A box-constrained optimization algorithm with negative curvature directions and spectral projected gradients

E. G. Birgin * J. M. Martínez †

February 13, 2001

Abstract

A practical algorithm for box-constrained optimization is introduced. The algorithm combines an active-set strategy with spectral projected gradient iterations. In the interior of each face a strategy that deals efficiently with negative curvature is employed. Global convergence results are given. Numerical results are presented.

Keywords: box constrained minimization, active set methods, spectral projected gradients, dogleg path methods.

AMS Subject Classification: 49M07, 49M10, 65K, 90C06, 90C20.

1 Introduction

The problem considered in this paper is

$$\text{Minimize } f(x) \text{ subject to } x \in \Omega, \tag{1}$$

where

$$\Omega = \{x \in I\!R^n \mid \ell \leq x \leq u\}, \tag{2}$$

$\ell, u \in I\!R^n$, $\ell < u$ and $f : I\!R^n \to I\!R$ has continuous second partial derivatives.

*Department of Computer Science IME-USP, University of São Paulo, Rua do Matão 1010, Cidade Universitária, 05508-900, São Paulo SP, Brazil. This author was supported by PRONEX-Optimization 76.79.1008-00 and FAPESP (Grant 99/08029-9). e-mail: egbirgin@ime.usp.br

†Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil. This author was supported by PRONEX-Optimization 76.79.1008-00, FAPESP (Grant 90-3724-6), CNPq and FAEP-UNICAMP. e-mail: martinez@ime.unicamp.br

In a recent paper [9] a spectral projected gradient method (SPG) was introduced for solving optimization problems with convex constraints. This algorithm is easily implementable when the constraints are simple, as it is the case of the box (2). Numerical tests shown in [9] reveal that SPG can be quite efficient in very large and practical problems. See, also, [7, 8, 32].

In this research, SPG iterations are combined with an active set strategy that allows one to use negative curvature directions in the active face. Essentially, what we do here is to use SPG when an adequate test reveals that the current face must be abandoned and to use the algorithm recently introduced in [35] when the test recommends to stay in the current face.

The motivation for this research came from the observation of the behavior of some numerical algorithms. The algorithm described in [25] is a box-constrained method for large-scale problems that uses trust regions and, at each iteration, solves a box-constrained quadratic subproblem using a specific quadratic solver. The quadratic solver was introduced in [22] and updated in [6, 14, 24, 25]. A related algorithm was introduced in [16] where, also, interesting physical interpretations were given. See, also, [17, 18, 19, 20]. This box-constraint quadratic algorithm uses an active-set strategy by means of which one stays in a face or abandons it according to the relation between components of the gradient. The quadratic method solved successfully many applied problems [17, 18, 19] and was shown to be more efficient than other large-scale quadratic solvers in numerical studies [14, 15]. This motivated us to apply the principles of the quadratic solver directly to the original bound constrained problem (1), instead of using these principles on a subproblem. In fact, the philosophy of [6, 14, 22, 24, 25] seems to be independent of the fact of dealing with a quadratic or not.

The present research represents our first attempt of constructing an algorithm that solves (1) using the principles of [22]. For this implementation we used two recently introduced tools for unconstrained optimization and convex-constrained optimization. The unconstrained optimization algorithm used inside the faces is the dogleg method defined in [35], which is able to deal efficiently with negative curvatures. The algorithm used for leaving the faces is the spectral projected gradient introduced in [9] (see also [33, 34, 23]) which uses first order information in a very efficient and economical way. Although our final objective is a method for large-scale problems (we have in mind its association with augmented Lagrangian algorithms and other algorithms for general constrained optimization like the ones in [29, 30, 31, 11, 12]), we found it useful to begin our numerical research with an implementation that uses full factorization of matrices.

In practical numerical analysis, the efficiency of many algorithms for a class of problems is strongly linked to the efficiency of algorithms for solving simpler problems. The reason is that simpler problems play the role of subproblems that must be solved (perhaps many times) by the main algorithm. For example, the efficiency of trust region methods like the one described in [25] is related to the development of algorithms for minimizing box-constrained quadratics. On the other hand, the efficiency of active set methods is related to the development of unconstrained minimization algorithms. The development of new attractive unconstrained optimization methods like the one in [35] naturally stimulates its application to active-set strategies in constrained minimization. This was an independent motivation for our present research.

The organization of this paper is as follows. The algorithm is described in Section 2. In Section 3, basic global convergence theorems are proved. In Section 4 we present numerical experiments. Conclusions are given in Section 5.

2 The algorithm

As in [22], let us divide the feasible set Ω into disjoint open faces, as follows. For all $I \subset \{1, 2, \ldots, n, n+1, n+2, \ldots, 2n\}$, we define

$$F_I = \{x \in \Omega \mid x_i = \ell_i \text{ if } i \in I,\ x_i = u_i \text{ if } n+i \in I,\ \ell_i < x_i < u_i \ \text{ otherwise}\}.$$

We also define V_I the smallest affine subspace that contains F_I and S_I the parallel linear subspace to V_I.

Throughout this paper, $\|\cdot\|$ will be the Euclidean norm, although many times it can be replaced by an arbitrary norm on $I\!R^n$. For all $y \in I\!R^n$, the orthogonal projection of y onto a convex set S will be denoted $P_S(y)$. The (continuous) projected gradient at $x \in \Omega$ is defined as

$$g_P(x) = P_\Omega(x - \nabla f(x)) - x.$$

For all $x \in F_I$, we define

$$g_I(x) = P_{S_I}[g_P(x)].$$

The main algorithm considered in this paper is described below.

Algorithm 2.1
Assume that $x^0 \in \Omega$ is an arbitrary initial point, $\eta \in (0, 1)$ and $0 < \sigma_{min} \leq \sigma_{max} < \infty$. Let F_I be the face that contains the current iterate x^k. Assume that $g_P(x^k) \neq 0$ (otherwise the algorithm terminates). At the main iteration of the algorithm we perform the test:

$$\|g_I(x^k)\|/\|g_P(x^k)\| \geq \eta. \tag{3}$$

If (3) takes place, we judge that it is convenient that the new iterate belongs to $\bar{F}_I$ (the closure of F_I) and, so, we compute x^{k+1} using Algorithm 2.2 below. If (3) does not hold, we decide that some constraints should be abandoned and, so, the new iterate x^{k+1} is computed using the SPG iteration described by Algorithm 2.3. In this case, before the computation of x^{k+1} we compute the spectral gradient coefficient σ_k in the following way. If $k > 0$ then

$$\sigma'_k = \frac{(x^k - x^{k-1})^T(\nabla f(x^k) - \nabla f(x^{k-1}))}{(x^k - x^{k-1})^T(x^k - x^{k-1})}.$$

Otherwise, define $\kappa = \frac{\nabla f(x^k)^T \nabla f(x^k)}{\nabla f(x^k)^T H(x^k) \nabla f(x^k)}$. If $\kappa > 0$ set $\sigma'_k = \kappa$, else define $\sigma'_k = 1/||g_P(x^k)||$. Finally,

$$\sigma_k = \begin{cases} \sigma_{max}, & \text{if } \sigma'_k < 0; \\ \min\{\sigma_{max}, \max\{\sigma_{min}, \sigma'_k\}\}, & \text{otherwise.} \end{cases}$$

The algorithm used inside the faces is the following:

Algorithm 2.2

Assume, without loss of generality, that the first m variables are free ($\ell_j < x_j^k < u_j \ \forall\, i = 1, \ldots, m$) at face F_I and $x_j = x_j^k$ for all $x \in F_I$. Define

$$\varphi(x_1, \ldots, x_m) = f(x_1, \ldots, x_m, x_{m+1}^k, \ldots, x_n^k).$$

So, $\varphi : I\!R^m \to I\!R$. Let Γ be the piecewise linear path defined by the Zhang-Xu algorithm [35] for the minimization of φ, using $(x_1^k, \ldots, x_m^k)$ as current point. If $\Gamma \subset \Omega$, compute $y \in I\!R^m$ as the "next iterate" given by the algorithm [35] and define $x^{k+1} = (y_1, \ldots, y_m, x_{m+1}^k, \ldots, x_n^k)$. Otherwise, proceed in one of the following ways:

(i) Let y be the trial point defined by the algorithm [35]. If

$$f(P_\Omega(y_1, \ldots, y_m, x_{m+1}^k, \ldots, x_n^k)) < f(x^k),$$

define $x^{k+1} = P_\Omega(y_1, \ldots, y_m, x_{m+1}^k, \ldots, x_n^k)$. Otherwise, proceed as in (ii).

(ii) Replace Γ by the connected component of $\Gamma \cap \Omega$ that contains x^k and compute y as the next iterate of the dogleg algorithm [35], taking the point of Γ that belongs to the boundary of F_I as first trial point. Define

$$x^{k+1} = (y_1, \ldots, y_m, x_{m+1}^k, \ldots, x_n^k).$$

Algorithm 2.3 is the algorithm used when it is necessary to leave the current face, according to the test (3).

Algorithm 2.3
Compute x^{k+1} as the next iterate of a monotone SPG2 iteration [9] with the spectral step σ_k.

Remark. Observe that $x^{k+1} \notin \bar{F}_I$ if $x^k \in F_I$ and x^{k+1} is computed by Algorithm 2.3. In this case, (3) does not hold, so $\|g_P(x^k)\| > \|g_I(x^k)\|$. Since the components corresponding to the free variables of $g_I(x^k)$ and $g_P(x^k)$ are the same, this means that $g_P(x^k)$ has nonnull components corresponding fixed variables. Therefore, $x^k + \alpha g_P(x^k) \notin \bar{F}_I$ for all $\alpha > 0$. So, $P_\Omega(x^k + \alpha g_P(x^k)) \notin \bar{F}_I$ for all $\alpha > 0$. But, according to the SPG2 iteration,

$$x^{k+1} = x^k + \alpha'[P_\Omega(x^k + \alpha g_P(x^k)) - x^k]$$

for some $\alpha > 0, \alpha' > 0$. This implies that $x^{k+1} \notin \bar{F}_I$.

3 Convergence

Theorem 3.1. *Algorithm 2.1 is well defined.*

Proof. This is a trivial consequence of the fact that the algorithms defined in [35] and [9] are well defined. □

Theorem 3.2. *Suppose that there exists $\bar{k} \in \{0, 1, 2, \ldots\}$ such that $x^k \in F_I$ for all $k \geq \bar{k}$. Then, every limit point of $\{x^k\}$ is first-order stationary and, moreover, the Hessian submatrix corresponding to the free variables is positive semidefinite at each limit point.*

Proof. In this case, x^{k+1} is computed by Algorithm 2.2 for all $k \geq \bar{k}$. This corresponds, essentially, to use the algorithm [35] for the unconstrained minimization of φ. Therefore, with slight modifications, the convergence theorems of [35] can be applied. Thus, the gradient $\nabla\varphi(x^k)$ tends to zero and the Hessian of φ is positive semidefinite at the limit point. But, by a straightforward projection argument, it follows that $\|g_I(x^k)\| \leq \|\nabla\varphi(x^k)\|$, therefore $\|g_I(x^k)\| \to 0$. Since (3) holds, this implies that $\|g_P(x^k)\| \to 0$. So, every limit point is first-order stationary. □

Theorem 3.3. *Suppose that for all $k \in \{0, 1, 2, \ldots\}$, $x^k \in F_I$, there exists $k' > k$ such that $x^{k'} \notin F_I$. Then, there exists a limit point of $\{x^k\}$ that is* *first-order stationary.*

Proof. If $x^{k+1} \notin F_I$, then

$$x^{k+1} \in \bar{F}_I - F_I \tag{4}$$

or

$$x^{k+1} \notin \bar{F}_I. \tag{5}$$

But, since the number of constraints is finite, (4) cannot take place infinitely many times unless (5) also holds infinitely many times. Therefore, the hypothesis of the theorem implies that (5) holds for infinitely many iterations. This implies that there exists a subsequence $\{x^k, k \in K\}$ such that x^{k+1} is computed by Algorithm 2.3 for all $k \in K$. Therefore, the whole sequence can be considered as monotone decreasing f-sequence ($f(x^{k+1}) \leq f(x^k)\ \forall\ k$) with infinitely many spectral gradient iterations. Therefore, updating the proof of convergence of SPG2 given in [9], we conclude that all the limit points of $\{x^k, k \in K\}$ are stationary. This completes the proof. □

Theorem 3.4. *Suppose that all the stationary points of (1) are nondegenerate. ($\frac{\partial f}{\partial x_i}(x) = 0$ only if $\ell_i < x_i < u_i$.) Then, the hypothesis of Theorem 3.2 (and, hence, its thesis) must hold.*

Proof. Suppose that the thesis is not true. Therefore, the hypothesis of Theorem 3.3 holds. Since the number of faces is finite, there exists a face F_I and an infinite set K such that $x^k \in F_I$ and $x^{k+1} \notin \bar{F}_I$ for all $k \in K$. Let x^* be a limit point of $\{x^k, k \in K\}$. By Theorem 3.3, this point is first-order stationary. Without loss of generality, assume that in F_I the first m variables are free and the remaining $n-m$ are on its lower bound. Therefore,

$$\frac{\partial f}{\partial x_j}(x^*) > 0\ \ \forall\ \ j = m+1, \ldots, n.$$

By continuity this implies that, for k large enough, $k \in K$,

$$\frac{\partial f}{\partial x_j}(x^k) > 0\ \ \forall\ \ j = m+1, \ldots, n.$$

Therefore, for all $k \in K$, k large enough, the nonnull components of $g_P(x^k)$ are the nonnull components of $g_I(x^k)$. So, $g_P(x^k) = g_I(x^k)$ and, hence, the test (3) is satisfied. So, $x^{k+1} \in \bar{F}_I$, which is a contradiction. □

4 Numerical experiments

In order to assess the reliability of the new algorithm, we tested it against the well known package LANCELOT [11] using a set of bound constrained problems with more than 10 variables and less than 500 variables from the CUTE [10] collection (version of May 28^{th}, 1998). We present the results in two tables. The first corresponds to non-quadratic problems with less than 100 variables and the second to problems with more than 100 variables. As a whole, we have 60 problems. In both cases, our method used $\eta = 0.1$. In the small-dimensional set Algorithm 2.1 used the strategy (ii). The strategy (i) was used in the large-dimensional tests.

In our numerical experiments we used all except one of the defalut options for LANCELOT, we chose a full-matrix preconditioner (bandsolver with band $= n$). This is the adequate option that must be tested against an algorithm that uses full factorization of matrices, as the one presented in this paper. Therefore, at a LANCELOT iteration, the sequence of procedures to obtain the trial point are:

(i) Obtaining the "Cauchy point" [11]. If the Cauchy point is a reasonable approximate solution of the quadratic model, then it is chosen as the trial point. Otherwise, the step (ii) is executed.

(ii) Perform preconditioned conjugate gradient iterations in the face determined by the Cauchy point. If a face of lower dimension is encountered, preconditioned conjugate gradient iterations continue in the smaller face. This finishes when the norm of the internal gradient to the face is small enough. No constraints are abandoned in this process.

If the Cauchy point is interior, the Hessian is positive definite, and the first conjugate gradient iteration is interior too, then the trial point is a full Newton step and coincides with the first conjugate gradient iteration. More conjugate gradient iterations can be necessary if the conjugate-gradient iterate falls outside the trust region or if the Hessian is not positive definite.

For deciding when to stop the execution of the algorithms declaring convergence we used the criterion $\|g_P(x_k)\|_\infty \leq 10^{-5}$. We also stopped the execution when 1000 function evaluations were completed without achieving convergence.

All the experiments were run in a SPARCstation Sun Ultra 1, with an UltraSPARC 64 bits processor, 167-MHz clock and 128-MBytes of RAM memory. All codes are in Fortran and were compiled with f77 compiler (SC 4.0 Fortran 77 4.0) using the -O optimization compiler option.

In the tables, GE means "gradient evaluations", FE means "function evaluations" and TIME is CPU time in seconds. For the new method, we report HE (Hessian evaluations) and in LANCELOT we report CG (conjugate

	New method					LANCELOT				
Problem (n)	GE	FE	HE	TIME	f	GE	FE	CG	TIME	f
CHEBYQAD (50)	138	186	137	26.24	5.386D−03	70	88	176	6.85	5.386D−03
DECONVB (61)	51	60	46	1.08	8.638D−03	16	20	108	0.31	5.720D−03
EXPLIN (12)	22	22	19	0.01	−6.850D+03	11	13	11	0.03	−6.850D+03
EXPLIN2 (12)	22	22	19	0.01	−7.092D+03	12	13	15	0.03	−7.092D+03
EXPQUAD (12)	18	18	16	0.01	−4.201D+03	12	14	20	0.04	−4.201D+03
HADAMALS (36)	40	41	38	0.19	1.303D+01	17	18	129	0.15	1.308D+01
HATFLDC (25)	5	5	4	0.01	8.264D−14	5	5	3	0.03	7.770D−19
HS110 (50)	3	3	2	0.02	−9.990D+09	2	2	0	0.01	−9.990D+09
LINVERSE (19)	7	7	6	0.01	7.000D+00	15	18	29	0.06	6.000D+00
MCCORMCK (50)	6	6	5	0.02	−4.613D+01	6	7	5	0.05	−4.613D+01
NONSCOMP (50)	10	10	9	0.02	1.180D−18	9	9	8	0.06	8.207D−17
PROBPENL (50)	174	211	172	0.99	−2.331D+05	2	3	0	0.03	3.918D−06
QR3DLS (40)	31	39	30	0.12	6.707D−13	28	33	32	0.17	8.376D−13
QRTQUAD (12)	20	31	17	0.01	−3.608D+03	36	44	42	0.08	−3.608D+03
S368 (50)	20	20	19	0.92	−3.063D+01	7	8	9	0.46	−3.400D+01
SCON1LS (52)	629	882	628	2.17	1.086D−16	339	397	1011	1.95	2.169D−10
SINEALI (20)	9	9	8	0.01	−1.896D+03	872	1000	1211	1.81	−1.901D+03

Table 1: Non-quadratic problems with dimension between 10 and 99.

gradient iterations). The number of Hessian evaluations in LANCELOT is GE−1.

Looking at Tables 1 and 2 we observe that:

1. In 9 problems the algorithms clearly arrived to different solutions. In 6 cases (SINEALI(20), S368(50), LINVERSE(19), S368(100), HADAMALS(400) and NCVXBQP3(100)) the solution obtained by LANCELOT was better than the one obtained by the new method.

2. In 8 problems, it is not clear whether different functional values at the final point correspond to different solutions or to different final precisions. The new method was better than LANCELOT in 5 of these cases.

3. Concerning the 43 problems where both methods arrived to the same solution, we observed that our method used less computer time than LANCELOT in 35 problems. However, it must be warned that in 19 of these problems the CPU time used by LANCELOT cannot be justified in terms of GE, FE or CG. Considering that the same compilation options have been used for both methods, these time differences must be due to implementation details. It is worth noting that, to compute the objective function as well as its gradient and Hessian, LANCELOT exploits structure in a more efficient way than the one provided by the interface tools of the CUTE collection [10].

	New method					LANCELOT				
Problem (n)	GE	FE	HE	TIME	f	GE	FE	CG	TIME	f
BDEXP (100)	13	13	12	0.09	3.471D−05	11	11	10	0.22	3.918D−05
BIGGSB1 (100)	103	287	52	0.31	1.500D−02	52	51	50	0.35	1.500D−02
CHENHARK (100)	6	6	5	0.04	−2.000D+00	25	24	61	0.25	−2.000D+00
CVXBQP1 (100)	6	6	5	0.04	2.273D+02	2	1	0	0.07	2.273D+02
EXPLIN (120)	40	40	37	0.07	−7.238D+05	14	13	50	0.11	−7.238D+05
EXPLIN2 (120)	41	42	38	0.07	−7.245D+05	12	11	24	0.10	−7.245D+05
EXPQUAD (120)	24	28	21	0.58	−3.626D+06	16	18	46	0.53	−3.626D+06
GRIDGENA (170)	4	4	3	0.08	4.800D+02	4	3	3	0.14	4.800D+02
HADAMALS (400)	93	114	84	101.71	2.098D+02	29	28	1989	35.79	1.121D+02
HARKERP2 (100)	37	37	36	14.74	−5.000D−01	2	1	2	0.14	−5.000D−01
JNLBRNG1 (100)	5	6	3	0.03	−1.790D−01	2	1	1	0.06	−1.790D−01
JNLBRNG2 (100)	5	5	3	0.03	−3.953D+00	3	2	2	0.07	−3.953D+00
JNLBRNGA (100)	6	6	2	0.02	−3.612D−01	3	2	2	0.07	−3.612D−01
JNLBRNGB (100)	6	6	3	0.02	−7.255D+00	4	3	3	0.07	−7.255D+00
LINVERSE (199)	12	16	11	0.28	6.800D+01	25	30	473	5.49	6.800D+01
MCCORMCK (100)	6	6	5	0.05	−9.179D+01	6	7	5	0.13	−9.179D+01
NCVXBQP1 (100)	52	52	51	0.34	−1.996D+06	2	1	0	0.07	−1.996D+06
NCVXBQP2 (100)	72	72	70	0.49	−1.333D+06	3	2	4	0.07	−1.333D+06
NCVXBQP3 (100)	56	56	52	0.40	−6.601D+05	4	3	4	0.08	−6.708D+05
NOBNDTOR (484)	14	26	7	1.14	−4.980D−01	7	6	6	2.58	−4.980D−01
NONSCOMP (100)	10	10	9	0.06	3.652D−18	9	9	8	0.18	2.387D−16
OBSTCLAE (100)	6	6	4	0.04	1.398D+00	3	2	29	0.12	1.398D+00
OBSTCLAL (100)	8	9	3	0.02	1.398D+00	4	3	3	0.07	1.398D+00
OBSTCLBL (100)	7	7	3	0.02	2.875D+00	3	2	6	0.07	2.875D+00
OBSTCLBM (100)	3	3	2	0.02	2.875D+00	2	1	3	0.07	2.875D+00
OBSTCLBU (100)	4	4	1	0.01	2.875D+00	2	1	1	0.06	2.875D+00
PROBPENL (500)	701	1000	700	2438.96	−9.215D+04	2	1	0	4.44	3.992D−07
QR3DLS (155)	163	215	162	17.44	7.084D−12	112	142	286	10.36	2.719D−11
QRTQUAD (120)	21	31	18	0.49	−3.625D+06	100	126	160	2.26	−3.625D+06
S368 (100)	19	20	18	4.39	−1.260D+02	7	7	8	2.21	−1.337D+02
SCON1LS (102)	568	772	567	7.26	6.906D−16	684	820	755	12.40	1.945D−16
TORSION1 (484)	16	28	7	0.76	−4.561D−01	8	7	7	1.09	−4.561D−01
TORSION2 (484)	10	13	6	1.02	−4.561D−01	5	4	112	8.58	−4.561D−01
TORSION3 (484)	8	9	3	0.28	−1.242D+00	4	3	3	0.33	−1.242D+00
TORSION4 (484)	11	11	7	0.75	−1.242D+00	5	4	115	2.49	−1.242D+00
TORSION5 (484)	4	7	1	0.09	−2.885D+00	2	1	1	0.28	−2.885D+00
TORSION6 (484)	7	7	5	0.61	−2.885D+00	4	3	24	0.52	−2.885D+00
TORSIONA (484)	14	23	6	0.68	−4.161D−01	7	6	6	1.03	−4.161D−01
TORSIONB (484)	10	15	6	1.08	−4.161D−01	5	4	114	10.39	−4.161D−01
TORSIONC (484)	8	10	3	0.24	−1.199D+00	4	3	3	0.34	−1.199D+00
TORSIOND (484)	11	11	7	0.78	−1.199D+00	6	5	166	6.97	−1.199D+00
TORSIONE (484)	4	7	1	0.09	−2.841D+00	2	1	1	0.28	−2.841D+00
TORSIONF (484)	7	7	5	0.63	−2.841D+00	4	3	34	0.62	−2.841D+00

Table 2: Bound constrained problems with dimension between 100 and 500.

4. The number of SPG iterations is, in our method, equal to the difference between GE-1 and HE. It is perhaps surprising that, in most cases, this number is very small. This means that, in many cases, the new method works on faces of decreasing dimensions. This may be due to a characteristic of the initial points of the test set [10]. Therefore, the internal algorithm [35] have more influence in the overall behavior of the method than the strategy for leaving the faces.

5 Conclusions

We have introduced a reliable algorithm for solving small to medium dimensional box constrained problems. The algorithm is of active-set type and inside the faces it uses the method introduced in [35]. Since the algorithm [35] is able to deal with negative curvature regions and, in fact, cannot converge to a stationary point where the Hessian is indefinite, we expected that its combination with SPG could have the property of finding better minimizers than alternative first-order methods. Numerical experiments did not confirm this conjecture. Whereas in some cases we obtained lower functional values than LANCELOT, the opposite situation occurred in at least as many cases. In the cases in which our method and LANCELOT found the same solution, the computer time of the new method tends to be smaller, but in many cases this seems to be due to implementation features of LANCELOT that are not easy to detect.

In these set of problems it has been quite remarkable the monotone decrease property of the dimension of the current face. We do not think that such a property will be maintained in large-scale problems (the kind of situation where the quadratic algorithm [22] and its updatings showed to be effective). Therefore, the objective of finding an efficient algorithm for general box constraints that uses the philosophy of [22] remains valid. The present research shows that, perhaps, adapting the algorithm [35] to large-scale optimization by means of sparse Cholesky and Bunch-Parlett factorizations is not as promising as we expected. However, much research should be expected along these lines. In any case, we think that the work in [35] opened the possibility of dealing very cheaply with trust-region paths. Refinements of the main ideas of [35] will possibly give rise of even more efficient implementations of the unconstrained algorithm and can motivate further adaptations to active-set strategies in constrained optimization. The dialog between unconstrained methods and active-set strategies will continue, as far as interesting unconstrained research continues to be done.

With respect to small and medium-scale problems, it is probably interesting to test the algorithm introduced in [13] as internal-face solver. Its comparison with [35] in the unconstrained case is, of course, of independent interest. However, the large-scale situation is the most challenging one to be addressed in the near future. Large-scale bound constrained (or even unconstrained) problems lie in the nucleus of many efficient algorithms for complementarity problems and variational inequalities (see [1]–[5],[21, 26, 27, 28] and many others) which are largely used in modern Engineering.

Acknowledgement. We are indebted to an anonymous referee whose comments helped us to improve the paper.

References

[1] R. Andreani, A. Friedlander and J. M. Martínez. On the solution of finite-dimensional variational inequalities using smooth optimization with simple bounds. *Journal of Optimization Theory and Applications* 94, pp. 635–657 (1997).

[2] R. Andreani and J. M. Martínez. On the solution of the extended linear complementarity problem. *Linear Algebra and its Applications* 281, pp. 247–257 (1998).

[3] R. Andreani and J. M. Martínez. Solving complementarity problems by means of a new smooth constrained nonlinear solver. In *Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing methods* . Edited by M. Fukushima and L. Qi. Kluwer, pp.1–24 (1999).

[4] R. Andreani and J. M. Martínez. On the reformulation of nonlinear complementarity problems using the Fischer-Burmeister function. *Applied Mathematics Letters* 12, pp. 7–12 (1999).

[5] R. Andreani and J. M. Martínez. Reformulation of variational inequalities on a simplex and compactification of complementarity problems. *SIAM Journal on Optimization* 10, pp. 878–895 (2000).

[6] R. H. Bielschowsky, A. Friedlander, F. M. Gomes, J. M. Martínez and M. Raydan. An adaptive algorithm for bound constrained quadratic minimization. *Investigación Operativa* 7, pp. 67–102 (1998).

[7] E. G. Birgin, R. Biloti, M. Tygel and L.T. Santos. Restricted optimization: a clue to a fast and accurate implementation of the Common Reflection Surface stack method. *Journal of Applied Geophysics* 42, pp. 143–155 (1999).

[8] E. G. Birgin, I. Chambouleyron and J. M. Martínez. Estimation of the optical constants and the thickness of thin films using unconstrained optimization. *Journal of Computational Physics* 151, pp. 862–880 (1999).

[9] E. G. Birgin, J. M. Martínez and M. Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM Journal on Optimization* 10, pp. 1196–1211 (2000).

[10] I. Bongartz, A. R. Conn, N. I. M. Gould and Ph. L. Toint. CUTE: constrained and unconstrained testing environment. *ACM Transactions on Mathematical Software* 21, pp. 123–160 (1995).

[11] A. R. Conn, N. I. M. Gould and Ph. L. Toint. Global convergence of a class of trust region algorithms for optimization with simple bounds. *SIAM Journal on Numerical Analysis* 25, pp. 433–460 (1988).

[12] A. R. Conn, N. I. M. Gould and Ph. L. Toint. A globally convergent augmented Lagrangean algorithm for optimization with general constraints and simple bounds, *SIAM Journal on Numerical Analysis* 28, pp. 545–572 (1991).

[13] J. E. Dennis, N. Echebest, M. Guardarucci, J. M. Martínez, H. Scolnik and C. Vacchino. A curvilinear search using tridiagonal secant updates for unconstrained optimization. *SIAM Journal on Optimization* 1, pp. 352–372 (1991).

[14] M. A. Diniz-Ehrhardt, Z. Dostál, M. A. Gomes-Ruggiero, J. M. Martínez and S. A. Santos. Nonmonotone strategy for minimization of quadratics with simple constraints. To appear in *Applications of Mathematics*.

[15] M. A. Diniz-Ehrhardt, M. A. Gomes-Ruggiero and S. A. Santos. Comparing the numerical performance of two trust-region algorithms for large-scale bound-constrained minimization, in *International Workshop of Numerical Linear Algebra and Optimization*, edited by R. J. B. Sampaio and J. Y. Yuan, Department of Mathematics, Universidade Federal do Paraná, Brazil, pp. 23–24 (1997).

[16] Z. Dostál. Box constrained quadratic programming with proportioning and projections. *SIAM Journal on Optimization* 7, pp. 871–887 (1997).

[17] Z. Dostál, A. Friedlander and S. A. Santos. Solution of coercive and semicoercive contact problems by FETI domain decomposition. *Contemporary Mathematics* 218, pp. 82–93 (1998).

[18] Z. Dostál, A. Friedlander and S. A. Santos. Augmented Lagrangians with adaptive precision control for quadratic programming with equality constraints. *Computational Optimization and Applications* 14, pp.1–17 (1999).

[19] Z. Dostál, A. Friedlander and S. A. Santos. Adaptive Precision Control in Quadratic Programming with Simple Bounds and/or Equalities. To appear in *High Performance Algorithms and Software in Nonlinear Optimization: Applied Optimization*, edited by R. De Leone, A. Murli, P. Pardalos, G. Toraldo, editors, Kluwer.

[20] Z. Dostál, F. A. M. Gomes and S. A. Santos. Solution of contact problems by FETI domain decomposition with natural coarse space projections. To appear in *Computational Methods in Applied Mechanics and Engineering.*

[21] L. Fernandes, A. Friedlander, M. Guedes and J. Júdice. Solution of General Linear Complementarity Problem using smooth optimization and its application to bilinear programming and LCP. To appear in *Applied Mathematics and Optimization.*

[22] A. Friedlander and J. M. Martínez. On the maximization of a concave quadratic function with box constraints. *SIAM Journal on Optimization* 4, pp. 177–192 (1994).

[23] A. Friedlander, J. M. Martínez, B. Molina and M. Raydan. Gradient methods with retards and generalizations. *SIAM Journal on Numerical Analysis* 36, pp. 275–289 (1999).

[24] A. Friedlander, J. M. Martínez and M. Raydan. A new method for large-scale box constrained convex quadratic minimization problems. *Optimization Methods and Software* 5, pp. 57–74 (1995).

[25] A. Friedlander, J. M. Martínez and S. A. Santos. A new trust region algorithm for bound constrained minimization. *Applied Mathematics and Optimization* 30, pp. 235–266 (1994).

[26] A. Friedlander, J. M. Martínez and S. A. Santos. On the resolution of large scale linearly constrained convex minimization problems. *SIAM Journal on Optimization* 4, pp. 331–339 (1994).

[27] A. Friedlander, J. M. Martínez and S. A. Santos. Solution of linear complementarity problems using minimization with simple bounds. *Journal of Global Optimization* 6, pp. 1–15 (1995).

[28] A. Friedlander, J. M. Martínez and S. A. Santos. A new strategy for solving variational inequalities on bounded polytopes. *Numerical Functional Analysis and Optimization* 16, pp. 653–668 (1995).

[29] F. M. Gomes, M. C. Maciel and J. M. Martínez. Nonlinear programming algorithms using trust regions and augmented Lagrangians with nonmonotone penalty parameters. *Mathematical Programming* 84, pp. 161–200 (1999).

[30] N. Krejić, J. M. Martínez, M. P. Mello and E. A. Pilotta. Validation of an augmented Lagrangian algorithm with a Gauss-Newton Hessian approximation using a set of hard-spheres problems. *Computational Optimization and Applications* 16, pp. 247–263 (2000).

[31] J. M. Martínez. BOX-QUACAN and the implementation of Augmented Lagrangian algorithms for minimization with inequality constraints. To appear in *Computational and Applied Mathematics*.

[32] M. Mulato, I. Chambouleyron, E.G. Birgin and J.M. Martínez. Determination of thickness and optical constants of a-Si:H films from transmittance data. *Applied Physics Letters* 77, pp. 2133–2135 (2000).

[33] M. Raydan. On the Barzilai and Borwein choice of steplength for the gradient method, *IMA Journal of Numerical Analysis* 13, pp. 321–326 (1993).

[34] M. Raydan. The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. *SIAM Journal on Optimization* 7, pp. 26–33 (1997).

[35] J. Zhang and C. Xu. A class of indefinite dogleg path methods for unconstrained minimization. *SIAM Journal on Optimization* 9, pp. 646–667 (1999).