An inertial forward-backward method for solving vector optimization problems

Sorin-Mihai Grad

Chemnitz University of Technology
www.tu-chemnitz.de/∼gsor
research supported by the DFG project GR 3367/4-1
joint work with Radu Ioan Boţ

Splitting Algorithms, Modern Operator Theory, and Applications Workshop
Oaxaca, September 19, 2017

Outline

- Preliminaries
- An inertial forward-backward proximal method
- Alternative constructions and hypotheses
- Numerical experiments

Preliminaries

- X - Hilbert space, Y - separable Banach space
- $C \subseteq Y$ pointed (i.e. $C \cap (-C) = \{0\}$) closed convex cone ($\lambda C \subseteq C\ \forall \lambda \geq 0$) $\rightarrow$ partial ordering "$\leqq_C$" on Y
- ∞_C - greatest element with respect to "$\leqq_C$", $\infty_C \notin Y$, denote $Y^\bullet = Y \cup \{\infty_C\}$
- we write $x \leq_C y$ if $x \leqq_C y$ and $x \neq y$
- $C^* = \{y^* \in Y^* : \langle y^*, y\rangle \geq 0\ \forall y \in C\}$ - *dual cone* to C

A vector function $F : X \rightarrow Y^\bullet$ is

- *proper*: $\operatorname{dom} F = \{x \in X : F(x) \in Y\} \neq \emptyset$
- *C-convex*: $F(tx + (1-t)y) \leqq_C tF(x) + (1-t)F(y)$ for all $x, y \in Y$ and all $t \in [0, 1]$
- *positively C-lsc*: $\langle z^*, F(\cdot)\rangle$ is lower semicontinuous for all $z^* \in C^* \setminus \{0\}$

Solution concepts

Consider the vector optimization problem

$$(P) \qquad \underset{x\in X}{\mathrm{WMin}}\, F(x),$$

where $F : X \to Y^\bullet$ proper and $\mathrm{int}\, C \neq \emptyset$. An $\bar{x} \in X$ is

- a *weakly efficient solution* to (P): $(\bar{x} - \mathrm{int}\, C) \cap F(X) = \emptyset$
  notation: $\bar{x} \in \mathcal{WE}(P)$
- an *efficient solution* to (P): $\nexists x \in X$ s.t. $F(x) \leq_C F(\bar{x})$
  notation: $\bar{x} \in \mathcal{E}(P)$

Proposition. F is C-convex $\Rightarrow$

$$\bar{x} \in \mathcal{WE}(P) \qquad \Leftrightarrow$$

$$\exists z^* \in C^* \setminus \{0\} \text{ s.t. } \langle z^*, F(\bar{x})\rangle \leq \langle z^*, F(x)\rangle\ \forall x \in X.$$

Example

Consider the vector optimization problem

$$(P) \qquad \underset{x_1,x_2\in\mathbb{R}}{\text{WMin}} \begin{pmatrix} x_1^2 - x_2 \\ x_2 \end{pmatrix},$$

where the vector-minimization is considered with respect to $\mathbb{R}^2_+$.

For all $\lambda = (\lambda_1, \lambda_2)^\top \in \mathbb{R}^2_+ \setminus \{0\}$ with $\lambda_1 \neq \lambda_2$, one has

$$\inf\{\lambda_1(x_1^2 - x_2) + \lambda_2 x_2 : x_1, x_2 \in \mathbb{R}\} = -\infty$$

$\Rightarrow$ one **cannot** identify the weakly efficient solutions of (P)

Only for $\bar{\lambda} = (\bar{\lambda}_1, \bar{\lambda}_2)^\top \in \mathbb{R}^2_+ \setminus \{0\}$ with $\bar{\lambda}_1 = \bar{\lambda}_2 > 0$ one gets

$\bar{\lambda}_1(\bar{x}_1^2 - \bar{x}_2) + \bar{\lambda}_2\bar{x}_2 \leq \bar{\lambda}_1(x_1^2 - x_2) + \bar{\lambda}_2 x_2 \; \forall x_1, x_2 \in \mathbb{R}$ for $\bar{x}_1 = 0, \bar{x}_2 \in \mathbb{R}$.

$\Rightarrow$ an unfortunate choice of the scalarization may lead nowhere

(even for simple vector optimization problems)

Iterative methods for solving multiobjective/vector optimization problems

- algorithms for delivering one (weakly) efficient solution
  - *Newton*: Fliege, Graña Drummond & Svaiter; Graña Drummond, Raupp & Svaiter
  - *steepest descent*: Fliege & Svaiter; Graña Drummond & Svaiter
  - *projected gradient*: Fukuda & Graña Drummond; Graña Drummond & Iusem
  - *proximal point*: Bonel, Iusem & Svaiter; Villacorta & Oliveira
- algorithms for finding (approximating) the whole (weakly) efficient set
  - *Benson type*: Löhne & Weißing
  - *adaptive scalarization*: Berger & Büskens

Problem formulation

Consider the vector optimization problem

$$(VP) \qquad \underset{x \in X}{\text{WMin}} \left[F(x) + G(x)\right],$$

where

- $F : X \to Y$ - Fréchet differentiable with an L-Lipschitz-continuous gradient ∇
- $G : X \to Y^\bullet$ - proper
- $\text{int}\, C \neq \emptyset$

Inertial forward-backward proximal algorithm

choose $x_0, x_1 \in X$

$\frac{1}{9} > \beta \geq \beta_n \geq 0 \ \forall n \geq 0$: $(\beta_n)_n$ is nondecreasing

$(z_n^*)_n \in C^* \setminus \{0\}$: $\|z_n^*\| = 1$

$e_n \in \operatorname{int} C \ \forall n \geq 0$: $\langle z_n^*, e_n \rangle = 1 \ \forall n \geq 0$

1. let $n = 1$
2. if $x_n \in \mathcal{WE}(VP)$: STOP
3. find $x_{n+1} \in \mathcal{WE}\left\{ G(x) + \frac{L}{2}\left\|x - (x_n + \beta_n(x_n - x_{n-1}) - \frac{1}{L}\nabla(z_n^* F)(x_n))\right\|^2 e_n : x \in \Omega_n \right\}$, where $\Omega_n = \{x \in X : (F + G)(x) \leqq_C (F + G)(x_n)\}$
4. let $n := n + 1$ and go to 2

Remarks

- $F \equiv 0$ & $\beta_n = 0\ \forall n \geq 0 \Rightarrow$ proximal point method [Bonel, Iusem & Svaiter]
- $Y = \mathbb{R}$ & $C = \mathbb{R}_+ \Rightarrow$ inertial forward-backward proximal point method [Moudafi & Oliny]
- $Y = \mathbb{R}$ & $C = \mathbb{R}_+$ & $F \equiv 0 \Rightarrow$ inertial proximal point method [Alvarez & Attouch]
- $Y = \mathbb{R}$ & $C = \mathbb{R}_+$ & $F \equiv 0$ & $\beta_n = 0\ \forall n \geq 0 \Rightarrow$ ISTA [Beck & Teboulle]
- it is not necessary to impose the existence of a weakly efficient solution to (P)
- any $z^* \in C^* \setminus \{0\}$ provides a good scalarization function for the vector optimization problems in Step 3
- under additional hypotheses (e.g. $\exists \delta > 0 : \{z^* \in Y^* : \langle z^*, x\rangle \geq \delta \|x\| \|z^*\| \forall x \in C\} \neq \emptyset$) the method delivers efficient solutions to (VP)

Convergence statement

If

- F *and* G *are* C*-convex*
- G *is positively* C*-lsc*
- $(F+G)(X) \cap (F(x_0)+G(x_0)-C)$ *is* C*-complete* *(i.e.* $\forall (a_n)_n \in X$ *with* $a_0 = x_0$ *s.t.* $(F+G)(a_{n+1}) \leqq_C (F+G)(a_n)\ \forall n \geq 0\ \exists a \in X$: $(F+G)(a) \leqq_C (F+G)(a_n)\ \forall n \geq 0$*)*

then $x_n \rightharpoonup \bar{x} \in \mathcal{WE}(VP)$.

Proof steps

- the strong convexity of the scalarized intermediate problems guarantees the existence of new iterates
- descent lemma for $(z^* F)$ (convex and Fréchet differentiable with an $L\|z^*\|$-Lipschitz continuous gradient $\forall z^* \in C^*$)
- existence of a weak cluster point of $(x_n)_n$ that is weakly efficient to (VP)
- Opial's Lemma guarantees the weak convergence of $(x_n)_n$ to a point of $\{x \in X : F(x) \leqq_C F(x_n) \ \forall n \geq 0\}$
- uniqueness of the weak cluster point of $(x_n)_n$
- **main challenges**
  - the necessity of using constrained intermediate vector optimization problems
  - the considered function changes at each step
  - we deal with two vector functions with different properties

Alternative hypotheses/stopping rule/inexact version

- imposing the condition (cf. [Alvares & Attouch])

$$\sum_{k=1}^{+\infty} \beta_k \|x_k - x_{k-1}\|^2 < +\infty$$

  $(\beta_n)_n$ needs not be nondecreasing and $\beta \in [0,1[$

- considering instead of

  2 if $x_n \in \mathcal{WE}(VP)$: STOP

  the following stopping rule

  2' if $x_{n+1} = x_n = x_{n-1}$: STOP

- or replacing 3 with

  3' find $x_{n+1} \in X$ such that $0 \in \partial_{\varepsilon_n}(\langle z_n^*, G(\cdot) + \frac{L}{2}\| \cdot - x_n - \beta_n(x_n - x_{n-1}) + \frac{1}{L}\nabla(z_n^* F)(x_n)\|^2 e_n\rangle + \delta_{\Omega_n}(\cdot))(x_{n+1})$

  with $(\varepsilon_n)_n$ fulfilling e.g. $\sum_{n\geq 1} \varepsilon_n < +\infty$,

the converge statement remains valid

Forward-backward proximal algorithm

choose $x_0 \in X$, $(z_n^*)_n \in C^* \setminus \{0\}$: $\|z_n^*\| = 1$
$e_n \in \operatorname{int} C \ \forall n \geq 0$: $\langle z_n^*, e_n \rangle = 1 \ \forall n \geq 0$

1 let $n = 1$
2 if $x_n \in \mathcal{WE}(VP)$: STOP
3 find
$x_{n+1} \in \mathcal{WE}\left\{ G(x) + \frac{L}{2}\left\|x - (x_n - \frac{1}{L}\nabla(z_n^* F)(x_n))\right\|^2 e_n : x \in \Omega_n \right\}$
4 let $n := n + 1$ and go to 2

If

- F *and* G *are* C*-convex*
- G *is positively* C*-lsc*
- $(F+G)(X) \cap (F(x_0) + G(x_0) - C)$ *is* C*-complete*
- $z_n^* = z^* \in C^* \setminus \{0\} \ \forall n \geq 1$

then for any $n \geq 0$ *and* $\tilde{x} \in \Omega = \cap_{n \geq 0} \Omega_n$ *one has*

$$\langle z^*, F(x_n) + G(x_n) - F(\tilde{x}) - G(\tilde{x}) \rangle \leq \frac{L\|\tilde{x} - x_0\|^2}{2n}.$$

Second inertial forward-backward proximal algorithm

choose $x_0, x_1 \in X$

$0 \leq \beta_n \leq \beta \in \mathbb{R}\ \forall n \geq 0$: $(\beta_n)_n$ is nondecreasing

$(z_n^*)_n \in C^* \setminus \{0\}$: $\|z_n^*\| = 1$

$e_n \in \operatorname{int} C\ \forall n \geq 0$: $\langle z_n^*, e_n \rangle = 1\ \forall n \geq 0$

1. let $n = 1$
2. if $x_n \in \mathcal{WE}(VP)$: STOP
3. find $x_{n+1} \in \mathcal{WE}\left\{G(x) + \frac{L}{2}\|x - (x_n + \beta_n(x_n - x_{n-1}) - \frac{1}{L}\nabla(z_n^* F)(x_n + \beta_n(x_n - x_{n-1})))\|^2 e_n : x \in \Omega_n\right\}$
4. let $n := n + 1$ and go to 2

If

- $\beta < \frac{1}{9}$
- *F and G are C-convex*
- *G is positively C-lsc*
- $(F + G)(X) \cap (F(x_0) + G(x_0) - C)$ *is C-complete*

then $x_n \rightharpoonup \bar{x} \in \mathcal{WE}(VP)$.

Convergence rate

If

- *F and G are C-convex*
- *G is positively C-lsc*
- *$(F+G)(X) \cap (F(x_0)+G(x_0)-C)$ is C-complete*
- $z_n^* = z^* \in C^* \setminus \{0\}\ \forall n \geq 1$
- $\beta_n = \frac{t_n - 1}{t_n}$, *where* $t_1 = 1,\ t_n = \frac{1+\sqrt{1+4t_n^2}}{2}\ \forall n \geq 0$

then for any $n \geq 0$ and $\tilde{x} \in \Omega$ one has

$$\langle z^*, F(x_n) + G(x_n) - F(\tilde{x}) - G(\tilde{x}) \rangle \leq 2\frac{L\|\tilde{x} - x_0\|^2}{(n+1)^2}.$$

Open problems

- derive convergence rates without taking $(z_n^*)_n$ constant
- alternative hypotheses to the C-completeness
- avoid using Ω_n without losing the convergence

Numerical experiments

The portfolio vector optimization problem

$$(EP) \qquad \underset{\substack{x=(x_1,\ldots,x_d)\in\mathbb{R}^d_+,\\ \sum_{i=1}^d x_i=1}}{\text{WMin}} \begin{pmatrix} -x^\top u \\ x^\top Vx \end{pmatrix},$$

where $u \in \mathbb{R}^d$ and $V \in \mathbb{R}^{d\times d}$ is symmetric positive semidefinite, can be recast as a special case of (VP) by taking $X = \mathbb{R}^d$, $Y = \mathbb{R}^2$, $C = \mathbb{R}^2_+$, $F(x) = (-x^\top u, x^\top Vx)^\top$ and $G(x) = (\delta_{\mathbb{R}^d_+\cap T}(x), \delta_{\mathbb{R}^d_+\cap T}(x))^\top$, where $T = \{x = (x^1, \ldots, x^d) \in \mathbb{R}^d_+ : \sum_{i=1}^d x^i = 1\}$.

- x - portfolio vector for d given assets
- short sales are excluded: $x \in \mathbb{R}^d_+$
- expected return: $x^\top u$
- variance of the portfolio: $x^\top Vx$

Matlab implementation

- real data collected in [Duan, 2007]
- stocks: IBM, Microsoft, Apple, Quest Diagnostics, and Bank of America, between 02/01/2002-02/01/2007
- $d = 5$, $u = (0.4, 0.513, 4.085, 1.006, 1.236)^\top$ and

$$V = \begin{pmatrix} 0.006461 & 0.002983 & 0.00235487 & 0.00235487 & 0.00096889 \\ 0.002983 & 0.0039 & 0.00095937 & -0.0001987 & 0.00063459 \\ 0.002355 & 0.000959 & 0.01267778 & 0.00135712 & 0.00134481 \\ 0.002355 & -0.0002 & 0.00135712 & 0.00559836 & 0.00041942 \\ 0.000969 & 0.000635 & 0.00134481 & 0.00041942 & 0.0016229 \end{pmatrix}$$

- $z_n^* = (1/\sqrt{2}, 1/\sqrt{2})^\top$, $e_n = (1,1)^\top \ \forall n \geq 0$
- $x_0 = (0.25, 0.25, 0, 0.25, 0.25)$, $x_1 = (0.15, 0.25, 0.25, 0.2, 0.15)$
- stopping rule: $\|x_{n+1} - x_n\| \leq \varepsilon = 0.00001 \geq \|x_n - x_{n-1}\|$
- the intermediate scalar problems solved with $\mathrm{fmincon}$ (interior point methods)
- approximate weakly efficient solution $\bar{x} = (0.00000015603, 0.0718, 0.3189, 0.1317, 0.4777)$

β_n	iterations	time (s)
0	625	33.288016
$1/10$	281	15.517354
$1/11$	579	31.590278
$1/12$	707	38.557769
$1/15$	603	32.456625
$1/20$	507	28.100981
$1/30$	571	30.952835
$1/100$	813	43.925126
$1/10 - 1/10n$	13	1.080089
$1/10 - 1/(n+10)$	117	11.592547
$1/15 - 1/n + 15$	117	11.692819
$1/20 - 1/(n+20)$	15	1.721906
$1/30 - 1/(n+30)$	13	0.988701
$1/50 - 1/(n+50)$	15	1.245741