Journal of Global Optimization **13:** 389–406, 1998.

Proximal Point Methods and Nonconvex Optimization

A. KAPLAN and R. TICHATSCHKE
Department of Mathematics, University of Trier, Trier, Germany

(Accepted in final form 30 June 1998)

Abstract. The goal of this paper is to discover some possibilities for applying the proximal point method to nonconvex problems. It can be proved that – for a wide class of problems – proximal regularization performed with appropriate regularization parameters ensures convexity of the auxiliary problems and each accumulation point of the method satisfies the necessary optimality conditions.

Key words: Ill-posed problems, Nonconvex optimization, Proximal point methods, Regularization

1. Introduction

The proximal point method was first suggested by Martinet [24] for solving variational problems of the form

$$\min\{f(u) : u \in K\}, \tag{1.1}$$

where $f : V \to \bar{\mathbb{R}} \equiv \mathbb{R} \cup \{+\infty\}$ is a proper convex lower semicontinuous functional and K is a convex closed subset of a Hilbert space V.

The method is described as

$$u^{i+1} \approx \arg\min_{u \in K}\left\{f(u) + \frac{\chi_i}{2}\|u - u^i\|^2\right\}, \tag{1.2}$$

with $u^0 \in V$ an arbitrary point and $\{\chi_i\}$ a given sequence, $0 < \chi_i \le \bar{\chi} < \infty$.

Obviously, to deal with a real numerical algorithm, Method (1.2) has to be combined with an optimization algorithm for solving the auxiliary problems

$$\min\{f(u) + \frac{\chi_i}{2}\|u - u^i\|^2 : u \in K\}. \tag{1.3}$$

If the optimal set U^* of Problem (1.1) is non-empty, then weak convergence of the iterates in (1.2) to some $u^* \in U^*$, as well as convergence of the objective values $\{f(u^i)\}$ to $f(u^*)$, are guaranteed under the condition that, in each prox-step i, the point u^{i+1} approximates

$$\mathcal{P}_{f,K,\chi_i}u^i = \arg\min_{u \in K}\left\{f(u) + \frac{\chi_i}{2}\|u - u^i\|^2\right\}$$

with sufficient accuracy, namely,

$$\|u^{i+1} - \mathcal{P}_{f,K,\chi_i} u^i\| \le \epsilon_i, \qquad \sum_i \frac{\epsilon_i}{\chi_i} < \infty. \tag{1.4}$$

Concerning the conditions providing strong convergence of $\{u^i\}$, see [18, 23]. Some of these conditions do not exclude that $dim\ U^* = \infty$.

Proximal point methods are a fundamental tool for solving ill-posed or ill-conditioned problems. Because, in distinction to the Tikhonov regularization, $\chi_i \to 0$ is not required for ensuring convergence, the auxiliary problems (1.3) possess better properties than the corresponding auxiliary problems in Tikhonov's method. In particular, if f is sufficiently smooth, then by means of an appropriate choice of χ_i, one attains at all iterations a proper conditionality of the Hessians of the objective functions in (1.3) (or their approximations, in case V is infinite dimensional).

Thus, we deal with well-posed auxiliary problems (1.3), and the behaviour of the sequence $\{u^i\}$ is defined by the properties of the *proximal mapping*

$$\mathcal{P} = \mathcal{P}_{f,K,\chi} : a \in V \to \arg\min_{u \in K} \left\{ f(u) + \frac{\chi}{2}\|u - a\|^2 \right\}, \quad (\chi > 0), \tag{1.5}$$

which generates the iterates (1.2).

Important properties of the proximal mapping are, in particular (cf. [17], Section 8):

- $\mathcal{P}$ is a firmly non-expansive operator, i.e. it holds
$$\|\mathcal{P}v - \mathcal{P}v'\|^2 \le \|v - v'\|^2 - \|(\mathcal{I} - \mathcal{P})v - (\mathcal{I} - \mathcal{P})v'\|^2 \quad \forall v, v' \in V;$$
- $\mathcal{P}u = u \quad \Leftrightarrow \quad u \in U^*$;
- the functional $\eta(a) = \min_{u \in K} \left\{ f(u) + \frac{\chi}{2}\|u - a\|^2 \right\}$ is convex and continuously differentiable on V, and $\nabla\eta(a) = \chi(a - \mathcal{P}a)$ (differentiability of f is not supposed).

Starting with the papers of Rockafellar [27], the proximal point method was extended to problems of finding a zero of a maximal monotone (in general, multi-valued) operator, including as special cases monotone variational inequalities, convex–concave games etc. It is well-known that the *Hestenes–Powell multiplier method* (see [26]) for convex programming problems as well as the *Douglas–Rachford splitting method* (see [11]) for finding a zero of the sum of two monotone operators are special cases of the proximal point method.

– In order to clarify this relationship for the Hestenes–Powell multiplier method, let us consider the convex problem
$$\min\{f_0(u) : \ f_j(u) \le 0, \ j = 1, ..., m\}, \quad V = \mathbb{R}^n.$$

The current vector λ^{i+1} in the exact multiplier method, performed with the augmented Lagrangian

$$\mathcal{L}(u,\lambda,r) = f_0(u) + \frac{1}{2r}\sum_{j=1}^{m}\left\{\max^2[0,\lambda_j + rf_j(u)] - \lambda_j^2\right\},$$

coincides with

$$\tilde{\mathcal{P}}_{g,\mathbb{R}^n_+,\chi}\lambda^i = \arg\max_{\lambda\in\mathbb{R}^m_+}\left\{g(\lambda) - \frac{\chi}{2}\|\lambda-\lambda^i\|^2\right\}, \quad \chi = \frac{1}{r},$$

where

$$g(\lambda) = \inf_{u\in\mathbb{R}^n}\left\{f_0(u) + \sum_{j=1}^{m}\lambda_j f_j(u)\right\}. \tag{1.6}$$

Here the operator $\tilde{\mathcal{P}}_{g,\mathbb{R}^n_+,\chi}$ is an analogue of the proximal mapping (1.5), corresponding to the maximization problem

$$\max\{g(\lambda):\ \lambda\in\mathbb{R}^m_+\}.$$

Obviously, g is a concave function.

Besides the applications of the classical proximal point method, there are a lot of methods using proximal regularization to stabilize standard optimization or discretization techniques. A part of these methods can be briefly described as follows: When the standard algorithm applied to Problem (1.1) constructs (formally) a sequence of auxiliary problems

$$\min\{f_i(u):\ u\in K_i\subset V\}, \tag{1.7}$$

then the corresponding ‘regularized’ method takes the form

$$\tilde{u}^{i+1} \approx arg\min_{u\in K_{i+1}}\{f_{i+1}(u) + \frac{\chi_i}{2}\|u-u^i\|^2\}, \tag{1.8}$$

$$u^{i+1} = u^i + \alpha_i(\tilde{u}^{i+1} - u^i), \quad \alpha_i\in[0,1], \tag{1.9}$$

i.e. ‘iterative’ regularization of a sequence of auxiliary problems is performed. Usually $\alpha_i\equiv 1$, i.e. $u^{i+1} = \tilde{u}^{i+1}$, but line-search can also be met in the literature.

We refer to [1, 4, 17] for proximal penalty methods, to [2, 26] for proximal multiplier methods and to [16] for proximal methods in the framework of discretization of elliptic variational inequalities and convex semi-infinite programming problems. A modification of the scheme (1.8), (1.9) is developed in [17, 19]: proximal iterations for the i-th auxiliary problem (1.7) are repeated until they provide a ‘significant’ decrease of the objective function f_i.

In the last fifteen years the number of papers dealing with proximal-like methods is growing rapidly. One can distinguish some main directions in the development of this technique:

- Modifications of the standard methods for convex optimization in order to ensure a more qualified convergence of a minimizing sequence (for instance, to prevent oscillation) and a better stability of the auxiliary problems [cf. 1–4, 17, 26];
- Stable successive approximation or discretization of ill-posed monotone variational inequalities [cf. 17, 19, 22];
- Decomposition methods for convex minimization problems, without any assumption on strong or strict convexity of the objective functional [cf. 7, 11, 14];
- Proximal-like methods using Bregman's distance function [cf. 5, 6, 15].

However, convex problems have been considered almost exclusively. Even for multiplier methods, intensively studied in the nonconvex case, the 'proximal' aspect has been analyzed only for convex problems, although the function g, defined by (1.6), remains concave under nonconvex f_0, f_j, too.

Reasons for this distrust might be:

- The proximal mapping for a nonconvex problem may be no more nonexpansive, even in an arbitrary small neighborhood of a local or global solution. This mapping does not necessarily possess the Fejer-property w.r.t. U^*, too.

  EXAMPLE 1. $K = V = \mathbb{R}^2$, $f(u) = \min\{u_1^2, u_2^2\}$. Obviously, the optimal set of this problem is

  $$U^* = \{u \in \mathbb{R}^2 : \ u_1 = 0\} \cup \{u \in \mathbb{R}^2 : \ u_2 = 0\}.$$

  Consider the analogue of the proximal mapping (1.5):

  $$a \in V \to \operatorname{Arg}\min_{u \in K} \left\{ f(u) + \frac{\chi}{2} \|u - a\|^2 \right\},$$

  with $\chi = 2$ and take two points $a^1 = (2\alpha, \alpha)$ and $a^2 = (\alpha, 2\alpha)$ with an arbitrary small $\alpha > 0$. Then

  $$\mathcal{P}a^1 = \left(2\alpha, \frac{\alpha}{2}\right), \quad \mathcal{P}a^2 = \left(\frac{\alpha}{2}, 2\alpha\right)$$

  and

  $$\|\mathcal{P}a^1 - \mathcal{P}a^2\| = \sqrt{\frac{9}{2}}\alpha > \sqrt{2}\alpha = \|a^1 - a^2\|.$$

- Using the proximal method in the nonconvex case, iterations may terminate in a point, which is not a local minimum.

EXAMPLE 2.

$$V = \mathbb{R}^2, \quad K = \{u = (u_1, u_2) : u_1 + u_2 \le 0, u_1 \le 0, -u_2 \le 1\},$$

$$f(u) = -u_1 - \frac{1}{2}u_2^2, \qquad u^0 = (-\frac{1}{4}, \frac{1}{4}).$$

After one step of the exact proximal method (1.2) performed with $\chi_i \equiv 2$ we obtain $u^1 = (0, 0)$, which is a fixed point of the corresponding proximal mapping, but $u^1 \notin U^* = \{(0, -1)\}$. Note that the corresponding auxiliary problems (1.3) are convex problems with strongly convex objective functions.

- In general, one cannot expect that the regularized problems (1.3) are easier solvable than the original one.

The main goal of this paper is to discover some possibilities for applying proximal methods to nonconvex problems. In order to emphasize the basic ideas, only the exact proximal point method is analyzed in Section 2. Section 3 is concerned with known applications of the proximal regularization to nonconvex problems and describes the connection between the proximal approach and respective ideas in nonconvex optimization.

2. Proximal point method for nonconvex problems

In the sequel, we deal with Problem (1.1), $V = \mathbb{R}$, without convexity of the objective function f. However, the application of the exact proximal point method will be studied under the hypothesis that the objective functions in (1.3) are strongly convex on some convex set Ω, which has to contain the sequence of proximal iterates $\{u^i\}$. Due to the evident non-increase of $\{f(u^i)\}$, Ω may be any convex set containing $\{u \in K : f(u) \le f(u^0)\}$. In such situation, usually we can handle with (1.3) as a convex programming problem, and this is a very essential argument for applying proximal methods. But, of course, the assumption that the function $f(\cdot) + \frac{\chi}{2}\|\cdot\|^2$ becomes convex under a suitable choice of $\chi > 0$ is restrictive (see for instance the problem $f(u) = \min\{|u|, |u + 1|\}$, $K = \mathbb{R}^1$).

Therefore, we start with a simple statement which establishes the property mentioned for an important class of nonconvex functions.

PROPOSITION 1. *Let $f(u) = \sup_{\tau \in T} \varphi(u, \tau)$, $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, and Ω be a convex set such that $\Omega \cap dom f \ne \emptyset$. Assume that all the functions $\varphi(\cdot, \tau)$, $\tau \in T \subset \mathbb{R}^l$ are differentiable on Ω and $\sup_{\tau \in T} \|\nabla_u \varphi(\bar{u}, \tau)\| < \infty$ for some $\bar{u} \in \Omega \cap dom\ f$. Assume moreover, that for each $\tau \in T$ the gradient $\nabla_u \varphi(\cdot, \tau)$ is Lipschitz-continuous on Ω with a constant L_τ and $L = \sup_{\tau \in T} L_\tau < \infty$.*

Then, for $\chi \geq L$*, the function* $f(\cdot) + \frac{\chi}{2}\|\cdot\|^2$ *is convex and finite on* Ω*.*

Proof. For each $\tau \in T$ and arbitrary $u, v \in \Omega$ one gets

$$(\nabla_u \varphi(u,\tau) + \chi u - \nabla_u \varphi(v,\tau) - \chi v,\ u - v) \geq -L\|u-v\|^2 + \chi\|u-v\|^2.$$

Hence, $\varphi(\cdot,\tau) + \frac{\chi}{2}\|\cdot\|^2$ is convex on Ω if $\chi \geq L$.

Due to the properties of $\varphi(\cdot,\tau)$, for any $v, u \in \Omega$, we obtain also

$$\begin{aligned}\varphi(u,\tau) &= \varphi(v,\tau) + \int_0^1 (\nabla_u \varphi(v + t(u-v),\tau), u - v)dt \\ &= \varphi(v,\tau) + \int_0^1 (\nabla_u \varphi(v + t(u-v),\tau) - \nabla_u \varphi(v,\tau), u - v)dt \\ &\quad + (\nabla_u \varphi(v,\tau), u - v) \\ &\leq \varphi(v,\tau) + \frac{L_\tau}{2}\|v-u\|^2 + (\nabla_u \varphi(v,\tau), u-v).\end{aligned}$$

Inserting in this inequality $v = \bar{u}$ and taking into account that

$$\sup_{\tau \in T} \|\nabla_u \varphi(\bar{u},\tau)\| \equiv c < \infty,$$

one can conclude that

$$\varphi(u,\tau) \leq \varphi(\bar{u},\tau) + \frac{L_\tau}{2}\|u-\bar{u}\|^2 + c\|u-\bar{u}\|,$$

hence,

$$\sup_{\tau \in T} \varphi(u,\tau) + \frac{\chi}{2}\|u\|^2 = \sup_{\tau \in T}\{\varphi(u,\tau) + \frac{\chi}{2}\|u\|^2\} < \infty \quad \forall u \in \Omega.$$

In view of the convexity of $\varphi(\cdot,\tau) + \frac{\chi}{2}\|\cdot\|^2$ for each $\tau \in T$, this ensures that the function $\Psi(u) = f(u) + \frac{\chi}{2}\|u\|^2$ is convex and finite on Ω. □

Now, let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ be a given lower semicontinuous function and we suppose that the functions

$$f(\cdot) + \frac{\chi_i}{2}\|\cdot\|^2 \ \ (\chi_i \in (\chi, \bar{\chi}],\ \chi > 0)$$

are strongly convex on $\Omega \supset \{u \in \mathbb{R}^n : f(u) \leq f(\hat{u})\}$. Then, starting the exact proximal point method

$$u^{i+1} = \arg\min_{u \in \mathbb{R}^n}\{f(u) + \frac{\chi_i}{2}\|u - u^i\|^2\} \tag{2.10}$$

with $u^0 :\ f(u^0) \leq f(\hat{u})$, we have the known facts that

– u^{i+1} is uniquely defined;

– $0 \in \partial(f(u^{i+1}) + (\chi_i/2)\|u^{i+1} - u^i\|^2)$;

– $f(u^{i+1}) + (\chi_i/2)\|u^{i+1} - u^i\|^2 \leq f(u^i)$.

If $\inf_{u\in\mathbb{R}^n} f(u) > -\infty$, then the latter inequality yields

$$f(u^i) \to \bar{f} > -\infty \quad \text{and} \quad \|u^{i+1} - u^i\| \to 0.$$

Above the term $\partial\xi$ denotes the subdifferential of a convex function ξ,

$$\partial\left(f(u^{i+1}) + \frac{\chi_i}{2}\|u^{i+1} - u^i\|^2\right) \equiv \partial\left(f(\cdot) + \frac{\chi_i}{2}\|\cdot - u^i\|^2\right)(u^{i+1}).$$

It is also well-known that a nonlinear programming problem

$$\min\{f_0(u) : \ f_j(u) \leq 0, \ j = 1, ..., m\}, \qquad V = \mathbb{R}^n, \tag{2.11}$$

whose Lagrangian possesses a saddle point, can be transformed into the unconstrained problem

$$\min\{f(u) : \ u \in \mathbb{R}^n\},$$

with $f(u) = \max_{0\leq j\leq m} \eta_j(u)$, $\eta_0 = f_0$, $\eta_j = f_0 + \alpha f_j$, $j = 1, ..., m$ and a sufficiently large $\alpha > 0$.

Along with other applications this motivates to consider the proximal point method (2.10) for $f(u) = \max_{j\in J} \varphi_j(u)$, $|J| < \infty$, with differentiable functions φ_j. Concerning φ_j we suppose also that their gradients are Lipschitz-continuous (with constants L_j) on some convex set $\Omega \supset \{u : f(u) \leq f(\hat{u})\}$. Then, due to the finiteness of J, the conditions of Proposition 1 are satisfied.

THEOREM 1. *Let the suppositions made on φ_j be valid and χ_i be chosen such that* $\max_{j\in J} L_j < \chi_i \leq \bar{\chi}$. *Moreover, let* $\inf_{u\in\mathbb{R}^n} f(u) > -\infty$.
Then, starting Method (2.10) with $u^0 : f(u^0) \leq f(\hat{u})$, we obtain:
either

(i) $u^i = u^{i+1}$ holds for some i, providing that u^i is a stationary point of f,

or

(ii) any accumulation point of $\{u^i\}$ is a stationary point of f.

Proof. Due to the differentiability of φ_j and the convexity of $f(\cdot) + \frac{\chi_i}{2}\|\cdot - u^i\|^2$ on Ω the subdifferential $\partial(f(u^{i+1}) + (\chi_i/2)\|u^{i+1} - u^i\|^2)$ is the convex hull of the gradients at u^{i+1} of the functions $\varphi_j(\cdot) + (\chi_i/2)\|\cdot - u^i\|^2$, $j \in J(u^{i+1})$, where

$J(u^{i+1}) = \{j \in J : \varphi_j(u^{i+1}) = f(u^{i+1})\}$.
Hence, for some $\lambda_j^{i+1} \geq 0,\ j \in J(u^{i+1})$, such that $\sum_{j \in J(u^{i+1})} \lambda_j^{i+1} = 1$, we obtain

$$0 = \sum_{j \in J(u^{i+1})} \lambda_j^{i+1} \nabla \varphi_j(u^{i+1}) + \chi_i(u^{i+1} - u^i). \tag{2.12}$$

If $u^{i+1} = u^i$, then

$$0 = \sum_{j \in J(u^{i+1})} \lambda_j^{i+1} \nabla \varphi_j(u^{i+1}),$$

hence, the point 0 is included in Clarke's subdifferential $\partial_{cl} f(u^{i+1})$, proving that u^{i+1} is a stationary point of f. If $\{u^i\}$ is an infinite sequence and $\bar{u}$ is its accumulation point, then due to the finiteness of J and $0 \leq \lambda_j^i \leq 1\ \forall j \in J(u^i), \forall i$, we are able to choose a subsequence $\{i_k\}$ such that

$$J(u^{i_1+1}) = J(u^{i_2+1}) = \cdots = \bar{J},$$

and for $k \to \infty$,

$$u^{i_k+1} \to \bar{u} \quad \text{and} \quad \lambda_j^{i_k+1} \to \bar{\lambda}_j, \quad \forall j \in \bar{J}.$$

Moreover, because of $\inf_{u \in \mathbb{R}^n} f(u) > -\infty$, the relation

$$\lim_{i \to \infty} \|u^{i+1} - u^i\| = 0$$

is valid. Now taking limit in (2.12) for $i = i_k,\ k \to \infty$, one gets

$$0 = \sum_{j \in \bar{J}} \bar{\lambda}_j \nabla \varphi_j(\bar{u}), \quad \text{with } \bar{\lambda}_j \geq 0, \quad \sum_{j \in \bar{J}} \bar{\lambda}_j = 1.$$

On account of $J(\bar{u}) \supset \bar{J}$ and the property of Clarke's subdifferential, this means that $0 \in \partial_{cl} f(\bar{u})$, i.e. $\bar{u}$ is a stationary point of f. $\square$

REMARK 1. If the set $\{u :\ f(u) \leq f(u^0)\}$ is unbounded, the existence of accumulation points, in general, is not guaranteed. Moreover, for the case that Problem (1.1) is a convex one, it is known that $\|u^i\| \to \infty$ if $U^* = \emptyset$.

REMARK 2. If $\bar{u} = \lim_{k \to \infty} u^{i_k+1}$ and f is convex on $\tilde{U} = \{u : \|u - \bar{u}\| \leq r\}$ with some $r > 0$, then of course, $\bar{u}$ is a local minimum of f. In this case, on account of $\|u^{i+1} - u^i\| \to 0$, for sufficient large k one can claim that $\|u^{i_k+1} - \bar{u}\| \leq r/2$ and

$$\|u^{i_k+l+1} - u^{i_k+l}\| \leq \frac{r}{2}, \quad l = 1, 2, \ldots$$

Thus, $u^{i_k+2} \in \tilde{U}$ and hence u^{i_k+2} is the minimum point of the sum of the convex functions f and $(\chi_{i_k+1}/2)\| \cdot - u^{i_k+1}\|^2$ on $\tilde{U}$. From Proposition II.2.2 in [12] we obtain

$$f(\bar{u}) - f(u^{i_k+2}) + \chi_{i_k+1}\left(u^{i_k+2} - u^{i_k+1}, \bar{u} - u^{i_k+2}\right) \geq 0.$$

Together with the evident identity

$$\|u^{i_k+2}-\bar{u}\|^2-\|u^{i_k+1}-\bar{u}\|^2= -\|u^{i_k+2}- u^{i_k+1}\|^2 +2\left(u^{i_k+2}-u^{i_k+1}, u^{i_k+2}-\bar{u}\right)$$

this leads to

$$\|u^{i_k+2}-\bar{u}\|^2-\|u^{i_k+1}-\bar{u}\|^2\leq -\|u^{i_k+2}- u^{i_k+1}\|^2 +\frac{2}{\chi_{i_k+1}}\left(f(\bar{u})-f(u^{i_k+2})\right)$$

providing that $\|u^{i_k+2}-\bar{u}\|\leq\|u^{i_k+1}-\bar{u}\|$ and, in general,

$$\|u^{i_k+l+1}-\bar{u}\|\leq\|u^{i_k+l}-\bar{u}\|,\quad l=1,2,...$$

Hence, the sequence $\{\|u^i - \bar{u}\|\}$ converges, and in view of $\lim_{k\to\infty} u^{i_k+1}=\bar{u}$, one can conclude that $\lim_{i\to\infty} u^i=\bar{u}$.

It should be noted that we did not suppose that $\bar{u}$ is the unique minimum in $\tilde{U}$.

The following result is of interest, in particular, for semi-infinite programming problems, which often can be reduced to the unconstrained minimization of functions of the type $f(u)=\sup_{\tau\in T}\varphi(u,\tau)$ with T a compact set.

THEOREM 2. *Let the function f be defined by $f(u)=\sup_{\tau\in T}\varphi(u,\tau)$, where $T\subset\mathbb{R}^l$ is a compact set and φ is continuous on $\mathbb{R}^n\times T$. Moreover, suppose that $\inf_{u\in\mathbb{R}^n} f(u)>-\infty$ and that $\nabla_u\varphi$ is continuous on $\Omega\times T$, where Ω is an open convex set containing $\{u:\ f(u)\leq f(\hat{u})\}$. Finally, let*

$$\|\nabla_u\varphi(u',\tau)-\nabla_u\varphi(u'',\tau)\|\leq L\|u'-u''\|\quad\forall u',\quad u''\in\Omega,\quad\forall\tau\in T.$$

Then the function $f(\cdot)+\frac{\chi}{2}\|\cdot\|^2$ is strongly convex on Ω if $\chi>L$, and for the proximal point method (2.10) with $L<\chi_i\leq\bar{\chi}$ and the starting point $u^0: f(u^0)\leq f(\hat{u})$ the conclusion of Theorem 1 remains true.

Proof. Due to the compactness of T and the continuity of $\nabla_u\varphi$ on $\Omega\times T$, finiteness of $\sup_{\tau\in T}\|\nabla_u\varphi(u,\tau)\|$ is guaranteed for each $u\in\Omega$. Therefore, Proposition 1 implies strong convexity of $f(\cdot)+\frac{\chi}{2}\|\cdot\|^2$ on Ω for $\chi>L$.

Moreover, f is a locally Lipschitzian function and its upper Clarke's derivative coincides with the directional derivative (see [9], Example 2.1.4). Therefore, the theorem about Clarke's subdifferential for the sum of functions (cf. [8]) permits to conclude that, for $u\in\Omega$,

$$\partial_{cl}\left(f(u)+\frac{\chi_i}{2}\|u-u^i\|^2\right)=\partial_{cl}f(u)+\partial_{cl}\left(\frac{\chi_i}{2}\|u-u^i\|^2\right),$$

and with regard to the convexity of $f(\cdot)+(\chi/2)\|\cdot-u^i\|^2$ and $(\chi/2)\|\cdot-u^i\|^2$, this leads to

$$\partial\left(f(u)+\frac{\chi_i}{2}\|u-u^i\|^2\right)=\partial_{cl}f(u)+\chi_i(u-u^i).$$

Hence,

$$0\in\partial\left(f(u^{i+1})+\frac{\chi_i}{2}\|u^{i+1}-u^i\|^2\right)=\partial_{cl}f(u^{i+1})+\chi_i(u^{i+1}-u^i), \qquad (2.13)$$

and if $u^i=u^{i+1}$, then u^{i+1} is a stationary point of f.

If $\{u^i\}$ is an infinite sequence and $\bar{u}=\lim_{k\to\infty}u^{i_k}$, then the inclusion $0\in\partial_{cl}(f(\bar{u}))$ follows from (2.13), relation $\|u^{i+1}-u^i\|\to 0$ as well as from the closedness of the mapping $u\to\partial_{cl}f(u)$. □

REMARK 3. In principle, Theorem 1 could be considered as a particular case of Theorem 2, however, we preferred to give a direct proof.

Now, let us consider a very special case in nonconvex programming where it is possible to guarantee that the proximal point method generates a sequence $\{u^i\}$ converging to $\bar{u}\in \operatorname{Arg\,min}\{f(u):\ u\in\mathbb{R}^n\}$.
Let $f:\mathbb{R}^n\to\bar{\mathbb{R}}$ be a lower semicontinuous function and $\operatorname{dom} f\neq\emptyset$. With a given $c>\inf_{u\in\mathbb{R}^n}f(u)$, define

$$\Omega_c=\{u:\ f(u)\leq c\}.$$

We want to solve the problem

$$\min\{f(u):\ u\in\mathbb{R}^n\},$$

under the following

ASSUMPTION 1.

(i) $U^*=\{u:\ f(u)=\inf_{v\in\mathbb{R}^n}f(v)\equiv f^*\}\neq\emptyset$;

(ii) Ω_c is contained in a convex set Ω, and for some $\chi>0$ the function $\Psi(u)=f(u)+\frac{\chi}{2}\|u\|^2$ is convex on Ω;

(iii) for some $c_0\in[f^*,c)$ and χ chosen as in (ii) one has

$$\|y(u)-\chi u\|>d>0\quad \forall u\in\Omega_c\backslash\Omega_{c_0},\quad \forall y(u)\in\Lambda(u), \qquad (2.14)$$

where $\Lambda(u)=\partial\Psi(u)-\chi u$;

(iv) Ω_{c_0} is convex, and f is convex on Ω_{c_0}.

THEOREM 3. *The proximal iterates in (2.10) with $\chi<\chi_i\leq\bar{\chi}$ and arbitrary $u^0\in\Omega_c$ converge to a solution point $u^*\in U^*$.*

Proof. Let $u^0 \in \Omega_c \backslash \Omega_{c_0}$. Then it is clear that $u^i \in \Omega_c$ for all proximal steps i. Assume now that $u^{i+1} \in \Omega_c \backslash \Omega_{c_0}$. Due to

$$f(u) + \frac{\chi_i}{2}\|u - u^i\|^2 = \Psi(u) + \frac{\chi_i - \chi}{2}\|u\|^2 - \chi_i(u, u^i) + \frac{\chi_i}{2}\|u^i\|^2$$

and the definition of u^{i+1}, we obtain

$$0 \in \partial\Psi(u^{i+1}) - \chi u^{i+1} + \chi_i(u^{i+1} - u^i).$$

Observing (2.14), this leads to

$$\|u^{i+1} - u^i\| > \frac{d}{\bar{\chi}}.$$

Now, taking into account the inequality

$$f(u^{i+1}) + \frac{\chi_i}{2}\|u^{i+1} - u^i\|^2 \leq f(u^i),$$

one can conclude that

$$f(u^{i+1}) < f(u^i) - \frac{\chi}{2\bar{\chi}^2}d^2. \tag{2.15}$$

Estimate (2.15) shows that, after a finite number of steps, the prox-iterates fall into the set Ω_{c_0}. With regard to Assumption 1(iv) and the non-increase of $\{f(u^i)\}$, the use of the convergence results for convex problems (see the introduction part) ensures convergence of $\{u^i\}$ to a solution $u^* \in U^*$. The same happens if $u^0 \in \Omega_{c_0}$. □

EXAMPLE 3. (Showing the fulfilment of Assumption 1):

$$V = \mathbb{R}^2, \qquad f(u) = u_1^2 + u_2^2 + 15u_1^2u_2^2.$$

Obviously, Assumption 1(i) is valid: $u^* = (0, 0)$ is the unique minimum. The function f is nonconvex, but it is strongly convex on the sphere $\{u \in \mathbb{R}^2 : \|u\| \leq 1/\sqrt{15}\}$. Outside of this sphere we have $\|\nabla f(u)\| > 2/\sqrt{15}$. Therefore, the Assumptions 1(iii) and (iv) are satisfied, for instance with $c = 2$ and $c_0 = 1/15$. As it follows from Proposition 1, Assumption 1(ii) can also be satisfied for any convex compact set Ω containing Ω_c with $c = 2$.

REMARK 4. Insignificant modifications in the proofs of the Theorems 1-3 permit to establish analogous statements for the inexact version of the proximal point method (1.2) with stopping rules (1.4).

We finish this section by considering the constrained problem

$$\min\{f(u) : u \in K\}, \tag{2.16}$$

where $K \subset \mathbb{R}^n$ is a convex closed set and f is twice differentiable on K. Let us introduce the operator

$$\mathcal{T}u := \begin{cases} \nabla f(u) + N_K(u) & \text{if } u \in K \\ \emptyset & \text{if } u \notin K, \end{cases}$$

with $N_K(u) = \{v \in \mathbb{R}^n : (v, z - u) \leq 0 \ \forall z \in K\}$ the normal cone of the set K at the point $u \in K$.

ASSUMPTION 2.
- (i) For a given $c > \inf\{f(u) : u \in K\}$ the set $Q = \{u : 0 \in \mathcal{T}u, \ f(u) \leq c\}$ is nonempty, and if $\bar{u} \in Q$, then $\bar{u}$ is a local minimum of f on K;
- (ii) For some $\bar{d} > 0$ the set $\{u : f(u) \leq c, \ \inf_{y \in \mathcal{T}u} \|y\| < \bar{d}\}$ is bounded.

PROPOSITION 2. *Suppose that Assumption 2 is satisfied for Problem (2.16). Then for each $\delta > 0$ there exists $d \in (0, \bar{d})$ such that u belongs to the δ-neighborhood $U_\delta(\bar{u})$ of some local minimum $\bar{u} \in Q$, whenever $u \in K$, $f(u) \leq c$ and $\inf_{y \in \mathcal{T}} \|y\| < d$.*

Proof. Suppose that for some $\delta > 0$ such a constant d does not exist. Denoting by U_δ the union of δ-neighborhoods of all local minima in Q, we choose a sequence $\{d_i\} \to +0$, $d_i < \bar{d}$ and define $w^i \in K \backslash U_\delta$, $y^i \in \mathcal{T}w^i$ such that

$$f(w^i) \leq c, \qquad \|y^i\| < d_i.$$

Due to Assumption 2(ii), without loss of generality, one can assume that $\{w^i\}$ converges to a point $\bar{w}$. For this point we infer that

$$\bar{w} \in K \quad \text{and} \quad f(\bar{w}) \leq c. \tag{2.17}$$

Because of

$$y^i = \nabla f(w^i) + v^i \quad \text{for some } v^i \in N_K(u^i) \quad \text{and}$$
$$\nabla f(w^i) \to \nabla f(\bar{w}), \quad \|y^i\| \to 0,$$

one gets $v^i \to \bar{v} = -\nabla f(\bar{w})$. But from the definition of the normal cone $N_K(w^i)$, the inequality

$$(v^i, w^i - z) \geq 0 \quad \forall z \in K$$

holds true, and taking limit, we infer

$$(\bar{v}, \bar{w} - z) \geq 0 \quad \forall z \in K,$$

i.e. $\bar{v} \in N_K(\bar{w})$. Therefore, $0 = \nabla f(\bar{w}) + \bar{v} \in \mathcal{T}\bar{w}$, and due to (2.17) and Assumption 2(i), $\bar{w}$ is a local minimum, contradicting the fact that $\{w^i\} \cap U_\delta = \emptyset$. □

Suppose now that for Problem (2.16) the set $\{u \in K : f(u) = \inf_{v \in K} f(v)\}$ is nonempty and that for some $\chi > 0$ the function $\Psi(u) = f(u) +$

$\frac{\chi}{2}\|u\|^2$ is convex on K. (Due to Proposition 1, these conditions are fulfilled, in particular, if K is a non-empty and bounded set). Then, obviously, the function

$$\zeta(u) := \begin{cases} \Psi(u) & \text{if } u \in K \\ +\infty & \text{if } u \notin K \end{cases}$$

is convex, lower semicontinuous and $\partial\zeta(u) - \chi u = \mathcal{T}u$. If, moreover, Assumption 2 is valid, then using Proposition 2 and the proof of Theorem 3, one can easily show that, for arbitrary $\delta > 0$, the iterates u^i, generated by the exact method (1.2) with $\chi < \chi_i \leq \bar{\chi}$ and the starting point $u^0 \in K: \ f(u^0) \leq c$, belong to U_δ for sufficiently large i $(i \geq i(\delta))$.

EXAMPLE 4. $V = \mathbb{R}^2, \ K = \{u \in \mathbb{R}^2 : -1 \leq u_1 \leq 1, \ -1 \leq u_2 \leq 1\}$, $f(u) = (1 - u_1^2)u_2^2$. Evidently, we have

$$U^* = \{(u_1, 0) : -1 \leq u_1 \leq 1\} \cup \{(1, u_2) : -1 \leq u_2 \leq 1\} \cup \{(-1, u_2) : -1 \leq u_2 \leq 1\}$$

and

$$\{u \in K : \nabla f(u) = 0\} = \{(u_1, 0) : -1 \leq u_1 \leq 1\} \subset U^*.$$

Now, calculating ∇f on the boundary of K, it is easy to see that Assumption 2(i) is fulfilled with any $c > 0$, and Assumption 2(ii) follows from the boundedness of K.

Of course, it should be emphasized that the verification of the Assumptions 1(iii), (iv) and 2(i) causes difficulties. However, these conditions are not binding for treating proximal point methods. Nevertheless, the complete validity of Assumption 2 or 1 ensures a much more comfortable result, namely, according to Theorem 1 and 2, convergence to a local or global minimum.

3. Proximal point ideas in methods of nonconvex optimization

A retrospective analysis of some algorithms for nonconvex minimization shows their relationship to proximal methods. This concernes, in particular, a series of linearization methods. Let us consider briefly the linearization method suggested by Pshenichnyi [25]. Its application to Problem (2.11) with continuously differentiable functions $f_j, \ j = 0, ..., m$, can be described as follows.

Given the numbers $\delta \geq 0, \ \nu > 0, \ \alpha \in (0, 1)$ and a starting point u^0, define

$$F(u) = \max\{0, f_1(u), ..., f_m(u)\},$$

$$J_\delta(u) = \{j \geq 1 : \ f_j(u) \geq F(u) - \delta\},$$

$$\Omega_\nu = \{u : \ f_0(u) + \nu F(u) \le f_0(u^0) + \nu F(u^0)\}.$$

We suppose that the set Ω_ν is bounded and the gradients $\nabla f_0, \nabla f_1, \dots,$ ∇f_m are Lipschitz-continuous on Ω_ν. Moreover, let for each $u \in \Omega_\nu$ the quadratic programming problem

$$\mathcal{P}(u) \ \min \left\{ (\nabla f_0(u), d) + \frac{1}{2}\|d\|^2 : (\nabla f_j(u), d) + f_j(u) \le 0, \ \forall j \in J_\delta(u) \right\},$$

be solvable, and $\sum_{j \in J_\delta(u)} \lambda_j(u) \le \nu$ hold for some Lagrange multipliers $\lambda_j(u)$ of $\mathcal{P}(u)$.

Now, let the iterate u^i be given, then the $(i+1)$-th step of the method consists of the following substeps:

1. Solve the quadratic problem

$$\begin{aligned} &\min \{(\nabla f_0(u^i), u) + \frac{1}{2}\|u - u^i\|^2\} \\ &\text{s.t. } (\nabla f_j(u^i), u - u^i) + f_j(u^i) \le 0, \quad j \in J_\delta(u^i); \end{aligned} \tag{3.18}$$

2. With the solution $\tilde{u}^{i+1}$ of Problem (3.18) define the smallest integer $k \ge 0$ satisfying

$$\begin{aligned} &f_0\left(u^i + \left(\frac{1}{2}\right)^k (\tilde{u}^{i+1} - u^i)\right) + \nu F\left(u^i + \left(\frac{1}{2}\right)^k (\tilde{u}^{i+1} - u^i)\right) \\ &\le f_0(u^i) + \nu F(u^i) - \left(\frac{1}{2}\right)^k \alpha \|\tilde{u}^{i+1} - u^i\|^2; \end{aligned} \tag{3.19}$$

3. Put $u^{i+1} = u^i + \left(\frac{1}{2}\right)^k (\tilde{u}^{i+1} - u^i)$.

Obviously, calculation of the current point $\tilde{u}^{i+1}$ is nothing else than one step of the proximal point method applied to a linear model of Problem (2.11) in a neighborhood of u^i. Substep 2 prevents from a ‘non-local’ application of this approximation.

Under the assumptions given above, each accumulation point of $\{u^i\}$ satisfies the necessary optimality conditions for Problem (2.11).

As far as we know, the first direct generalization of the proximal point method for certain nonconvex minimization problems was performed by Fukushima and Mine [13]. They considered the unconstrained problem

$$\min \{f(u) + \phi(u) : \ u \in \mathbb{R}^n\},$$

where $\phi : \mathbb{R}^n \to \bar{\mathbb{R}}$ is a proper convex, lower semicontinuous function, and $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is continuously differentiable on an open set including dom ϕ. Given a starting point $u^0 \in \operatorname{dom} \phi$ and regularization parameters $\chi_i \in (\chi, \bar{\chi}),\ \chi > 0$, the $(i+1)$-th step of the method calculates

$$\tilde{u}^{i+1} = \arg\min \left\{ (\nabla f(u^i), u) + \phi(u) + \frac{\chi_i}{2} \|u - u^i\|^2 :\ u \in \mathbb{R}^n \right\},$$

and

$$u^{i+1} = u^i + \alpha_i (\tilde{u}^{i+1} - u^i),$$

where the step-size α_i is defined by means of the Armijo rule, too.

As a particular case, if $\phi \equiv 0$, a gradient type method is obtained, and if $f \equiv 0$ then the choice $\alpha_i \equiv 1$ is possible leading to the usual proximal point method for convex problems.

In [13] it is proved that each accumulation point of $\{u^i\}$ is a stationary point of the function $f + \phi$ if, in particular, the set

$$D = \{u :\ f(u) + \phi(u) \le f(u^0) + \phi(u^0)\}$$

is bounded, $D \subset ri \operatorname{dom} \phi$ and $f + \phi$ is Lipschitz-continuous on D.

Spingarn [29] has developed the proximal point method for finding a zero of a maximal strictly hypomonotone operator $\mathcal{T} : \mathbb{R}^n \to 2^{\mathbb{R}^n}$. In this method, proximal iterations are performed not with the operator $\mathcal{T}$, but with an auxiliary Lipschitz continuous operator constructed in a convex neighborhood U of a point $\bar{u} : 0 \in \mathcal{T}\bar{u}$. The method, adapted in [29] to the problem of the local minimization of a function $f = g - h$, with g lower semicontinuous, convex and h of class C^2, turns into the 'usual' proximal point method

$$u^{i+1} = \arg\min_{u \in U} \left\{ f(u) + \frac{\chi}{2} \|u - u^i\|^2 \right\}.$$

If $\chi > 0$ is chosen such that $f(\cdot) + (\chi/2)\|\cdot\|^2$ is convex on U, and u^0 is chosen close enough to a local minimum $\bar{u}$ of f, and if the mapping ∂f^{-1} has a monotone derivative at $(0, \bar{u})$, then $\{u^i\}$ converges to $\bar{u}$ linearly.

A very important result in [29] is: The assumption that ∂f^{-1} has a monotone derivative at $(0, \bar{u})$ is generically fulfilled (more precisely, this assertion is proved at classes of functions

$$f_v(u) = f(u) - (v, u),$$

where f is as above and $v \in \mathbb{R}^n$ is a parameter).

Finally, let us observe Kiwiel's proximal bundle method [21] destined for the nonconvex nonsmooth problem

$$\min \{f(u) :\ u \in \mathbb{R}^n\}. \tag{3.20}$$

Here it is supposed that $f : \mathbb{R}^n \to \mathbb{R}$ is a locally Lipschitzian function and that we are able to compute $f(u)$ and an arbitrary element $g(u) \in \partial_{cl} f(u)$. For locally Lipschitzian functions it holds

$$\partial_{cl} f(u) \equiv co\,\{\, l : \ v^i \to u, \ \nabla f(v^i) \text{ exists and } \nabla f(v^i) \to l\},$$

co denotes the convex hull.

This bundle method can be described conceptually as follows: In the $(i+1)$-th step let the points $u^k, \ y^k, \ k = 0, ..., i$, and the set $J^i \subset \{1, ..., i\}$ be given, where $u^0 = y^0$ are arbitrarily chosen and $J^0 = \{0\}$.
A polyhedral model of f is defined by

$$\check{f}^{i+1}(u) = \max\,\{f(u^i) - \alpha(u^i, y^j) + (g(y^j), u - u^i) : \ j \in J^i\}, \tag{3.21}$$

with $\alpha(u, y) = |f(u) - f(y) - (g(y), u - y)|$. Choosing $\chi_i > 0$, at $(i+1)$-th step a descent direction d^{i+1} has to be calculated by means of

$$d^{i+1} = \arg\min \left\{ \check{f}^{i+1}(u^i + d) + \frac{\chi_i}{2}\|d\|^2 : d \in \mathbb{R}^n \right\}. \tag{3.22}$$

Then calculate

$$u^{i+1} = u^i + t_L^i d^i, \qquad y^{i+1} = u^i + t_R^i d^i,$$

with $0 < t_L^i = t_R^i \le 1$, if a linear search allows to find u^{i+1} such that $f(u^{i+1})$ is 'substantially less' than $f(u^i)$; otherwise set $t_L^i = 0$ (i.e. $u^{i+1} = u^i$), and the choice of the step-size $t_R^i \in (0, 1]$ has to ensure that the next model $\check{f}^{i+2}$ with $i + 1 \in J^{i+1}$ approximates f 'substantially better' than $\check{f}^{i+1}$, at least at the point $u^i + d^{i+1}$. The new set J^{i+1} is chosen such that $J^{i+1} \subset J^i \cup \{i + 1\}$.

Due to the nonconvexity of f we have to take into account:

- in general, $\check{f}^{i+1}$ is a useful local approximation of f only when the points $y^j, \ j \in J^i$, are close to u^i. This enforces to choose for J^i only a part $\check{J}^{i-1}$ of J^{i-1} ($J^i = \check{J}^{i-1} \cup \{i\}$) such that $y^j, \ j \in \check{J}^{i-1}$, are close to u^i;
- the choice χ_i has to provide that $u^i + d^{i+1}$ belongs to some trust region $\{u : \|u - u^i\| \le \rho_i\}$, where $\check{f}^{i+1}$ is close to f. Concerning a strategy for varying ρ_i in dependence on the properties of f, see, for instance [10].

Avoiding details, the method considered can be interpreted as the following modification of the proximal point method:

$$\tilde{u}^{i+1} = \arg\min_{u \in \mathbb{R}^n} \left\{ \check{f}^{i+1}(u) + \frac{\chi_i}{2}\|u - u^i\|^2 \right\}, \tag{3.23}$$

$$u^{i+1} = u^i + t_L^i(\tilde{u}^{i+1} - u^i). \tag{3.24}$$

Thus, formally this method follows the scheme of iterative proximal regularization (1.8), (1.9) developed for convex problems.

Note that the polyhedral functions $\check{f}^i$ are convex, and $\tilde{u}^{i+1}$ can be found by solving the quadratic programming problem

$$\begin{aligned}&\zeta + \frac{\chi_i}{2}\|u - u^i\|^2 \to \min \\ &\text{s.t.} - \alpha(u^i, y^j) + (g(y^j), u - u^i) \le \zeta, \quad \forall j \in J^i.\end{aligned}$$

If the initially chosen χ_i is not suitable, i.e. $\|\tilde{u}^{i+1} - u^i\| > \rho_i$, one can use a simple procedure to vary χ_i until $\|\tilde{u}^{i+1} - u^i\| \approx \rho_i$ is achieved (cf. [20]). This procedure is based on the important property that a trajectory

$$u(\chi) = \arg\min \left\{ \check{f}^{i+1}(u) + \frac{\chi}{2}\|u - u^i\|^2 : \; u \in \mathbb{R}^n \right\}$$

describing the dependence of the proximal point $u(\chi)$ on χ is continuous and piecewise linear in $1/\chi$ [28].

Kiwiel's method can be interpreted as a *trust region method* performed with polyhedral models for the objective function, the Euclidean norm to set up the trust region and a special procedure for minimizing a polyhedral function on this trust region.

In [21] it is proved that each accumulation point of $\{u^i\}$ is a stationary point of Problem (3.20). Moreover, in case the function f is convex, Kiwiel's algorithm possesses the characteristic property of the proximal point method:

$$\text{either } u^i \to \bar{u} \in \operatorname{Arg\,min} f, \quad \text{or} \quad \operatorname{Arg\,min} f = \emptyset \quad \text{and} \quad \|u^i\| \to \infty,$$

and $f(u^i) \to \inf_{u\in\mathbb{R}^n} f(u)$ takes place.

Acknowledgments

We are indebted to an anonymous referee for very helpful remarks.

References

1. Alart, P. and Lemaire, B. (1991), Penalization in non-classical convex programming via variational convergence, *Math. Programming* 51: 307–331.
2. Antipin, A.S. (1976), On a method for convex programs using a symmetrical modification of the Lagrange function, *Ekonomika i Mat. Metody* 12: 1164–1173.
3. Auslender, A. (1987), Numerical methods for nondifferentiable convex optimization, *Math. Programming Study* 30: 102–126.
4. Auslender, A., Crouzeix, J.P. and Fedit, P. (1987), Penalty proximal methods in convex programming, *J. Opt. Theory Appl.* 55: 1–21.
5. Censor, Y. and Zenios, S.A. (1992), Proximal minimization algorithm with D-functions, *J. Optim. Theory Appl.* 73: 451–464.
6. Chen, G. and Teboulle, M. (1993), Convergence analysis of a proximal-like minimization algorithm using Bregman functions, *SIAM J. Optim.* 3: 538–543.

7. Chen, G. and Teboulle, M. (1994), A proximal-based decomposition method for convex minimization problems, *Math. Programming* 64: 81–101.
8. Clarke, F.H. (1983), *Optimization and Nonsmooth Analysis*, Wiley, New York.
9. Demyanov, V.F. and Rubinov, A.M. (1990), *Fundations of Nonsmooth Analysis and Quasidifferential Calculus*, Nauka, Moscow (in Russian).
10. Dennis, J.E. and Schnabel, R.B. (1989), A view of unconstrained optimization, in G.L. Nemhauser, A.H.G. Rinnooy-Kan and M.J. Todd (eds.), *Optimization – Handbooks of Operations Research and Management Science*, Vol. 1, pp. 1–72, North-Holland, Amsterdam.
11. Eckstein, J. and Bertsekas, D. (1992), On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, *Math. Programming* 55: 293–318.
12. Ekeland, I. and Temam, R. (1976), *Convex Analysis and Variational Problems*, North-Holland, Amsterdam.
13. Fukushima, M. and Mine, H. (1981), A generalized proximal point algorithm for certain non-convex minimization problems, *Int. J. Systems Sci.* 12: 989–1000.
14. Ibaraki, S., Fukushima, M. and Ibaraki, T. (1992), Primal-dual proximal point algorithm for linearly constrained convex programming problems, *Comput. Opt. Appl.* 1: 207–226.
15. Iusem, A.N., Svaiter, B.F. and Teboulle, M. (1994) Entropy-like proximal methods in convex programming, *Math. Oper. Res.* 19: 790–814.
16. Kaplan, A. and Tichatschke, R. (1992) A regularized penalty method for solving convex semi-infinite programs, *Optimization* 26: 215–228.
17. Kaplan, A. and Tichatschke, R. (1994), *Stable Methods for Ill-Posed Variational Problems - Prox-Regularization of Elliptical Variational Inequalities and Semi-Infinite Optimization Problems*, Akademie Verlag, Berlin.
18. Kaplan, A. and Tichatschke, R. (1996), A note on the paper 'Multi-step-prox regularization methods for solving convex variational problems', *Optimization* 37: 149–152.
19. Kaplan, A. and Tichatschke, R. (1997), Prox-regularization and solution of ill-posed elliptic variational inequalities, *Applications of Mathematics* 42: 111–145.
20. Kiwiel, K. (1989), An ellipsoid trust region bundle method for nonsmooth convex minimization, *SIAM J. Control Optim.* 27: 737–757.
21. Kiwiel, K. (1996), Restricted step and Levenberg–Marquardt techniques in proximal bundle methods for non-convex nondifferentiable optimization, *SIAM J. Optimization* 6: 227–249.
22. Lemaire, B. (1988), Coupling optimization methods and variational convergence, *ISNM* 84: 163–179.
23. Lemaire, B. (1995), On the convergence of some iterative methods for convex minimization, in R. Durier and Ch. Michelot (eds.), *Recent Developments in Optimization*, pp. 252–268 [Lecture Notes in Econ. and Mathem. Systems, vol. 429], Springer Verlag, Berlin.
24. Martinet, B. (1970), Régularisation d'inéquations variationelles par approximations successives, *RIRO* 4: 154–159.
25. Pshenichnyi, B.N. (1983), *Methods of Linearization*, Nauka, Moscow (in Russian).
26. Rockafellar, R.T. (1976), Augmented Lagrange multiplier functions and applications of the proximal point algorithm in convex programming, *Math. Oper. Res.* 1: 97–116.
27. Rockafellar, R.T. (1976), Monotone operators and the proximal point algorithm, *SIAM J. Control and Optimization* 14: 877–898.
28. Schramm, H. and Zowe, J. (1988), A combination of the bundle approach and the trust region concept, in J. Guddat et al. (eds.), *Advances in Mathematical Optimization*, Vol. 45, pp. 196–209, Akademie-Verlag, Berlin.
29. Spingarn, J.E. (1981–1982), Submonotone mappings and the proximal point algorithms, *Numer. Funct. Anal. and Optimiz.* 4: 123–150.