

Using exact penalties to derive a new equation reformulation of KKT systems associated to variational inequalities

Thiago A. de André* Paulo J. S. Silva†

March 24, 2007

Abstract

In this paper, we present a new reformulation of the KKT system associated to a variational inequality as a semismooth equation. The reformulation is derived from the concept of differentiable exact penalties for nonlinear programming. The best results are presented for nonlinear complementarity problems, where simple, verifiable, conditions ensure that the penalty is exact. We also develop a semismooth Newton method for complementarity problems based on the reformulation. We close the paper showing some preliminary computational tests comparing the proposed method with classical reformulations, based on the minimum or on the Fischer-Burmeister function.

1 Introduction

Consider a constrained nonlinear programming problem

$$\begin{aligned} \min f(x) \\ \text{s.t. } g(x) \leq 0 \\ h(x) = 0, \end{aligned} \tag{NLP}$$

where x lies in \mathbb{R}^n and $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$, and $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ are C^2 functions.

*Instituto de Matemática e Estatística – University of São Paulo, Brazil. Partially supported by FAPESP, grant 02/10942-9

†Instituto de Matemática e Estatística – University of São Paulo, Brazil (pjssilva@ime.usp.br). Partially supported by CNPq, grants PQ 304133/2004-3, 477083/2006-4, and PRONEX - Optimization.

Penalty methods are a very popular framework to solve such problems. In these methods, the constrained problem is replaced by a sequence of unconstrained ones. A good example is the augmented Lagrangian algorithm, that can be derived from a proximal point method applied to the Lagrangian dual problem [3, 25].

Another possibility is the exact penalty approach, where a special penalty function is used to transform (NLP) into a single unconstrained problem. For example, it is easy to see that, under reasonable assumptions, the solutions to (NLP) are exactly the unconstrained minima of

$$\min_{x \in \mathbb{R}^n} \phi(x, \mu) \stackrel{\text{def}}{=} f(x) + \mu \max \{0, g_1(x), g_2(x), \dots, g_m(x), |h_1(x)|, |h_2(x)|, \dots, |h_p(x)|\},$$

if μ is chosen big enough [4]. However, this unconstrained problem is naturally nonsmooth and special methods should be used to solve it. Moreover, it is not easy to estimate how large μ must be to ensure the equivalence of the minima.

To overcome the lack of differentiability of the maximum, many authors proposed differentiable formulas for exact penalties. The history of differentiable exact penalties starts with Fletcher in 1970, when he published a series of three articles proposing automatic updates for the multipliers in the augmented Lagrangian method for equality constrained problems [13, 16, 14]. The idea was to estimate the multipliers as a function of the primal variables, denoted by $\lambda(x)$, followed by the minimization of the associated augmented Lagrangian

$$f(x) + \langle \lambda(x), h(x) \rangle + c_k \|h(x)\|^2.$$

However, the multiplier function was not easy to compute and it was not clear how to choose good values for the penalty parameter c_k . Later on, in 1975, Mukai and Polak proposed a new formula for $\lambda(x)$ and showed that there is a threshold for c_k that once achieved would allow the modified augmented Lagrangian to recover the solutions of the original problem after a single minimization [21].

In 1979, Di Pillo and Grippo presented a new formulation for exact penalties that simplified the analysis of the associated problems [6]. In this work, they propose to further extend the augmented Lagrangian function, penalizing deviations from the first order conditions:

$$f(x) + \langle \lambda(x), h(x) \rangle + c_k \|h(x)\|^2 + \|M(x) (\nabla f(x) + Jh(x)' \lambda(x))\|^2,$$

where $Jh(x)$ denotes the Jacobian of h at x . Special choices for $M(x)$ resulted in modified augmented Lagrangians that are quadratic in λ . In this case, it is possible to isolate the dual variable in terms of x . One of such choices for $M(x)$ recovered the method proposed by Fletcher and the results from Mukai and Polak.

This last formulation is also important because it is able to deal with inequality constraints using slack variables and the classic transformation $h_i(x) = g_i(x) + s_i^2$. With an

appropriate choice for $M(x)$, one obtains a quadratic problem in the slacks. Then, the slacks can be written as an explicit function of the original variables x . However, in this case it is not possible to isolate the multipliers λ as a function of x .

In 1973, Fletcher had already extended his ideas to deal with inequality constraints [15], but the proposed function lacked good differentiability properties. In 1979, Glad and Polak proposed a new formula for $\lambda(x)$ in inequality constrained problems and showed how to control the parameter c_k [17].

Finally, in 1985 and 1989, Di Pillo and Grippo reworked the results from Glad and Polak and created a differentiable exact penalty for inequality constrained problems that depends only on the primal variables [7, 8]. These papers are the base of our work. In particular, from now on we focus exclusively on inequality constrained problems.

In this paper we extend the ideas of Di Pillo and Grippo to variational inequalities with functional constraints and the related KKT system. The remaining of the paper is organized as follows: Section 2 presents the formula for the penalty, Section 3 derives the exactness results, Section 4 specializes the results for Nonlinear Complementarity Problems (NCP), and Section 5 uses the proposed penalty to develop a semismooth Newton method for complementarity. This last section is closed with some preliminary computational results comparing the new penalty with other classical NCP functions.

2 Extending exact penalties

As described above, it is possible to build differentiable exact penalties for constrained optimization problems using an augmented Lagrangian function coupled with a multiplier estimate computed from the primal point. A natural multiplier estimate for inequality constrained problems was given by Glad and Polak. It is computed solving, in the least-squares sense, the equations involving the multipliers in the KKT conditions

$$\min_{\lambda \in \mathbb{R}^m} \|\nabla_x L(x, \lambda)\|^2 + \zeta^2 \|G(x)\lambda\|^2, \quad (1)$$

where L is the usual Lagrangian function, $\zeta > 0$, and $G(x) \in \mathbb{R}^{m \times m}$ is a diagonal matrix with $G(x)_{ii} = g_i(x)$. The first term tries to find a multiplier for which the fixed primal point is a minimum of the Lagrangian function.¹ The second term tries to enforce the complementarity conditions.

This problem is convex and quadratic in λ and can be easily solved if the point x conforms to the Linear Independence Constraint Qualification (LICQ), that is, if the gradients of the constraints that are active at x are linearly independent. The results concerning (1) are summarized in the following proposition.

¹Actually, a first order stationary point.

Proposition 2.1. [17, Proposition 1] Assume that $x \in \mathbb{R}^n$ conforms to LICQ and define the matrix $N(x) \in \mathbb{R}^{m \times m}$ by

$$N(x) \stackrel{\text{def}}{=} Jg(x)Jg(x)' + \zeta^2 G(x)^2.$$

Then,

1. $N(x)$ is positive definite.
2. The solution $\lambda(x)$ of (1) is

$$\lambda(x) = -N^{-1}(x)Jg(x)\nabla f(x).$$

3. If $(\bar{x}, \bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^m$ is a KKT pair where LICQ holds then $\bar{\lambda} = \lambda(\bar{x})$, i.e. $\bar{\lambda}$ solves (1) for $x = \bar{x}$.
4. [7, Proposition 4] If LICQ holds in a neighborhood of x , then $\lambda(\cdot)$ is continuously differentiable at x and its Jacobian is given by

$$J\lambda(x) = -N^{-1}(x) \left[Jg(x)\nabla_{xx}^2 L(x, \lambda(x)) \sum_{i=1}^m e^i \nabla_x L(x, \lambda(x))' \nabla^2 g_i(x) + \right. \\ \left. + 2\zeta^2 \Lambda(x)G(x)Jg(x) \right], \quad (2)$$

where e^i is the i -th element of the canonical base of \mathbb{R}^m and $\Lambda(x) \in \mathbb{R}^{m \times m}$ is a diagonal matrix with $\Lambda(x)_{ii} = \lambda(x)_i$.

Using such estimate, one can build a differentiable exact penalty from the standard augmented Lagrangian function,

$$L_c(x, \lambda) \stackrel{\text{def}}{=} f(x) + \frac{1}{2c} \sum_{i=1}^m (\max\{0, \lambda_i + cg_i(x)\}^2 - \lambda_i^2) \\ = f(x) + \langle \lambda, g(x) \rangle + \frac{c}{2} \|g(x)\|^2 - \frac{1}{2c} \sum_{i=1}^m \max\{0, -\lambda_i - cg_i(x)\}^2.$$

The resulting exact penalty function, that we call $w_c(\cdot)$, is obtained plugging the multiplier estimate in the augmented Lagrangian

$$w_c(x) \stackrel{\text{def}}{=} L_c(x, \lambda(x)). \quad (3)$$

Our aim is to extend the definition of w_c to the context of variational inequalities. Let $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a continuous function and $\mathcal{F} \subset \mathbb{R}^n$ a non-empty closed set. The variational inequality problem is to find $x \in \mathcal{F}$ such that

$$\forall y \in \mathcal{F}, \quad \langle F(x), y - x \rangle \geq 0. \quad (\text{VIP})$$

If \mathcal{F} is convex it is easy to see that (VIP) is an extension of the geometrical optimality conditions for (NLP) where the gradient of the objective function is replaced by a general continuous function. In this paper we focus on problems where the feasible set can be described as a system of twice differentiable convex inequalities $\mathcal{F} = \{x \mid g(x) \leq 0\}$.

In 1999, Eckstein and Ferris proposed an extension of multiplier methods to nonlinear mixed complementarity problems [9], variational inequalities where \mathcal{F} is a box. Afterwards, Auslender and Teboulle proposed an extension of the notion of augmented Lagrangians to (VIP) with general functional constraints [1, 2]. Their results are related to the rich duality theory for generalized equations developed by Pennanen [22].

Following these developments, a natural way to extend the notion of exact penalty to (VIP) is to use the gradient of $w_c(\cdot)$, replacing ∇f by F . However, such gradient involves the Jacobian of $\lambda(\cdot)$ given in Proposition 2.1. This first choice for exact penalty would have a very complicated formula, depending on the Hessians of the constraints and the Jacobian of F which may not be easily available.

To overcome such difficulty, we start with the classical augmented Lagrangian for variational inequality, which is the gradient of $L_c(\cdot, \cdot)$ with respect to the first variable. We then plug into it the multiplier estimate $\lambda(\cdot)$:

$$\lambda(x) \stackrel{\text{def}}{=} -N^{-1}(x)Jg(x)F(x), \quad (4)$$

$$W_c(x) \stackrel{\text{def}}{=} F(x) + Jg(x)'\lambda(x) + cJg(x)'g(x) + cJg(x)'\max\{0, -\lambda(x)/c - g(x)\} \quad (5)$$

$$= F(x) + Jg(x)'\lambda(x) + cJg(x)'\max\{g(x), -\lambda(x)/c\}. \quad (6)$$

In the next sections we will show the relation between the zeros of W_c , for c large enough, and the KKT system associated to (VIP).

Definition 2.2. *The Karush-Kuhn-Tucker (KKT) system associated to (VIP) is*

$$F(x) + Jg(x)'\lambda = 0, \quad (\text{Zero Condition})$$

$$g(x) \leq 0, \quad (\text{Primal Feasibility})$$

$$\lambda \geq 0, \quad (\text{Dual Feasibility})$$

$$\forall i = 1, \dots, m, \quad \lambda_i g_i(x) = 0. \quad (\text{Complementarity})$$

A pair $(x, \lambda) \in \mathbb{R}^{n+m}$ that conforms to these equations is called a KKT pair. The primal variable x is called a KKT (stationary) point.

This system is known to be equivalent to (VIP) whenever the feasible set \mathcal{F} is defined by convex inequalities and conforms to a constraint qualification [10].

Some comments must be made before presenting the exactness properties for W_c . Note that since W_c is not the gradient of w_c , its zeros are not clearly related to the solutions of an unconstrained optimization problem. In this sense, the proposed exact penalty approach is not equivalent to the penalties usually proposed in the optimization literature. *In particular, it has the major advantage of not depending on the Jacobian of F and on second order information of the constraints.*

As for the differentiability properties of W_c , the maximum present in its definition clearly make it nonsmooth. This is a direct heritage of the classical augmented Lagrangian used to derive it. Even though, it is (strongly) semismooth if F is (LC^1) C^1 and g is (LC^2) C^2 . Therefore, its zeros can be found by an extension of the Newton method to semismooth equations [23, 24]. We present such a method in Section 5. In this sense, W_c can be viewed as a NCP function, like the classical $\min(x, F(x))$ or the Fischer-Burmeister function. However, W_c incorporates dual information through the multiplier estimates.

3 Exactness results

Let us present the exactness results for W_c . Here we follow closely the results presented in the nonlinear programming case by Di Pillo and Grippo [7, 8]. First, we show that the proposed penalty has zeros whenever the original KKT system has solutions.

In order to define W_c in the whole space we will need the following assumption, that we assume valid throughout this section:

Assumption 3.1. *LICQ holds on the whole \mathbb{R}^n , so that $\lambda(\cdot)$ and, hence, W_c is well-defined everywhere.*

This assumption is restrictive, but was present already in the original papers on (differentiable) exact penalties [17, 6, 7, 8]. Fortunately, in many cases it is easily verifiable. For example, it holds trivially in nonlinear and mixed complementarity problems.

Proposition 3.2. *Let (x, λ) be a KKT pair. Then, for all $c > 0$, $W_c(x) = 0$.*

Proof. The LICQ assumptions ensures that $\lambda = \lambda(x)$. Then,

$$\begin{aligned} W_c(x) &= F(x) + Jg(x)' \lambda(x) + cJg(x)' \max \{g(x), -\lambda(x)/c\} \\ &= 0 + cJg(x)' \max \{g(x), -\lambda/c\} \\ &= 0, \end{aligned}$$

where the last equality follows from primal and dual feasibility and the complementary condition. \square

Next, we show that for c large enough the zeroes of W_c are nearly feasible. Then, we show that if a zero is nearly feasible it will be a KKT point associated to (VIP).

Proposition 3.3. *Let $\{x^k\} \subset \mathbb{R}^n$ and $\{c_k\} \subset \mathbb{R}_+$ be sequences such that $x^k \rightarrow \bar{x}$, $c_k \rightarrow +\infty$, and $W_{c_k}(x^k) = 0$. Then, $\bar{x} \in \mathcal{F}$.*

Proof. We have,

$$0 = W_{c_k}(x^k) = F(x^k) + Jg(x^k)' \lambda(x^k) + c_k Jg(x^k)' \max \{g(x^k), -\lambda(x^k)/c_k\}.$$

Now recall that, under LICQ, $\lambda(\cdot)$ is continuous. Moreover, F is assumed continuous and g continuously differentiable. Hence, as $x^k \rightarrow \bar{x}$ and $c_k \rightarrow +\infty$, we may divide the equation above by c_k and take limits to conclude that

$$0 = \sum_{i=1}^m \max \{g_i(\bar{x}), 0\} \nabla g_i(\bar{x}).$$

Once again we use LICQ to see that $\max \{g_i(\bar{x}), 0\} = 0$ for all $i = 1, \dots, m$, that is, \bar{x} is feasible. \square

Proposition 3.4. *Let $\bar{x} \in \mathcal{F}$. Then, there are $c_{\bar{x}}, \delta_{\bar{x}} > 0$ such that if $\|x - \bar{x}\| \leq \delta_{\bar{x}}$, $c > c_{\bar{x}}$, and $W_c(x) = 0$ imply that $(x, \lambda(x))$ is a KKT pair associated to (VIP).*

Proof. Let us introduce some notation

$$y(x) \stackrel{\text{def}}{=} \max \{0, -\lambda(x)/c - g(x)\}.$$

We will use capital letters to denote the usual diagonal matrix build from vectors. For example, $Y(x)$ denotes the diagonal matrix with $y(x)$ in the diagonal.

It is easy to show that,

$$Y(x)\lambda(x) = -cY(x)(g(x) + y(x)).$$

Hence,

$$\begin{aligned} Jg(x)(F(x) + Jg(x)'\lambda(x)) &= Jg(x)F(x) + Jg(x)Jg(x)'\lambda(x) \\ &= -N(x)\lambda(x) + Jg(x)Jg(x)'\lambda(x) \\ &= -\zeta^2 G(x)^2 \lambda(x) \\ &= -\zeta^2 G(x)(G(x) + Y(x))\lambda(x) + \zeta^2 G(x)Y(x)\lambda(x) \\ &= -\zeta^2 G(x)\Lambda(x)(g(x) + y(x)) + \zeta^2 G(x)Y(x)\lambda(x). \end{aligned} \tag{7}$$

We combine the last results to get

$$\frac{1}{c}Jg(x)(F(x) + Jg(x)'\lambda(x)) = -\zeta^2G(x) \left(\frac{1}{c}\Lambda(x) + Y(x) \right) (g(x) + y(x)).$$

Recalling the definition of W_c , we have

$$\frac{1}{c}Jg(x)W_c(x) = \frac{1}{c}Jg(x)(F(x) + Jg(x)'\lambda(x)) + Jg(x)Jg(x)'(g(x) + y(x)) \quad (8)$$

$$= K(x, c)(g(x) + y(x)), \quad (9)$$

where $K(x, c) \stackrel{\text{def}}{=} Jg(x)Jg(x)' - \zeta^2G(x)Y(x) - \frac{1}{c}\zeta^2G(x)\Lambda(x)$.

For $x = \bar{x}$, recalling that it is feasible, if $c \rightarrow +\infty$ then $y(\bar{x}) \rightarrow -g(\bar{x})$ and therefore $K(\bar{x}, c) \rightarrow N(\bar{x})$. As $N(\bar{x})$ is nonsingular due to LICQ, we can conclude that there must be $c_{\bar{x}}, \delta_{\bar{x}} > 0$ such that if $\|x - \bar{x}\| \leq \delta_{\bar{x}}$, $c > c_{\bar{x}}$ then $K(x, c)$ is also nonsingular.

Let x, c be such that $\|x - \bar{x}\| \leq \delta_{\bar{x}}$, $c > c_{\bar{x}}$ and $W_c(x) = 0$. Then, (9) imply that $g(x) + y(x) = 0$. Plugging this into (8) and using LICQ once again, we see that $F(x) + Jg(x)'\lambda(x) = 0$, the zero condition. Moreover, $g(x) + y(x) = 0$ is equivalent to

$$\begin{aligned} \max \{g(x), -\lambda(x)/c\} = 0 &\implies \\ g(x) \leq 0, \quad \lambda(x) \geq 0, & \end{aligned}$$

primal and dual feasibility. Finally, using (7) and the zero condition, complementarity slackness holds. \square

These two results may be combined in the following exactness theorem:

Theorem 3.5. *Let $\{x^k, c_k\} \subset \mathbb{R}^{n+1}$ be a sequence such that $W_{c_k}(x^k) = 0$, $c_k \rightarrow +\infty$, and $\{x^k\}$ is bounded. Then, there is a finite index K such that for $k > K$, $(x^k, \lambda(x^k))$ is a KKT solution associated to (VIP).*

Proof. Suppose, by contradiction, that we can extract a sub-sequence $\{x^{k_j}\}$ of points that are not KKT. Since $\{x^k\}$ is bounded, we can assume without loss of generality that $\{x^{k_j}\}$ converges to some \bar{x} . Using Proposition 3.3 we conclude that \bar{x} is feasible. Then, Proposition 3.4 ensures that when x^{k_j} is close enough to \bar{x} , $(x^{k_j}, \lambda(x^{k_j}))$ will be a solution to the KKT system. \square

Corollary 3.6. *If there is a $\bar{c} \in \mathbb{R}$ such that the set $\{x \mid W_c(x) = 0, c > \bar{c}\}$ is bounded, then there is a $\tilde{c} > 0$ such that $W_c(x) = 0, c > \tilde{c}$, implies that $(x, \lambda(x))$ is a KKT solution associated to (VIP).*

Proof. Suppose by contradiction that the result is false. Then, there must be $c_k \rightarrow +\infty$ and a sequence $\{x^k\} \subset \mathbb{R}^n$ such that $W_{c_k}(x^k) = 0$, and such that $(x^k, \lambda(x^k))$ is not a KKT solution. But for $c_k > \bar{c}$ we have that x^k belongs to the bounded set $\{x \mid W_c(x) = 0, c > \bar{c}\}$ and then $\{x^k\}$ is bounded. This is not possible, as Theorem 3.5 ensures that for big enough k , $(x^k, \lambda(x^k))$ is a KKT solution. \square

Note that the exactness results above depend on the boundedness of the zeroes. This property may not be easily verified from the problem data for general variational inequalities. In the optimization literature such boundedness is forced by the minimization on an extraneous compact set that should contain the feasible set or at least a minimum [7, 8]. We believe, on the other hand, that by exploring the coerciveness properties of a (VIP) it is possible to drop the necessity of such compact. We give the first results in this direction in the next section, that deals with a special (VIP).

4 Nonlinear complementarity problems

We specialize the proposed exact penalty to nonlinear complementarity problems:

$$F(x) \geq 0, \quad x \geq 0, \quad \langle F(x), x \rangle = 0. \quad (\text{NCP})$$

It is easy to see that (NCP) is a Variational Inequality with $\mathcal{F} = \mathbb{R}_+^n$ and, as stated before, that LICQ holds everywhere.

After some algebra, we may see that the proposed exact penalty $W_c(\cdot)$, simplifies to

$$W_c(x)_i = \min \left\{ \frac{\zeta^2 x_i^2}{1 + \zeta^2 x_i^2} F(x)_i + c x_i, F(x)_i \right\}, \quad i = 1, \dots, n. \quad (10)$$

In particular, the multiplier estimate can be computed explicitly.

In this case we can derive a reasonable assumption that ensures that, for large c , the zeros of $W_c(\cdot)$ are solutions to (NCP).

Theorem 4.1. *Assume that there are $\rho, M > 0$ such that $\langle F(x), x \rangle \geq -M$ for $\|x\| > \rho$ or that F is monotone and (NCP) has a solution. Then, there is a $\bar{c} > 0$ such that $W_c(\cdot)$ is exact for $c > \bar{c}$, i.e. any zero of $W_c(x)$ for $c > \bar{c}$ is a solution to (NCP).*

Proof. Suppose, by contradiction, that the result does not hold. Then there are $c_k \rightarrow +\infty$, and a sequence $\{x^k\}$ such that $W_{c_k}(x^k) = 0$ and x^k is not a solution to (NCP). Theorem 3.5 asserts that $\|x^k\| \rightarrow +\infty$. Proposition 3.4 says that x^k is never feasible.

For each x^k and each of its coordinates, (10) allows only three possibilities:

1. If $x_i^k > 0$, then $F(x^k)_i = 0$.

Observe that (10) implies $F(x^k)_i \geq W_{c_k}(x^k)_i = 0$. If $F(x^k)_i > 0$, $W_{c_k}(x^k)_i$ would be the minimum of two strictly positive numbers, which contradicts the fact that it is zero.

2. If $x_i^k = 0$, then $F(x^k)_i \geq 0$.

It follows from $W_{c_k}(x^k)_i = 0$, that $\min\{0, F(x^k)_i\} = 0$. This is equivalent to $F(x^k)_i \geq 0$.

3. If $x_i^k < 0$, then $F(x^k)_i = -c_k \frac{1 + \zeta^2(x_i^k)^2}{\zeta^2 x_i^k}$.

First, if $F(x^k)_i \leq 0$, $W_{c_k}(x^k)_i$ would be the minimum of a strictly negative number and a negative number. This contradicts $W_{c_k}(x^k)_i = 0$. Now, as $F(x^k)_i > 0$, it is clear that the minimum is achieved in the first term, leading to

$$0 = \frac{\zeta^2(x_i^k)^2}{1 + \zeta^2(x_i^k)^2} F(x^k)_i + c_k x_i^k,$$

which gives the desired result.

Note that the cases above show that $F(x^k) \geq 0$.

Now consider that there are $\rho, M > 0$ such that $\langle F(x), x \rangle \geq -M$ for $\|x\| > \rho$. On the other hand, we have just proved that

$$\begin{aligned} \langle F(x^k), x^k \rangle &= \sum_{x_i^k < 0} -c_k \frac{1 + \zeta^2(x_i^k)^2}{\zeta^2} \\ &\rightarrow -\infty, \end{aligned} \tag{11} \quad [x^k \text{ is not feasible}]$$

a contradiction.

Finally consider the case where F is monotone and where (NCP) has a solution \bar{x} . We have

$$\begin{aligned} 0 &\leq \langle F(x^k) - F(\bar{x}), x^k - \bar{x} \rangle \\ &= \langle F(x^k), x^k \rangle - \langle F(x^k), \bar{x} \rangle - \langle F(\bar{x}), x^k \rangle + \langle F(\bar{x}), \bar{x} \rangle \\ &\leq \langle F(x^k), x^k \rangle - \langle F(\bar{x}), x^k \rangle \\ &\leq \sum_{x_i^k < 0} -c_k \frac{1 + \zeta^2(x_i^k)^2}{\zeta^2} - F(\bar{x})_i x_i^k \\ &= \sum_{x_i^k < 0} \frac{-c_k - \zeta^2 x_i^k (c_k x_i^k + F(\bar{x})_i)}{\zeta^2}, \end{aligned}$$

where the second inequality follows from the fact that $F(x^k), \bar{x} \geq 0$ and $\langle F(\bar{x}), \bar{x} \rangle = 0$, and the third follows from Equation (11) and $F(\bar{x}) \geq 0$.

If, for some x^k , $c_k x_i^k + F(\bar{x})_i \leq 0$ whenever $x_i^k < 0$, the last equation already shows a contradiction as it must be strictly smaller than 0. Hence we conclude that for at least one coordinate, $c_k x_i^k + F(\bar{x})_i > 0$ and we can write

$$\begin{aligned}
0 &\leq \sum_{x_i^k < 0, c_k x_i^k + F(\bar{x})_i > 0} \frac{-c_k - \zeta^2 x_i^k (c_k x_i^k + F(\bar{x})_i)}{\zeta^2} \\
&\leq \sum_{x_i^k < 0, c_k x_i^k + F(\bar{x})_i > 0} \frac{-c_k - \zeta^2 x_i^k F(\bar{x})_i}{\zeta^2} \\
&\leq \sum_{x_i^k < 0, c_k x_i^k + F(\bar{x})_i > 0} \frac{-c_k + \zeta^2 F(\bar{x})_i^2 / c_k}{\zeta^2} \quad [c_k x_i^k + F(\bar{x})_i > 0] \\
&\rightarrow -\infty,
\end{aligned}$$

a contradiction □

The coerciveness assumption on $F(\cdot)$ that appears in Theorem 4.1 is not very restrictive. In particular, it holds whenever F presents a weak coercive property that is associated to the compactness of the solution set [10, Proposition 2.2.7]:

Proposition 4.2. *Let F conform to the following coerciveness property:*

$$\liminf_{\|x\| \rightarrow \infty} \frac{\langle F(x), x \rangle}{\|x\|^\eta} > 0,$$

for some $\eta \geq 0$. Then, there is a $\rho > 0$ such that $\langle F(x), x \rangle \geq 0$ for $\|x\| > \rho$. In particular, the coercive assumption of Theorem 4.1 holds with $M = 0$.

Proof. There must be an $\epsilon > 0$ such that

$$\liminf_{\|x\| \rightarrow \infty} \frac{\langle F(x), x \rangle}{\|x\|^\eta} > 2\epsilon.$$

This implies that there is a $\rho > 0$ such that if $\|x\| > \rho$, $\frac{\langle F(x), x \rangle}{\|x\|^\eta} \geq \epsilon$, which implies that $\langle F(x), x \rangle \geq 0$. □

5 Numerical Methods

In this section, we develop a semismooth Newton method for nonlinear complementarity problems based on W_c . We focus on the complementarity case as it has the most complete

exactness results and because the computation of the respective multiplier estimates is trivial, see (10).

The idea is to use the exact penalty to compute the Newton direction in a semismooth Newton method that will be globalized using the Fischer-Burmeister function $\varphi_{FB}(a, b) \stackrel{\text{def}}{=} \sqrt{a^2 + b^2} - a - b$ [12]. This function has the important property that, whenever $\varphi_{FB}(a, b) = 0$, both a and b are positive and complementary. Such functions are called NCP functions. Hence, the nonlinear complementarity problem can be rewritten as

$$\Phi_{FB}(x) \stackrel{\text{def}}{=} \begin{bmatrix} \varphi_{FB}(x_1, F(x)_1) \\ \vdots \\ \varphi_{FB}(x_n, F(x)_n) \end{bmatrix} = 0.$$

Under reasonable assumptions, the above system of equations is semismooth and can be solved using a semismooth Newton algorithm [23]. Moreover, $\Psi_{FB}(x) \stackrel{\text{def}}{=} 1/2 \|\Phi_{FB}(x)\|^2$ is differentiable and can be used to globalize the Newton method.

However, there are other important NCP functions whose least square reformulation is not differentiable. They do not have a natural globalization strategy. In this case, it is usual to build hybrid methods, where the local fast convergence is obtained by a Newton algorithm based on the desired NCP function, but the globalization is achieved using a differentiable merit function like Ψ_{FB} . Such globalization ideas are described in the recent books of Facchinei and Pang [10, 11]. A typical choice is the combination of the NCP function based on the minimum, $\Phi_{\min}(x) \stackrel{\text{def}}{=} \min(x, F(x))$, with a merit function based on Fischer-Burmeister. Such combination gives rise to many practical algorithms, see for example [5].

Before presenting the variant of the semismooth Newton method used in this paper, it is natural to search for regularity conditions that can ensure fast local convergence. The semismooth Newton method can be shown to converge superlinearly if all the elements of the B -subdifferential at the desired zero x^* are nonsingular [23]. Such zeroes are called BD -regular. In complementarity problems, the BD -regularity of the zeroes of a reformulation is usually connected to the concepts of b and R -regularity of the solutions:

Definition 5.1. *Let x^* be a solution to the (NCP). Define the index sets*

$$\begin{aligned} \alpha &\stackrel{\text{def}}{=} \{i \mid x_i^* > 0 = F(x^*)_i\}, \\ \beta &\stackrel{\text{def}}{=} \{i \mid x_i^* = 0 = F(x^*)_i\}, \\ \gamma &\stackrel{\text{def}}{=} \{i \mid x_i^* = 0 < F(x^*)_i\}. \end{aligned}$$

The solution x^ is said to be b -regular if the principal submatrices $JF(x^*)_{\alpha \cup \delta, \alpha \cup \delta}$ are nonsingular for all subsets $\delta \subset \beta$. It is called R -regular if $JF(x^*)_{\alpha, \alpha}$ is nonsingular and the*

Schur complement of this matrix in

$$\begin{bmatrix} JF(x^*)_{\alpha,\alpha} & JF(x^*)_{\alpha,\beta} \\ JF(x^*)_{\beta,\alpha} & JF(x^*)_{\beta,\beta} \end{bmatrix}$$

is a P -matrix.

In [5], the authors show that b -regularity is weaker than R -regularity. However, both conditions are equivalent in important cases, like when x^* is a nondegenerate solution or if F is a P_0 function, in particular if it is monotone.

We are ready to show the main regularity result for the penalty W_c .

Proposition 5.2. *Let F be a C^1 function and suppose that x^* is a b -regular solution of (NCP). Then, x^* is a BD -regular solution of the system $W_c(x) = 0$, where W_c is defined in (10) and $c > 0$.*

Proof. The regularity is actually inherited from the minimum function, and hence we follow the proof of [5, Proposition 2.10].

First, define

$$\begin{aligned} \xi_1(x) &\stackrel{\text{def}}{=} \{i \mid W_c(x)_i = (\zeta^2 x_i^2)/(1 + \zeta^2 x_i^2)F(x)_i + cx_i\}, \\ \xi_2(x) &\stackrel{\text{def}}{=} \{i \mid W_c(x)_i = F(x)_i\}. \end{aligned}$$

It is clear that W_c is differentiable at x if, and only if, $\xi_1(x) \cap \xi_2(x) = \emptyset$. In this case,

$$JW_c(x)_i = \begin{cases} \frac{\zeta^2 x_i^2}{1+\zeta^2 x_i^2} JF(x)_i + \left(\frac{2\zeta^2 x_i}{(1+\zeta^2 x_i^2)^2} F(x)_i + c \right) e^i, & \text{if } i \in \xi_1(x) \\ JF(x)_i, & \text{if } i \in \xi_2(x), \end{cases}$$

where $JW_c(x)_i$ denotes the i -th line of the Jacobian of W_c at x , $JF(x)_i$ is the analogous for F , and e^i is again the i -th canonical vector.

Let $H \in \partial_B W_c(x)$ and denote its i -th row H_i . Using the index sets from Definition 5.1, it is easy to see that:

1. If $i \in \alpha$, $H_i = JF(x^*)_i$.
2. If $i \in \gamma$, $H_i = \frac{\zeta^2 (x_i^*)^2}{1+\zeta^2 (x_i^*)^2} JF(x^*)_i + \left(\frac{2\zeta^2 x_i^*}{(1+\zeta^2 (x_i^*)^2)^2} F(x^*)_i + c \right) e^i = ce^i$.
3. If $i \in \beta$, $H_i = JF(x^*)_i$ or $H_i = ce^i$.

Hence, as in the proof of [5, Proposition 2.10], there is an index set $\delta \subset \beta$ such that

$$H = \begin{bmatrix} JF(x^*)_{\alpha \cup \delta, \alpha \cup \delta} & JF(x^*)_{\alpha \cup \delta, \alpha \cup \bar{\delta}} \\ 0_{\gamma \cup \bar{\delta}, \alpha \cup \delta} & cI_{\gamma \cup \bar{\delta}, \alpha \cup \bar{\delta}} \end{bmatrix},$$

where $\bar{\delta} = \beta \setminus \delta$. The result follows from the definition of b -regularity. \square

This proposition guarantees the fast local convergence of a semismooth Newton method that starts in a neighborhood of a BD-solution to (NCP), see Theorem 5.5.

We can now present a variant of the Newton method based on W_c and globalized by Ψ_{FB} . We choose to use a Levenberg-Marquardt method, as we used the LMMCP implementation of Kanzow and Petra [20] as a base for our code. Their program can be downloaded from Kanzow web page [18]. In order to make the test more interesting, we also consider the hybrid algorithm when the Newton direction is computed using Φ_{\min} instead of W_c . More formally, we follow the ‘‘General Line Search Algorithm’’ from [5], and propose the following modification to Algorithm 3.1 in [20]:

Semismooth Levenberg-Marquardt method with alternative search directions

(LMAD) Let $\Psi_{FB}(x) \stackrel{\text{def}}{=} 1/2\|\Phi_{FB}(x)\|^2$, and let G denote either W_c for a fixed $c > 0$ or Φ_{\min} . Choose $\epsilon_1, \epsilon_2 \geq 0$, $\bar{\mu} > 0$, $\alpha_1 > 0$. Choose $\alpha_2, \beta, \sigma_1 \in (0, 1)$, $\sigma_2 \in (0, \frac{1}{2})$. Set $k = 0$.

1. If $\Psi_{FB}(x) \leq \epsilon_1$ or $\|\nabla\Psi_{FB}(x^k)\| \leq \epsilon_2$, stop.

2. Compute the search direction:

(a) Compute $G(x^k)$, $H_k \in \partial_B G(x^k)$, and choose the Levenberg-Marquardt parameter $\mu_k \in (0, \bar{\mu}]$.

(b) Find d^k such that

$$(H'_k H_k + \mu_k I) d^k = -H'_k G(x^k).$$

(c) If

$$\Psi_{FB}(x^k + d^k) \leq \sigma_1 \Psi_{FB}(x^k),$$

set $x^{k+1} = x^k + d^k$, $k = k + 1$, and go to Step 1.

(d) If

$$\|d^k\| < \alpha_1 \|\nabla\Psi_{FB}(x^k)\|$$

or if

$$\langle d^k, \nabla\Psi_{FB}(x^k) \rangle > -\alpha_2 \|d^k\| \|\nabla\Psi_{FB}(x^k)\|,$$

change d^k to $-\nabla\Psi_{FB}(x^k) / \|\nabla\Psi_{FB}(x^k)\|$.

3. Find the largest value t_k in $\{\beta^l \mid l = 0, 1, 2, \dots\}$ such that

$$\Psi_{FB}(x^k + t_k d^k) \leq \Psi_{FB}(x^k) + \sigma_2 t_k \langle \nabla\Psi_{FB}(x^k), d^k \rangle.$$

Set $x^{k+1} = x^k + t_k d^k$, $k = k + 1$ and go to Step 1.

Note that the conditions in Step 2d ensure that the Armijo search in Step 3 is well defined and will stop in a finite number of steps.

To prove the global convergence of the LMAD we will use the spacer steps result presented in [4, Proposition 1.2.6]. In order to do so, let us recall the definition of a gradient related sequence:

Definition 5.3. *A direction sequence $\{d^k\}$ is gradient related to $\{x^k\}$ with respect to Ψ_{FB} if, for any subsequence $\{x^{k_j}\}$ that converges to a nonstationary point, the corresponding subsequence $\{d^{k_j}\}$ is bounded and satisfies*

$$\limsup_{k_j \rightarrow \infty} \langle d^{k_j}, \nabla \Psi_{FB}(x^{k_j}) \rangle < 0.$$

We can present now the convergence results:

Theorem 5.4. *Let $\{x^k\}$ be a sequence computed by the LMAD method. Then, every accumulation point is a stationary point of Ψ_{FB} .*

Proof. First, let us recall that Ψ_{FB} is continuously differentiable [20, Theorem 2.7].

Now, let \mathcal{K} be the set of indexes where the condition of Step 2c failed. That is, the set of indexes where the Armijo line search took place. We show that $\{d^k\}_{k \in \mathcal{K}}$ is gradient related as defined above.

Let $x^{k_j} \rightarrow x^*$, with $k_j \in \mathcal{K}$ converging to a nonstationary point of Ψ_{FB} . Observe that $\{d^{k_j}\}$ is bounded due to the boundedness of the B -subdifferential of G on bounded sets and the boundedness of the Levenberg-Marquardt parameters μ_{k_j} . Hence, without loss of generality we can assume that there is a $d^* \in \mathbb{R}^n$ such that

$$\begin{aligned} d^{k_j} &\rightarrow d^*, \\ \limsup_{k_j \rightarrow \infty} \langle d^{k_j}, \nabla \Psi_{FB}(x^{k_j}) \rangle &= \langle d^*, \nabla \Psi_{FB}(x^*) \rangle. \end{aligned}$$

We need to prove that $\langle d^*, \nabla \Psi_{FB}(x^*) \rangle < 0$.

If, in a given iteration k_j , the condition described in Step 2d was true, we would have $d^{k_j} = -\nabla \Psi_{FB}(x^{k_j}) / \|\nabla \Psi_{FB}(x^{k_j})\|$. If this happens in an infinite number of iterations, we get $d^* = -\nabla \Psi_{FB}(x^*) / \|\nabla \Psi_{FB}(x^*)\|$, and the desired inequality holds trivially. Therefore, we may assume that for big enough k_j ,

$$\langle d^{k_j}, \nabla \Psi_{FB}(x^{k_j}) \rangle \leq -\alpha_2 \|d^{k_j}\| \|\nabla \Psi_{FB}(x^{k_j})\| \leq -\alpha_1 \alpha_2 \|\nabla \Psi_{FB}(x^{k_j})\|^2.$$

Taking limits, it follows that $\langle d^*, \nabla \Psi_{FB}(x^*) \rangle < -\alpha_1 \alpha_2 \|\nabla \Psi_{FB}(x^*)\|^2 < 0$. Hence $\{d^k\}_{k \in \mathcal{K}}$ is gradient related to $\{x^k\}_{k \in \mathcal{K}}$. It follows from [4, Proposition 1.2.6], that every limit point of $\{x^k\}_{x \in \mathcal{K}}$ is stationary.

Finally, consider an arbitrary convergent subsequence $x^{k_j} \rightarrow x^*$, where it is not always true that $k_j \in \mathcal{K}$. If there is still an infinite subset of the indexes k_j that belong to \mathcal{K} , we can easily reduce the subsequence to this indexes to see that x^* is stationary. On the other hand, if $k_j \notin \mathcal{K}$ for all big enough k_j , we use the definition of \mathcal{K} to see that for these indexes

$$0 \leq \Psi_{FB}(x^{k_{j+1}}) \leq \Psi_{FB}(x^{k_j+1}) = \Psi_{FB}(x^{k_j} + d^{k_j}) \leq \sigma_1 \Psi_{FB}(x^{k_j}),$$

where the first inequality follows from the monotonicity of $\Psi_{FB}(x^k)$. Hence, it is trivial to see that $\Psi_{FB}(x^{k_j}) \rightarrow 0$, and then the monotonicity of $\Psi_{FB}(x^k)$ ensures that the whole sequence goes to zero. In particular, x^* minimizes Ψ_{FB} and $\nabla \Psi_{FB}(x^*) = 0$. \square

We can also present a standard result for local convergence rate.

Theorem 5.5. *Let $\{x^k\}$ be a sequence computed by the LMAD method. Assume that it converges to a b -regular solution to (NCP). If $\mu_k \rightarrow 0$, then eventually the condition in Step 2c will be satisfied and $\{x^k\}$ will converge Q -superlinearly to x^* . Moreover, if F is a LC^1 -function and $\mu_k = O(\|H'_k G(x^k)\|)$, we have that the convergence is Q -quadratic.*

Proof. Remember that G in LMAD is either Φ_{\min} or W_c for some $c > 0$. Using the BD -regularity at x^* of these functions, given by [5, Proposition 2.10] and Proposition 5.2 above, it follows that x^* is an isolated solution of the equation $G(x) = 0$. Moreover, as x^* is b -regular, it is an isolated solution to (NCP) [10, Corollary 3.3.9]. Hence it is also an isolated solution to the equation $\Psi_{FB}(x) = 0$.

As G is continuous, it follows that there is a neighborhood of x^* and constants $c_1 > 0$ and $c_2 > 0$ such that

$$c_1 \|G(x)\|^2 \leq \Psi_{FB}(x) \leq c_2 \|G(x)\|^2.$$

Note that this is a local version of Lemma 3.4 in [5].

The result now follows as in the proof of Theorem 3.2 and Theorem 4.4(b) in [5]. \square

5.1 Computational Tests

We have implemented the LMAD method in MATLAB starting from the LMMCP code from Kanzow and Petra. In the current version of the code, the authors have incorporated a filter trust region method as a preprocessor before starting the main algorithm [19]. Since we do not want to use the preprocessor in our tests, we have turn it off. Moreover, the code uses a combination of the Fischer-Burmeister function with an extra term to force complementarity. As we wanted to use the pure Fischer-Burmeister function we have adapted the code. This can be achieved, basically, setting a parameter to 1.

Let us describe the choice of the parameters for LMAD. Following the LMMCP we used $\epsilon_1 = 10^{-10}$, $\epsilon_2 = 0$, $\beta = 0.55$, and $\sigma_2 = 10^{-4}$. The Levenberg-Marquardt parameter is

chosen to be 10^{-16} if the estimated condition number for $H_k' H_k$ is less than 10^{25} . Otherwise, we use $\mu_k = 10^{-1}/(k + 1)$.

As for the constants that control the choice of the alternative direction we have $\sigma_1 = 0.5$, $\alpha_1 = \alpha_2 = \sqrt{\epsilon_{mac}}$, where ϵ_{mac} denotes that machine epsilon that stands approximately for $2.2204 \cdot 10^{-16}$. The value of the constants that define the exact penalty W_c were $c = 5$, and $\gamma = 0.2$.

The test set is composed by all the nonlinear complementarity problems in the MATLAB version of the MCPLIB test suite. It has 83 problems when we consider different starting points. The full list of problems is presented in the Appendix A.

Figure 1 presents the performance profile of this first test. The label “FB” stands for the pure Fischer-Burmeister method, while “Exact” represents the LMAD with $G = W_c$ and “Min” is the same method with $G = \Phi_{min}$. As the iterations of all three methods are dominated by the computation of the Newton step, we use the number of iterations as the performance metric.

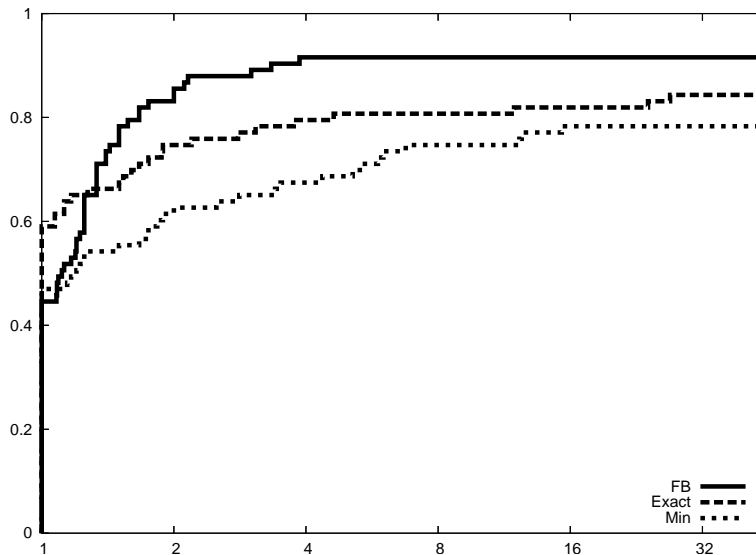


Figure 1: Performance profile of the LMAD variations and the pure Fischer-Burmeister method.

We note that the LMAD variation based on the exact penalty seems to be the fastest method by a small margin and a real improvement when compared to the LMAD based on the Φ_{min} function. However, it is less reliable than FB.

If we analyse the reason for the failures of the LMAD variants in more problems than the FB method, we identify that in some cases the direction computed is not a good descent direction for the merit function based on the Fischer-Burmeister reformulation. This force

the LMAD to use the Cauchy step for the merit function as search direction, resulting in a very small improvement.

Hence, it is natural to ask if it is possible to predict, *before the solution of the linear equation described in the Step 2b*, that the resulting direction may not be a good descent direction. In such case, we could try to use a Newton step based on the original merit function instead. With this objective, we propose the following modification of Step 2b:

Modified Alternative Direction *Let $\theta \in (0, \pi/2)$.*

If the angle between $H'_k G(x^k)$ and $\nabla \Psi_{FB}(x^k)$ is smaller than θ , find d^k such that

$$(H'_k H_k + \mu_k I) d^k = -H'_k G(x^k). \quad (12)$$

Otherwise, compute $\tilde{H}_k \in \partial_B \Psi_{FB}(x^k)$ and find d^k solving

$$(\tilde{H}'_k \tilde{H}_k + \mu_k I) d^k = -\nabla \Psi_{FB}(x^k). \quad (13)$$

The idea behind the angle criterion to choose between the alternative search direction and the direction based on the original merit function is simple to explain. The solution of (12) can bend its right hand side, $-H'_k G(x^k)$ by a maximum angle of $\pi/2$. Hence, if $-H'_k G(x^k)$ makes a small angle with $-\nabla \Psi_{FB}(x^k)$ the direction computed by the first linear system will be likely a good search direction. On the other hand, if this angle is large, the direction computed by the first linear system can only be a good descent direction if it is bent by the system towards $-\nabla \Psi_{FB}(x^k)$. But there is no guarantee that this will happen. To avoid taking chances, we use the direction based on the merit function itself, given by (13).

The convergence of the modified algorithm can be proved following the same lines of the proofs of Theorems 5.4 and 5.5. In particular, the inequalities that ensure that the search directions are gradient related remains untouched. As for the rate of convergence result, it would require R -regularity of the solution, instead of b -regularity, like in [5, Theorem 4.3]. This is a consequence that the Newton steps can be taken with respect to Ψ_{FB} and not G .

Figure 2 presents the performance profiles of the variations of the LMAD when we change Step 2b by the Modified Alternative Direction presented above. The parameter θ was set to $\pi/6$. Both variations, based on the exact penalty W_c and in Φ_{\min} , clearly benefit from the new directions, with a better improvement for the Φ_{\min} version. Note that both methods became more robust and a little faster.

Finally, Figure 3 shows the profile of the three methods together. Here, we can see that the method based on the exact penalty practically dominates the others in numbers of iterations. It also displays the same robustness of the FB version.

Acknowledgement. We would like to thank Ellen Hidemi Fukuda for carefully reading the manuscript.

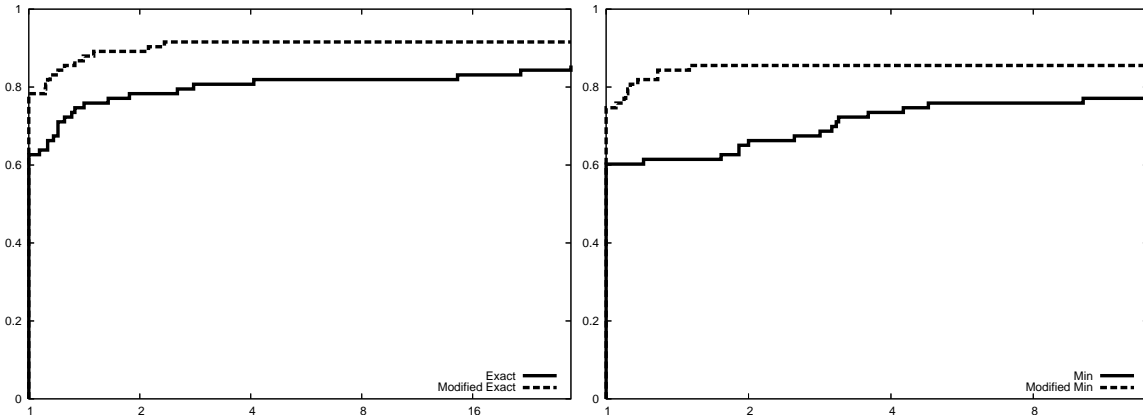


Figure 2: Performance profiles of the LMAD variations with and without the Modified Alternative Direction.

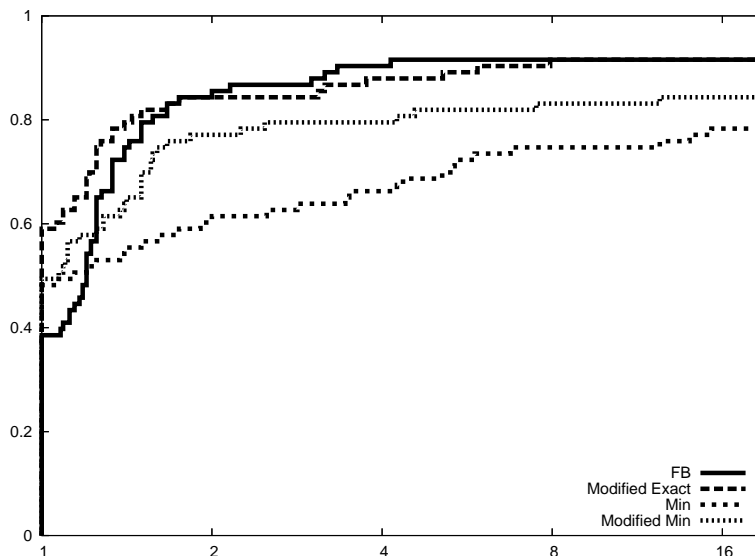


Figure 3: Performance profile of the LMAD variations using the Modified Alternative Direction and the pure Fischer-Burmeister method.

A Tables with numerical results

We present here the full tables that were used to draw the performance profiles. The first column shows the problem name, the next five columns presents the number of Newton steps for the original LMMCP method, for the LMAD using the exact penalty W_c and

Φ_{\min} , and, finally, for the LMAD with the Modified Alternative Direction based on W_c and Φ_{\min} functions respectively.

Problem	FB	Exact	Min.	Mod. Exact	Mod. Min
bertsekas1	28	30	Fail	25	39
bertsekas2	27	25	Fail	25	62
bertsekas3	16	24	Fail	17	68
bertsekas4	28	30	Fail	25	39
bertsekas5	12	9	46	9	15
bertsekas6	37	31	Fail	29	41
colvdual1	14	Fail	Fail	Fail	Fail
colvdual2	Fail	Fail	Fail	223	Fail
colvdual3	1	1	1	1	1
colvdual4	14	Fail	Fail	Fail	Fail
colvnlp1	14	14	Fail	16	22
colvnlp2	14	43	61	17	64
colvnlp3	1	1	1	1	1
colvnlp4	13	12	22	12	11
colvnlp5	13	12	21	12	11
colvnlp6	9	8	48	9	10
cycle1	3	3	3	3	3
explcp1	19	9	17	19	6
hanskoop10	1	1	1	1	1
hanskoop2	1	1	1	1	1
hanskoop4	1	1	1	1	1
hanskoop6	1	1	1	1	1
hanskoop8	1	1	1	1	1
josephy1	6	8	5	6	5
josephy2	6	4	5	5	5
josephy3	58	23	15	14	15
josephy4	5	5	3	4	3
josephy5	4	3	3	3	3
josephy6	6	6	7	5	7
josephy7	5	14	17	5	4
josephy8	3	2	2	2	2

Problem	FB	Exact	Min.	Mod. Exact	Mod. Min
kojshin1	8	15	98	8	7
kojshin2	7	4	6	4	6
kojshin3	6	266	35	9	45
kojshin4	4	3	3	3	3
kojshin5	5	3	3	3	3
kojshin6	5	5	6	5	5
kojshin7	6	Fail	Fail	5	4
kojshin8	3	2	2	2	2
mathinum1	4	15	4	15	6
mathinum2	5	4	4	4	4
mathinum3	12	7	4	7	4
mathinum4	7	5	5	5	5
mathinum5	10	10	10	10	10
mathinum6	7	7	7	7	7
mathisum1	5	4	4	4	4
mathisum2	5	4	4	4	4
mathisum3	6	3	5	7	5
mathisum4	6	5	5	5	5
mathisum5	1	1	1	1	1
mathisum6	7	6	6	6	6
mathisum7	10	3	3	4	3
nash1	8	6	6	6	6
nash2	10	9	11	8	11
nash3	7	5	62	5	62
nash4	5	4	5	4	5
pgvon1054	41	Fail	Fail	Fail	Fail
pgvon1055	Fail	Fail	59	98	Fail
pgvon1056	Fail	Fail	297	Fail	Fail
pgvon1064	Fail	Fail	Fail	Fail	Fail
pgvon1065	Fail	Fail	Fail	Fail	Fail
pgvon1066	Fail	Fail	Fail	Fail	Fail
powell1	8	9	20	8	8
powell2	8	6	6	6	7
powell3	11	7	8	7	9
powell4	10	7	7	7	9

Problem	FB	Exact	Min.	Mod. Exact	Mod. Min
powell5	Fail	79	36	111	36
powell6	5	4	7	4	4
scarfanum1	11	17	9	13	10
scarfanum2	12	14	11	14	12
scarfanum3	10	10	10	12	11
scarfanum4	4	2	2	2	2
scarfbnum1	24	111	Fail	123	Fail
scarfbnum2	20	481	Fail	118	380
scarfbsum1	15	Fail	Fail	119	Fail
scarfbsum2	28	13	Fail	13	Fail
sppe1	8	9	43	10	12
sppe2	6	7	21	6	11
sppe3	5	4	8	6	8
tobin1	11	131	31	9	10
tobin2	8	216	54	10	18
tobin3	10	Fail	153	10	15
tobin4	2	2	2	2	2
scarfanum1	11	17	9	13	10
scarfanum2	12	14	11	14	12
scarfanum3	10	10	10	12	11
scarfanum4	4	2	2	2	2
scarfbnum1	24	111	Fail	123	Fail
scarfbnum2	20	481	Fail	118	380
scarfbsum1	15	Fail	Fail	119	Fail
scarfbsum2	28	13	Fail	13	Fail
sppe1	8	9	43	10	12
sppe2	6	7	21	6	11
sppe3	5	4	8	6	8
tobin1	11	131	31	9	10
tobin2	8	216	54	10	18
tobin3	10	Fail	153	10	15
tobin4	2	2	2	2	2

References

- [1] A. Auslender and M. Teboulle. Lagrangian duality and related multiplier methods for variational inequality problems. *SIAM Journal on Optimization*, 10(4):1097–1115, 2000.
- [2] A. Auslender and M. Teboulle. *Asymptotic cones and functions in optimization and variational inequalities*. Springer Monographs in Mathematics. Springer-Verlag, New York, 2003.
- [3] D. Bertsekas. *Constrained Optimization and Lagrange Multipliers*. Academic Press, New York, 1982.
- [4] D. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1995.
- [5] T. De Luca, F. Facchinei, and C. Kanzow. A theoretical and numerical comparison of some semismooth algorithms for complementarity problems. *Computational Optimization and Applications*, 16:173–205, 2000.
- [6] G. di Pillo and L. Grippo. A new class of augmented Lagrangians in nonlinear programming. *SIAM Journal on Control and Optimization*, 17(5):618–628, 1979.
- [7] G. Di Pillo and L. Grippo. A continuously differentiable exact penalty function for nonlinear programming problems with inequality constraints. *SIAM Journal on Control and Optimization*, 23(1):72–84, 1985.
- [8] G. Di Pillo and L. Grippo. Exact penalty functions in constrained optimization. *SIAM Journal on Control and Optimization*, 27(6):1333–1360, 1989.
- [9] J. Eckstein and M. C. Ferris. Smooth methods of multipliers for complementarity problems. *Mathematical Programming*, 86(1, Ser. A):65–90, 1999.
- [10] F. Facchinei and J.-S. Pang. *Finite-dimensional variational inequalities and complementarity problems, Vol. I*. Springer Series in Operations Research. Springer-Verlag, New York, 2003.
- [11] F. Facchinei and J.-S. Pang. *Finite-dimensional variational inequalities and complementarity problems, Vol. II*. Springer Series in Operations Research. Springer-Verlag, New York, 2003.
- [12] A. Fischer. A special newton-type optimization method. *Optimization*, 24:269–284, 1992.

- [13] R. Fletcher. A class of methods for nonlinear programming with termination and convergence properties. In *Integer and Nonlinear Programming*, pages 157–175. North-Holland, Amsterdam, 1970.
- [14] R. Fletcher. A class of methods for nonlinear programming. III. Rates of convergence. In *Numerical Methods for Nonlinear Optimization (Conf., Dundee, 1971)*, pages 371–381. Academic Press, London, 1972.
- [15] R. Fletcher. An exact penalty function for nonlinear programming with inequalities. *Mathematical Programming*, 5:129–150, 1973.
- [16] R. Fletcher and S. A. Lill. A class of methods for nonlinear programming. II. Computational experience. In *Nonlinear Programming (Proc. Sympos., Univ. of Wisconsin, Madison, Wis., 1970)*, pages 67–92. Academic Press, New York, 1970.
- [17] T. Glad and E. Polak. A multiplier method with automatic limitation of penalty growth. *Mathematical Programming*, 17(2):140–155, 1979.
- [18] C. Kanzow. C. Kanzow web site, <http://www.mathematik.uni-wuerzburg.de/~kanzow/>. Accessed in January 2007.
- [19] C. Kanzow and S. Petra. Projected filter trust region methods for a semismooth least squares formulation of mixed complementarity problems. to appear in *Optimization Methods & Software*.
- [20] C. Kanzow and S. Petra. On a semismooth least squares formulation of complementarity problems with gap reduction. *Optimization Methods & Software*, 19(5):507–525, 2004.
- [21] H. Mukai and E. Polak. A quadratically convergent primal-dual algorithm with global convergence properties for solving optimization problems with equality constraints. *Mathematical Programming*, 9(3):336–349, 1975.
- [22] T. Pennanen. Dualization of generalized equations of maximal monotone type. *SIAM Journal on Optimization*, 10(3):809–835, 2000.
- [23] L. Qi. Convergence analysis of some algorithms for solving nonsmooth equations. *Mathematics of Operations Research*, 18(1):227–244, 1993.
- [24] L. Qi and J. Sun. A nonsmooth version of Newton’s method. *Mathematical Programming*, 58(3, Ser. A):353–367, 1993.
- [25] R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1:97–116, 1976.