J. Appl. Math. & Informatics Vol. **28***(2010), No. 5 - 6, pp. 1249-1262*
Website: http://www.kcam.biz

AN APPROXIMATE ALTERNATING LINEARIZATION DECOMPOSITION METHOD†

DAN LI, LI-PING PANG* AND ZUN-QUAN XIA

Abstract. An approximate alternating linearization decomposition method, for minimizing the sum of two convex functions with some separable structures, is presented in this paper. It can be viewed as an extension of the method with exact solutions proposed by Kiwiel, Rosa and Ruszczynski(1999). In this paper we use inexact optimal solutions instead of the exact ones that are not easily computed to construct the linear models and get the inexact solutions of both subproblems, and also we prove that the inexact optimal solution tends to proximal point, i. e., the inexact optimal solution tends to optimal solution.

AMS Mathematics Subject Classification : 90C33, 90C30.
Key words and phrases : convex programming, decomposition method, proximal point method, approximate subdifferential.

1. Introduction

The decomposition method is applied frequently for solving optimization problems with some separable structures and also can be used to convert a complicated and large scale problem to simpler and smaller subproblems.

The first decomposition scheme was proposed by Dantzig and Wolfe for solving primal block-angular structured linear problems, see [3]. A similar one is Benders' Decomposition which is applying Dantzig-Wolfe decomposition to the dual problems for solving mixed integer programming problems, see [1]. Many researchers are interested in the progress of the decomposition method and it is developped rapidly, for example, there are the augmented Lagrangian decomposition, see [15] for instance, operator splitting methods, see [11, 17, 18, 19, 4, 12] for instance, and alternating direction methods, see [8, 2, 7, 5, 10] for instance. Recently, decomposition methods plays an active role in stochastic programming problems since stochastic programming problems have very large dimension and

Received November 21, 2009. Revised November 26, 2009. Accepted December 18, 2009.
*Corresponding author. †This work was supported by the NSFC, Grant 10771026; the Foundation of DUT, MXDUT73008; the Foundation of DUT, MXDUT98009.

characteristic structures, see [16]. The augmented Lagrangian decomposition method is equivalent to the proximal point method applied to the dual problem and is a useful tool for solving convex multistage stochastic programs. The convergence mechanism of the operator splitting method is rather involved and non-intuitive, however, it is difficult to monitor the progress of these methods. The alternating direction methods has been proved to be an efficient tool for solving convex multistages stochastic programming [10].

Many decomposition methods are explicitly or implicitly derived from the proximal point algorithm which draws on a large volume of prior work by various authors for maximal monotone operator, see [14]. The proximal point method for convex program in exact form generates a sequence $\{z^k\}$:

$$z^{k+1} = \arg\min_{z} f(z) + \frac{1}{2}\mu_k \parallel z - z^k \parallel^2,$$

where $\mu_k > 0$ for a convex and closed function f. In this algorithm exact minimization at each iteration is weakened and the subdifferential ∂f is replaced by an arbitrary maximal monotone operator T. However, this method is often impractical since the exact iteration in many cases requires a computation which is difficult as the same as solving the original problem $0 \in T(z)$, see [6].

In this paper we propose a decomposition method to minimize the sum of two convex functions with the following form

$$\min_{x} F(x) = h(x) + f(x), \tag{1}$$

where $h : R^n \to (-\infty, +\infty]$ and $f : R^n \to R$ are closed proper convex functions.

In alternating linearization method, see [10], each iteration involves solving two subproblems of the following form:

$$\begin{array}{ll} h\text{-subproblem} & \min_{x} h(x) + \widetilde{f}_k(x) + \frac{1}{2}\mu_k \parallel x - x^k \parallel^2 \\ f\text{-subproblem} & \min_{x} \widetilde{h}_k(x) + f(x) + \frac{1}{2}\mu_k \parallel x - x^k \parallel^2 . \end{array}$$

where $\widetilde{f}_k$ and $\widetilde{h}_k$ are linearization of f and h respectively. However, the construction of $\widetilde{f}_k$ and $\widetilde{h}_k$ need the information of ∂f and ∂h, and generally it is difficult to calculate the subdifferentials of f and h. Moreover, minimizing the subproblems and compute the exact solutions are also difficult to be implemented in practice.

For these reasons one try to use the elements of approximate subdifferentials to construct $\widetilde{f}_k$ and $\widetilde{h}_k$ and at the same time calculate the approximate solution of both subproblems instead of exact ones in the algorithms. The reason of employing approximate subgradients at a given point is that when a subgradient $g_f(x) \in \partial f(x)$ is expensive to compute then one may take an already computed subgradient $g_f(x)$ of f at some $\widetilde{x}$ near x, thus $g_f(\widetilde{x}) \in \partial_{\varepsilon_f} f(x)$ with

$$0 \leq \varepsilon_f = f(x) - f(\widetilde{x}) - \langle g_f(\widetilde{x}), x - \widetilde{x}\rangle.$$

In this paper, we use ε_k-subgradients and inexact solutions to construct $\widetilde{f}_k$ and $\widetilde{h}_k$ in order to make the alternating linearization decomposition algorithm more implementable and more applicable.

The paper is organized as follows. In Section 2, we outline briefly the approximate alternating linearization decomposition method and how the method convert to two subproblems. In Section 3, we show the implementable algorithm by approximate proximal decomposition via alternating linearization. According to the properties of strongly convexity we complete the convergence analysis of this algorithm in Section 4.

2. Approximate Alternating Linearization Decomposition Algorithm.

It is impossible in general to evaluate exactly the *proximal point* of objective function F

$$p(\bar{x}) = \arg\min_x h(x) + f(x) + w(x)$$

where $w(x) = \frac{1}{2}\mu \parallel x - \bar{x} \parallel^2$, $\mu > 0$ is fixed and $\bar{x} \in R^n$ is the *proximal center.*

For implementation we shall make use of ε-approximate subdifferentials to compute the approximation of the proximal point. Specially, suppose that, for each $\bar{x} \in R^n$ and for each $\varepsilon_k > 0$, we can find approximate proximal points $p_h(\bar{x}, \varepsilon_k)$, $p_f(\bar{x}, \varepsilon_k) \in R^n$ to the unique minimizer $p(\bar{x})$ such that $p_h(\bar{x}, \varepsilon_k) = z_h^k$ and $p_f(\bar{x}, \varepsilon_k) = z_f^k$ are the ε_k-approximate solutions of h-subproblem and f-subproblem respectively. We denote the ε_k-approximate solutions of the two subproblems by ε_k- $\arg\min_x h(x) + \widetilde{f}_k(x) + w(x)$ and ε_k- $\arg\min_x \widetilde{h}_k(x) + f(x) + w(x)$. The following notations will be used in the rest of this section.

Notations

k	Iteration counter
μ	Prox center
$w(\cdot)$	$w(x) = \frac{1}{2}\mu \parallel x - \bar{x} \parallel^2$
h-subproblem	$\min_x h(x) + \widetilde{f}_k(x) + w(x)$, where $\widetilde{f}_k$ is a linear model of f
f-subproblem	$\min_x \widetilde{h}_k(x) + f(x) + w(x)$, where $\widetilde{h}_k$ is a linear model of h
z_h^k	An ε_k-approximate solution of h-subproblem
z_f^k	An ε_k-approximate solution of f-subproblem
$\widetilde{f}_k$	$\widetilde{f}_k(\cdot) = f(z_f^{k-1}) + \langle g_f^{k-1}, \cdot - z_f^{k-1} \rangle$, where $g_f^0 \in \partial_{\varepsilon_0} f(z_f^0), z_f^0 \in R^n$ and $g_f^{k-1} = -g_h^{k-1} - \mu(z_f^{k-1} - \bar{x})$
$\widetilde{h}_k$	$\widetilde{h}_k(\cdot) = h(z_h^{k-1}) + \langle g_h^k, \cdot - z_h^k \rangle$, where $g_h^k = -g_f^{k-1} - \mu(z_h^k - \bar{x})$

Next, we present the structure of approximate alternating linearization decomposition algorithm.

Algorithm I: approximate alternating linearization decomposition algorithm.

Step 0: `Initiation`
Let $z_f^0 \in R^n$, $g_f^0 \in \partial_{\varepsilon_0} f(z_f^0)$, $\varepsilon_0 \in (0,1)$ and set $k=1$.

Step 1: `Solving the h-subproblem`
Compute $z_h^k = p_h(\bar{x}, \varepsilon_k)$ and $g_h^k = -g_f^{k-1} - \mu(z_h^k - \bar{x})$.

Step 2: `Solving the f-subproblem`
Compute $z_f^k = p_f(\bar{x}, \varepsilon_k)$ and $g_f^k = -g_h^k - \mu(z_f^k - \bar{x})$.

Step 3: `Update` ε_k and k
Let $\varepsilon_k = \gamma\varepsilon_{k-1}$, $\gamma \in (0,1)$. 2 Set $k = k+1$ and loop at Step 1.

End of the algorithm

Obviously the difference between the algorithm given above and the one given in [10] is that we use inexact solutions in iteration but not exact ones here. In this section the main work is to prove that $p_h(\bar{x}, \varepsilon_k) \to p(\bar{x})$ as $k \to \infty$.

Remarks

(i) When $\varepsilon_k = 0$, the Algorithm I is just the alternating linearization algorithm given in [10].

(ii) Suppose f is convex and $g_f \in \partial_\varepsilon f(x)$. Then one has

$$f(z) \geq f(x) + \langle g_f, z - x \rangle - \varepsilon,$$

and x^*, an $\varepsilon-$approximate solution of f, is an optimal solution if and only if $0 \in \partial_\varepsilon f(x^*)$ in other words for any x we have $f(x) \geq f(x^*) + \langle 0, x - x^* \rangle - \varepsilon$, i. e. $f(x) \geq f(x^*) - \varepsilon$, $\forall x$. It can be written by $x^* \in \varepsilon\text{-}\arg\min f(x)$.

(iii) The necessary and sufficient condition of optimality for the approximate solutions of h-subproblem has the form

$$0 \in \partial\varepsilon_k h(z_h^k) + g_f^{k-1} + \mu(z_h^k - \bar{x})$$

and $g_h^k = -g_f^{k-1} - \mu(z_h^k - \bar{x}) \in \partial_{\varepsilon_k} h(z_h^k)$, so the vector g_h^k is one element of $\partial_{\varepsilon_k} h(z_h^k)$. Hence $\widetilde{h}_k \leq h + \varepsilon_k$ by the ε_k-subgradient inequality.

Similarly, the vector $g_f^k = -g_h^k - \mu(z_f^k - \bar{x})$ is the element of $\partial_{\varepsilon_k} f(x)$ that satisfies the optimality condition for the approximate solutions of f-subproblem:

$$0 \in g_h^k + \partial_{\varepsilon_k} f(z_f^k) + \mu(z_f^k - \bar{x}).$$

Therefore, $\widetilde{f}_{k+1} \leq f + \varepsilon_k$.

(iv) For convenient we denote $\widetilde{F}_k := h + \widetilde{f}_k$, $\breve{F}_k = \widetilde{h}_k + f$, and $\overline{F}_k = \widetilde{h}_k + \widetilde{f}_k$. By the construction of linear model and approximate subgradient inequality we have

$$\widetilde{F}_k \leq F + \varepsilon_k, \quad \breve{F}_k \leq F + \varepsilon_k, \quad \overline{F}_k \leq F + 2\varepsilon_k.$$

Hence, $\widetilde{F}_k$, $\breve{F}_k$, and $\overline{F}_k$ are lower approximation of the objective function $F = h + f$ as $k \to \infty$ and $\lim\limits_{k\to\infty} \varepsilon_k = 0$.

Let ξ^k and $\xi^{k+1/2}$ denote

$$\xi^k = h(z_h^k) + \widetilde{f}_k(z_h^k) + w(z_h^k), \;\; \xi^{k+1/2} = \widetilde{h}_k(z_f^k) + f(z_f^k) + w(z_f^k),$$

where z_h^k and z_f^k are the ε_k-approximate solutions of h-subproblem and f-subproblem respectively.

The sequences $\{\xi^k\}$ and $\{\xi^{k+1/2}\}$, which are generated by this approximate alternating linearization decomposition algorithm are different from the sequence $\{\eta_k\}$ that is monotone, see [10]. Obviously, the sequences can't be nondecreasing.

Lemma 1. *The sequences $\{\xi^k\}$ and $\{\xi^{k+1/2}\}$ have the property that*

$$\xi^k \;\le \xi^{k+1/2} + \varepsilon_k \le \xi^{k+1} + 2\varepsilon_k$$

Proof. Since $z_h^k \in \varepsilon_k\text{-}\arg\min h(x) + \widetilde{f}_k(x) + w(x)$, the optimality condition of ε_k-approximate solutions

$$0 \in \partial_{\varepsilon_k} h(z_h^k) + g_f^{k-1} + \mu(z_h^k - \bar{x})$$

holds. Then it follows that there exists $g_h^k \in \partial_{\varepsilon_k} h(z_h^k)$ such that $0 = g_h^k + g_f^{k-1} + \mu(z_h^k - \bar{x})$. By construction of $\widetilde{h}_k$ we have $\nabla\widetilde{h}_k(z_h^k) = g_h^k$ and $h(z_h^k) = \widetilde{h}_k(z_h^k)$. Hence

$$0 = \nabla\widetilde{h}_k(z_h^k) + \nabla\widetilde{f}_k(z_h^k) + \mu(z_h^k - \bar{x}),$$

and so we have $z_h^k = \arg\min_x \widetilde{h}_k(x) + \widetilde{f}_k(x) + w(x)$ and

$$\xi^k \le \widetilde{h}_k(x) + \widetilde{f}_k(x) + w(x) \le \widetilde{h}_k(x) + f(x) + w(x) + \varepsilon_k.$$

It leads to

$$\xi^k \le \widetilde{h}_k(z_f^k) + f(z_f^k) + w(z_f^k) + \varepsilon_k,$$

i. e., $\xi^k \le \xi^{k+1/2} + \varepsilon_k$.

Similarly, $z_f^k \in \varepsilon_k\text{-}\arg\min \widetilde{h}_k + f(x) + w(x)$ because there exists $g_f^k = \nabla\widetilde{f}_{k+1}(z_f^k) \in \partial_{\varepsilon_k} f(z_f^k)$ such that $0 = \nabla\widetilde{h}_k(z_f^k) + \nabla\widetilde{f}_{k+1}(z_f^k) + \mu(z_f^k - \bar{x})$, one has

$$\xi^{k+1/2} = \widetilde{h}_k(z_f^k) + \widetilde{f}_{k+1}(z_f^k) + w(z_f^k) \le h(x) + \widetilde{f}_{k+1}(x) + w(x) + \varepsilon_k.$$

Therefore, $\xi^{k+1/2} \le h(z_h^{k+1}) + \widetilde{f}_{k+1}(z_h^{k+1}) + w(z_h^{k+1}) + \varepsilon_k = \xi^{k+1} + \varepsilon_k$. □

For estimating $\xi^{k+1/2} - \xi^k$ following the structure of the family of relaxations of f-subproblem at iteration k given in [10] we construct $G_k(x, \lambda)$ as follows,

$$\min_x G_k(x, \lambda) = \widetilde{h}_k(x) + (1-\lambda)\widetilde{f}_k(x) + \lambda\widehat{f}_k(x) + w(x), \tag{2}$$

where $\lambda \in [0, 1]$ and $\widehat{f}_k(x) = f(z_h^k) + \langle g_f(z_h^k), x - z_h^k\rangle$ for $g_f(z_h^k) \in \partial_{\varepsilon_k} f(z_h^k)$. According to the ε_k-subgradient inequality, one has $\widehat{f} \le f + \varepsilon_k$. Since $\widetilde{f}_k$ and $\widehat{f}_k$ are lower approximations of f as $k \to \infty$, (2) is a relaxation of f-subproblem for all $\lambda \in [0, 1]$.

Denoting the optimal value of (2) by $\hat{G}_k(\lambda)$, we have

$$\max_{\lambda\in[0,1]} \hat{G}_k(\lambda) - \hat{G}_k(0) \le \xi^{k+1/2} - \xi^k + \varepsilon_k.$$

In fact,

$$\begin{aligned} G_k(x,\lambda) &\le \widetilde{h}_k(x) + (1-\lambda)[f(x)+\varepsilon_k] + \lambda[f(x)+\varepsilon_k] + w(x) \\ &= \widetilde{h}_k(x) + f(x) + w(x) + \varepsilon_k. \end{aligned}$$

Thus, by virtue of the relations given above we have

$$\max_{\lambda\in[0,1]} \hat{G}_k(\lambda) \le \xi^{k+1/2} + \varepsilon_k, \tag{3}$$

and $\hat{G}_k(0) = \xi^k$ since $\widetilde{h}_k(z_h^k) = h(z_h^k)$. We minus the equal value of the two hands of the above equality (3) to complete the result.

For the continuity and differentiability of $G_k(x,\lambda)$, we have the following equation by directly calculating partial derivation for x, i. e.

$$0 = \frac{\partial}{\partial x} G_k(\hat{x}(\lambda),\lambda) = g_h^k + (1-\lambda)g_f^{k-1} + \lambda g_f(z_h^k) + \mu[\hat{x}(\lambda) - \bar{x}].$$

$$\hat{x}(\lambda) = \bar{x} - \frac{1}{\mu}\{g_h^k + g_f^{k-1} + \lambda[g_f(z_h^k) - g_f^{k-1}]\}, \tag{4}$$

$\hat{x}(0) = \bar{x} - \frac{1}{\mu}(g_h^k + g_f^{k-1})$ and $z_h^k = \arg\min_x G_k(x(0)) = \hat{x}(0)$, since the construction of $G_k(x,\lambda)$. Therefore $\hat{x}(\lambda) - \hat{x}(0) = -\frac{\lambda}{\mu}[g_f(z_h^k) - g_f^{k-1}]$.

Computing the derivative of $\hat{G}_k(\lambda)$ one has:

$$\begin{aligned} \hat{G}'_k(\lambda) &= \widehat{f}_k(\hat{x}(\lambda)) - \widetilde{f}_k(\hat{x}(\lambda)) \\ &= f(z_h^k) + \langle g_f(z_h^k), \hat{x}(\lambda) - z_h^k\rangle - [f(z_f^{k-1}) - \langle g_f^{k-1}, \hat{x}(\lambda) - z_f^{k-1}\rangle] \\ &= \langle g_f(z_h^k) - g_f^{k-1}, \hat{x}(\lambda)\rangle + f(z_h^k) - \langle g_f(z_h^k), z_h^k\rangle - [f(z_f^{k-1}) - \langle g_f^{k-1}, z_f^{k-1}\rangle] \\ &= \langle g_f(z_h^k) - g_f^{k-1}, \hat{x}(\lambda) - \hat{x}(0)\rangle + f(z_h^k) - [f(z_f^{k-1}) + \langle g_f^{k-1}, z_h^k - z_f^{k-1}\rangle] \\ &= -\frac{\lambda}{\mu} \parallel g_f(z_h^k) - g_f^{k-1} \parallel^2 + F(z_h^k) - \widetilde{F}(z_h^k). \end{aligned}$$

Lemma 2. *The following inequalities hold for any* $g_f(z_h^k) \in \partial_{\varepsilon_k} f(z_h^k)$, $g_f^{k-1} \in \partial_{\varepsilon_k} f(z_f^{k-1})$, *and* $\delta_k = F(z_h^k) - \widetilde{F}(z_h^k)$:

(a) $\max_{\lambda\in[0,1]} \hat{Q}_k(\lambda) - \hat{Q}_k(0) \ge \hat{Q}_k(\bar{\lambda}_k) - \hat{Q}_k(0) \ge \dfrac{1}{2}\bar{\lambda}_k\delta_k$;

(b) $\xi^{k+1} + \varepsilon_k \ge \xi^{k+1/2} \ge \xi^k + \frac{1}{2}\bar{\lambda}_k\delta_k - \varepsilon_k$,

where $\bar{\lambda}_k = \max\{0, \min\{1, \|g_f(z_h^k) - g_f^{k-1}\|^{-2}\delta_k\mu\}$.

Proof. According to Newton-Leibniz formula and the bound of λ_k, one has

$$\begin{aligned}\hat{G}_k(\bar{\lambda}_k) - \hat{G}_k(0) &= \int_0^{\bar{\lambda}_k} \hat{G}'_k(\lambda)d\lambda \\ &= \bar{\lambda}_k[\delta_k - \frac{\bar{\lambda}_k}{2\mu} \parallel g_f(z_h^k) - g_f^{k-1} \parallel^2] \\ &\geq \bar{\lambda}_k[\delta_k - \frac{1}{2\mu} \cdot \frac{\mu\delta_k}{\parallel g_f(z_h^k) - g_f^{k-1} \parallel} \cdot \parallel g_f(z_h^k) - g_f^{k-1} \parallel^2] \\ &= \frac{1}{2}\bar{\lambda}_k\delta_k.\end{aligned}$$

The item (b) follows from (a) and Lemma 1. □

The following results are similar to the ones given in Section 2 of [10] and the argument is also similar in form but essentially different.

Theorem 1. *The approximate proximal points $p_h(\bar{x}, \varepsilon_k)\}$ and approximations $\{\widetilde{F}_k\}$ generated by Algorithm I have the following properties when $\lim_{k\to\infty} \varepsilon_k = 0$:*

(a) $\parallel p_h(\bar{x}, \varepsilon_k) - p(\bar{x}) + 2\varepsilon_k \parallel \leq \{\frac{1}{\mu}[F(p_h(\bar{x}, \varepsilon_k)) - \widetilde{F}_k(p_h(\bar{x}, \varepsilon_k))]\}^{1/2}$;

(b) $\lim_{k\to\infty} [F(p_h(\bar{x}, \varepsilon_k)) - \widetilde{F}_k(p_h(\bar{x}, \varepsilon_k))] = 0$;

(c) $\lim_{k\to\infty} p_h(\bar{x}, \varepsilon_k) = p(\bar{x})$.

Proof. According to $F \geq \widetilde{F}_k - \varepsilon_k$ and the definition of $p_h(\bar{x}, \varepsilon_k)$ that is the ε_k-approximate solution of the strongly convex h-subproblem, we have [14]

$$\begin{aligned}F(p(\bar{x})) + w(p(\bar{x})) &\geq \widetilde{F}_k(p(\bar{x})) + w(p(\bar{x})) - \varepsilon_k \\ &\geq \widetilde{F}_k(p_h(\bar{x}, \varepsilon_k) + w(p_h(\bar{x}, \varepsilon_k)) \\ &\quad + \langle 0, p(\bar{x}) - p_h(\bar{x}, \varepsilon_k)\rangle + \frac{1}{2}\mu \parallel p(\bar{x}) - p_h(\bar{x}, \varepsilon_k) \parallel^2 - 2\varepsilon_k,\end{aligned}$$

i. e. ,

$$\begin{aligned}&F(p(\bar{x})) + w(p(\bar{x})) \\ &\geq \widetilde{F}_k(p_h(\bar{x}, \varepsilon_k)) + w(p_h(\bar{x}, \varepsilon_k)) + \frac{1}{2}\mu \parallel p(\bar{x}) - p_h(\bar{x}, \varepsilon_k) \parallel^2 - 2\varepsilon_k.\end{aligned} \tag{5}$$

Similarly, $p(\bar{x})$ solves the strongly convex f-subproblem, so

$$\begin{aligned}&F(p_h(\bar{x}, \varepsilon_k)) + w(p_h(\bar{x}, \varepsilon_k)) \\ &\geq F(p(\bar{x})) + w(p(\bar{x})) + \frac{1}{2}\mu \parallel p(\bar{x}) - p_h(\bar{x}, \varepsilon_k) \parallel^2\end{aligned} \tag{6}$$

Summing the two equalities (5) and (6) above, one obtains

$$F(p_h(\bar{x}, \varepsilon_k)) - \widetilde{F}_k(p_h(\bar{x}, \varepsilon_k)) \geq \mu \parallel p(\bar{x}) - p_h(\bar{x}, \varepsilon_k) \parallel^2 - 2\varepsilon_k,$$

that leads to (a). Next, (5) can be equivalently written as

$$\frac{1}{2}\mu \parallel p(\bar{x}) - p_h(\bar{x}, \varepsilon_k) \parallel^2 \leq F(p(\bar{x})) + w(p(\bar{x})) - \xi^k + 2\varepsilon_k. \tag{7}$$

By Lemma 2, ξ^k is nondecreasing as $\varepsilon_k \to 0$, so (7) implies that $p_h(\bar{x}, \varepsilon_k)$ is bounded. Then applying the Theorem 24.7 in [13] to $\{g_f(z_h^k)\}$ one obtains its boundedness. Similarly, according to the property of strongly convexity

$$F(p(\bar{x})) + w(p(\bar{x})) \geq \breve{F}(p(\bar{x})) + w(p(\bar{x})) - \varepsilon_k$$
$$\geq \breve{F}(p_f(\bar{x}, \varepsilon_k)) + w(p_f(\bar{x}, \varepsilon_k)) + \frac{1}{2}\mu \parallel p(\bar{x}) - p_f(\bar{x}, \varepsilon_k) \parallel^2 - 2\varepsilon_k,$$

one has

$$\frac{1}{2}\mu \parallel p(\bar{x}) - p_f(\bar{x}, \varepsilon_k) \parallel^2 \leq F(p(\bar{x})) + w(p(\bar{x})) - \xi^{k+1/2} + 2\varepsilon_k,$$

Therefore z_f^k and $g_f^{k-1} \in \partial_{\varepsilon_k} f(z_f^{k-1})$ are bounded. By (7), as $k \to \infty$ the sequence ξ^k is bounded from above, so Lemma 2 implies that it converges and $\lim_{k\to\infty} \bar{\lambda}_k \delta_k \to 0$. Since $\parallel g_f(z_h^k) - g_f^{k-1} \parallel$ is bounded, (b) follows from the definition of $\bar{\lambda}_k$ by Lemma 2. The final assertion is a consequence of (a) and (b). $\square$

Theorem 1 ensures that for every $\epsilon > 0$ and after finite many steps an approximate proximal point $p_h(\bar{x}, \varepsilon_k)$ satisfying $\parallel p_h(\bar{x}, \varepsilon_k) - p(\bar{x}) \parallel \leq \epsilon$. An implemental scheme will be presented in the next section.

3. The Implemental Approximate Alternating Linearization Decomposition Method.

The approximate alternating linearization method can be implemented under a simple descent test for terminating the loop of Algorithm I in order to update the prox center.

In this section the notations are modified as follows.

Notations

x^k	Current iteration point
μ_k	Prox center
$w_k(\cdot)$	$w_k(x) = \frac{1}{2}\mu_k \parallel x - x^k \parallel^2$
h-subproblem	$\min_x h(x) + \widetilde{f}_k(x) + w_k(x)$, where $\widetilde{f}_k$ is a linear model of f
f-subproblem	$\min_x \widetilde{h}_k(x) + f(x) + w_{k+1}(x)$, where $\widetilde{h}_k$ is a linear model of h
z_h^k	An ε_k-approximate solution of h-subproblem
z_f^k	An ε_k-approximate solution of f-subproblem
$\widetilde{f}_k$	$\widetilde{f}_k(\cdot) = f(z_f^{k-1}) + \langle g_f^{k-1}, \cdot - z_f^{k-1}\rangle$, where $g_f^0 \in \partial_{\varepsilon_0} f(z_f^0), z_f^0 \in R^n$ and $g_f^k = -g_h^k - \mu_{k+1}(z_f^k - x^{k+1})$
$\widetilde{h}_k$	$\widetilde{h}_k(\cdot) = h(z_h^{k-1}) + \langle g_h^k, \cdot - z_h^k\rangle$, where $g_h^k = -g_f^{k-1} - \sigma(z_h^k - x^k)$

Next the implemental approximate alternating linearization decomposition algorithm is displayed.

Algorithm II: implemental approximate alternating linearization decomposition algorithm.

Step 0: `Initiation`
Let $x^1 \in \operatorname{dom} h$, $z_f^0 \in R^n$, $g_f^0 \in \partial_{\varepsilon_0} f(z_f^0)$, $\varepsilon_0 \in (0,1)$ and choose parameters $\mu_1 \geq \mu_{\min} > 0$, $\kappa > 1$, $\beta_0 > 0$, $\beta_1 \in (0,1)$. Set $k = 1$.

Step 1: `Solving the h-subproblem`
Compute $z_h^k = p_h(\bar{x}, \varepsilon_k)$ and $g_h^k = -g_f^{k-1} - \mu_k(z_h^k - x^k)$.

Step 2: `Stopping test`
Let $\widetilde{F}_k = h + \widetilde{f}_k$. Set $v_k = F(x^k) - \widetilde{F}_k(z_h^k)$. If $F(z_h^k) \leq F(x^k) - \beta_1 v_k$, then set $x^{k+1} = z_h^k$ (descent step);
otherwise set $x^{k+1} = x^k$ (null step).

Step 3: `Update prox center`
If $x^{k+1} = z_h^k$, then choose $\mu_{k+1} \in [\max\{\mu_{\min}, \mu_k/\kappa\}, \mu_k]$.
If $x^{k+1} = x^k$ and

$$\delta_k := F(z_h^k) - \widetilde{F}_k(z_h^k) \geq \beta_0 \frac{v_k}{\|z_h^k - x^k\|};$$

then choose $\mu_{k+1} \geq \mu_k$; else set $\mu_{k+1} = \mu_k$.

Step 4: `Solving the f-subproblem`
Find the ε_k-approximate solution z_f^k of the following f-problem.
Set $g_f^k = -g_h^k - \mu_{k+1}(z_f^k - x^{k+1})$.

Step 5: `Update` ε_k `and` k.
Let $\varepsilon_k = \gamma \varepsilon_{k-1}$, $\gamma \in (0,1)$ and set $k = k+1$ and loop at Step 1.

End of the Algorithm II

In Algorithm II, we denote

$$\xi_{\mu_k}^k = h(z_h^k) + \widetilde{f}_k(z_h^k) + w_k(z_h^k) = \widetilde{F}_k(z_h^k) + w_k(z_h^k), \tag{8}$$

$$\xi_{\mu_{k+1}}^{k+1/2} = \widetilde{h}_k(z_f^k) + f(z_f^k) + w_{k+1}(z_f^k) = \widehat{F}_k(z_f^k) + w_{k+1}(z_f^k) \tag{9}$$

respectively. Let us first make a simple observation concerning the ε_k-optimal values of h-subproblem and f-subproblem.

Since $\widetilde{h}_k(z_h^k) = h(z_h^k)$ and

$$\begin{aligned} \xi_{\mu_k}^k &= \widetilde{h}_k(z_h^k) + \widetilde{f}_k(z_h^k) + w_k(z_h^k) \\ &\leq \widetilde{h}_k(x) + \widetilde{f}_k(x) + w_k(x) \leq \widetilde{h}_k(x) + f(x) + w_k(x) + \varepsilon_k, \end{aligned}$$

one has

$$\xi_{\mu_k}^k \leq \widetilde{h}_k(z_f^k) + f(z_f^k) + w_k(z_f^k) + \varepsilon_k \leq \xi_{\mu_{k+1}}^{k+1/2} + w_k(z_f^k) - w_{k+1}(z_f^k) + \varepsilon_k.$$

Similarly, we gain another result and hence the inequality $\xi_{\mu_{k+1}}^{k+1/2} \leq \xi_{\mu_{k+1}}^{k+1} + \varepsilon_k$ holds.

In short, the relations of $\xi^k_{\mu_k}$, $\xi^{k+1/2}_{\mu_{k+1}}$ and $\xi^{k+1}_{\mu_{k+1}}$ are given below:

$$\xi^{k+1}_{\mu_{k+1}} + \varepsilon_k \geq \xi^{k+1/2}_{\mu_{k+1}} \geq \xi^k_{\mu_k} - w_k(z^k_f) + w_{k+1}(z^k_f) - \varepsilon_k.$$

According to $g^k_f \in \partial_{\varepsilon_k} f(z^k_f)$ and $\widetilde{F}_k \leq F + \varepsilon_k$ we get $\xi^k_{\mu_k} \leq F(x^k) + \varepsilon_k$ and $v_k \geq -\varepsilon_k$. Thus it follows from Step 2

$$F(z^k_h) \leq F(x^k) - \beta_1 v_k \leq F(x^k) - \beta_1 \varepsilon_k \leq F(x^k)$$

and hence $\{F(x^k)\}$ is nonincreasing and $\{x^k\} \subset \mathrm{dom}F$. These inequalities show that if $v_k = 0$ or $\xi^k_{\mu_k} = F(x^k)$, then x^k tends to a cluster of $\arg\min F$.

4. Convergence analysis

First of all, in this section we make use of some results related to the sequences $\{\xi^k_{\mu_k}\}$ and $\{\xi^{k+1/2}_{\mu_{k+1}}\}$ generated by Algorithm II to investigate the stopping test. Secondly we consider the value of them in both null steps and descent steps so as to conclude that iterative point $x^k \to \widetilde{x}$, where $\widetilde{x} \in \mathrm{Arg}\min\limits_x F$.

Lemma 3. *The following inequalities are true for all $k = 1, 2, \ldots$ and $\varepsilon_k \to 0$ as $k \to \infty$:*
(a) $w_k(z^k_h) - 2\varepsilon_k \leq \frac{1}{2} v_k - \varepsilon_k \leq F(x^k) - \xi^k_{\mu_k} \leq v_k$;
(b) $w_{k+1}(z^k_f) - 2\varepsilon_k \leq F(x^{k+1}) - \xi^{k+1/2}_{\mu_{k+1}}$.

Proof. The right inequality of (a) comes from definitions of v_k and $\{\xi^k_{\mu_k}\}$ which imply $F(x^k) - v_k = \widetilde{F}_k(z^k_h) \leq \xi^k_{\mu_k}$. On the other hand by the ε_k-subgradient inequality and

$$-\mu_k(z^k_h - x^k) = g^k_h + g^{k-1}_f \in \partial_{\varepsilon_k} \widetilde{F}_k(z^k_h),$$

it follows from

$$\widetilde{F}_k(x^k) \geq \widetilde{F}_k(z^k_h) + \langle g^k_h + g^{k-1}_f, x^k - z^k_h \rangle - \varepsilon_k = \widetilde{F}_k(z^k_h) + \mu_k \parallel z^k_h - x^k \parallel^2 - \varepsilon_k.$$

Then one has

$$v_k = F(x^k) - \widetilde{F}_k(z^k_h) \geq \widetilde{F}_k(x^k) - \widetilde{F}_k(z^k_h) - \varepsilon_k \geq 2w_k(z^k_h) - 2\varepsilon_k.$$

In consequence v_k can be used to express $\xi^k_{\mu_k}$ as follows

$$\begin{aligned} \xi^k_{\mu_k} &= F(x^k) - v_k + w_k(z^k_h) \\ &\leq F(x^k) - v_k + \frac{1}{2}(v_k + 2\varepsilon_k) = F(x_k) - \frac{1}{2} v_k + \varepsilon_k. \end{aligned} \tag{10}$$

The assertion (a) is proved. Similarly, the following inequalities are true

$$F(x^{k+1}) - \widehat{F}_k(z^k_f) \geq \widehat{F}_k(x^{k+1}) - \widehat{F}_k(z^k_f) - \varepsilon_k \geq 2w_{k+1}(z^k_f) - 2\varepsilon_k,$$

since $-\mu_{k+1}(z^k_f - x^{k+1}) = g^k_f + g^k_h \in \partial_{\varepsilon_k} \widehat{F}_k(z^k_f)$. Therefore the inequality of (b)

$$F(x^{k+1}) - \xi^{k+1/2}_{\mu_{k+1}} \geq w_{k+1}(z^k_f) - 2\varepsilon_k.$$

is correct. □

Corollary 1. *If $v_k = 0$, then $x^k \to \widetilde{x}$, where $\widetilde{x} \in Arg\min F$.*

Proof. It is clear that $z_h^k = x^k$ and $\widetilde{F}_k(z_h^k) = F(z_h^k) = F(x^k)$ from Lemma 3 (a). Moreover $x^k = z_h^k \to \arg\min\{F + w_k(\cdot)\}$ as $k \to \infty$ since $\lim_{k\to\infty} \varepsilon_k = 0$. Thus we have $x^k \to \widetilde{x}$, where $\widetilde{x} \in \operatorname{Arg}\min F$[14]. □

To establish the convergence of the Algorithm II, it is necessary to consider the ε_k-optimal values of h-subproblem and f-subproblem in both null steps and descent steps.

Lemma 4. *In a null step at iteration k, we estimate the increase from $\xi_{\mu_k}^k$ to $\xi_{\mu_{k+1}}^{k+1}$ as follows:*

$$\xi_{\mu_{k+1}}^{k+1} \geq \xi_{\mu_k}^k + \frac{1}{2}(1-\beta_1)\bar{\lambda}_k v_k - 2\varepsilon_k,$$

where $\bar{\lambda}_k \geq \max\{\min\{1, (1-\beta_1)v_k\mu_k/\|g_f(z_h^k) - g_f^{k-1}\|^2\}\}$ *for any* $g_f(z_h^k) \in \partial_{\varepsilon_k} f(z_h^k)$.

Proof. If a null step is occurring at iteration k, we have $F(z_h^k) > F(x^k) - \beta_1 v_k$, hence

$$\delta_k = F(z_h^k) - \widetilde{F}_k(z_h^k) = F(z_h^k) - F(x^k) + v_k > (1-\beta_1)v_k.$$

If $\mu_{k+1} = \mu_k$, then from Lemma 2 (b), it follows that

$$\xi_{\mu_{k+1}}^{k+1} \geq \xi_{\mu_k}^k + \frac{1}{2}(1-\beta_1)\bar{\lambda}_k v_k - 2\varepsilon_k.$$

In consequence in a null step $\mu_{k+1} > \mu_k$, $\xi_{\mu_{k+1}}^{k+1} \geq \xi_{\mu_k}^{k+1}$ holds. Then proof is completed. □

Lemma 5. *Assume the set $K = \{x^{k+1} \neq x^k\}$ which contains the descent points generated by Algorithm II and $\varepsilon_k = 0$ as $k \to \infty$.*

(i) *If K is finite, then $v_k \to 0$.*

(ii) *If K is infinite and* $\inf F > -\infty$, *then*

(a) $\sum_{k\in K} v_k < \infty$; (b) $\lim_{k\to\infty} v_k = 0$;

(c) $\lim_{k\to\infty}[F(x^k) - \xi_{\mu_k}^k] = 0$; (d) $\lim_{k\to\infty}[F(x^{k+1}) - \xi_{\mu_{k+1}}^{k+1/2} = 0$.

Proof. (i) Suppose $k_0 \in \{k | x^k = x^{k_0}, k \geq k_0\}$. By virtue of (8) and Lemma 4, one has $\xi_{\mu_k}^k \leq F(x^{k_0}) + \varepsilon_{k_0}$ for $k \geq k_0$, and hence the sequence $\xi_{\mu_k}^k$ is convergent and $\xi_{\mu_{k+1}}^{k+1} - \xi_{\mu_k}^k \to 0$. Since $\mu_k \geq \mu_{\min} > 0$ for all k and $\{x^k\}$ is bounded, so are z_h^k and z_f^k from Lemma 3. Therefore $g_f(z_h^k) \in \partial_{\varepsilon_k} f(z_h^k)$ and $g_f^k \in \partial_{\varepsilon_k} f(z_f^k)$ are bounded as well, since f is a proper convex function coming from Theorem 23.4 [13]. Hence according to (4) one has $\bar{\lambda}_k v_k \to 0$ and $v_k \to 0$ as $k \to \infty$.

(ii) The items (a)-(d) can be easily proved from [10]. The only different proofs from [10] are only in which ε_k-optimal solutions are used here. Note that, $\varepsilon_k \to 0$

as $k \to \infty$. Hence there is rarely difference except the constant ε_k.
We prove the rest of this assertion. For a selected $k \in K$, we have

$$\begin{aligned}\| x^{k+1} - \tilde{x} \|^2 &= \| (x^{k+1} - x^k) + (x^k - \tilde{x} \|^2 \\ &= \| x^{k+1} - x^k \|^2 + 2\langle x^{k+1} - x^k, x^{k+1} - \tilde{x} - (x^{k+1} - x^k)\rangle + \| x^k - \tilde{x} \|^2 \\ &= \| x^k - \tilde{x} \|^2 + 2\langle x^{k+1} - x^k, x^{k+1} - \tilde{x}\rangle - \| x^{k+1} - x^k \|^2 .\end{aligned}$$

Since, $g^k_{\widetilde{F}} = g^k_h + g^{k-1}_f = -\mu_k(x_{k+1} - x_k) \in \partial_{\varepsilon_k} \widetilde{F}_k(x^{k+1})$, we have

$$\begin{aligned}\mu_k\langle x^{k+1} - x^k, x^{k+1} - \tilde{x}\rangle &= \langle \tilde{x} - x^{k+1}, g^k_{\widetilde{F}}\rangle \\ &\leq \widetilde{F}_k(\tilde{x}) - \widetilde{F}_k(x^{k+1}) + \varepsilon_k \\ &\leq F(\tilde{x}) - F(x^k) + v_k + 2\varepsilon_k.\end{aligned}$$

according to Step 2 of Algorithm II. It leads to

$$\begin{aligned}\| x^{k+1} - \tilde{x} \|^2 &= \| x^k - \tilde{x} \|^2 + 2\langle x^{k+1} - x^k, x^{k+1} - \tilde{x}\rangle - \| x^{k+1} - x^k \|^2 \\ &\leq \| x^k - \tilde{x} \|^2 + 2\langle x^{k+1} - x^k, x^{k+1} - \tilde{x}\rangle \\ &\leq \| x^k - \tilde{x} \|^2 + \frac{2}{\mu_k}(F(\tilde{x}) - F(x^k) + v_k + 2\varepsilon_k) \\ &\leq \| x^k - \tilde{x} \|^2 + \frac{2}{\mu_k}(v_k + 2\varepsilon_k) \leq \| x^k - \tilde{x} \|^2 + \frac{2}{\mu_k}v_k + \frac{4}{\mu_k}\varepsilon_k.\end{aligned}$$

It is easy to see that the accumulation set of $\{x^k\}$ is bounded from the last inequality above and Lemma 5 because $\{\mu_k\}$ is bounded away from zero by construction. Hence, $\{x^k\}$ is not empty, say, $\bar{x}$ is a cluster one of this set. This lead to the conclusion that $\bar{x}$ is the unique accumulation point of $\{x^k\}$. □

Lemma 6. *If there exists a point $\tilde{x}$ such that $F(x^k) \geq F(\tilde{x})$ for all k, then*
(1) $v_k \to 0$, $F(x^k) - \xi^k_{\mu_k} \to 0$, *and* $F(x^k) - \xi^{k+1/2}_{\mu_{k+1}} \to 0$, *as* $k \to \infty$;
(2) *the sequence $\{x^k\}$ converges to a point $\bar{x} \in arg\min F$, and $F(x^k) \downarrow F(\bar{x})$.*

Proof. Similarly, these results can be testified by [10] Lemma 4.7 because $\lim_{k\to\infty} \varepsilon_k = 0$ □

Our results can be summarized as the following theorem.

Theorem 2. *Algorithm II generates a sequence $\{x^k\}$ with the following properties:*
(1) $F(x^k) \downarrow \inf F$;
(2) *If $Arg\min F \neq \emptyset$, then $\{x^k\}$ converges to a point $\hat{x} \in Arg\min F$;*
(3) *If $Arg\min F = \emptyset$, then $\| x^k \| \to \infty$;*
(4) *If $Arg\min F \neq \emptyset$ and the sequence $\{\mu_k\}$ is bounded, then the sequences $\{g^k_f\}$ and $\{g^k_h\}$ are bounded, $g^k_h + g^{k-1}_f \to 0$, $g^k_h + g^k_f \to 0$, and every accumulation point $(\hat{g}_f, \hat{g}_h)$ of $\{(g^k_f, g^k_h)\}$ satisfies the relations $\hat{g}_f \in \partial_\varepsilon f(\hat{x})$, $\hat{g}_h \in \partial_\varepsilon h(\hat{x})$, and $\hat{g}_f + \hat{g}_h = 0$.*

Proof. These results resemble Theorem 4.8 in [10] and are easily proved similar to it. □

References

1. J. F. Benders, *Partitioning procedures for solving mixed-variable programming problems*, Numerische Mathematik **4** (1962) 238-252.
2. D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical-Methods*, Prentice-Hall, Englewood Cliffs, NJ, 1989.
3. G. B. Dantzig, P. Wolfe, *The decomposition algorithm for linear programs*, Econometrica, Vol. 29 No. **4** (1961) 767-778.
4. J. Eckstein and D. P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Programming **55** (1992) 293-318.
5. J. Eckstein and M. Fukushima, *Some reformulations and applications of the alternating direction method of multipliers*, in Large Scale Optimization: State of the Art, W. W. Hager, D. W. Hearn, and P. M. Pardalos, eds., Kluwer, Dordrecht, the Netherlands, 1994, 115-134.
6. J. Eckstein, *Approximate iterations in Bregman-function-based proximal algorithms, Mathematical. Programming* **83** 1998, 113-123.
7. M. Fukushima, *Application of the alternating direction method of multipliers to separable convex programming problems*, Comput. Optim. Appl. **1** (1992) 93-111.
8. D. Gabay, *Applications of the method of multipliers to variational inequalities, in Augmented Lagrangian Methods: Applications to the Solution of Boundary Value Problems*, M. Fortin and R. Glowinski, eds., North-Holland, Amsterdam, 1983, 299-331.
9. Hiriart-Urruty, Jean-Baptiste and Lemarchal.Claude, *Convex Analysis and Minimization Algorithms*, Two volumes, Springer Verlag, Heidelberg, 1993.
10. K. C. Kiwiel, C.H. Rosa and A. Ruszczynski, *Proximal decomposition via alternating linearization*, SIAM Journal on Optimization **9** (1999) 153–172.
11. P. L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM Journal on Numerical Analysis **16** (1979) 964-979.
12. P. Mahey and P. D. Tao, *Partial regularization of the sum of two maximal monotoneoperators*, RAIRO Model. Math. Anal. Number **27** (1993) 375-392.
13. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, NJ, 1970.
14. R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM. Control Optim. **14** (1976), 877-898.
15. R. T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res. **1** (1976) 97-116.
16. A. Ruszczynski, *Some advances in decomposition methods for stochastic linear programming Annals of Operations Research* **85**(1999) 153-172.
17. J. E. Spingarn, *Applications of the method of partial inverses to convex programming: Decomposition*, Math. Programming **32** (1985) 199-223.
18. P. Tseng, *Further applications of a splitting algorithm to decomposition in variational inequalities and convex programming*, Math. Programming **48** (1990) 249-263.
19. P. Tseng, *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities*, SIAM J. Control Optim. **29** (1991) 119-138.

Li Dan is a Ph.D candidate of Prof. Z.-Q. Xia in Institute of Operations Research and Control, School of Mathematical Sciences, Dalian University of Technology (DUT), Dalian, China. Her research interests focus on (smooth, nonsmooth, discrete and numerical) optimization and applications (in science and engineering), OR methods and applications.

Institute of Operations Research and Control, School of Mathematical Sciences, Dalian University of Technology (DUT), Dalian, 116024, China
e-mail: lidan@mail.dlut.edu.cn

Pang Li-Ping is an associate professor in Institute of Operations Research and Control, School of Mathematical Sciences, Dalian University of Technology (DUT), Dalian, China. She received her M.Sc in Polytechnic University of Jilin, Changchun, in 1993 and received her Ph.D in Dalian University of Technology, Dalian, 2004. Her research areas are (smooth, nonsmooth, discrete and numerical) optimization and applications (in science and engineering), OR methods and applications.

Institute of Operations Research and Control, School of Mathematical Sciences, Dalian University of Technology (DUT), Dalian, 116024, China
e-mail: lppang@dlut.edu.cn

Xia Zun-Quan is a professor in Institute of Operations Research and Control, School of Mathematical Sciences, Dalian University of Technology (DUT), Dalian, China. He graduated in Institute of Mathematica, Fudan University, Shanghai, as a graduate student in 1968. His research areas are (smooth, nonsmooth, discrete and numerical) optimization and applications (in science and engineering), OR methods and applications.

Institute of Operations Research and Control, School of Mathematical Sciences, Dalian University of Technology (DUT), Dalian, 116024, China
e-mail: zqxiazhh@dlut.edu.cn