A Proximal Average Suitable for Nonconvex Functions

Warren L. Hare

UBC OKANAGAN

Nov 5th, 2009

Abstract

Recently, researchers have examine the question of how to smoothly transform one function into another. This is, given functions f_0 and f_1, how can we build a "well-behaved" parameterized function $F(x, p)$ such that $F(x, 0) = f_0(x)$ and $F(x, 1) = f_1(x)$? For convex functions the idea of a "proximal average" has been shown to be highly effective. We explore the proximal average, provide some previous results regarding convex functions, and develop a method to extend these results to non-convex functions. In doing so we develop a new version of the proximal average, which is more complicated but provides stronger stability results.

Outline

1. Introduction, Convex Proximal Average
2. Proximal Envelopes, NC Proximal Average
3. Parametric Prox-Regularity
4. Conclusions

Introduction

Our question

Given f_0 and f_1 how can we create

$$F(x, \lambda)$$

such that

$$F(x,0) = f_0(x) \quad F(x,1) = f_1(x)$$

and F is well behaved in λ?

One Reason and Approach

One reason is multi-objective optimization

$$\min_x \{ f_0(x) \quad \text{and} \quad f_1(x) \}$$

One Reason and Approach

One reason is multi-objective optimization

$$\min_x \{f_0(x) \;\text{ and }\; f_1(x)\}$$

The common method in this case to set

$$F(x,\lambda) = (1-\lambda)f_0(x) + \lambda f_1(x)$$

One Reason and Approach

One reason is multi-objective optimization

$$\min_x \{ f_0(x) \quad \text{and} \quad f_1(x) \}$$

The common method in this case to set

$$F(x, \lambda) = (1 - \lambda) f_0(x) + \lambda f_1(x)$$

But if

$$\operatorname{dom} f_0 \cap \operatorname{dom} f_1 \neq \operatorname{dom} f_i$$

this does not “behave well”

An alternate approach

In 2008, Bauschke, Lucet, & Trienis proposed the Proximal Average

$$\widehat{\mathcal{PA}}(x,\lambda) = \left((1-\lambda)\left(f_0 + \frac{1}{2}q\right)^* + \lambda\left(f_1 + \frac{1}{2}q\right)^*\right)^*(x) - \frac{1}{2}q(x)$$

where

$$f^*(y) := \sup_x \{\langle x, y\rangle - f(x)\}$$

and

$$q(x) = ||x||^2$$

$\widehat{\mathcal{PA}}$ Behaves Well

Theorem:
Let f_0 and f_1 be proper **convex** lsc.
Then for $\lambda \in [0, 1]$,

1. $\widehat{\mathcal{PA}}$ is a well-defined and convex
2. $\widehat{\mathcal{PA}}(x, 0) = f_0(x)$ and $\widehat{\mathcal{PA}}(x, 1) = f_1(x)$
3. $\widehat{\mathcal{PA}}$ is epi-continuous in λ

F is Epi-cont in λ if
$\mathrm{epi} F(\cdot, \lambda_k) = \{(x, \alpha) : \alpha \geq F(x, \lambda_k)\}$
converges setwise to $\mathrm{epi} F(\cdot, \lambda)$ as $\lambda_k \to \lambda$.

Proximal Average for nonconvex functions

Can a Proximal Average work if f_i are nonconvex?

Proximal Envelopes and the NC-Proximal Average

Relation to Proximal Envelopes

The Proximal Envelope* and Proximal Point Mapping

$$e_r f(x) := \inf_y \left\{ f(y) + \frac{r}{2}|y-x|^2 \right\}$$

$$\mathcal{P}_r f(x) := \mathrm{argmin}_y \left\{ f(y) + \frac{r}{2}|y-x|^2 \right\}$$

We call r the **prox-parameter** and x the **prox-center**

f is **prox-bounded** if $e_r f$ is well-defined for some $r > 0$

threshold of prox-boundedness $=$ greatest lower bound on such r

Relation to Proximal Envelopes

The Proximal Envelope* and Proximal Point Mapping

$$e_r f(x) := \inf_y \left\{ f(y) + \frac{r}{2}|y-x|^2 \right\}$$

$$\mathcal{P}_r f(x) := \mathrm{argmin}_y \left\{ f(y) + \frac{r}{2}|y-x|^2 \right\}$$

We call r the **prox-parameter** and x the **prox-center**

f is **prox-bounded** if $e_r f$ is well-defined for some $r > 0$

threshold of prox-boundedness $=$ greatest lower bound on such r

* Moreau Envelope, Yosida Regularization, Tikhonov Regularization, etc...

Relation to Proximal Envelopes

Proximal Envelopes are related to conjugate functions via

$$\left(f + \frac{r}{2}q\right)^* (rx) = \left(-e_r f + \frac{r}{2}q\right)(x)$$

Using this, we see

$$\begin{aligned}
\widehat{\mathcal{PA}}(x,\lambda) &= \left((1-\lambda)\left(f_0 + \tfrac{1}{2}q\right)^* + \lambda\left(f_1 + \tfrac{1}{2}q\right)^*\right)^*(x) - \tfrac{1}{2}q(x) \\
&= -e_1(-(1-\lambda)e_1 f_0 - \lambda e_1 f_1)(x)
\end{aligned}$$

So

$$\widehat{\mathcal{PA}}(x,0) = -e_1(-e_1 f_0)(x) \text{ and } \widehat{\mathcal{PA}}(x,1) = -e_1(-e_1 f_1)(x)$$

Insight

$$\widehat{\mathcal{PA}}(x,0) = -e_1(-e_1 f_0)(x) \quad \text{and} \quad \widehat{\mathcal{PA}}(x,1) = -e_1(-e_1 f_1)(x)$$

Why/when does $f_i = -e_1(-e_1 f_i)$?

Insight

$$\widehat{\mathcal{PA}}(x,0) = -e_1(-e_1 f_0)(x) \text{ and } \widehat{\mathcal{PA}}(x,1) = -e_1(-e_1 f_1)(x)$$

Why/when does $f_i = -e_1(-e_1 f_i)$?

[Rockafellar & Wets, '98]

$$-e_r(-e_r f) = f \quad \Leftrightarrow \quad f + \frac{r}{2}q \text{ is convex and lsc}$$

Proximal-Average

This suggests a new, broader, form for the proximal average

$$\mathcal{PA}(x,\lambda) = -e_r \ \ (-(1-\lambda)e_r f_0 - \lambda e_r f_1)(x)$$

Another interesting fact

[Rockafellar & Wets, '98]

Let f be proper, lsc, and prox-bounded with threshold r
If $r_2 > r_1 > r$, then

$$-e_{r_2}(-e_{r_1}f) \in \mathcal{C}^{1+}$$

NC Proximal-Average

These two insights suggest a new NC-Proximal Average:

$$\mathcal{PA}(x,\lambda) = -e_{r+\lambda(1-\lambda)}(-(1-\lambda)e_r f_0 - \lambda e_r f_1)(x)$$

Basic Results

$$\mathcal{PA}(x,\lambda) = -e_{r+\lambda(1-\lambda)}(-(1-\lambda)e_r f_0 - \lambda e_r f_1)(x)$$

Implies

$$\mathcal{PA}(x,i) = -e_r(-e_r f_i) \quad \text{for } i = 0,1$$

so

$$\mathcal{PA}(x,i) = f_i \quad \text{if } f_i + \frac{r}{2}q \text{ is convex and lsc}$$

Basic Results

$$\mathcal{PA}(x,\lambda) = -e_{r+\lambda(1-\lambda)}(-(1-\lambda)e_r f_0 - \lambda e_r f_1)(x)$$

has

$$\text{“}r_2 > r_1\text{”} \quad \text{when} \quad \lambda \in (0,1)$$

so
if f_i is proper, lsc, and prox-bounded, then for fixed $\lambda \in (0,1)$

$$\mathcal{PA}(x,\lambda) \in \mathcal{C}^{1+} \text{ as a function of } x$$

Basic Results

$$\mathcal{PA}(x,\lambda) = -e_{r+\lambda(1-\lambda)}(-(1-\lambda)e_r f_0 - \lambda e_r f_1)(x)$$

has

$$r + \lambda(1-\lambda) > r \quad \text{when} \quad \lambda \in (0,1)$$

so
if f_i is proper, lsc, and prox-bounded, then for fixed $\lambda \in (0,1)$

$$\mathcal{PA}(x,\lambda) \in \mathcal{C}^{1+} \text{ as a function of } x$$

Lower-$\mathcal{C}^2$

f is **lower-$\mathcal{C}^2$** on $\mathcal{O}$ if

for all $x \in \mathcal{O}$ there exists $\rho > 0$ such that $f + \frac{\rho}{2} q$ is convex on $\mathcal{O}$

Basic Results Revisited

Theorem:
f_0, f_1 lsc, proper, prox-bounded
r be greater than the threshold of prox-boundedness for f_0 and f_1
Then

1. for all $\lambda \in [0,1]$ $\mathcal{PA}$ is proper (in x)
2. for all $\lambda \in (0,1)$ $\mathcal{PA}$ is lower-$\mathcal{C}^2$ and $\mathcal{C}^{1+}$ (in x)
3. if $f_i + \frac{r}{2}q$ is convex, then $\mathcal{PA}_r(x,i) = f_i(x)$

Parametric Prox-regularity

Sub-differentials

In this talk we will assume regularity, so

$$\partial f(\bar{x}) := \{w : f(x) \geq f(\bar{x}) + \langle w, x - \bar{x}\rangle + o(|x - \bar{x}|)$$

Prox-regular

f is **prox-regular** (**PR**) at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ if
f is locally lsc at $\bar{x}$ and there exist $\rho > 0$ such that

$$f(x') \geq f(x) + \langle v, x' - x \rangle - \frac{\rho}{2}|x' - x|^2$$

whenever x, x' near $\bar{x}$, $f(x)$ near $f(\bar{x})$, $v \in \partial f(x)$ near $\bar{v}$

Convex $\Rightarrow$ lower-$\mathcal{C}^2$ $\Rightarrow$ prox-regular

Parametric Prox-regular

$f(x,\lambda)$ is parametrically prox-regular (para-**PR**) at $\bar{x}$
with **compatible parameterization in** λ at $\bar{\lambda} \in \mathrm{dom} f(\bar{x},\cdot)$
for $\bar{v} \in \partial_x f(\bar{x},\bar{\lambda})$
if there exits $\rho > 0$ such that

$$f(x',\lambda) \geq f(x,\lambda) + \langle v, x'-x \rangle - \frac{\rho}{2}|x'-x|^2$$

whenever x, x' near $\bar{x}$, λ near $\bar{\lambda}$, $f(x,\lambda)$ near $f(\bar{x},\bar{\lambda})$, $v \in \partial f(x)$ near $\bar{v}$

Example of Parametric PR

Lemma:

f_0, f_1 lower-$\mathcal{C}^2$ near $\bar{x}$

Define

$$F(x, \lambda) := (1 - \lambda) f_0(x) + \lambda f_1(x)$$

Then F is para-**PR** at $\bar{x}$ with compatible parameterization in λ at any $\bar{\lambda} \in [0, 1]$

Example of Parametric PR

Lemma:

f_0, f_1 lower-$\mathcal{C}^2$ near $\bar{x}$

Define

$$F(x,\lambda) := (1-\lambda)f_0(x) + \lambda f_1(x)$$

Then F is para-**PR** at $\bar{x}$ with compatible parameterization in λ at any $\bar{\lambda} \in [0,1]$

Question: True for f, g **PR**?

Corollary

Lemma:

f_0, f_1 lsc, proper, prox-bounded

r be greater than the threshold of prox-boundedness

Then

$$F(x, \lambda) := -(1-\lambda)e_r f_0(x) - \lambda e_r f_1(x)$$

is para-**PR** at any $\bar{x}$ with compatible parameterization in λ at any $\bar{\lambda} \in [0, 1]$

Tool

Theorem [H. & Poliquin, '07]
$F(x,\lambda)$ prox-bounded and para-**PR** at $\bar{x}$ with compatible parameterization in λ at $\bar{\lambda}$
Suppose:

1. $(0,y) \in \partial^\infty F\left(\bar{x},\bar{\lambda}\right) \Rightarrow y = 0$,
2. $(0,\lambda') \in D^*(\partial_x F)\left(\bar{x},\bar{\lambda}|\bar{v}\right)(0) \Rightarrow \lambda' = 0$,
3. there exists $\rho > 0$ such that
   $(x',\lambda') \in D^*(\partial_x F)\left(\bar{x},\bar{\lambda}|\bar{v}\right)(v'),\ \ v' \neq 0 \Rightarrow \langle x', v'\rangle > -\rho|v'|^2$,
4. $\partial_x F(\bar{x},\cdot)$ has a continuous selection near $\bar{\lambda}$,

Then for $\bar{r}$ and K sufficiently large
$\mathcal{P}_r F_\lambda(x)$ is single-valued, with

$$\left|\mathcal{P}_r F_\lambda(x) - \mathcal{P}_{r'} F_{\lambda'}(x')\right| \leq K\left|(r(x-\bar{x}) - r'(x'-\bar{x}), \lambda - \lambda', r - r')\right|,$$

near $(\bar{x} + (1/r)\bar{v}, \bar{\lambda}, \bar{r})$
$(F_\lambda(x) = F(x,\lambda))$

Stability of NC-Proximal Average

f_0, f_1 lsc, proper, prox-bounded
Suppose $\mathcal{P}_r f_0$ and $\mathcal{P}_r f_1$ are Lipschitz with

$$lip\{r(\lambda \mathcal{P}_r f_0 + (1-\lambda)\mathcal{P}_r f_1 - I)\} \leq r,$$

then, for r sufficiently large and $\lambda \in (0,1)$ we have

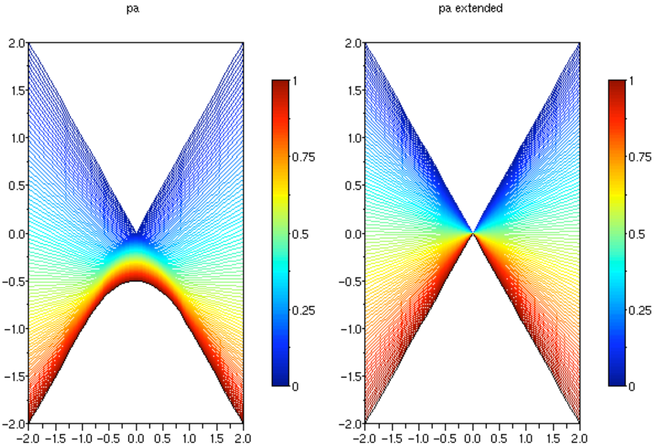
i. $\mathcal{PA}_r$ is $\mathcal{C}^{1+}$ in x,
ii. $\mathcal{PA}_r$ is locally Lipschitz continuous in λ, and
iii. $\nabla_x \mathcal{PA}_r$ is locally Lipschitz continuous in λ.

If for either $i = 0$ or $i = 1$ one has that $f_i + \frac{r}{2}q$ is convex, then $\mathcal{PA}_r(x, i) = f_i(x)$ for all x

Cororllary: If $f_0 + \frac{r}{2}q$ and $f_1 + \frac{r}{2}q$ are convex then all of the above holds

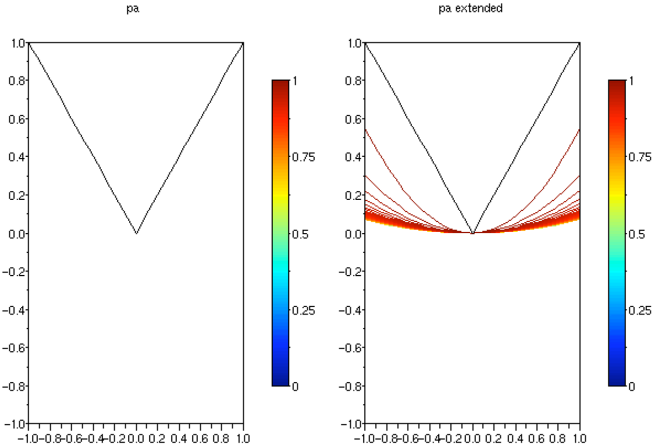
Pretty Pictures

$\widehat{\mathcal{PA}}$ and $\mathcal{PA}$ of $f_0(x) = |x|$, $f_1(x) = -|x|$



* Figures thanks to Yves Lucet

$\widehat{\mathcal{PA}}$ and $\mathcal{PA}$ of $f_0(x) = |x|, f_1(x) = |x|$



* Figures thanks to Yves Lucet

Conclusions and Future Directions

Conclusions

- Smoothly transforming one function into another is more challenging than it looks.
- One method is the Proximal Average

$$\widehat{\mathcal{PA}} = -e_1(-(1-\lambda)e_1 f_0 - \lambda e_1 f_1)(x)$$

- Another method is the NC-Proximal Average

$$\mathcal{PA}(x,\lambda) = -e_{r+\lambda(1-\lambda)}(-(1-\lambda)e_r f_0 - \lambda e_r f_1)(x)$$

- This NC-Proximal Average enjoys stronger stability, but may have other drawbacks

One research direction

What is

$$\min \mathcal{P}\mathcal{A}_r? \qquad \mathrm{argmin}\mathcal{P}\mathcal{A}?$$

One research direction

What is

$$\min \mathcal{PA}_r? \quad \operatorname{argmin} \mathcal{PA}?$$

Example:
Let $f_0 = i_A$ and $f_1 = i_B$, where A and B are convex sets. Then

$$\min_y \mathcal{PA}_r = \min_y \left\{ \frac{r}{2} \left((1-\lambda)\mathrm{dist}^2(y, A) + \lambda \mathrm{dist}^2(y, B) \right) \right\}$$

Results

Results

CMS: Windsor Dec 6th, 2009

Thank You

References

- Rockafellar & Wets, "Variational Analysis", *Springer*, New York, (1998)
- Bauschke, Lucet, & Trienis, "How to transform one convex function continuously into another," *SIAM Review*, 50(1), 115–132, (2008)
- Hare, "The Proximal Average of Nonconvex Functions: A Proximal Stability Perspective" *SIAM J. Optimization*, 20(2), 650–666, (2009)