17 PROXIMAL-LIKE METHODS FOR CONVEX MINIMIZATION PROBLEMS

Christian Kanzow

University of Würzburg
Institute of Applied Mathematics and Statistics
Am Hubland
97074 Würzburg
Germany
e-mail: kanzow@mathematik.uni-wuerzburg.de

Abstract: This paper gives a brief survey of some proximal-like methods for the solution of convex minimization problems. Apart from the classical proximal-point method, it gives an introduction to several proximal-like methods using Bregman functions, φ-divergences etc. and discusses a couple of recent developments in this area. Some numerical results for optimal control problems are also included in order to illustrate the numerical behaviour of these proximal-like methods.

Key words: Convex minimization, proximal-point method, Bregman functions, φ-divergences, global convergence.

1 INTRODUCTION

This paper gives a brief survey of some proximal-like methods for the solution of convex minimization problems. To this end, let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a closed, proper, convex function, and consider the associated optimization problem

$$\min f(x), \quad x \in \mathbb{R}^n. \tag{1.1}$$

Formally, this is an unconstrained problem. However, since f is allowed to be extended-valued, any constrained problem of the form

$$\min \tilde{f}(x) \quad \text{subject to} \quad x \in X$$

for some convex function $\tilde{f} : \mathbb{R}^n \to \mathbb{R}$ and a closed, nonempty and convex set $X \subseteq \mathbb{R}^n$ can easily be transformed into a minimization problem of the form (1.1) by defining

$$f(x) := \begin{cases} \tilde{f}(x), & \text{if } x \in X, \\ +\infty, & \text{if } x \notin X. \end{cases}$$

Hence, theoretically, there is no loss of generality by considering the unconstrained problem (1.1). In fact, many theoretical results can be obtained in this (unifying) way for both unconstrained and constrained optimization problems. The interested reader is referred to the classical book by Rockafellar (1970) for further details.

Despite the fact that extended-valued functions allow such a unified treatment of both unconstrained and constrained problems, they are typically not tractable from a numerical point of view. Therefore, numerical algorithms for the solution of a problem like (1.1) have to take into account the constraints explicitly or, at least, some of these constraints. This can be done quite elegantly by so-called proximal-like methods. Similar to interior-point algorithms, these methods generate strictly feasible iterates and are usually applied to the problem with nonnegativity constraints

$$\min f(x) \quad \text{s.t.} \quad x \geq 0 \tag{1.2}$$

or to the linearly constrained problem

$$\min f(x) \quad \text{s.t.} \quad A^T x \leq b, \tag{1.3}$$

where, in the latter case, $A \in \mathbb{R}^{n \times m}$ and $b \in \mathbb{R}^m$ are the given data.

While we concentrate on the application of proximal-like methods to optimization problems, we should at least mention that proximal-like methods may also be applied to several other problem classes like nonlinear systems of equations, complementarity problems, variational inequalities and generalized equations. The interested reader is referred to Lemaire (1989); Lemaire (1992); Eckstein (1998); Censor et al. (1998); Auslender et al. (1999b); Solodov and Svaiter (2000b) and the corresponding references in these papers for more details.

This paper is organized in the following way: Section 2 first reviews the classical proximal-point method and then describes several proximal-like methods for the solution of the constrained optimization problems (1.2) and (1.3). In Section 3, we then present some numerical results obtained with some of these proximal-like methods when applied to some classes of optimal control problems. We then conclude with some final remarks in Section 4.

The notation used in this paper is quite standard: $\mathbb{R}^n$ denotes the n-dimensional Euclidean space, inequalities like $x \geq 0$ or $x > 0$ for a vector $x \in \mathbb{R}^n$ are defined componentwise, $\mathbb{R}^n_{++} := \{x \in \mathbb{R}^n \mid x > 0\}$ denotes the strictly positive orthant, and $\overline{S}$ is the closure of a subset $S \subseteq \mathbb{R}^n$.

2 PROXIMAL-LIKE METHODS

Throughout this section, we make the blanket assumption that $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a closed, proper and convex function.

2.1 Classical Proximal-Point Method

The classical proximal-point method was introduced by Martinet (1970) and further developped by Rockafellar (1976) and others, see, e.g., the survey by Lemaire (1989). Being applied to the minimization problem (1.1), it generates a sequence $\{x^k\} \subseteq \mathbb{R}^n$ such that x^{k+1} is a solution of the subproblem

$$\min f(x) + \frac{1}{2\lambda_k}\|x - x^k\|^2 \tag{2.1}$$

for $k = 0, 1, \ldots$; here, λ_k denotes a positive number. The objective function of this subproblem is strictly convex since it is the sum of the original (convex) objective function f and a strictly convex quadratic term. This term is usually called the regularization term.

This strictly convex regularization term guarantees that the subproblem (2.1) has a unique minimizer for each $k \in \mathbb{N}$. Hence the classical proximal-point method is well-defined. Furthermore, it has the following global convergence properties, see, e.g., Güler (1991) for a proof of this result.

Theorem 2.1 *Let $\{x^k\}$ and $\{\lambda_k\}$ be two sequences generated by the classical proximal-point method (2.1), define $\sigma_k := \sum_{j=0}^{k} \lambda_j$, let $f_* := \inf\{f(x) \mid x \in \mathbb{R}^n\}$ be the optimal value and $\mathcal{S} := \{x^* \in \mathbb{R}^n \mid f(x^*) = f_*\}$ be the solution set of (1.1). Assume that $\sigma_k \to \infty$. Then the following statements hold:*

(a) The sequence of function values $\{f(x^k)\}$ converges to the optimal value f_*.

(b) If $\mathcal{S} \neq \emptyset$, then the entire sequence $\{x^k\}$ converges to an element of $\mathcal{S}$.

Theorem 2.1 states some very strong convergence properties under rather weak conditions. In particular, it guarantees the convergence of the entire sequence $\{x^k\}$ even if the solution set $\mathcal{S}$ contains more than one element; in fact, this statement also holds for an unbounded solution set $\mathcal{S}$. Note that the assumption $\sigma_k \to \infty$ holds, for example, if the sequence $\{\lambda_k\}$ is constant, i.e., if $\lambda_k = \lambda$ for all $k \in \mathbb{N}$ and some positive number λ.

We note that many variations of Theorem 2.1 are available in the literature. For example, it is not necessary to compute the exact minimizer of the subproblems (2.1) at each step, see Rockafellar (1976) for some criteria under which inexact solutions still provide similar global convergence properties. It should be noted, however, that the criteria of inexactness in Rockafellar (1976) are not implementable in general since they assume some knowledge regarding the exact solution of (2.1). On the other hand, Solodov and Svaiter (1999) recently gave a more constructive criterion in a slightly different framework.

Furthermore, some rate of convergence results can be shown for the classical proximal-point method under a certain error bound condition, cf. Luque (1984). This error bound condition holds, for example, for linearly constrained problems due to Hoffman's error bound Hoffman (1952). Moreover, being applied to linear programs, it is known that the classical proximal-point method has a finite termination property, see Ferris (1991) for details.

We also note that the classical proximal-point method can be extended to infinite-dimensional Hilbert spaces for which weak convergence of the iterates $\{x^k\}$ can be shown, see Rockafellar (1976); Güler (1991) once again. Strong convergence does not hold without any further modifications as noted by Güler (1991) who provides a counterexample. On the other hand, using a rather simple modification of the classical proximal-point method, Solodov and Svaiter (2000a) were able to present a strongly convergent version of the classical proximal-point method in Hilbert spaces.

In contrast to the classical proximal-point method, the proximal-like methods to be presented in our subsequent discussion have only been presented in the finite dimensional setting; this is mainly due to the fact that they typically involve some logarithmic functions.

2.2 Proximal-like Methods Using Bregman Functions

The simple idea which is behind each proximal-like method for the solution of convex minimization problems is, more or less, to replace the strictly quadratic term in the regularized subproblem (2.1) by a more general strictly convex function. Later on, we will see that this might be a very useful idea when solving constrained problems.

There are quite a few different possibilities to replace the term $\frac{1}{2}\|x - x^k\|^2$ by another strictly convex distance-like function. The one we discuss in this subsection is defined by

$$D_\psi(x, y) := \psi(x) - \psi(y) - \nabla\psi(y)^T(x - y),$$

where ψ is a so-called Bregman function. According to Solodov and Svaiter (2000a), a Bregman function may be defined in the following way.

Definition 2.1 *Let $S \subseteq \mathbb{R}^n$ be an open and convex set. A mapping $\psi : \overline{S} \to \mathbb{R}$ is called a* Bregman function *with zone S if it has the following properties:*

(i) ψ is strictly convex and continuous on $\overline{S}$;

(ii) ψ is continuously differentiable in S;

(iii) The partial level set

$$\mathcal{L}_\alpha(x) := \left\{y \in \overline{S} \,|\, D_\psi(x, y) \leq \alpha\right\}$$

is bounded for every $x \in \overline{S}$;

(iv) If $\{y^k\} \subseteq S$ *converges to* x, *then* $\lim_{k\to\infty} D_\psi(x, y^k) = 0$.

Earlier papers on Bregman functions require some additional properties, see De Pierro and Iusem (1986); Censor and Zenios (1992); Eckstein (1993); Güler (1994); Censor et al. (1998). However, as noted in Solodov and Svaiter (2000a), all these additional properties follow from those mentioned in Definition 2.1.

Two simple examples of Bregman functions are:

$$\begin{aligned} \psi_1(x) &:= \frac{1}{2}\|x\|^2 \quad \text{on } S = \mathbb{R}^n, \\ \psi_2(x) &:= \sum_{i=1}^{n} x_i \log x_i - x_i \quad \text{on } S = \mathbb{R}^n_{++} \quad \text{(convention: } 0 \log 0 = 0\text{)}. \end{aligned}$$

Using $\psi = \psi_1$ in the definition of D_ψ, we reobtain the subproblem (2.1) of the classical proximal-point method. On the other hand, using $\psi = \psi_2$ in the definition of D_ψ gives the so-called *Kullback-Leibler relative entropy function*

$$D_{\psi_2}(x, y) = \sum_{i=1}^{n} x_i \log \frac{x_i}{y_i} + y_i - x_i. \tag{2.2}$$

This function may be used in order to solve the constrained optimization problem (1.2) by generating a sequence $\{x^k\}$ in such a way that x^{k+1} is a solution of the subproblem

$$\min f(x) + \frac{1}{\lambda_k} D_\psi(x, x^k), \quad x > 0 \tag{2.3}$$

for $k = 0, 1, 2, \ldots$, where $x^0 > 0$ is a strictly feasible starting point. Then a convergence result completely identical to Theorem 2.1 can be shown for this method, see Chen and Teboulle (1993) for details. Some related papers dealing with Bregman functions in the context of proximal-like methods are Eckstein (1993); Güler (1994); Eckstein (1998); Censor et al. (1998), where the interested reader will find rules which allow inexact solutions of the subproblems (2.3) and where he will also find some rate of convergence results. In contrast to the classical proximal-point method, however, the proximal-like methods do not possess any finite termination properties for linear programs.

On the other hand, a major advantage of the proximal-like methods is that a subproblem like (2.3) is essentially unconstrained and can therefore be solved by unconstrained minimization techniques.

2.3 *Proximal-like Methods Using φ-Divergences*

Another variant of the classical proximal-point method is a proximal-like method based on so-called φ-divergences. These φ-divergences will be used in order to replace the strictly convex quadratic term in the subproblem (2.1) of the classical proximal-point method.

Definition 2.2 *Let Φ denote the class of closed, proper and convex functions $\varphi : \mathbb{R} \to (-\infty, +\infty]$ with $dom(\varphi) \subseteq [0, +\infty)$ having the following properties:*

(i) φ is twice continuously differentiable on $int(dom\varphi) = (0, +\infty)$;

(ii) φ is strictly convex on its domain;

(iii) $\lim_{t \to 0+} \varphi'(t) = -\infty$;

(iv) $\varphi(1) = \varphi'(1) = 0$ and $\varphi''(1) > 0$;

(v) There exists $\nu \in (\frac{1}{2}\varphi''(1), \varphi''(1))$ such that

$$(1 - 1/t)(\varphi''(1) + \nu(t-1)) \leq \varphi'(t) \leq \varphi''(1)(t-1) \quad \forall t > 0.$$

Then the φ-divergence *corresponding to a mapping $\varphi \in \Phi$ is defined by*

$$d_\varphi(x, y) := \sum_{i=1}^{n} y_i \varphi\left(\frac{x_i}{y_i}\right)$$

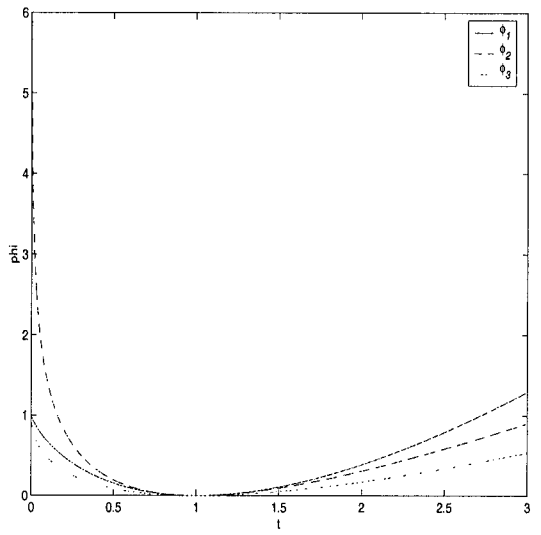
for $x, y \in \mathbb{R}^n_{++}$.

Sometimes condition (v) is not required in Definition 2.2. However, the corresponding convergence results are weaker without this additional property.

Some examples of functions $\varphi \in \Phi$ are

$$\begin{aligned} \varphi_1(t) &:= t \log t - t + 1, \\ \varphi_2(t) &:= -\log t + t - 1, \\ \varphi_3(t) &:= \left(\sqrt{t} - 1\right)^2. \end{aligned}$$

The graphs of these three functions are shown in the following figure.



In particular, if we choose $\varphi = \varphi_1$ in the definition of d_φ, we obtain

$$d_{\varphi_1}(x, y) = \sum_{i=1}^{n} x_i \log \frac{x_i}{y_i} + y_i - x_i,$$

and this is precisely the Kullback-Leibler relative entropy function from (2.2).

Using any φ-divergence, we may try to solve the constrained minimization problem (1.2) by generating a sequence $\{x^k\}$ in such a way that x^{k+1} solves the subproblem

$$\min f(x) + \frac{1}{\lambda_k} d_\varphi(x, x^k), \quad x > 0 \tag{2.4}$$

for $k = 0, 1, \ldots$, where $x^0 > 0$ is any given starting point. For this method, Teboulle (1997) shows that it has the same global convergence properties as those mentioned in Theorem 2.1 for the classical proximal-point method. In addition, Teboulle (1997) also allows inexact solutions of the subproblems (2.4). The method may also be applied to the linearly constrained problem (1.3), and, once again, superlinear convergence can be shown under a certain error bound assumption, see Auslender and Haddou (1995) for details. Further references on φ-divergences include Csiszár (1967); Eggermont (1990); Teboulle (1992); Iusem et al. (1994); Iusem, Teboulle (1995).

2.4 *Proximal-like Methods Using Quadratic Kernels and Regularization*

The method we describe in this subsection is taken from Auslender et al. (1999a) and based on earlier work by Tseng and Bertsekas (1993) and Ben Tal and Zibulevsky (1997). It can be motivated by using the ideas from the previous subsection. To this end, assume for a while that the mapping f from the linearly constrained optimization problem (1.3) is twice continuously differentiable. Then it is most natural to use Newton's method in order to minimize the objective function from the (essentially unconstrained) subproblem in (2.4) in order to obtain the next iterate in this way. Using Newton's method, however, we need second order derivatives of these functions. In particular, we need second order derivatives of the function d_φ from Definition 2.2. Calculating the first derivative with respect to x, we obtain

$$\nabla_x d_\varphi(x, y) = \sum_{i=1}^{n} \varphi'\left(\frac{x_i}{y_i}\right) e_i,$$

and calculating the second derivative gives

$$\nabla^2_{xx} d_\varphi(x, y) = \sum_{i=1}^{n} \frac{1}{y_i} \varphi''\left(\frac{x_i}{y_i}\right) e_i e_i^T,$$

where e_i denotes the i-th unit vector in $\mathbb{R}^n$. Hence, in each sum, we have the factor $\frac{1}{y_i}$ which increases to infinity during the iteration process for all indices i for which a constraint like $x_i \geq 0$ is active at a solution. Consequently, we therefore get Hessian matrices which are very ill-conditioned.

In order to avoid this drawback, it is quite natural to modify the idea from the previous subsection in the following way: Let Φ be the class of functions from Definition 2.2 and set

$$d_\varphi(x, y) := \sum_{i=1}^{n} y_i^2 \varphi\left(\frac{x_i}{y_i}\right) \tag{2.5}$$

The difference to the φ-divergence from the previous subsection is that we use y_i^2 instead of y_i. Calculating the second order derivative of the mapping d_φ from (2.5) gives

$$\nabla^2_{xx} d_\varphi(x, y) = \sum_{i=1}^{n} \varphi''\left(\frac{x_i}{y_i}\right) e_i e_i^T,$$

i.e., the crucial factor $\frac{1}{y_i}$ vanishes completely.

Now consider the linearly constrained optimization problem (1.3), i.e., consider the problem

$$\min f(x) \quad \text{s.t.} \quad g(x) \geq 0, \tag{2.6}$$

with $g(x) := b - A^T x$. Auslender et al. (1999a) consider an algorithm which defines a sequence $\{x^k\}$ in such a way that, given a strictly feasible starting point x^0, the next iterate x^{k+1} is computed as a solution of the subproblem

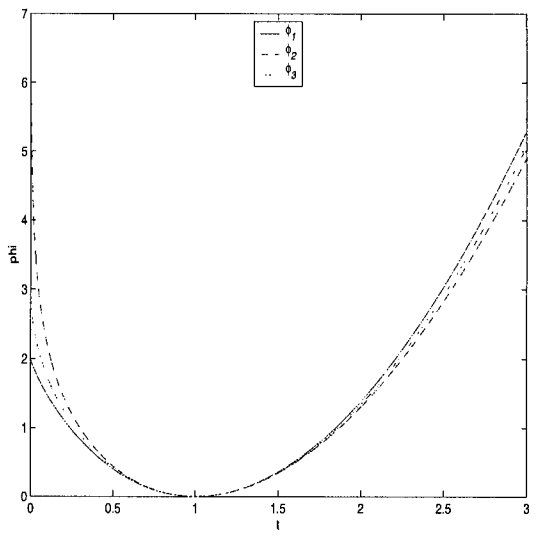
$$\min f(x) + \frac{1}{\lambda_k} d_\varphi\big(g(x), g(x^k)\big), \quad g(x) > 0,$$

where, of course, d_φ denotes the mapping from (2.5). However, it turns out that this method has weaker global convergence properties than all other methods discussed so far, see Auslender et al. (1999a).

In order to overcome this problem, Auslender et al. (1999a) suggest to use a regularization technique. More precisely, they suggest to add a quadratic penalty term to a function $\varphi \in \Phi$ (with Φ being the set from Definition 2.2) in order to obtain a function $\tilde{\varphi}$ in this way. For example, using the three mappings $\varphi_1, \varphi_2, \varphi_3$ from Subsection 2.3, we obtain

$$\begin{aligned}
\tilde{\varphi}_1(t) &:= t \log t - t + 1 + \frac{\nu}{2}(t-1)^2, \\
\tilde{\varphi}_2(t) &:= -\log t + t - 1 + \frac{\nu}{2}(t-1)^2, \\
\tilde{\varphi}_3(t) &:= 2\big(\sqrt{t} - 1\big)^2 + \frac{\nu}{2}(t-1)^2
\end{aligned}$$

for some constant $\nu > 1$. These functions are plotted in the following figure.



Now, let $\tilde{\varphi}$ denote any of these functions and set (similar to (2.5))

$$d_{\tilde{\varphi}}(x,y) := \sum_{i=1}^{n} y_i^2 \tilde{\varphi}(\frac{x_i}{y_i}).$$

The regularized method from Auslender et al. (1999a) then generates a sequence $\{x^k\}$ by starting with a strictly feasible point x^0 for problem (1.3) and by computing x^{k+1} as the solution of the subproblem

$$\min f(x) + \frac{1}{\lambda_k} d_{\tilde{\varphi}}\big(g(x), g(x^k)\big), \quad g(x) > 0. \tag{2.7}$$

Then the following result was shown in Auslender et al. (1999a).

Theorem 2.2 *Let $\{x^k\}$ be a sequence generated by the above method. Assume that the following assumptions are satisfied:*

(A.1) There exist constants $\lambda_{\max} \geq \lambda_{\min} > 0$ such that $\lambda_k \in [\lambda_{\min}, \lambda_{\max}]$ for all $k \in \mathbb{N}$.

(A.2) The optimal value $f_* := \inf\{f(x) \mid A^T x \leq b\}$ is finite.

(A.3) $dom(f) \cap int\{x \mid A^T x \leq b\}$ is nonempty.

(A.4) The matrix A has rank n.

Then the following statements hold:

(a) The sequence of function values $\{f(x^k)\}$ converges to the optimal value f_*.

(b) If $\mathcal{S} \neq \emptyset$, then the entire sequence $\{x^k\}$ converges to an element of $\mathcal{S}$.

The rank assumption (A.4) is satisfied, e.g., if $A = I$, i.e., if the feasible set is the nonnegative orthant. Moreover, this rank condition can be assumed to hold without loss of generality for linear programs if we view (1.3) as the dual of a standard form linear program.

2.5 *Infeasible Proximal-like Methods*

Consider again the linearly constrained minimization problem (1.3). The methods described in the previous subsections all assume, among other things, that

the interior of the feasible set is nonempty. Moreover, it is assumed that we can find a strictly feasible point in order to start the algorithm. However, for linear constraints, it is usually not easy to find such a starting point. Furthermore, there do exist convex optimization problems which are solvable but whose interior of the feasible region is empty. In this case, it is not possible to apply one of the methods from the previous subsections.

Yamashita et al. (2001) therefore describe an infeasible proximal-like method which can be started from an arbitrary point and which avoids the assumption that the interior of the feasible set is nonempty. The idea behind the method from Yamashita et al. (2001) is to enlarge the feasible set

$$X := \{x \in \mathbb{R}^n \mid A^T x \leq b\}$$

of the original minimization problem (1.3) by introducing a perturbation vector $\delta^k > 0$ and by replacing X at each iteration k by an enlarged region of the form

$$X_k := \{x \in \mathbb{R}^n \mid A^T x \leq b + \delta^k\}.$$

Note that $X \neq \emptyset$ then implies $\operatorname{int}(X_k) \neq \emptyset$. Hence we can apply the previous method to the enlarged problem

$$\min f(x) \quad \text{s.t.} \quad x \in X_k.$$

The fact that the method from the previous subsection uses a quadratic penalty term actually fits perfectly into our situation where we allow infeasible iterates.

Obviously, we can hope that we obtain a solution of the original problem (1.3) by letting $\delta^k \to 0$. To be more precise, let us define

$$g_{\delta^k}(x) := b + \delta^k - A^T x,$$

and let x^{k+1} be a solution of the subproblem

$$\min f(x) + \frac{1}{\lambda_k} d_{\tilde{\varphi}}\big(g_{\delta^k}(x), g_{\delta^k}(x^k)\big), \quad g_{\delta^k}(x) > 0$$

for some perturbation vector $\delta^k > 0$. Then is has been shown in Yamashita et al. (2001) that the statements of Theorem 2.2 remain true under a certain set of assumptions. However, without going into the details, we stress that these assumptions include the condition $\operatorname{dom}(f) \cap \{x \mid A^T x \leq b\} \neq \emptyset$ (in contrast to (A.3) which assumes that the domain of f intersected with the *interior* of the feasible set is nonempty). On the other hand, the main convergence result in

Yamashita et al. (2001) has to impose another condition which eventually guarantees that the iterates $\{x^k\}$ generated by the infeasible proximal-like method become feasible in the limit point.

3 NUMERICAL RESULTS FOR SOME OPTIMAL CONTROL PROBLEMS

3.1 Description of Test Problems

We consider two classes of optimal control problems. The first class contains control constraints, the second one involves state constraints.

The class of control constrained problems is as follows: Let $\Omega \subseteq \mathbb{R}^n$ be an open and bounded domain and consider the minimization problem

$$\begin{array}{ll} \min & J(u) := \frac{1}{2}\|y(u) - y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u - u_d\|^2_{L^2(\Omega)} \\ \text{s.t.} & u \in \mathcal{F} := \{u \in L^2(\Omega) \,|\, u \leq \psi \text{ on } \Omega\}, \end{array} \tag{3.1}$$

where $y = y(u) \in H^1_0(\Omega)$ denotes the weak solution of Poisson's equation $-\Delta y = u$ on Ω and $y_d, u_d, \psi \in L^2(\Omega)$ are given square-integrable functions. Here y is the state and u is the control variable. The meaning of (3.1) is that we want to minimize the distance to a desired state y_d (hence the subscript 'd') subject to some constraints on the control u. The penalty term in the objective function J is a standard regularization term multiplied by a small constant $\alpha > 0$.

The following two particular instances of the class of problems (3.1) will be used in our numerical tests; they are taken from Bergounioux et al. (2001).

Example 3.1 Consider the following two instances:

(a) Consider problem (3.1) with the following data:

$$\begin{aligned} \Omega &:= (0,1)^2 \subset \mathbb{R}^2, \\ y_d(x_1, x_2) &:= \tfrac{1}{6}\sin(2\pi x_1)\sin(2\pi x_2)\exp(2x_1), \\ u_d &\equiv 0, \\ \psi &\equiv 0, \\ \alpha &:= 10^{-2}. \end{aligned}$$

(b) Consider problem (3.1) with the following data:

$$\begin{aligned} \Omega &:= (0,1)^2 \subset \mathbb{R}^2, \\ y_d(x_1,x_2) &:= \begin{cases} 200x_1x_2(x_1-\frac{1}{2})^2(1-x_2), & \text{if } 0 < x_1 \leq \frac{1}{2}, \\ 200x_2(x_1-1)(x_1-\frac{1}{2})^2(1-x_2), & \text{if } \frac{1}{2} < x_1 \leq 1, \end{cases} \\ u_d &\equiv 0, \\ \psi &\equiv 1, \\ \alpha &:= 10^{-2} \end{aligned}$$

In order to deal with this problem numerically, we discretize the domain $\Omega = (0,1)^2$ by using an equidistant $(N \times N)$-grid. The Laplacian was approximated by using the standard 5-point finite difference scheme. After shifting the variables in order to get nonnegativity constraints, we then obtain an optimization problem of the form (1.2).

The second class of problems we will deal with is the state constrained problem

$$\begin{aligned} \min \quad & J(u) := \tfrac{1}{2}\|y(u) - y_d\|^2_{L^2(\Omega)} + \tfrac{\alpha}{2}\|u - u_d\|^2_{L^2(\Omega)} \\ \text{s.t.} \quad & u \in \mathcal{F} := \{u \in L^2(\Omega) \mid y \leq \phi \text{ on } \Omega\}, \end{aligned} \tag{3.2}$$

where we used the same notation as for the control constrained problem, i.e., $y = y(u) \in H^1_0(\Omega)$ denotes the weak solution of Poisson's equation $-\Delta y = u$ on Ω, $y_d, u_d, \phi \in L^2(\Omega)$ are given functions, and $\alpha > 0$ is a small regularization parameter. Note that the only difference between problems (3.1) and (3.2) is that we have different constraints.

The particular instance of problem (3.2) we are interested in is also taken from Bergounioux et al. (2001).

Example 3.2 Consider problem (3.2) with the following data:

$$\begin{aligned} \Omega &:= (0,1)^2 \subset \mathbb{R}^2, \\ y_d(x_1,x_2) &:= \sin(2\pi x_1 x_2), \\ u_d &\equiv 0, \\ \phi &\equiv 0.1, \\ \alpha &:= 10^{-3} \end{aligned}$$

In order to deal with the state constrained problem (3.2), we use the same discretization scheme as for the control constrained problem (3.1). In this case, however, this results in a linearly constrained optimization problem of the form (1.3), and it is in general not easy to find a strictly feasible starting point.

3.2 *Numerical Results for Control Constrained Problems*

We begin with a word of caution: The numerical results presented in this and the next subsection are not intended to show that proximal-point methods are the best methods for solving the two classes of optimal control problems from the previous subsection. The only thing we want to do is to provide a brief comparison between some of the different proximal-like methods for the solution of these problems in order to get some hints which proximal-like methods seem to work best.

All methods were implemented in MATLAB and use the same parameter setting whenever this was possible. The particular methods we consider in this subsection for the solution of the control constrained problem (3.1) are the proximal-like methods from Subsections 2.3 (φ-divergences) and 2.4 (quadratic kernels with regularization term). The unconstrained minimization is always carried out by applying Newton's method. For reasons explained earlier, we took the function φ_2 in combination with the proximal-like method from Subsection 2.3, and the corresponding mapping $\tilde{\varphi}_2$ for the regularized method from Subsection 2.4.

Table 3.2 contains the numerical results for Example 3.1 (a) for different sizes of N (the dimension of the discretized problem is $n = N^2$). This table contains the cumulated number of inner iterations, i.e., we present the total number of Newton steps and therefore the total number of linear system solves for each test problem. Both methods seem to work reasonably well, and the number of iterations is more or less independent from the mesh size. However, the number of iterations using the φ-divergence approach is significantly higher than the number of iterations for the regularized approach. The resulting optimal control and state for Example 3.1 (a) are given in Figures 3.1 and 3.2, respectively.

The observation is similar for Example 3.1 (b) as shown in Table 3.2, although this time the number of iterations needed by the two methods is pretty much the same. The resulting optimal control and states for these two examples are given in Figures 3.3 and 3.4, respectively.

We also tested both methods on Example 3.1 (a) using smaller values of α. Due to the quadratic penalty term in the regularized method, we do expect a better behaviour for this method. This is indeed reflected by the numerical results shown in Table 3.4.

Table 3.1 Number of iterations for Example 3.1 (a) ($n = N^2$)

N	φ-divergence	regularized φ function
20	25	17
30	25	17
40	26	17
50	26	16

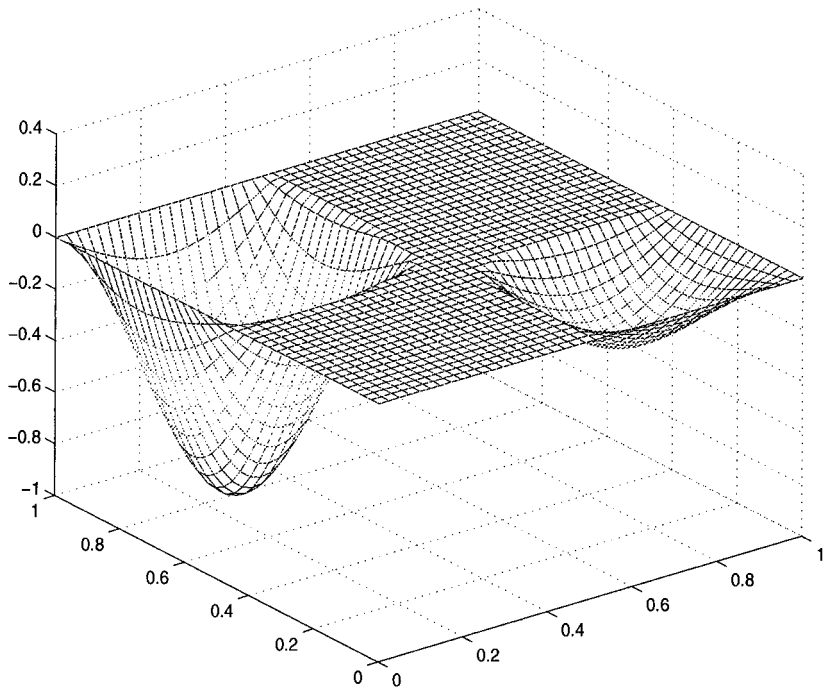


Figure 3.1 Resulting optimal control for Example 3.1 (a)

3.3 Numerical Results for State Constrained Problems

Since a strictly feasible starting point is usually not at hand for state constrained problems, we only applied the infeasible proximal-like method from Subsection 2.5 to the test problem from Example 3.2.

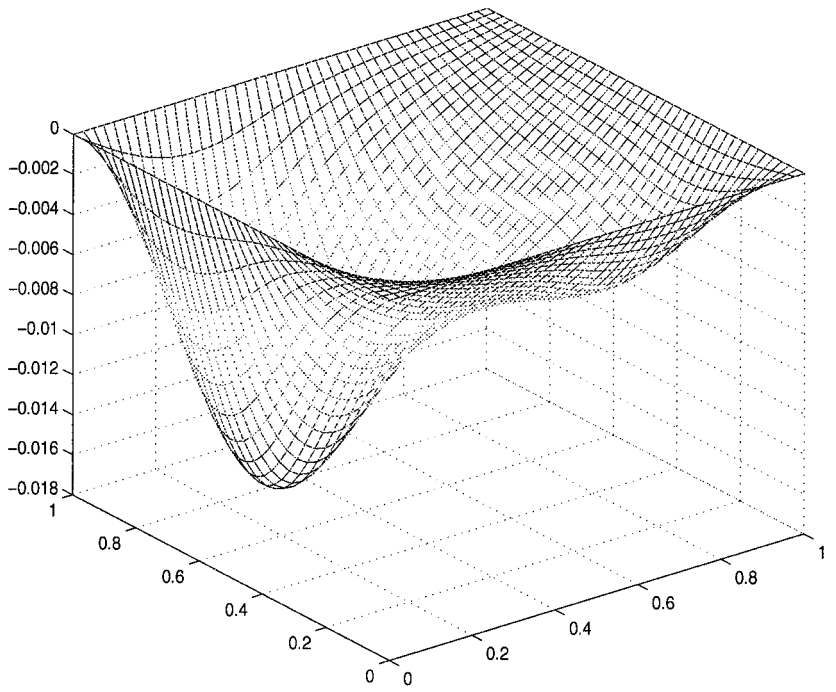


Figure 3.2 Resulting optimal state for Example 3.1 (b)

Table 3.2 Number of iterations for Example 3.1 (b) ($n = N^2$)

N	φ-divergence	regularized φ function
20	23	28
30	30	28
40	30	28
50	31	28

Similar to our description of the methods from the previous subsection, we use Newton's method in order to carry out the inner iterations. The cumulated number of inner iterations for different dimensions of this problem are reported in Table 3.3.

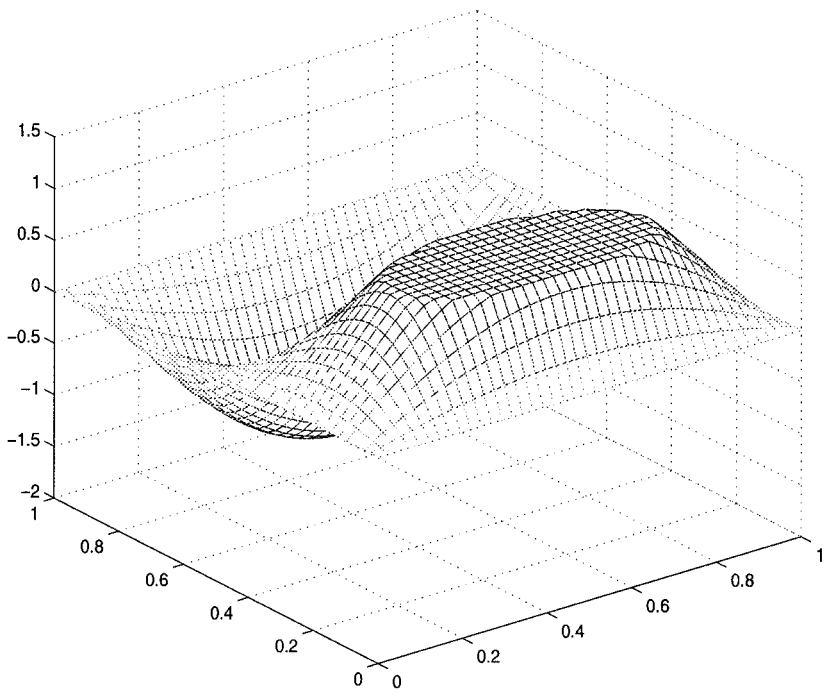


Figure 3.3 Resulting optimal control for Example 3.1 (b)

Table 3.3 Number of iterations for Example 3.1 (a) using different α ($N = 30$)

α	φ-divergence	regularized φ function
$\alpha = 10^{-2}$	25	17
$\alpha = 10^{-3}$	31	18
$\alpha = 10^{-4}$	34	18
$\alpha = 10^{-5}$	47	18
$\alpha = 10^{-6}$	60	18
$\alpha = 10^{-7}$	71	18
$\alpha = 10^{-8}$	62	18

The results indicate that the infeasible method works quite well. Similar to the results from the previous subsection, we can see from Table 3.4 that the number of iterations is again (more or less) independent of the mesh size.

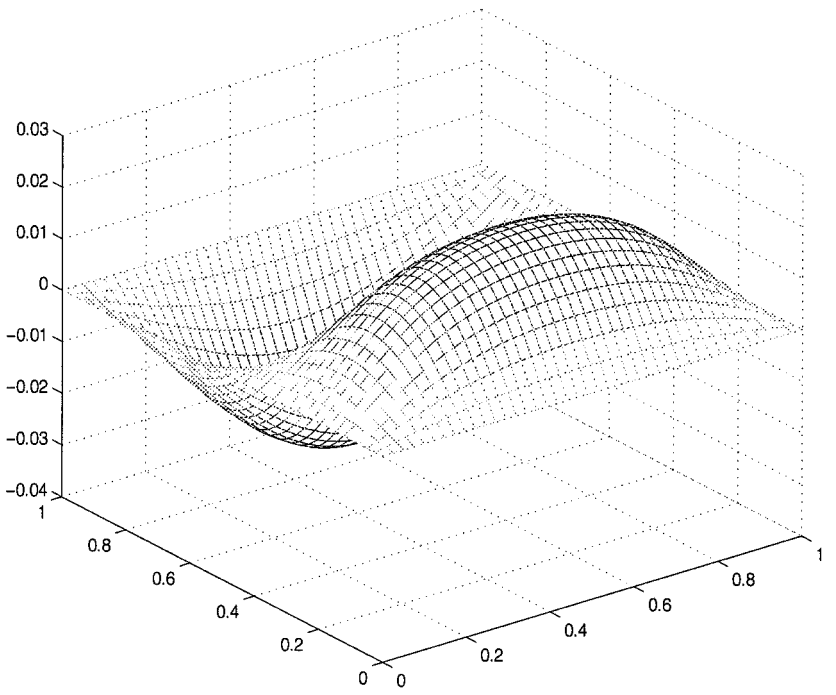


Figure 3.4 Resulting optimal state for Example 3.1 (b)

Table 3.4 Number of iterations for Example 3.2 ($n = N^2$)

N	iterations
20	16
30	14
40	16
50	14
60	14

The resulting optimal control and state for Example 3.2 are shown in Figures 3.5 and 3.6, respectively.

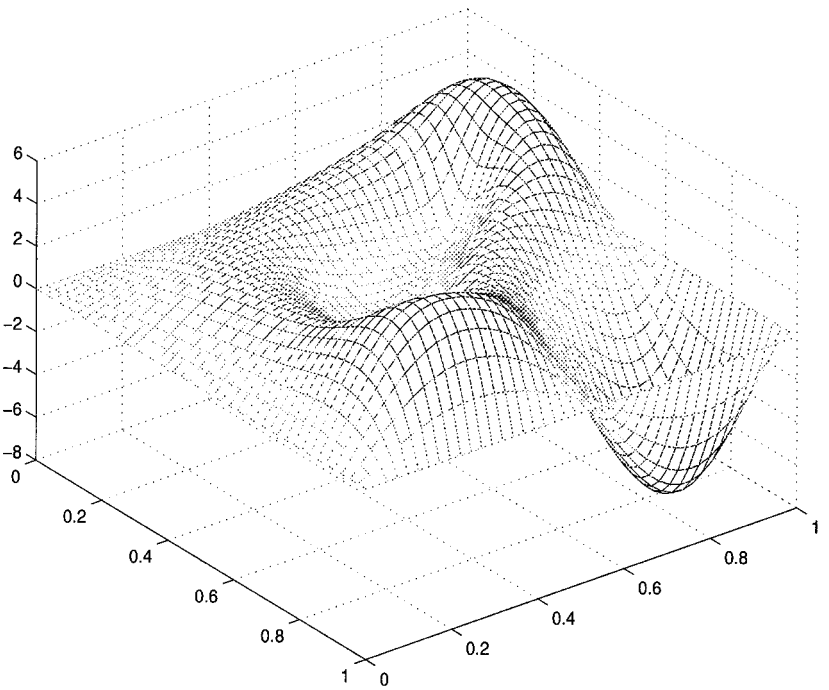


Figure 3.5 Resulting optimal control for Example 3.2

4 FINAL REMARKS

In this paper, we presented several proximal-like methods for the solution of convex minimization problems. A particular class of convex problems are the so-called semi-definite programs, see Nesterov and Nemirovskii (1994) for example. The main difference between semi-definite programs and a convex minimization problem of the form (1.1), say, is the fact that the variables in semi-definite programs are matrices, more precisely, symmetric positive semi-definite matrices. As far as the author is aware of, proximal-like methods have not been applied to semi-definite programs so far. However, due to the recent interest in semi-definite programs (mainly in the field of interior-point analysis), it might be a very useful research topic to see how proximal-like methods can be extended to semi-definite programs and how these methods behave for this class of convex problems.

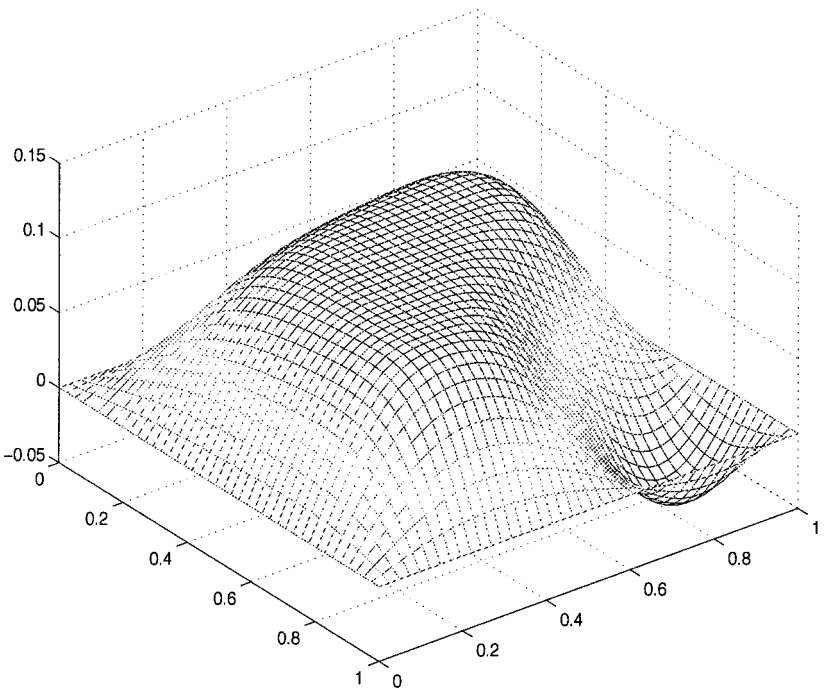


Figure 3.6 Resulting optimal state for Example 3.2

References

Auslender, A., and Haddou, M. (1995), An interior-proximal method for convex linearly constrained problems and its extension to variational inequalities, *Mathematical Programming*, Vol. 71, pp. 77-100.

Auslender, A., Teboulle, M., and Ben-Tiba, S. (1999a), Interior proximal and multiplier methods based on second order homogeneous kernels, *Mathematics of Operations Research*, Vol. 24, pp. 645-668.

Auslender, A., Teboulle, M., and Ben-Tiba, S. (1999b), A logarithmic-quadratic proximal method for variational inequalities, *Computational Optimization and Applications*, Vol. 12, pp. 31-40.

Ben-Tal, A., and Zibulevsky, M. (1997), Penalty-barrier methods for convex programming problems, *SIAM Journal on Optimization*, Vol. 7, pp. 347-366.

Bergounioux, M., Haddou, M., Hintermüller, M., and Kunisch, K. (2001): A comparison of a Moreau-Yosida-based active set strategy and interior point

methods for constrained optimal control problems, *SIAM Journal on Optimization*, Vol. 11, pp. 495-521.

Censor, Y., Iusem, A.N., and Zenios, S.A. (1998), An interior point method with Bregman functions for the variational inequality problem with paramonotone functions, *Mathematical Programming*, Vol. 81, pp. 373-400.

Censor, Y., and Zenios, S.A. (1992), Proximal minimization algorithm with D-functions, *Journal of Optimization Theory and Applications*, Vol. 73, pp. 451-464.

Chen, G., and Teboulle, M. (1993), Convergence analysis of a proximal-like minimization algorithm using Bregman functions, *SIAM Journal on Optimization*, Vol. 3, pp. 538-543.

Csiszár, I. (1967), Information-type measures of difference of probability distributions and indirect observations, *Studia Scientiarium Mathematicarum Hungarica*, Vol. 2, pp. 299-318.

De Pierro, A.R., and Iusem, A.N. (1986), A relaxed version of Bregman's method for convex programming, *Journal of Optimization Theory and Applications*, Vol. 51, pp. 421-440.

Eckstein, J. (1993), Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming, *Mathematics of Operations Research*, Vol. 18, pp. 202-226.

Eckstein, J. (1998), Approximate iterations in Bregman-function-based proximal algorithms. *Mathematical Programming*, Vol. 83, pp. 113-123.

Eggermont, P.P.B., Multiplicative iterative algorithms for convex programming, *Linear Algebra and its Applications*, Vol. 130, pp. 25-42.

Ferris, M.C. (1991), Finite termination of the proximal-point algorithm, *Mathematical Programming*, Vol. 50, pp. 359-366.

Güler, O. (1991), On the convergence of the proximal-point algorithm for convex minimization, *SIAM Journal on Control and Optimization*, Vol. 29, pp. 403-419.

Güler, O. (1994), Ergodic convergence in proximal point algorithms with Bregman functions, in Du, D.-Z. and Sun, J., *Advances in Optimization and Approximation*, Kluwer Academic Publishers, pp. 155-165.

Hoffman, A.J. (1952), On approximate solutions of systems of linear inequalities, *Journal of Research of the National Bureau of Standards*, Vol. 49, pp. 263-265.

Iusem, A.N., Svaiter, B.F., and Teboulle, M. (1994), Entropy-like proximal methods in convex programming. *Mathematics of Operations Research*, Vol. 19, pp. 790-814.

Iusem, A.N., and Teboulle, M. (1995), Convergence rate analysis of nonquadratic proximal methods for convex and linear programming, *Mathematics of Operations Research*, Vol. 20, pp. 657-677.

Lemaire, B. (1989), The proximal algorithm, in Penot. J.P. (ed.): *New Methods in Optimization and their Industrial Uses*, Birkhäuser-Verlag, Basel, pp. 73-87.

Lemaire, B. (1992), About the convergence of the proximal method, in *Advances in Optimization.* Lecture Notes in Economics and Mathematical Systems 382, Springer-Verlag, pp. 39-51.

Luque, F.J. (1984), Asymptotic convergence analysis of the proximal point algorithm. *SIAM Journal on Control and Optimization*, Vol. 22, pp. 277-293.

Martinet, B. (1970), Regularisation d'inéqations variationelles par approximations successives, *Revue Francaise d'Informatique et de Recherche Operationelle*, pp. 273-299.

Nesterov, Y., and Nemirovskii, A. (1994), *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, PA.

Rockafellar, R.T. (1970), *Convex Analysis.* Princeton University Press.

Rockafellar, R.T. (1976), Monotone operators and the proximal point algorithm, *SIAM Journal on Control and Optimization*, Vol. 14, pp. 877-898.

Solodov, M.V., and Svaiter, B.F. (1999), A hybrid projection-proximal point algorithm, *Journal of Convex Analysis*, Vol. 6, pp. 59-70.

Solodov, M.V., and Svaiter, B.F. (2000a), Forcing strong convergence of proximal point iterations in a Hilbert space, *Mathematical Programming*, Vol. 87, pp. 189-202.

Solodov, M.V., and Svaiter, B.F. (2000b), A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem, *SIAM Journal on Optimization*, Vol. 10, pp. 605-625.

Teboulle, M. (1992), Entropic proximal mappings with applications to nonlinear programming, *Mathematics of Operations Research*, Vol. 17, pp. 670-690.

Teboulle, M. (1997), Convergence of proximal-like algorithms, *SIAM Journal on Optimization* 7, pp. 1069-1083.

Tseng, P., and Bertsekas, D.P. (1993), On the convergence of the exponential multiplier method for convex programming, *Mathematical Programming*, Vol. 60, pp. 1-19.

Yamashita, N., Kanzow, C., Morimoto, T., and Fukushima, M. (2001), An infeasible interior proximal method for convex programming problems, *Journal of Nonlinear and Convex Analysis*, Vol. 2, 2001, pp. 139-156.