients at one point
as approximate-subgradients at another point

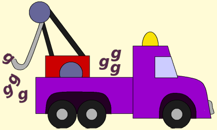
The transportation formula for convex f: from $\partial f(x_0)$ to $\partial_\varepsilon f(u_0)$



Or how to express subgradients at one point
as approximate-subgradients at another point

$f(x) \geq f(x_0) + \langle x_0^*, x - x_0 \rangle = f(u_0) + \langle x_0^*, x - u_0 \rangle - \varepsilon \quad \varepsilon = f(u_0) - f(x_0) - \langle x_0^*, u_0 - x_0 \rangle$

The transportation formula for convex f: from $\partial f(x_0)$ to $\partial_\varepsilon f(u_0)$



Or how to express subgradients at one point
as approximate-subgradients at another point

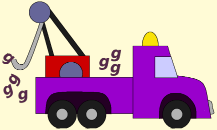
$$f(x) \geq f(x_0) + \langle x_0^*, x - x_0 \rangle = f(u_0) + \langle x_0^*, x - u_0 \rangle - \varepsilon \qquad \varepsilon = f(u_0) - f(x_0) - \langle x_0^*, u_0 - x_0 \rangle$$

Algorithmic optimality certificate

- checks if, for $x_\varepsilon^* \in \partial_\varepsilon f(x_\varepsilon)$, $\|x_\varepsilon^*\|$ and ε are small
- for some stepsize t, these objects are driven to 0 by a descent mechanism ensuring

$$0 \leq t\|x_\varepsilon^*\|^2 + \varepsilon \leq \text{fraction of}\Big(f(x^k) - f(x^{k+1})\Big)$$

The transportation formula for convex f: from $\partial f(x_0)$ to $\partial_\varepsilon f(u_0)$



Or how to express subgradients at one point
as approximate-subgradients at another point

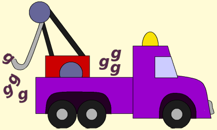
$f(x) \geq f(x_0) + \langle x_0^*, x - x_0 \rangle = f(u_0) + \langle x_0^*, x - u_0 \rangle - \varepsilon \quad \varepsilon = f(u_0) - f(x_0) - \langle x_0^*, u_0 - x_0 \rangle$

Algorithmic optimality certificate

- checks if, for $x_\varepsilon^* \in \partial_\varepsilon f(x_\varepsilon)$, $\|x_\varepsilon^*\|$ and ε are small
- for some stepsize t, these objects are driven to 0 by a descent mechanism ensuring

$$0 \leq t\|x_\varepsilon^*\|^2 + \varepsilon \leq \text{fraction of}\Big(f(x^k) - f(x^{k+1})\Big) \to 0$$

The transportation formula for convex f: from $\partial f(x_0)$ to $\partial_\varepsilon f(u_0)$



Or how to express subgradients at one point
as approximate-subgradients at another point

$f(x) \geq f(x_0) + \langle x_0^*, x - x_0 \rangle = f(u_0) + \langle x_0^*, x - u_0 \rangle - \varepsilon \quad \varepsilon = f(u_0) - f(x_0) - \langle x_0^*, u_0 - x_0 \rangle$

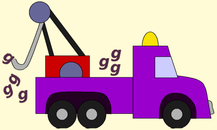
Algorithmic optimality certificate

- checks if, for $x_\varepsilon^* \in \partial_\varepsilon f(x_\varepsilon)$, $\|x_\varepsilon^*\|$ and ε are small
- for some stepsize t, these objects are driven to 0 by a descent mechanism ensuring

$$0 \leq t\|x_\varepsilon^*\|^2 + \varepsilon \leq \text{fraction of}\Big(f(x^k) - f(x^{k+1})\Big) \to 0$$

- **EUREKA**

The transportation formula for convex f: from $\partial f(x_0)$ to $\partial_\varepsilon f(u_0)$



Or how to express subgradients at one point
as approximate-subgradients at another point

$f(x) \geq f(x_0) + \langle x_0^*, x - x_0 \rangle = f(u_0) + \langle x_0^*, x - u_0 \rangle - \varepsilon \quad \varepsilon = f(u_0) - f(x_0) - \langle x_0^*, u_0 - x_0 \rangle$

Algorithmic optimality certificate

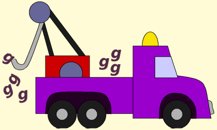
- checks if, for $x_\varepsilon^* \in \partial_\varepsilon f(x_\varepsilon)$, $\|x_\varepsilon^*\|$ and ε are small
- for some stepsize t, these objects are driven to 0 by a descent mechanism ensuring

$$0 \leq t\|x_\varepsilon^*\|^2 + \varepsilon \leq \text{fraction of}\Big(f(x^k) - f(x^{k+1})\Big) \to 0$$

- **EUREKA**

Transportation in the inverse direction, from $\partial_\varepsilon f(u_0)$ to $\partial f(x_0)$?

The transportation formula for convex f: from $\partial f(x_0)$ to $\partial_\varepsilon f(u_0)$



Or how to express subgradients at one point
as approximate-subgradients at another point

$$f(x) \geq f(x_0) + \langle x_0^*, x - x_0 \rangle = f(u_0) + \langle x_0^*, x - u_0 \rangle - \varepsilon \qquad \varepsilon = f(u_0) - f(x_0) - \langle x_0^*, u_0 - x_0 \rangle$$

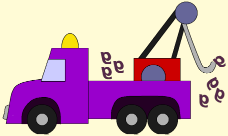
Algorithmic optimality certificate

- checks if, for $x_\varepsilon^* \in \partial_\varepsilon f(x_\varepsilon)$, $\|x_\varepsilon^*\|$ and ε are small
- for some stepsize t, these objects are driven to 0 by a descent mechanism ensuring

$$0 \leq t\|x_\varepsilon^*\|^2 + \varepsilon \leq \text{fraction of}\Big(f(x^k) - f(x^{k+1})\Big) \to 0$$

- **EUREKA**

Transportation in the inverse direction, from $\partial_\varepsilon f(u_0)$ to $\partial f(x_0)$?



$\Leftarrow$ Bröndsted-Rockafellar theorem

Bröndsted-Rockafellar's-like results

(I) For a closed proper convex function f

- given $u_0^\star \in \partial_\varepsilon f(u_0)$
- there exist $x_\varepsilon \in \mathbb{R}^n$, $x_\varepsilon^\star \in \partial f(x_\varepsilon)$ such that

$$\begin{aligned} \|x_\varepsilon - u_0\| &\leq \sqrt{\varepsilon} \\ \|x_\varepsilon^\star - u_0^\star\| &\leq \sqrt{\varepsilon} \end{aligned}$$

Bröndsted-Rockafellar's-like results

(I) For a closed proper convex function f

- given $u_0^\star \in \partial_\varepsilon f(u_0)$
- there exist $x_\varepsilon \in \mathbb{R}^n$, $x_\varepsilon^\star \in \partial f(x_\varepsilon)$ such that

$$\begin{aligned} \|x_\varepsilon - u_0\| &\leq \sqrt{\varepsilon} \\ \|x_\varepsilon^\star - u_0^\star\| &\leq \sqrt{\varepsilon} \end{aligned}$$

(II) There is a unique perturbation p such that

$$u_0^* - p \in \partial f(u_0 + p)\,, \quad \|p\| \leq \sqrt{\varepsilon}\,.$$

(useful for showing linear convergence rate)

Bröndsted-Rockafellar's-like results

(I) For a closed proper convex function f

- given $u_0^\star \in \partial_\varepsilon f(u_0)$
- there exist $x_\varepsilon \in \mathbb{R}^n$, $x_\varepsilon^\star \in \partial f(x_\varepsilon)$ such that

$$\begin{aligned} \|x_\varepsilon - u_0\| &\leq \sqrt{\varepsilon} \\ \|x_\varepsilon^\star - u_0^\star\| &\leq \sqrt{\varepsilon} \end{aligned}$$

(II) There is a unique perturbation p such that

$$u_0^* - p \in \partial f(u_0 + p), \quad \|p\| \leq \sqrt{\varepsilon}.$$

(useful for showing linear convergence rate)

(III) **a more detailed result:**

J.-P. Penot. "Subdifferential Calculus Without Qualification Assumption"

Journal of Convex Analysis 3 (1996), pp. 207-220

(III) For a closed proper convex function f,

- given $u_0^\star \in \partial_\varepsilon f(u_0)$
- there exist $x_\varepsilon \in \mathbb{R}^n$, $x_\varepsilon^\star \in \partial f(x_\varepsilon)$ and $\gamma \in [-1,1]$, such that

$$\begin{aligned}
\|x_\varepsilon - u_0\| + \frac{1}{\sqrt{\varepsilon}}|\langle u_0^\star, x_\varepsilon - u_0\rangle| &\leq \sqrt{\varepsilon} \\
\|x_\varepsilon^\star - (1+\gamma)u_0^\star\| &\leq \sqrt{\varepsilon} \\
|\langle x_\varepsilon^\star - u_0^\star, x_\varepsilon - u_0\rangle| &\leq \varepsilon \\
|\langle x_\varepsilon^\star, x_\varepsilon - u_0\rangle| &\leq 2\varepsilon \\
|f(x_\varepsilon) - f(u_0)| &\leq 2\varepsilon
\end{aligned}$$

(III) For a closed proper convex function f,

- given $u_0^\star \in \partial_\varepsilon f(u_0)$
- there exist $x_\varepsilon \in \mathbb{R}^n$, $x_\varepsilon^\star \in \partial f(x_\varepsilon)$ and $\gamma \in [-1,1]$, such that

$$\begin{aligned}
\|x_\varepsilon - u_0\| + \frac{1}{\sqrt{\varepsilon}}|\langle u_0^\star, x_\varepsilon - u_0\rangle| &\leq \sqrt{\varepsilon} \\
\|x_\varepsilon^\star - (1+\gamma)u_0^\star\| &\leq \sqrt{\varepsilon} \\
|\langle x_\varepsilon^\star - u_0^\star, x_\varepsilon - u_0\rangle| &\leq \varepsilon \\
|\langle x_\varepsilon^\star, x_\varepsilon - u_0\rangle| &\leq 2\varepsilon \\
|f(x_\varepsilon) - f(u_0)| &\leq 2\varepsilon
\end{aligned}$$

What if f is not convex?

(III) For a closed proper convex function f,

- given $u_0^\star \in \partial_\varepsilon f(u_0)$
- there exist $x_\varepsilon \in \mathbb{R}^n$, $x_\varepsilon^\star \in \partial f(x_\varepsilon)$ and $\gamma \in [-1,1]$, such that

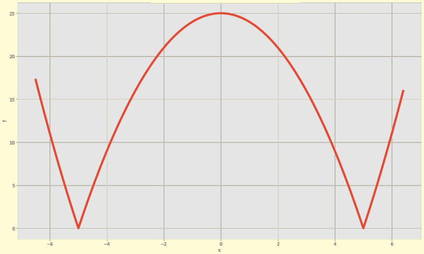
$$
\begin{aligned}
\|x_\varepsilon - u_0\| + \frac{1}{\sqrt{\varepsilon}}|\langle u_0^\star, x_\varepsilon - u_0\rangle| &\leq \sqrt{\varepsilon}\\
\|x_\varepsilon^\star - (1+\gamma)u_0^\star\| &\leq \sqrt{\varepsilon}\\
|\langle x_\varepsilon^\star - u_0^\star, x_\varepsilon - u_0\rangle| &\leq \varepsilon\\
|\langle x_\varepsilon^\star, x_\varepsilon - u_0\rangle| &\leq 2\varepsilon\\
|f(x_\varepsilon) - f(u_0)| &\leq 2\varepsilon
\end{aligned}
$$

What if f is not convex?

is there some form of benign nonconvexity that preserves BR's-like results?

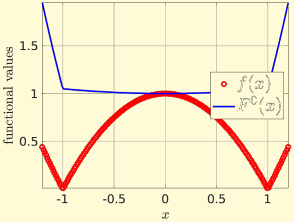
YES! Weakly convex functions

$$\exists \rho > 0 : \forall y, x \quad \text{the functions } F^y(x) := f(x) + \frac{\rho}{2}\|x - y\|^2 \text{ are convex}$$

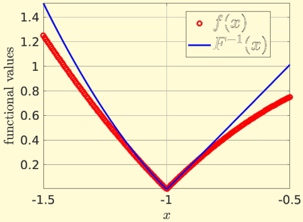


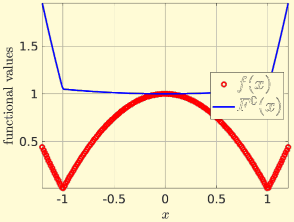
YES! Weakly convex functions

$\exists \rho > 0 : \forall y, x$ the functions $F^y(x) := f(x) + \frac{\rho}{2}\|x - y\|^2$ are convex



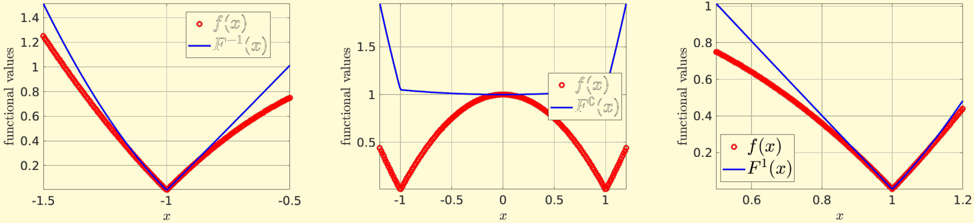
$$\exists \rho > 0 : \forall \hat{x}, x \quad \text{the functions } F^{\hat{x}}(x) := f(x) + \frac{\rho}{2}\|x - \hat{x}\|^2 \text{ are convex}$$



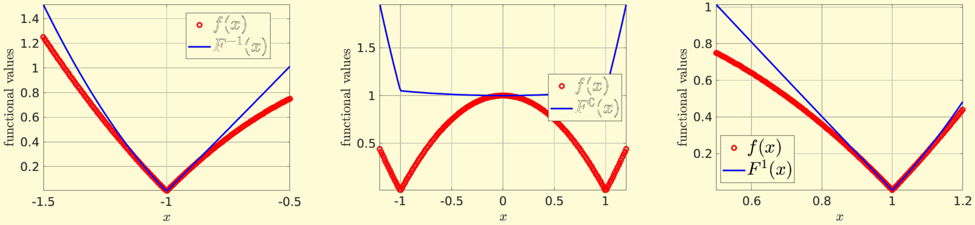


YES! Weakly convex functions

$$\exists \rho > 0 : \forall \hat{x}, x \quad \text{the functions } F^{\hat{x}}(x) := f(x) + \frac{\rho}{2}\|x - \hat{x}\|^2 \text{ are convex}$$

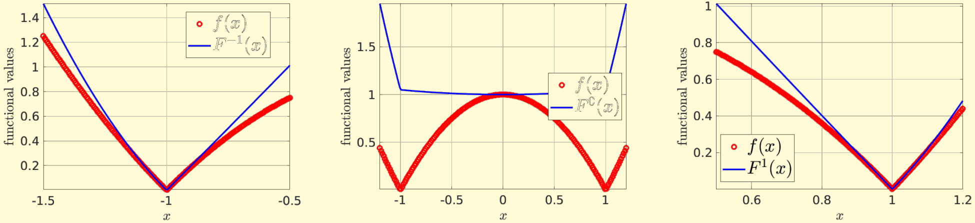


$$\exists \rho > 0 : \forall \hat{x}, x \quad \text{the functions } F^{\hat{x}}(x) := f(x) + \frac{\rho}{2}\|x - \hat{x}\|^2 \text{ are convex}$$



A w.c. *f* is a *benign* DC function:

$$\exists \rho > 0 : \forall \hat{x}, x \quad \text{the functions } F^{\hat{x}}(x) := f(x) + \frac{\rho}{2}\|x - \hat{x}\|^2 \text{ are convex}$$

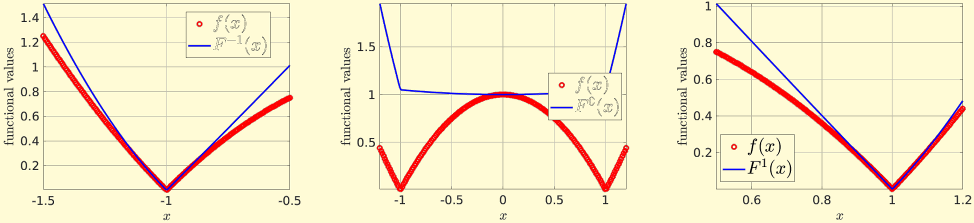


A w.c. f is a *benign* DC function:

- f is para-convex or prox-bounded *globally*

YES! Weakly convex functions

$\exists \rho > 0 : \forall \hat{x}, x \quad$ the functions $F^{\hat{x}}(x) := f(x) + \frac{\rho}{2}\|x - \hat{x}\|^2$ are convex

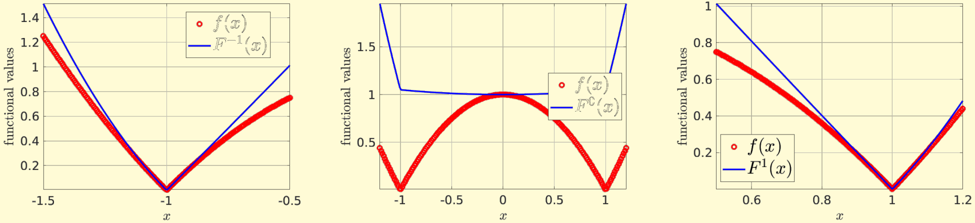


A w.c. f is a *benign* DC function:

- f is para-convex or prox-bounded *globally*
- f proximal subdifferential coincides with Clarke's

YES! Weakly convex functions

$\exists \rho > 0 : \forall \hat{x}, x$ the functions $F^{\hat{x}}(x) := f(x) + \frac{\rho}{2}\|x - \hat{x}\|^2$ are convex

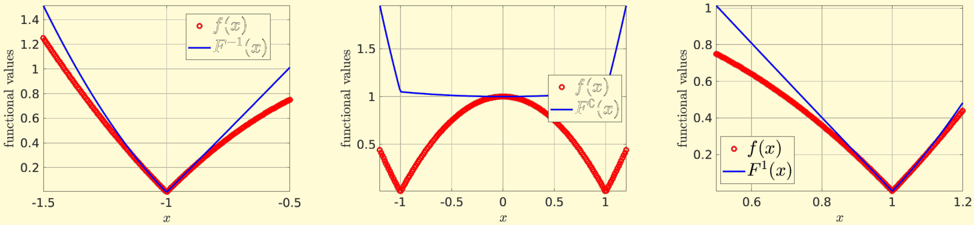


A w.c. f is a *benign* DC function:

- f is para-convex or prox-bounded *globally*
- f proximal subdifferential coincides with Clarke's
- $\partial f(\cdot) + \rho(\cdot - \hat{x})$ is the subdifferential of the convex function $F^{\hat{x}}(\cdot)$

YES! Weakly convex functions

$\exists \rho > 0 : \forall \hat{x}, x$ the functions $F^{\hat{x}}(x) := f(x) + \frac{\rho}{2}\|x - \hat{x}\|^2$ are convex



A w.c. f is a *benign* DC function:

- f is para-convex or prox-bounded *globally*
- f proximal subdifferential coincides with Clarke's
- $\partial f(\cdot) + \rho(\cdot - \hat{x})$ is the subdifferential of the convex function $F^{\hat{x}}(\cdot)$

$$\Longrightarrow \partial F^{\hat{x}}(\hat{x}) = \partial f(\hat{x})$$

There are many weakly convex functions

- Composite functions $f = h \circ c$
  - h is a convex function with Lipschitz constant L_h
  - c is a C^1 mapping, gradient has Lipschitz constant L_c

$$\rho = L_h L_c$$

There are many weakly convex functions

- Composite functions $f = h \circ c$
  - h is a convex function with Lipschitz constant L_h
  - c is a C^1 mapping, gradient has Lipschitz constant L_c

$$\rho = L_h L_c$$

- **Phase retrieval**

$$f(x) = \frac{1}{m} \sum_{i=1}^{m} |\langle a_i, x \rangle^2 - b_i|$$

There are many weakly convex functions

- Composite functions $f = h \circ c$
  - h is a convex function with Lipschitz constant L_h
  - c is a C^1 mapping, gradient has Lipschitz constant L_c

$$\rho = L_h L_c$$

- **Phase retrieval**

$$f(x) = \frac{1}{m} \sum_{i=1}^{m} |\langle a_i, x \rangle^2 - b_i|$$

- **Covariance matrix estimation** similar, but $b_i \approx a_i^\top X X^\top a_i$

There are many weakly convex functions

- Composite functions $f = h \circ c$
  - h is a convex function with Lipschitz constant L_h
  - c is a C^1 mapping, gradient has Lipschitz constant L_c

$$\rho = L_h L_c$$

- **Phase retrieval**

$$f(x) = \frac{1}{m} \sum_{i=1}^{m} |\langle a_i, x \rangle^2 - b_i|$$

- **Covariance matrix estimation** similar, but $b_i \approx a_i^\top X X^\top a_i$

In both cases ρ is independent of dimension and m

Approximate subgradients for w.c. functions

Exploit subdifferential structure of $F^{\hat{x}}(\cdot) = f(\cdot) + \frac{\rho}{2}\|\cdot - \hat{x}\|^2$

Approximate subgradients for w.c. functions

Exploit subdifferential structure of $F^{\hat{x}}(\cdot) = f(\cdot) + \frac{\rho}{2}\|\cdot - \hat{x}\|^2$

- $\partial f(\cdot) + \rho(\cdot - \hat{x})$ is the subdifferential of the convex function $F^{\hat{x}}(\cdot)$

Approximate subgradients for w.c. functions

Exploit subdifferential structure of $F^{\hat{x}}(\cdot) = f(\cdot) + \frac{\rho}{2}\|\cdot - \hat{x}\|^2$

- $\partial f(\cdot) + \rho(\cdot - \hat{x})$ is the subdifferential of the convex function $F^{\hat{x}}(\cdot)$
- $\Longrightarrow$ substract $\rho(\cdot - \hat{x})$ from $\partial_\varepsilon F^{\hat{x}}(\cdot)$

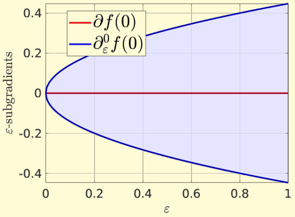
Approximate subgradients for w.c. functions

Exploit subdifferential structure of $F^{\hat{x}}(\cdot) = f(\cdot) + \frac{\rho}{2}\|\cdot - \hat{x}\|^2$

- $\partial f(\cdot) + \rho(\cdot - \hat{x})$ is the subdifferential of the convex function $F^{\hat{x}}(\cdot)$
- $\Longrightarrow$ substract $\rho(\cdot - \hat{x})$ from $\partial_\varepsilon F^{\hat{x}}(\cdot)$

  we define

$$\partial_\varepsilon^{\hat{x}} f(\cdot) := \partial_\varepsilon F^{\hat{x}}(\cdot) - \rho(\cdot - \hat{x})$$

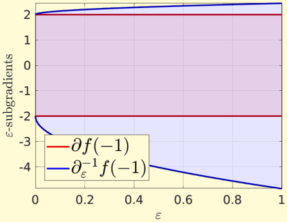


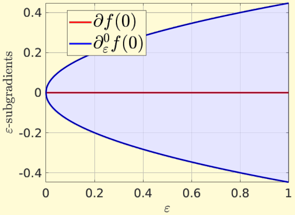
Exploit subdifferential structure of $F^{\hat{x}}(\cdot) = f(\cdot) + \frac{\rho}{2}\|\cdot - \hat{x}\|^2$

- $\partial f(\cdot) + \rho(\cdot - \hat{x})$ is the subdifferential of the convex function $F^{\hat{x}}(\cdot)$
- $\Longrightarrow$ substract $\rho(\cdot - \hat{x})$ from $\partial_\varepsilon F^{\hat{x}}(\cdot)$

we define

$$\partial_\varepsilon^{\hat{x}} f(\cdot) := \partial_\varepsilon F^{\hat{x}}(\cdot) - \rho(\cdot - \hat{x})$$





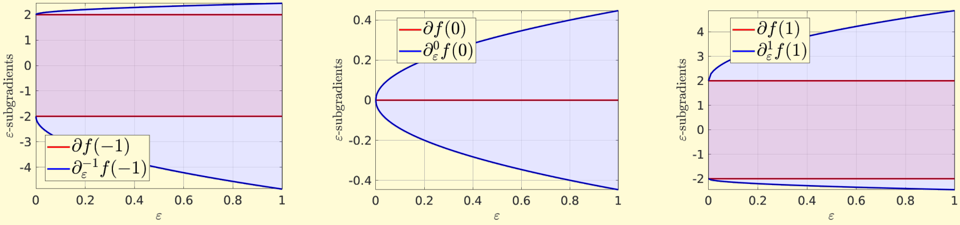
Approximate subgradients for w.c. functions

Exploit subdifferential structure of $F^{\hat{x}}(\cdot) = f(\cdot) + \frac{\rho}{2}\|\cdot - \hat{x}\|^2$

- $\partial f(\cdot) + \rho(\cdot - \hat{x})$ is the subdifferential of the convex function $F^{\hat{x}}(\cdot)$
- $\Longrightarrow$ substract $\rho(\cdot - \hat{x})$ from $\partial_\varepsilon F^{\hat{x}}(\cdot)$

we define

$$\partial_\varepsilon^{\hat{x}} f(\cdot) := \partial_\varepsilon F^{\hat{x}}(\cdot) - \rho(\cdot - \hat{x})$$

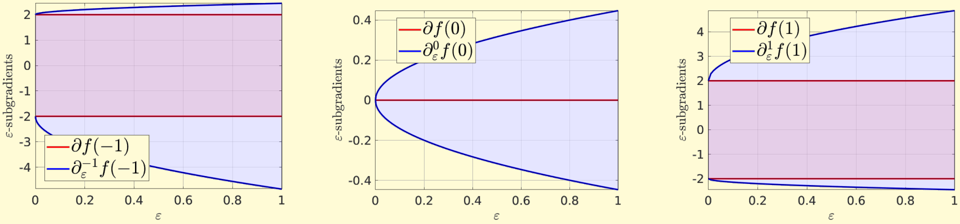


Exploit subdifferential structure of $F^{\hat{x}}(\cdot) = f(\cdot) + \frac{\rho}{2}\|\cdot - \hat{x}\|^2$

- $\partial f(\cdot) + \rho(\cdot - \hat{x})$ is the subdifferential of the convex function $F^{\hat{x}}(\cdot)$
- $\Longrightarrow$ substract $\rho(\cdot - \hat{x})$ from $\partial_\varepsilon F^{\hat{x}}(\cdot)$

we define

$$\partial_\varepsilon^{\hat{x}} f(\cdot) := \partial_\varepsilon F^{\hat{x}}(\cdot) - \rho(\cdot - \hat{x})$$



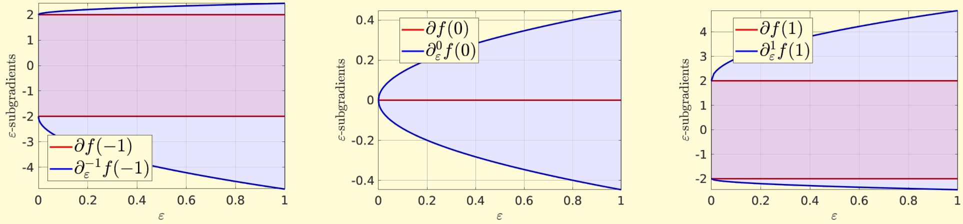
this localized continuous approximate subdifferential inherits **all** the Convex Analysis calculus for ε-subdifferentials!

Exploit subdifferential structure of $F^{\hat{x}}(\cdot) = f(\cdot) + \frac{\rho}{2}\|\cdot - \hat{x}\|^2$

- $\partial f(\cdot) + \rho(\cdot - \hat{x})$ is the subdifferential of the convex function $F^{\hat{x}}(\cdot)$
- $\Longrightarrow$ substract $\rho(\cdot - \hat{x})$ from $\partial_\varepsilon F^{\hat{x}}(\cdot)$

we define

$$\partial_\varepsilon^{\hat{x}} f(\cdot) := \partial_\varepsilon F^{\hat{x}}(\cdot) - \rho(\cdot - \hat{x})$$



this localized continuous approximate subdifferential inherits **all** the Convex Analysis calculus for ε-subdifferentials!

for instance,...

(III) For a closed proper **ρ-weakly** convex function f,

- given $u_0^\star \in \partial_\varepsilon^{u_1} f(u_0)$

(III) For a closed proper **ρ-weakly** convex function f,

- given $u_0^\star \in \partial_\varepsilon^{u_1} f(u_0)$
- there exist $x_\varepsilon \in \mathbb{R}^n$, $x_\varepsilon^\star \in \partial f(x_\varepsilon)$ and $\gamma \in [-1,1]$, such that

$$\begin{aligned}
\|x_\varepsilon - u_0\| + \frac{1}{\sqrt{\varepsilon}}|\langle u_0^\star, x_\varepsilon - u_0\rangle| &\leq \sqrt{\varepsilon} \\
\|x_\varepsilon^\star - (1+\gamma)u_0^\star\| &\leq \sqrt{\varepsilon} \\
|\langle x_\varepsilon^\star - u_0^\star, x_\varepsilon - u_0\rangle| &\leq \varepsilon \\
|\langle x_\varepsilon^\star, x_\varepsilon - u_0\rangle| &\leq 2\varepsilon
\end{aligned}$$

(III) For a closed proper **ρ-weakly** convex function f,

- given $u_0^\star \in \partial_\varepsilon^{u_1} f(u_0)$
- there exist $x_\varepsilon \in \mathbb{R}^n$, $x_\varepsilon^\star \in \partial f(x_\varepsilon)$ and $\gamma \in [-1,1]$, such that

$$\begin{aligned}
\|x_\varepsilon - u_0\| + \frac{1}{\sqrt{\varepsilon}}|\langle u_0^\star, x_\varepsilon - u_0\rangle| &\leq \sqrt{\varepsilon} \\
\|x_\varepsilon^\star - (1+\gamma)u_0^\star\| &\leq \sqrt{\varepsilon} \\
|\langle x_\varepsilon^\star - u_0^\star, x_\varepsilon - u_0\rangle| &\leq \varepsilon \\
|\langle x_\varepsilon^\star, x_\varepsilon - u_0\rangle| &\leq 2\varepsilon \\
|f(x_\varepsilon) - f(u_0)| &\leq (2+\rho)\varepsilon + \frac{3\rho}{2}\|u_0 - u_1\|^2
\end{aligned}$$

More results

- **Extension of JPP's Corollary 1.2** For a closed proper **ρ-weakly** convex function f,
  - given $u_0 \in dom\, f$
  - there exists a sequence $\left(x_k, x_k^* \in \partial f(x_k)\right)_k$ such that

$$\begin{aligned} x_k &\to u_0 \\ f(x_k) &\to f(u_0) \\ \langle x_k^\star, x_k - u_0 \rangle &\to 0. \end{aligned}$$

Any cluster point of $\{x_k^\star\}$ satisfying these properties, whenever it exists, belongs to $\partial f(u_0)$.

More results

- **Extension of JPP's Corollary 1.2** For a closed proper **ρ-weakly** convex function f,
  - given $u_0 \in dom\, f$
  - there exists a sequence $\left(x_k, x_k^* \in \partial f(x_k)\right)_k$ such that

$$\begin{array}{rcl} x_k & \to & u_0 \\ f(x_k) & \to & f(u_0) \\ \langle x_k^\star, x_k - u_0 \rangle & \to & 0. \end{array}$$

Any cluster point of $\{x_k^\star\}$ satisfying these properties, whenever it exists, belongs to $\partial f(u_0)$.

- **Algorithmic optimality certificate** checks if for $x_\varepsilon^* \in \partial_\varepsilon^{x_\varepsilon} f(x_\varepsilon)$, $\|x_\varepsilon^*\|$ and ε are small
  - since $\partial_\varepsilon^{x_\varepsilon} f(x_\varepsilon) = \partial_\varepsilon F^{x_\varepsilon}(x_\varepsilon)$

$$f(y) + \frac{\rho}{2}\|y - x_\varepsilon\|^2 \geq f(x_\varepsilon) + \langle x_\varepsilon^*, y - x_\varepsilon \rangle - \varepsilon$$

(on going work)

ε-subgradient descent schemes for w.c. optimization *à la* S. M. Robinson.

"Linear convergence of epsilon-subgradient descent methods for a class of convex functions"

Mathematical Programming 86.1 (1999)

Sequences

$$x_{k+1} = x_k - t_k d_k \quad \text{for} \quad d_k \in \partial_{\varepsilon_k}^{x_k} f(x_k) \quad \text{and } t_k \in [t_{\min}, t_{\max}]$$

(on going work)

ε-subgradient descent schemes for w.c. optimization *à la* S. M. Robinson.

"Linear convergence of epsilon-subgradient descent methods for a class of convex functions"

Mathematical Programming 86.1 (1999)

Sequences

$$x_{k+1} = x_k - t_k d_k \quad \text{for} \quad d_k \in \partial_{\varepsilon_k}^{x_k} f(x_k) \quad \text{and } t_k \in [t_{\min}, t_{\max}]$$

satisfying a descent condition

$$f(x_{k+1}) \leq f(x_k) - m\Big(\varepsilon_k + t_k \|d_k\|^2\Big) \qquad m \in (0,1)$$

(on going work)

ε-subgradient descent schemes for w.c. optimization *à la* S. M. Robinson.

"Linear convergence of epsilon-subgradient descent methods for a class of convex functions"

Mathematical Programming 86.1 (1999)

Sequences

$\subset$ nonconvex prox and redistributed bundle serious steps

$$x_{k+1} = x_k - t_k d_k \quad \text{for} \quad d_k \in \partial_{\varepsilon_k}^{x_k} f(x_k) \quad \text{and } t_k \in [t_{\min}, t_{\max}]$$

satisfying a descent condition

$$f(x_{k+1}) \leq f(x_k) - m\left(\varepsilon_k + t_k \|d_k\|^2\right) \qquad m \in (0,1)$$

(on going work)

ε-subgradient descent schemes for w.c. optimization *à la* S. M. Robinson.

"Linear convergence of epsilon-subgradient descent methods for a class of convex functions"

Mathematical Programming 86.1 (1999)

$\subset$ nonconvex prox and redistributed bundle serious steps

Sequences

$$x_{k+1} = x_k - t_k d_k \quad \text{for} \quad d_k \in \partial_{\varepsilon_k}^{x_k} f(x_k) \quad \text{and } t_k \in [t_{\min}, t_{\max}]$$

satisfying a descent condition

$$f(x_{k+1}) \leq f(x_k) - m\left(\varepsilon_k + t_k \|d_k\|^2\right) \qquad m \in (0,1)$$

can be shown to be

globally convergent

with linear speed, if KL condition and proper separation of isocost

What next?

- faster algorithms for w.c. optimization
  - stable VU-decompositions exploiting ε-subdifferential structure

What next?

- faster algorithms for w.c. optimization
  - stable VU-decompositions exploiting ε-subdifferential structure
- For composite functions $f = h \circ c$,

$$\partial f(x) = c'(x)^\top \partial h(C) \quad \text{for } C = c(x)$$

What next?

- faster algorithms for w.c. optimization
  - stable VU-decompositions exploiting ε-subdifferential structure
- For composite functions $f = h \circ c$,

$$\partial f(x) = c'(x)^\top \partial h(C) \quad \text{for } C = c(x)$$

  - compare $\partial_\varepsilon^x f(x)$ with the (global) approximate subdifferential

$$\partial_\varepsilon f(x) := c'(x)^\top \partial_\varepsilon h(C)$$

What next?

- faster algorithms for w.c. optimization
  - stable VU-decompositions exploiting ε-subdifferential structure
- For composite functions $f = h \circ c$,

$$\partial f(x) = c'(x)^\top \partial h(C) \quad \text{for } C = c(x)$$

  - compare $\partial_\varepsilon^x f(x)$ with the (global) approximate subdifferential

$$\partial_\varepsilon f(x) := c'(x)^\top \partial_\varepsilon h(C)$$

- theory
  - continuous time proximal methods, ODE's
  - results from Convex Analysis $\varepsilon-$subdifferential

What next?

- faster algorithms for w.c. optimization
  - stable VU-decompositions exploiting ε-subdifferential structure
- For composite functions $f = h \circ c$,

$$\partial f(x) = c'(x)^\top \partial h(C) \quad \text{for } C = c(x)$$

  - compare $\partial_\varepsilon^x f(x)$ with the (global) approximate subdifferential

$$\partial_\varepsilon f(x) := c'(x)^\top \partial_\varepsilon h(C)$$

- theory
  - continuous time proximal methods, ODE's
  - results from Convex Analysis $\varepsilon-$subdifferential

The end: merci et joyeux anniversaire