# Splitting methods for decomposing separable convex programs

Philippe Mahey
LIMOS - ISIMA - Université Blaise Pascal
PGMO, ENSTA 2013

October 4, 2013

# Plan

# Plan

# Proximal steps

$$x^* = \operatorname*{Argmin}_{x\in X} f(x) \iff 0 \in T(x^*)$$

where $f$ is closed convex so that $T = \partial f$ is maximal monotone.
Forward step (subgradient method) : $x^{t+1} \in (I - \lambda T)(x^t)$
Resolvent and proximal iteration (Backward step) :

- $J_\lambda = (I + \lambda T)^{-1}$ is the resolvent operator
- $x^* \in x^* + \lambda T(x^*) \implies x^* = J_\lambda(x^*)$
- The resolvent is firmly nonexpansive
- $x^{t+1} = J_\lambda(x^t) \iff x^{t+1} = \text{Argmin}_{x\in X} f(x) + \frac{1}{2\lambda}\|x - x^t\|^2$

Advantages and drawbacks :

- Implementable when $f$ is the dual function of a relaxed model $\rightarrow$ Augmented Lagrangian method
- $\lambda > 0$ is the implicit step size : smaller is sharper but numerically unstable
- Quadratic terms are not separable

# Proximal steps

$$x^* = \underset{x \in X}{\text{Argmin}} f(x) \iff 0 \in T(x^*)$$

where $f$ is closed convex so that $T = \partial f$ is maximal monotone.
Forward step (subgradient method) : $x^{t+1} \in (I - \lambda T)(x^t)$
Resolvent and proximal iteration (Backward step) :

- $J_\lambda = (I + \lambda T)^{-1}$ is the resolvent operator
- $x^* \in x^* + \lambda T(x^*) \implies x^* = J_\lambda(x^*)$
- The resolvent is firmly nonexpansive
- $x^{t+1} = J_\lambda(x^t) \iff x^{t+1} = \text{Argmin}_{x \in X} f(x) + \frac{1}{2\lambda} \|x - x^t\|^2$

Advantages and drawbacks :

- Implementable when $f$ is the dual function of a relaxed model $\rightarrow$ Augmented Lagrangian method
- $\lambda > 0$ is the implicit step size : smaller is sharper but numerically unstable
- Quadratic terms are not separable

# Sum of two monotone operators

If $A$ and $B$ are two maximal monotone operators, $A+B$ is monotone but may not be maximal

$$x^* = \operatorname*{Argmin}_{x \in X} f_1(x) + f_2(x) \iff 0 \in (A+B)(x^*)$$

where $f_i$ are closed convex functions with respective subdifferential operators $A$ and $B$

- Compute separately forward or backward steps on $A$ or $B$ but not on $A+B$
- Reformulation : $y \in A(x^*)$ and $-y \in B(x^*)$

Reformulation with coupling subspace :

- $$x^* = \operatorname*{Argmin}_{(x_1,x_2) \in X \times X} f_1(x_1) + f_2(x_2) \qquad \text{subject to } x_1 = x_2$$
- Copying variables yields the coupling subspace $\mathcal{A} = \{(x_1, x_2) \in X \times X \mid x_1 = x_2\}$
- Observe that, in $X \times X$, $\partial(f_1 + f_2) = \partial f_1 + \partial f_2$
- Splitting will now decouple $A = \partial(f_1) + \partial(f_2)$ from $B = (\mathcal{A})^{\perp}$

# Sum of two monotone operators

If $A$ and $B$ are two maximal monotone operators, $A+B$ is monotone but may not be maximal

$$x^* = \operatorname*{Argmin}_{x\in X} f_1(x) + f_2(x) \Longleftrightarrow 0 \in (A+B)(x^*)$$

where $f_i$ are closed convex functions with respective subdifferential operators $A$ and $B$

- Compute separately forward or backward steps on $A$ or $B$ but not on $A+B$
- Reformulation : $y \in A(x^*)$ and $-y \in B(x^*)$

Reformulation with coupling subspace :

- $$x^* = \operatorname*{Argmin}_{(x_1,x_2)\in X\times X} f_1(x_1) + f_2(x_2) \qquad \text{subject to } x_1 = x_2$$
- Copying variables yields the coupling subspace $\mathcal{A} = \{(x_1,x_2) \in X \times X \mid x_1 = x_2\}$
- Observe that, in $X \times X$, $\partial(f_1 + f_2) = \partial f_1 + \partial f_2$
- Splitting will now decouple $A = \partial(f_1) + \partial(f_2)$ from $B = (\mathcal{A})^{\perp}$

# Proximal decomposition on the sum of operators

$$x^* = \operatorname{Argmin} F(x) \qquad \text{subject to } x \in \mathcal{A}$$

Optimality condition :

$$(x^*, y^*) \in \mathcal{A} \times \mathcal{A}^\perp \qquad \text{such that} \qquad y^* \in T(x^*)$$

$$x^* + \lambda y^* \in (I + \lambda T)(x^*) \Longrightarrow x^* = (I + \lambda T)^{-1}(x^* + \lambda y^*)$$

Prox Decomp iteration

$$x^{t+1} = (I + \lambda T)^{-1}(x^t + \lambda y^t)$$
$$y^{t+1} = \lambda^{-1}(x^t + \lambda y^t - x^{t+1})$$

# Proximal decomposition on the sum of operators

$$x^* = \operatorname{Argmin} F(x) \qquad \text{subject to } x \in \mathcal{A}$$

Optimality condition :

$$(x^*, y^*) \in \mathcal{A} \times \mathcal{A}^\perp \qquad \text{such that} \qquad y^* \in T(x^*)$$

$$x^* + \lambda y^* \in (I + \lambda T)(x^*) \Longrightarrow x^* = (I + \lambda T)^{-1}(x^* + \lambda y^*)$$

Prox Decomp iteration

$$x^{t+1} = (I + \lambda T)^{-1}(x^t + \lambda y^t)$$
$$y^{t+1} = \lambda^{-1}(x^t + \lambda y^t - x^{t+1})$$

# Proximal decomposition on the sum of operators

$$x^* = \operatorname{Argmin} F(x) \qquad \text{subject to } x \in \mathcal{A}$$

Optimality condition :

$$(x^*, y^*) \in \mathcal{A} \times \mathcal{A}^\perp \qquad \text{such that} \qquad y^* \in T(x^*)$$

$$x^* + \lambda y^* \in (I + \lambda T)(x^*) \Longrightarrow x^* = (I + \lambda T)^{-1}(x^* + \lambda y^*)$$

Prox Decomp iteration

$$\begin{aligned} x^{t+1} &= (I + \lambda T)^{-1}(x^t + \lambda y^t) \\ y^{t+1} &= \lambda^{-1}(x^t + \lambda y^t - x^{t+1}) \end{aligned}$$

# Generalized reflections

Step size parameters are omitted below

- Generalized reflections : for $T$ maximal monotone, $N = (I+T)^{-1} - (I+T^{-1})^{-1} = 2(I+T)^{-1} - I$
- Generalized reflections are nonexpansive

$u \in Ty \iff d = N(s)$ with

$$\begin{cases} s &= y + u \\ d &= y - u \end{cases} \quad \text{or} \quad \begin{cases} y &= \frac{1}{2}(d+s) \\ u &= \frac{1}{2}(d-s) \end{cases}$$

# Generalized reflections

Step size parameters are omitted below

- Generalized reflections : for $T$ maximal monotone,
  $N = (I+T)^{-1} - (I+T^{-1})^{-1} = 2(I+T)^{-1} - I$
- Generalized reflections are nonexpansive

$u \in Ty \iff d = N(s)$ with

$$\begin{cases} s &= y+u \\ d &= y-u \end{cases} \quad \text{or} \quad \begin{cases} y &= \frac{1}{2}(d+s) \\ u &= \frac{1}{2}(d-s) \end{cases}$$

# Splitting iterations

- Forward-Backward : $x^{t+1} \in (I+B)^{-1}(I-A)(x^t)$ (ergodic convergence)
- Double-Backward : $x^{t+1} = (I+B)^{-1}(I+A)^{-1}(x^t)$ (ergodic convergence)
- Peaceman-Rachford : $s^{t+1} = (N_B \circ N_A)(s^t)$ with $s = x + y$ (convergence not guaranteed)
- Douglas-Rachford : $s^{t+1} = [\frac{1}{2}I + \frac{1}{2}(N_B \circ N_A)](s^t)$ (linear convergence : see Proximal Decomposition)

Introduction of a relaxation parameter :

$$\mathcal{J}_{\alpha^t} = (1-\alpha^t)I + \alpha^t N_B \circ N_A$$

Peaceman-Rachford for $\alpha = 1$
Douglas-Rachford for $\alpha = 1/2$

# Splitting iterations

- Forward-Backward : $x^{t+1} \in (I+B)^{-1}(I-A)(x^t)$ (ergodic convergence)
- Double-Backward : $x^{t+1} = (I+B)^{-1}(I+A)^{-1}(x^t)$ (ergodic convergence)
- Peaceman-Rachford : $s^{t+1} = (N_B \circ N_A)(s^t)$ with $s = x + y$ (convergence not guaranteed)
- Douglas-Rachford : $s^{t+1} = [\frac{1}{2}I + \frac{1}{2}(N_B \circ N_A)](s^t)$ (linear convergence : see Proximal Decomposition)

Introduction of a relaxation parameter :

$$\mathcal{J}_{\alpha^t} = (1-\alpha^t)I + \alpha^t N_B \circ N_A$$

Peaceman-Rachford for $\alpha = 1$
Douglas-Rachford for $\alpha = 1/2$

# Extensions and convergence results

- Inexact subproblem computations
- Conditions for linear rates of convergence
- Generalized frameworks beyond Peaceman-Rachford
- Projective techniques (Eckstein and Svaiter M.Prog. 2008)

Introduction of scaling parameters :

Scaled operators : $0 \in \mu A(x) + \mu B(x)$

Scaled norm : $\| x \|_H, H \succ 0$

# Extensions and convergence results

- Inexact subproblem computations
- Conditions for linear rates of convergence
- Generalized frameworks beyond Peaceman-Rachford
- Projective techniques (Eckstein and Svaiter M.Prog. 2008)

Introduction of scaling parameters :

$$\text{Scaled operators} : 0 \in \mu A(x) + \mu B(x)$$
$$\text{Scaled norm} : \| x\|_H, H \succ 0$$

# Plan

# Symmetric duality

$$\underset{(x_1,\ldots,x_p)}{\text{Minimize}} \quad \sum_{i=1}^{p} f_i(x_i)$$
$$\sum_{i=1}^{p} g_i(x_i) = 0$$

Primal decomposition : $(y_1, \ldots, y_p) \in \mathcal{A} = \{\sum_i y_i = 0\}$
Subproblems :

$$\underset{x_i}{\text{Minimize}} \quad f_i(x_i)$$
$$g_i(x_i) = y_i$$

Dual decomposition : $(u_1, \ldots, u_p) \in \mathcal{A}^\perp = \{u_1 = \cdots = u_p\}$
Subproblems :

$$\underset{x_i}{\text{Minimize}} \quad f_i(x_i) + \langle u_i, g_i(x_i)\rangle$$

# Symmetric duality

$$\underset{(x_1,\dots,x_p)}{\text{Minimize}} \quad \sum_{i=1}^{p} f_i(x_i)$$
$$\sum_{i=1}^{p} g_i(x_i) = 0$$

Primal decomposition : $(y_1, \dots, y_p) \in \mathcal{A} = \{\sum_i y_i = 0\}$
Subproblems :

$$\underset{x_i}{\text{Minimize}} \quad f_i(x_i)$$
$$g_i(x_i) = y_i$$

Dual decomposition : $(u_1, \dots, u_p) \in \mathcal{A}^\perp = \{u_1 = \cdots = u_p\}$
Subproblems :

$$\underset{x_i}{\text{Minimize}} \quad f_i(x_i) + \langle u_i, g_i(x_i)\rangle$$

# Symmetric duality

$$\underset{(x_1,\ldots,x_p)}{\text{Minimize}} \quad \sum_{i=1}^{p} f_i(x_i)$$
$$\sum_{i=1}^{p} g_i(x_i) = 0$$

Primal decomposition : $(y_1, \ldots, y_p) \in \mathcal{A} = \{\sum_i y_i = 0\}$
Subproblems :

$$\underset{x_i}{\text{Minimize}} \quad f_i(x_i)$$
$$g_i(x_i) = y_i$$

Dual decomposition : $(u_1, \ldots, u_p) \in \mathcal{A}^\perp = \{u_1 = \cdots = u_p\}$
Subproblems :

$$\underset{x_i}{\text{Minimize}} \quad f_i(x_i) + \langle u_i, g_i(x_i) \rangle$$

# Augmented Lagrangian

Back to the constrained optimization problem :

$$\underset{x_i}{\text{Minimize}} \qquad \sum_{i=1}^{p} f_i(x_i)$$
$$\sum_{i=1}^{p} g_i(x_i) = 0$$

The Augmented Lagrangian function associated with coupling constraints is not separable.
First duplicate local variables in the coupling constraints

$$\underset{x_i}{\text{Minimize}} \qquad \sum_{i=1}^{p} f_i(x_i)$$
$$\forall i = 1, \ldots p \qquad g_i(x_i) = y_i; \qquad \sum_{i=1}^{p} y_i = 0$$

and dualize the coupling subspace

# Augmented Lagrangian

Back to the constrained optimization problem :

$$\underset{x_i}{\text{Minimize}} \quad \sum_{i=1}^{p} f_i(x_i)$$
$$\sum_{i=1}^{p} g_i(x_i) = 0$$

The Augmented Lagrangian function associated with coupling constraints is not separable.
First duplicate local variables in the coupling constraints

$$\underset{x_i}{\text{Minimize}} \quad \sum_{i=1}^{p} f_i(x_i)$$
$$\forall i = 1, \dots p \qquad g_i(x_i) = y_i; \qquad \sum_{i=1}^{p} y_i = 0$$

and dualize the coupling subspace

# Separable Augmented Lagrangian

Use alternate computations on the separable piece of the Lagrangian and on the coupling subspace :

$$L_\lambda(x,y;u) = \sum_{i=1}^{p} \left\{ f_i(x_i) - \langle u_i, g_i(x_i)\rangle + \frac{\lambda}{2}\|g_i(x_i) - y_i\|^2 \right\} \text{ s.c } \sum_{i=1}^{p} y_i = 0$$

- Augmented Lagrangian algorithm

$$(x^{k+1}, y^{k+1}) = \arg\min L_\lambda(x,y;u^k) \qquad \text{non-separable}$$
$$u^{k+1} = u^k - \lambda(g(x^{k+1}) - y^{k+1})$$

- Separable Augmented Lagrangian

$$x^{k+1} = \arg\min L_\lambda(x, y^k; u^k) \qquad \text{separable}$$
$$y^{k+1} = \arg\min L_\lambda(x^{k+1}, y; u^k) \qquad \text{non-separable but easy}$$
$$u^{k+1} = u^k - \lambda(g(x^{k+1}) - y^{k+1})$$

# Separable Augmented Lagrangian

Use alternate computations on the separable piece of the Lagrangian and on the coupling subspace :

$$L_\lambda(x, y; u) = \sum_{i=1}^{p} \left\{ f_i(x_i) - \langle u_i, g_i(x_i) \rangle + \frac{\lambda}{2} \|g_i(x_i) - y_i\|^2 \right\} \text{ s.c } \sum_{i=1}^{p} y_i = 0$$

- Augmented Lagrangian algorithm

$$(x^{k+1}, y^{k+1}) = \arg\min L_\lambda(x, y; u^k) \qquad \text{non-separable}$$
$$u^{k+1} = u^k - \lambda(g(x^{k+1}) - y^{k+1})$$

- Separable Augmented Lagrangian

$$x^{k+1} = \arg\min L_\lambda(x, y^k; u^k) \qquad \text{separable}$$
$$y^{k+1} = \arg\min L_\lambda(x^{k+1}, y; u^k) \qquad \text{non-separable but easy}$$
$$u^{k+1} = u^k - \lambda(g(x^{k+1}) - y^{k+1})$$

# Separable Augmented Lagrangian

Use alternate computations on the separable piece of the Lagrangian and on the coupling subspace :

$$L_\lambda(x,y;u) = \sum_{i=1}^{p} \left\{ f_i(x_i) - \langle u_i, g_i(x_i) \rangle + \frac{\lambda}{2} \|g_i(x_i) - y_i\|^2 \right\} \text{ s.c} \sum_{i=1}^{p} y_i = 0$$

- Augmented Lagrangian algorithm

$$\begin{aligned} (x^{k+1}, y^{k+1}) &= \arg\min L_\lambda(x,y;u^k) && \text{non-separable} \\ u^{k+1} &= u^k - \lambda(g(x^{k+1}) - y^{k+1}) \end{aligned}$$

- Separable Augmented Lagrangian

$$\begin{aligned} x^{k+1} &= \arg\min L_\lambda(x,y^k;u^k) && \text{separable} \\ y^{k+1} &= \arg\min L_\lambda(x^{k+1},y;u^k) && \text{non-separable but easy} \\ u^{k+1} &= u^k - \lambda(g(x^{k+1}) - y^{k+1}) \end{aligned}$$

# Global scaling

Separable Augmented Lagrangian algorithm :

$$\Lambda = \text{Diag}\{\Lambda_1, \ldots, \Lambda_p\}$$

$$x_i^{k+1} = \arg\min f_i(x_i) - \langle u_i^k, g_i(x_i)\rangle + \frac{1}{2}\|g_i(x_i) - y_i^k\|^2_{\Lambda_i}$$

$$r^{k+1} = \sum_{i=1}^{p} g_i(x_i)^{k+1}; \qquad y_i^{k+1} = g_i(x_i^{k+1}) - \Lambda_i^{-1}(\sum_{j=1}^{p}\Lambda_j^{-1})^{-1} r^{k+1}$$

$$u_i^{k+1} = u_i^k - (\sum_{j=1}^{p}\Lambda_j^{-1})^{-1} r^{k+1}$$

# Variable scaling

**Proposition**

*Under assumptions :*

**a)** $\Lambda_k$ *is positive definite for all* $k$

**b)** $\Lambda_k \to \Lambda_\infty$ *positive definite*

**c)** $\sum_{k=0}^{+\infty} \|\Lambda_{k+1} - \Lambda_k\| < +\infty$

*proximal decomposition with parameters* $\Lambda_k$ *converges to a solution.*

- Adaptive scaling : $\Lambda_{k+1} = (1 - \beta_k)\Lambda_k + \beta_k \Gamma_k$
- $\gamma_i^{k+1} = \frac{\|u_i^{k+1} - u_i^k\|}{\|u_i^{k+1} - u_i^k\|}$

# Plan

# References


- Decomposition techniques : A. Geoffrion (1970), L. Lasdon (1970), D. Gabay (1983), G. Cohen (1980)
- Operator splitting methods : H. Brezis (1973), P.L. Lions and B. Mercier (1979), Bauschke and Combettes (2003)
- Proximal method of multipliers : R.T. Rockafellar (1976), Teboulle (1994)
- Proximal decomposition : J. Eckstein (1989), PM et al (1995)
- Extensions : A. Ouorou (1995), A. Lenoir (2008), O. Gueye, J.P. Dussault and PM (2005)