Augmented Lagrangians with adaptive precision control for quadratic programming with equality constraints *

Z. Dostál † A. Friedlander ‡ S. A. Santos §

January 9, 2001

Abstract

In this paper we introduce an augmented Lagrangian type algorithm for strictly convex quadratic programming problems with equality constraints. The new feature of the proposed algorithm is the adaptive precision control of the solution of auxiliary problems in the inner loop of the basic algorithm. Global convergence and boundedness of the penalty parameter are proved and an error estimate is given that does not have any term that accounts for the inexact solution of the auxiliary problems. Numerical experiments illustrate efficiency of the algorithm presented.

AMS classification: Primary 90C20; Secondary 65K05

Key words and phrases: Quadratic programming, equality constraints, augmented Lagrangians, adaptive precision control

* This research has been supported by CNPq, FAPESP 95/6574-9 and by grants GAČR 201/97/0421, 105/95/1273.

† VŠB-Technical University Ostrava, Tř 17 listopadu, CZ-70833 Ostrava, Czech Republic. E-mail: Zdenek.Dostal@vsb.cz

‡ Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil. E-mail: friedlan@ime.unicamp.br.

§ Department of Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil. E-mail: sandra@ime.unicamp.br.

1 Introduction

We shall be concerned with the problem of finding a minimizer of a quadratic function subject to linear equality constraints, that is

$$\begin{array}{ll} \text{minimize} & h(x) \\ \text{subject to} & x \in \Omega \end{array} \tag{1.1}$$

with $\Omega = \{x \in \mathbb{R}^p : Dx = d\}$, $h(x) = \frac{1}{2}x^T Bx - c^T x$, $c, x \in \mathbb{R}^p$, $d \in \mathbb{R}^m$, $B \in \mathbb{R}^{p\times p}$ symmetric positive definite, and $D \in \mathbb{R}^{m\times p}$ a full rank matrix. We shall be especially interested in problems with m much smaller than p and with the matrix B large and reasonably conditioned (or preconditioned), so that conjugate gradient based methods are directly applicable to the solution of unconstrained problems.

Applications that lead to problem (1.1) include numerical solution of elliptic partial differential equations with periodic boundary conditions (e.g. Dostál [4]) and implementation of domain decomposition methods to parallel solution of three-dimensional elasticity problems (e.g. Le Tallec and Sassi [14]). It may be advantageous to reduce the solution of some problems to a sequence of problems of type (1.1). As an example, let us mention an algorithm for numerical solution of contact problems of elasticity that was proposed by Simo and Laursen [20]. The results involving the solution of (1.1) are also useful for problems with simple bounds and inequalities [6] that may be used for efficient solution of semicoercive contact problems [5].

An efficient algorithm for the solution of (1.1) is the augmented Lagrangian method proposed independently by Powell [16] and Hestenes [13] for problems with general cost function subject to general equality constraints. Their algorithm generates approximations of the Lagrange multipliers in an outer loop while unconstrained auxiliary problems with well structured symmetric positive definite matrices are solved in an inner loop. Specifically, the auxiliary problems are of the type

$$\begin{array}{ll} \text{minimize} & L(x, \mu^k, \rho_k) \\ \text{subject to} & x \in \mathbb{R}^p \end{array} \tag{1.2}$$

where

$$L(x, \mu^k, \rho_k) = h(x) + (\mu^k)^T(Dx - d) + \frac{\rho_k}{2}\|Dx - d\|^2 \tag{1.3}$$

is known as the augmented Lagrangian function, $\mu^k = (\mu_1^k, \ldots, \mu_m^k)^T$ is the vector of Lagrange multipliers for the equality constraints, ρ_k is the penalty parameter, and $\|\cdot\|$ denotes the Euclidean norm. The precision of the approximate solution x^k of the auxiliary problem will be measured by the Euclidean norms of the error of feasibility and of the gradient of the augmented Lagrangian. The latter is always denoted by g, so that

$$g(x,\mu,\rho) \quad = \quad \nabla_x L(x,\mu,\rho) \quad = \quad \nabla h(x) + D^T\mu + \rho D^T(Dx-d). \tag{1.4}$$

Powell and Hestenes proved that their method converges without the hypothesis of the unboudedness of the penalty parameter ρ_k. Hence the augmented Lagrangian method compares favorably both with the Lagrange multiplier method that works with indefinite matrices and with the penalty method that may require very large values of the penalty parameter. Rockafellar obtained additional results for this type of method in [17, 18, 19]. In [2] the multiplier method is thoroughly analysed. Solution of problems with inequality constraints using this approach are considered in [3]. Results for problems formulated in Hilbert spaces appear in [9].

Let us mention that the structure of the Hessian matrix of the Lagrangian function L is usually simpler than the one of the matrix arising from elimination of dependent variables and that the convergence of the augmented Lagrangian methods is usually faster than that of variants of the Uzawa method [7].

Even though Hestenes and Powell assumed in their theory that the auxiliary problems are solved exactly, it has been proved later that the convergence of the algorithm may be preserved even when the auxiliary problems are solved only approximately with apriori prescribed precisions provided that these precisions converge to zero [2, 3]. The price paid for the inexact minimization is an additional term in the estimate of the rate of convergence. Hager in [10, 11] obtains global convergence results for an algorithm of this type using inexact minimization in the solution of the auxiliary problems. In both papers the size of the optimality error is compared with the size of the feasibility error of the solution of the auxiliary problems trying to balance this quantities throughout the whole process. In [10] this comparison is used to decide whether the penalty parameter will be increased or not. In [11] it is used as a stopping criterion for the minimization of the auxiliary problems. The rate of convergence is free of any term due to inexact minimization when the least squares estimate of the Lagrange multipliers is used. The main improvement on the algorithm of Powell and Hestenes that we propose here concerns the precision control of the solution of the auxiliary problems. Our approach arises from the simple observation that the precision of the solution x^k of the auxiliary problems should be related to the feasibility of x^k, i.e. $\|Dx^k - d\|$, since it does not seem reasonable to solve these problems with high precision when μ^k is far from the Lagrange multiplier of the solution of (1.1). In this aspect, our approach is similar to Hager's in [10, 11]. In our algorithm we decide to increase the penalty parameter as in [2, 3] but stop the minimization problems in the inner loop as in [11]. Our interest is to solve very large scale problems arising from the applications mentioned above, therefore a matrix-free approach is proposed and also just the first order estimate for updating the Lagrange multipliers is used.

In Section 2 we present the algorithm, prove that it is well defined and that the quite natural precision control of the solution of the auxiliary problems guarantees an improvement of the current estimate of the Lagrange multipliers in the classical method of multipliers. Unlike the classical results concerning inexact solution of the auxiliary problems (see Chapter 2 of [2]), Theorem 2.3 yields relevant information at each stage of the process. In Section 3 we prove the global convergence of the algorithm to the solution of problem (1.1). The main results in this section are the boundedness of the sequence of Lagrange multipliers, the convergence of the full sequences x^k and μ^k and the boundedness of the penalty parameters ρ_k. The choice of the stopping criterion results in an estimate of the rate of convergence of μ^k that does not have any term accounting for the inexact minimization. Computational implementation and numerical experiments are presented in Section 4. Finally, some conclusions are discussed in Section 5.

The following notation will be used throughout the whole paper:

- $\widehat{x}$ and $\widehat{\mu}$ are the Kuhn-Tucker pair of (1.1).
- β_1 and β_m are, respectively, the smallest and largest eigenvalues of $DB^{-1}D^T$.
- $\widetilde{\mu} = \mu + \rho(Dx - d)$.
- $r = g(x, \mu, \rho)$.

2 Algorithm for Equality Constraints with Adaptive Precision Control

The following algorithm is a modification of the classical augmented Lagrangian method for the solution of strictly convex quadratic programming problems with equality constraints that enables adaptive precision control of the solution of auxiliary problems.

Algorithm 2.1. Given $\eta_0 > 0, 0 < \alpha < 1, \beta > 1, M > 0, \rho_0 > 0, \nu > 0$ and $\mu^0 \in I\!R^m$, set $k = 0$.

Step 1. {Inner iteration with adaptive precision control.}
Find x^k such that

$$\|g(x^k, \mu^k, \rho_k)\| \leq M\|Dx^k - d\|. \tag{2.1}$$

Step 2. {Update μ.}

$$\mu^{k+1} = \mu^k + \rho_k(Dx^k - d). \tag{2.2}$$

Step 3. {Update ρ, η.}
If $\|Dx^k - d\| \leq \eta_k$ then

$$\rho_{k+1} = \rho_k, \quad \eta_{k+1} = \alpha\eta_k \tag{2.3}$$

else

$$\rho_{k+1} = \beta\rho_k, \quad \eta_{k+1} = \eta_k. \tag{2.4}$$

Step 4. Set $k = k + 1$ and return to the Step 1.

In Step 1 we can use any convergent algorithm for minimizing the strictly convex quadratic function such as a preconditioned conjugate gradient method [1]. A similar stopping criterion for the minimization process was also proposed by Hager in [10, 11]. Optimality and feasibility are both targets of the whole process, but solving a problem in Step 1 with high precision when the estimated multiplier μ is still very far from the correct one $\widehat{\mu}$, seems to be undesirable. That is the main motivation to use the adaptive precision control proposed here. In [11] an interesting discussion about this aspect is presented. The algorithm proposed in that paper executes two fundamental steps, one is called a constraint step that takes care of the feasibility and the other is called a Kuhn-Tucker step that handles the optimality. In this context a criterion similar to the stopping criterion in Step 1 of our algorithm is used to achieve a balanced reduction in the total error that results in a more efficient algorithm. Algorithm 2.1 is similar in structure to Algorithm 5.1 in [10]. The main difference is that the adaptive precision control is used there to decide whether to increase or not the penalty parameter and not as a stopping criterion for the minimization process. Another important difference is the rule for updating the multiplier μ in Step 2.

The next lemma shows that Algorithm 2.1 is well defined, that is, any convergent algorithm for the solution of the auxiliary problem required in Step 1 will generate either x^k that satisfies (2.1) in a finite number of steps or a sequence of approximations that converges to the solution of (1.1). It is also clear that there is no hidden enforcement of exact solution in (2.1) and consequently typically inexact solutions of the auxiliary unconstrained problems are obtained in Step 1.

Lemma 2.2. *Let $M > 0, \mu \in I\!R^m$ and $\rho \geq 0$ be given and let $\{x^k\}$ denote any sequence that converges to the unique solution $\overline{x}$ of the problem*

$$\text{minimize} \quad L(x, \mu, \rho). \tag{2.5}$$

Then $\{x^k\}$ either converges to the solution $\widehat{x}$ of problem (1.1) or there is an index k such that

$$\|g(x^k, \mu, \rho)\| \leq M\|Dx^k - d\|. \tag{2.6}$$

Proof: First observe that if (2.6) does not hold for any k, then we must have $D\overline{x} = d$. In this case, since $\overline{x}$ is the solution of (2.5), it follows that

$$B\overline{x} - d + D^T\mu + \rho D^T(D\overline{x} - d) = 0 \tag{2.7}$$

and after substituting $D\overline{x} = d$ into (2.7), we get

$$B\overline{x} - d + D^T\mu = 0. \tag{2.8}$$

However, conditions (2.8) and $D\overline{x} = d$ are sufficient conditions for $\overline{x}$ to be the unique solution of (1.1) so that $\overline{x} = \widehat{x}$. $\square$

The next theorem states the basic result that relates the adaptive precision control used in Step 1 of Algorithm 2.1 to the improvement on the multiplier estimation when it is updated as in Step 2.

Theorem 2.3. *Let $\rho > 0$ and $\tau \in [0, \rho)$. If*

$$\|r\| \leq \tau \frac{(\rho + \beta_m^{-1})(\beta_1 + \rho^{-1})}{\|D\| \, \|B^{-1}\|} \|Dx - d\| \tag{2.9}$$

then

$$\|\widetilde{\mu} - \widehat{\mu}\| \leq \delta(\tau, \rho) \|\mu - \widehat{\mu}\| \tag{2.10}$$

where

$$\delta(\tau, \rho) = \tau\rho/(\rho - \tau) + 1/(\rho\beta_1 + 1). \tag{2.11}$$

The proof of this theorem requires some inequalities that, for the sake of clarity of exposition, we obtain in the following lemmas.

Lemma 2.4. *For any vectors $x \in R^p$ and $\mu \in R^m$,*

$$\|\widetilde{\mu} - \widehat{\mu}\| \quad \leq \quad \frac{\|D\| \, \|B^{-1}\|}{\beta_1 + \rho^{-1}} \|r\| + \rho^{-1} \frac{1}{\beta_1 + \rho^{-1}} \|\mu - \widehat{\mu}\|. \tag{2.12}$$

Proof: The definition of $\widetilde{\mu}$ and r implies that

$$\begin{array}{rcccl} Bx & + & D^T\widetilde{\mu} & = & r + c \\ Dx & - & \rho^{-1}\widetilde{\mu} & = & -\rho^{-1}(\mu - \widehat{\mu}) - \rho^{-1}\widehat{\mu} + d, \end{array} \tag{2.13}$$

also $\widehat{\mu}$ and $\widehat{x}$ are completely determined by

$$\begin{array}{rcccl} B\widehat{x} & + & D^T\widehat{\mu} & = & c \\ D\widehat{x} & - & \rho^{-1}\widehat{\mu} & = & -\rho^{-1}\widehat{\mu} + d. \end{array} \tag{2.14}$$

Subtracting (2.13) from (2.14) and switching to matrix notation, we get

$$\begin{pmatrix} B & D^T \\ D & -\rho^{-1}I \end{pmatrix} \begin{pmatrix} x-\widehat{x} \\ \widetilde{\mu}-\widehat{\mu} \end{pmatrix} = \begin{pmatrix} r \\ \rho^{-1}(\widehat{\mu}-\mu) \end{pmatrix}. \tag{2.15}$$

The inverse of the matrix in (2.15) is given by

$$\begin{pmatrix} (B+\rho D^T D)^{-1} & B^{-1}D^T S_\rho^{-1} \\ S_\rho^{-1}DB^{-1} & -S_\rho^{-1} \end{pmatrix} \tag{2.16}$$

where $S_\rho = DB^{-1}D^T + \rho^{-1}I$ (see [12]).

It follows that

$$\widetilde{\mu}-\mu \;=\; S_\rho^{-1}DB^{-1}r - \rho^{-1}S_\rho^{-1}(\widehat{\mu}-\mu). \tag{2.17}$$

Inequality (2.12) results after taking norms in (2.17) and noting that

$$\|S_\rho^{-1}\| = 1/(\beta_1+\rho^{-1}) \; . \qquad \square$$

The previous lemma gives a bound on the distance between the updated multiplier and the correct one, proportional to the error due to inexact minimization in Step 1 and to the error in the previous multiplier estimate. This bound is related to the results of Proposition 2.4 in [2].

Lemma 2.5. *Let $\rho > 0$, for any vectors $x \in I\!R^p$ and $\mu \in I\!R^m$,*

$$\|\mu-\widehat{\mu}\| \geq (\rho+\beta_m^{-1})\|Dx-d\| - \frac{\rho^{-1}\|D\|\;\|B^{-1}\|}{\beta_1+\rho^{-1}}\|r\|. \tag{2.18}$$

Proof: By the definition of r we get

$$(B+\rho D^T D)(x-\widehat{x}) + D^T(\mu-\widehat{\mu}) + (B+\rho D^T D)\widehat{x} + D^T\widehat{\mu} = c + \rho D^T d + r, \tag{2.19}$$

and using the equations that determine $(\widehat{x},\widehat{\mu})$

$$\begin{aligned} B\widehat{x} + D^T\widehat{\mu} &= c \\ D\widehat{x} &= d \end{aligned} \tag{2.20}$$

equation (2.19) reduces to

$$(B+\rho D^T D)(x-\widehat{x}) + D^T(\mu-\widehat{\mu}) = r. \tag{2.21}$$

Equation (2.21) together with

$$D(x-\widehat{x}) = Dx - d \tag{2.22}$$

may be written in matrix form as

$$\begin{pmatrix} B+\rho D^T D & D^T \\ D & 0 \end{pmatrix} \begin{pmatrix} x-\widehat{x} \\ \mu-\widehat{\mu} \end{pmatrix} = \begin{pmatrix} r \\ Dx-d \end{pmatrix}. \tag{2.23}$$

The inverse of the matrix in (2.23) is given by (see [12])

$$\begin{pmatrix} B^{-1}-B^{-1}D^T S^{-1} D B^{-1} & B^{-1}D^T S^{-1} \\ S^{-1}DB^{-1} & -S^{-1} \end{pmatrix} \tag{2.24}$$

where $S = DB^{-1}D^T$. Then,

$$\mu-\widehat{\mu} = (D(B+\rho D^T D)^{-1}D^T)^{-1}(D(B+\rho D^T D)^{-1} r - (Dx-d)), \tag{2.25}$$

so that

$$\|\mu-\widehat{\mu}\| \geq \lambda_1((D(B+\rho D^T D)^{-1}D^T)^{-1})(\|Dx-d\| - \|D(B+\rho D^T D)^{-1}\|\ \|r\|), \tag{2.26}$$

where $\lambda_1(A)$ denotes the smallest eigenvalue of matrix A.

Applying the Sherman-Morrison-Woodbury formula (see Golub and Van Loan [8])

$$(B+\rho D^T D)^{-1} = B^{-1} - B^{-1}D^T(\rho^{-1}I + DB^{-1}D^T)^{-1}DB^{-1}, \tag{2.27}$$

we get

$$I-\rho D(B+\rho D^T D)^{-1}D^T \quad = \quad \rho^{-1}(\rho^{-1}I + DB^{-1}D^T)^{-1}.$$

Thus

$$I-\rho D(B+\rho D^T D)^{-1}D^T = \rho^{-1}(\rho^{-1}I + DB^{-1}D^T)^{-1}. \tag{2.28}$$

It follows that

$$D(B+\rho D^T D)^{-1}D^T = \rho^{-1}(I-(I+\rho DB^{-1}D^T)^{-1}) \tag{2.29}$$

and

$$\|D(B+\rho D^T D)^{-1}D^T\| = \rho^{-1}\beta_m/(\beta_m+\rho^{-1}). \tag{2.30}$$

Spectral properties and (2.30) imply

$$\lambda_1((D(B+\rho D^TD)^{-1}D^T)^{-1}) = \|D(B+\rho D^TD)^{-1}D^T\|^{-1} = \beta_m^{-1} + \rho. \tag{2.31}$$

Finally, substituting (2.31) into (2.26) and noting that by (2.29)

$$D(B+\rho D^TD)^{-1} = \rho^{-1}(\rho^{-1}I + DB^{-1}D^T)^{-1}DB^{-1} \tag{2.32}$$

and

$$\|D(B+\rho D^TD)^{-1}\| \leq \rho^{-1}\|D\|\|B^{-1}\|/(\beta_1+\rho^{-1}), \tag{2.33}$$

(2.18) is easily obtained. □

Lemma 2.5 gives us a computable lower bound of the norm of the error in the approximation of the Lagrange multipliers. Now we are ready to prove Theorem 2.3.

Proof of Theorem 2.3: The assumptions imply that inequalities (2.12) and (2.18) hold.

If $\tau = 0$, (2.9) implies $r = 0$ and substituting in (2.12), (2.10) is obtained.

If $\tau > 0$, (2.9) implies that

$$\tau^{-1}\frac{\|D\|\ \|B^{-1}\|}{\beta_1+\rho^{-1}}\|r\| \leq (\rho+\beta_m^{-1})\|Dx-d\|. \tag{2.34}$$

After substituting (2.34) in (2.18), we get

$$(\tau^{-1}-\rho^{-1})\frac{\|D\|\ \|B^{-1}\|}{\beta_1+\rho^{-1}}\|r\| \leq \|\mu-\widehat{\mu}\| \tag{2.35}$$

so that, for $\tau \in (0,\rho)$,

$$\frac{\|D\|\ \|B^{-1}\|}{\beta_1+\rho^{-1}}\|r\| \leq \frac{\tau\rho}{\rho-\tau}\|\mu-\widehat{\mu}\|. \tag{2.36}$$

To finish the proof, it is enough to substitute (2.36) in (2.12) and check that the resulting inequality is equivalent to (2.10). □

The problem of finding the bounds of the norm of gradients that yield estimates of the updated Lagrange multipliers is now reduced to obtaining a bound of $\delta(\tau,\rho)$. For this purpose the following properties of the latter function will be useful.

Lemma 2.6. *Let $\delta(\tau,\rho)$ be defined by (2.11) for $0 \le \tau < \rho$, let $K > 0$, and $\tau(\rho) = K/\rho$.*
(i) If

$$\overline{\rho} > \sqrt{K} \tag{2.37}$$

then, for any $\rho \ge \overline{\rho}$,

$$\tau(\rho) < \rho \quad \text{and} \quad \delta(\tau(\rho),\rho) \le \frac{\overline{\rho}}{\rho}\frac{K+\beta_1^{-1}}{\overline{\rho}-\sqrt{K}}. \tag{2.38}$$

(ii) If

$$\overline{\rho} > K + \sqrt{K} + \beta_1^{-1}, \tag{2.39}$$

then there exists $\alpha \in (0,1)$ such that, for $\rho \ge \overline{\rho}$,

$$\delta(\tau(\rho),\rho) \le \alpha\frac{\overline{\rho}}{\rho} \le \alpha. \tag{2.40}$$

Proof: Inequalities (2.37) and $\rho \ge \overline{\rho}$ imply

$$\tau(\rho) = K/\rho \le K/\overline{\rho} < \sqrt{K}\rho/\overline{\rho}$$

and

$$\rho - \tau(\rho) > \rho - \frac{\sqrt{K}\rho}{\overline{\rho}} = \rho\frac{\overline{\rho}-\sqrt{K}}{\overline{\rho}}$$

so that

$$\delta(\tau(\rho),\rho) < \frac{\overline{\rho}K}{\rho(\overline{\rho}-\sqrt{K})} + \frac{1}{\rho\beta_1} < \frac{\overline{\rho}}{\rho}\frac{K+\beta_1^{-1}}{\overline{\rho}-\sqrt{K}}$$

To prove (2.40), observe that $\delta(\tau(\rho),\rho)$ is a decreasing function of ρ for $\rho > \sqrt{K}$. If $\overline{\rho}$ satisfies (2.39), then by (i) also

$$\delta(\tau(\rho),\rho) \le \frac{\overline{\rho}}{\rho}\frac{K+\beta_1^{-1}}{\overline{\rho}-\sqrt{K}} = \alpha\frac{\overline{\rho}}{\rho}$$

with

$$\alpha = \frac{K+\beta_1^{-1}}{\overline{\rho}-\sqrt{K}} < 1. \qquad \square$$

Corollary 2.7. *Under the assumptions of Theorem 2.3, the following statements hold* *(i) If M_1 is a positive constant and*

$$\overline{\rho} > \sqrt{\|D\|\|B^{-1}\|M_1/\beta_1}, \tag{2.41}$$

then there is a positive constant M_2 such that for any $x \in I\!R^p, \mu \in I\!R^m$ and $\rho \geq \overline{\rho}$

$$\|g(x,\mu,\rho)\| \leq M_1\|Dx - d\| \tag{2.42}$$

implies

$$\|\widetilde{\mu} - \widehat{\mu}\| \leq \frac{1}{\rho}M_2\|\mu - \widehat{\mu}\|. \tag{2.43}$$

(ii) If M_1 is a positive constant and

$$\overline{\rho} > M_1 + \sqrt{M_1} + \beta_1^{-1}, \tag{2.44}$$

then for any $x \in I\!R^p, \mu \in I\!R^m$ and $\rho \geq \overline{\rho}$ the inequality (2.42) implies

$$\|\widetilde{\mu} - \widehat{\mu}\| \leq \frac{\overline{\rho}}{\rho}\|\mu - \widehat{\mu}\|. \tag{2.45}$$

Proof: (i) Part (i) of Lemma 2.6 and

$$K = \|D\|\|B^{-1}\|M_1/\beta_1 \tag{2.46}$$

imply that for $\rho \geq \overline{\rho} > \sqrt{M}$

$$\delta(\tau(\rho),\rho) \leq M_2/\rho \quad \text{and} \quad \tau(\rho) < \rho$$

where

$$M_2 = \overline{\rho}\frac{K + \beta_1^{-1}}{\overline{\rho} - \sqrt{K}}.$$

Thus, by Theorem 2.3,

$$\|g(x,\mu,\rho)\| \leq \frac{\tau(\rho)\beta_1\rho}{\|D\|\|B^{-1}\|}\|Dx - d\| = M_1\|Dx - d\|$$

implies

$$\|\widetilde{\mu} - \widehat{\mu}\| \leq \delta(\tau(\rho),\rho)\|\mu - \widehat{\mu}\| \leq \frac{M_2}{\rho}\|\mu - \widehat{\mu}\|.$$

(ii) If $DD^T = I$, then

$$\lambda_1(DB^{-1}D^T) = \min\left\{\frac{z^T DB^{-1}D^T z}{z^T DD^T z} : z \neq 0\right\} \geq \lambda_1(DB^{-1}D^T) = \|B\|^{-1}$$

so that

$$\beta_1^{-1} \leq \|B\|. \tag{2.47}$$

Thus, assumption (2.44) implies (2.39) and by Lemma 2.6(ii)

$$\delta(\tau(\rho), \rho) \leq \frac{\overline{\rho}}{\rho}.$$

Moreover, using DD^T=I and (2.46) it is easy to verify that

$$\tau\frac{(\rho + \beta_m^{-1})(\beta_1 + \rho^{-1})}{\|D\|\ \|B^{-1}\|} \geq K,$$

so that (2.45) follows from Theorem 2.3. □

3 Global Convergence

In this section we prove the global convergence of Algorithm 2.1. In Theorem 3.2. we prove the convergence of the whole sequences x^k and μ^k to the Kuhn-Tucker pair $(\widehat{x},\ \widehat{\mu})$, respectively. Theorem 3.3. states a local convergence result for the sequence μ^k, that does not have any term accounting for the inexact minimization in Step 1 of the algorithm. The final result of this section is Theorem 3.4, where the boundedness of the penalty parameter ρ_k is obtained.

Lemma 3.1. *Let $\{\mu^k\}$ be a sequence generated by Algorithm 2.1. Then $\{\mu^k\}$ is bounded.*

Proof: Let $\{\mu^k\}, \{x^k\}$ and $\{\rho_k\}$ be generated by Algorithm 2.1. In particular, it follows that $\{\rho_k\}$ is non-decreasing.

Let us first assume that $\{\rho_k\}$ is not bounded and observe that μ^{k+1} is assigned by (2.2) in Step 2 of Algorithm 2.1 so that

$$\mu^{k+1} = \mu^k + \rho_k(Dx^k - d). \tag{3.1}$$

Let $\delta \in (0,1)$, applying Corollary 2.7(ii), there is a positive ρ such that for $\rho_k \geq \rho$, if

$$\|g(x^k, \mu^k, \rho_k)\| \leq M\|Dx^k - d\| \tag{3.2}$$

then

$$\|\mu^{k+1} - \widehat{\mu}\| \leq \delta\|\mu^k - \widehat{\mu}\|. \tag{3.3}$$

Therefore, as (2.1) holds, (3.2) is true and we obtain (3.3).

Now, if $\{\rho_k\}$ is bounded, there is k_0 such that for $k \geq k_0$ the values of ρ_k and η_k are updated by (2.3) in Step 3. It follows that for any $\ell \geq 0$,

$$\|Dx^{k_0+\ell} - d\| \leq \eta_{k_0+\ell} = \alpha^\ell \eta_{k_0}$$

and

$$\mu^{k_0+\ell} - \mu^{k_0} = \rho_{k_0} \sum_{i=0}^{\ell-1}(Dx^{k_0+i} - d),$$

so that

$$\begin{aligned} \|\mu^{k_0+\ell}\| &\leq \|\mu^{k_0}\| + \rho_{k_0} \sum_{i=0}^{\ell-1} \|Dx^{k_0+i} - d\| \\ &\leq \|\mu^{k_0}\| + \rho_{k_0}(1 + \ldots + \alpha^{\ell-1})\eta_{k_0} \\ &\leq \|\mu^{k_0}\| + \frac{\rho_{k_0}\eta_{k_0}}{1-\alpha}. \end{aligned}$$

Hence $\{\mu^k\}$ is also bounded in this case. □

Theorem 3.2. *The sequences $\{x^k\}$ and $\{\mu^k\}$ generated by Algorithm 2.1. converge to $\widehat{x}$ and $\widehat{\mu}$, respectively.*

Proof: Since all μ^k are generated by (2.2) in Step 2, we have

$$\|Dx^k - d\| = \rho_k^{-1}\|\mu^{k+1} - \mu^k\| \leq \rho_k^{-1}(\|\mu^{k+1}\| + \|\mu^k\|). \tag{3.4}$$

If $\{\rho_k\}$ is not bounded, then, as it is monotonous, it follows by Lemma 3.1 that $\|Dx^k - d\|$ converges to zero.

On the other hand, if $\{\rho_k\}$ is bounded, it follows that there is k_0 such that for $k \geq k_0$, ρ_k and η_k are generated by (2.3) in Step 3 and

$$\|Dx^k - d\| \leq \eta_k = \alpha^{k-k_0}\eta_{k_0}. \tag{3.5}$$

Hence we can conclude that $\|Dx^k - d\|$ converges to zero. However, since at each iteration x^k satisfies (2.1), it follows that $\|g(x^k, \mu^k, \rho_k)\|$ converges to zero, too.

Let $r^k = g(x^k, \mu^k, \rho_k)$, then

$$x^k = (B + \rho_k D^T D)^{-1} c + (B + \rho_k D^T D)^{-1}(r^k - D^T \mu^k) + \rho_k (B + \rho_k D^T D)^{-1} D^T d. \quad (3.6)$$

Taking into account that

$$\|(B + \rho_k D^T D)^{-1}\| \leq \|B^{-1}\|,$$

we get

$$\|x^k\| \leq \|B^{-1}\|(\|c\| + \|r^k\| + \|D^T\| \; \|\mu^k\|) + \rho_k \|(B + \rho_k D^T D)^{-1} D^T\| \; \|d\|. \quad (3.7)$$

By Lemma 3.1. and (3.7) we obtain that $\{x^k\}$ is bounded. Since both sequences $\{x^k\}$ and $\{\mu^k\}$ are bounded, they have limit points $\overline{x}$ and $\overline{\mu}$, respectively. As $\|Dx^k - d\|$ converges to zero, $\overline{x}$ is feasible, i.e.

$$D\overline{x} = d. \quad (3.8)$$

As μ^{k+1} is generated by (2.2) in Step 2 and

$$g(x^k, \mu^k, \rho_k) = g(x^k, \mu^{k+1}, 0), \quad (3.9)$$

from the fact that $\|g(x^k, \mu^k, \rho_k)\|$ converges to zero, it follows that

$$g(\overline{x}, \overline{\mu}, 0) = 0. \quad (3.10)$$

Equations (3.8) and (3.10) are the sufficient conditions for $\overline{x}$ to be the unique solution of problem (1.1), with corresponding vector of Lagrange multipliers $\overline{\mu}$. Therefore, $\overline{x} = \widehat{x}$, $\overline{\mu} = \widehat{\mu}$ and both sequences $\{x^k\}$, $\{\mu^k\}$ are convergent. □

Theorem 3.3. *Let $\{x^k\}$ and $\{\mu^k\}$ be generated by Algorithm 2.1 and M be the constant used in Step 1. Then the following statements hold:*
(i) If

$$\overline{\rho} > \sqrt{\|D\| \|B^{-1}\| M / \beta_1}, \quad (3.11)$$

then there is a positive constant M_2 such that $\rho_k \geq \overline{\rho}$ implies

$$\|\mu^{k+1} - \widehat{\mu}\| \leq \frac{1}{\rho_k} M_2 \|\mu^k - \widehat{\mu}\|. \quad (3.12)$$

(ii) If

$$DD^T = I, \quad \overline{\rho} > M + \sqrt{M} + \|B\|, \tag{3.13}$$

then $\rho_k \geq \overline{\rho}$ *implies*

$$\|\mu^{k+1} - \widehat{\mu}\| \leq \frac{\overline{\rho}}{\rho_k}\|\mu^k - \widehat{\mu}\|. \tag{3.14}$$

Proof: If $DD^T = I$, then

$$\beta_1 = \min\left\{\frac{z^T DB^{-1}D^T z}{z^T DD^T z} : z \neq 0\right\} = \lambda_1(B^{-1}) = \|B\|^{-1}. \tag{3.15}$$

The rest is an easy consequence of Corollary 2.7. □

Theorem 3.4 Let the sequences $\{x^k\}, \{\mu^k\}$ and $\{\rho_k\}$ be generated by Algorithm 2.1. Then $\{\rho_k\}$ is bounded.

Proof: Let us assume that $\{\rho_k\}$ is not bounded. Then by (2.3) in Step 3, there exists a subsequence $I\!\!K$ such that $\|Dx^k - d\| > \eta_k$ for $k \in I\!\!K$. It follows that $I\!\!N \setminus I\!\!K$ is also infinite by (2.4) and the fact that $\|Dx^k - d\|$ converges to zero. Thus, there exists a subsequence $I\!\!K_1$ of $I\!\!K$ such that for $k \in I\!\!K_1$, $k+1 \in I\!\!K$ and $k \notin I\!\!K$. For $k \in I\!\!K_1$, $\rho_{k+1} = \rho_k$. Let

$$r^{k+1} = g(x^{k+1}, \mu^{k+1}, \rho_{k+1}), \qquad r^k = g(x^k, \mu^k, \rho_k),$$

then, for $k \in I\!\!K_1$

$$x^{k+1} = (B + \rho_k D^T D)^{-1}(c - D^T\mu^k + \rho_k D^T d - \rho_k D^T(Dx^k - d) + r^{k+1}) \tag{3.16}$$

$$x^k = (B + \rho_k D^T D)^{-1}(c - D^T\mu^k + \rho_k D^T d + r^k). \tag{3.17}$$

Thus

$$\begin{aligned} Dx^{k+1} - d &= D(B + \rho_k D^T D)^{-1}((c - D^T\mu^k + \rho_k D^T d + r^k) \\ &\quad - \rho_k D^T(Dx^k - d) + r^{k+1} - r^k) - d \\ &= (I - \rho_k D(B + \rho_k D^T D)^{-1}D^T)(Dx^k - d) \\ &\quad + D(B + \rho_k D^T D)^{-1}(r^{k+1} - r^k). \end{aligned} \tag{3.18}$$

Using (2.28), (2.33) and properties of norms, we obtain

$$\|Dx^{k+1} - d\| \leq \rho_k^{-1}\left(\frac{\|D\|\ \|B^{-1}\|}{\beta_1 + \rho_k^{-1}}\|r^{k+1} - r^k\| + \frac{1}{\beta_1 + \rho_k^{-1}}\|Dx^k - d\|\right). \tag{3.19}$$

Let $M_1 = \dfrac{\|D\|\|B^{-1}\|}{\beta_1 + \rho_k^{-1}}$ and $M_2 = \dfrac{1}{\beta_1 + \rho_k^{-1}}$. Then

$$\|Dx^{k+1} - d\| \leq \rho_k^{-1}(M_1(\|r^k\| + \|r^{k+1}\|) + M_2\|Dx^k - d\|). \tag{3.20}$$

From (2.1), we get

$$\|Dx^{k+1} - d\| \leq \rho_k^{-1}(MM_1\|Dx^{k+1} - d\| + (MM_1 + M_2)\|Dx^k - d\|), \tag{3.21}$$

and for sufficiently large $k \in I\!\!K_1$ we can find M_0 such that

$$\|Dx^{k+1} - d\| \leq \rho_k^{-1} M_0 \|Dx^k - d\|.$$

On the other hand, for $k \in I\!\!K_1$

$$\alpha\eta_k = \eta_{k+1} \leq \|Dx^{k+1} - d\| \leq \rho_k^{-1} M_0 \|Dx^k - d\| \leq \rho_k^{-1} M_0 \eta_k.$$

It follows that for arbitrarily large k, $\alpha \leq \rho_k^{-1} M_0$. This contradicts the assumption that $\{\rho_k\}$ is unbounded. $\square$

4 Numerical experiments

We are especially interested in the use of the proposed algorithm for the solution of problems arising in mechanical engineering that were mentioned in the introduction. The algorithm has been tested on the solution of a model problem resulting from the finite difference discretization of the following continuous problem:

$$\text{minimize} \ \sum_{i=1}^{2} \left(\int_{\Omega_i} |\nabla u_i|^2 d\Omega - \int P u_i d\Omega \right)$$

$$\text{subject to} \ \ u_1(0,y) = u_2(y) = 0 \quad \text{and } u_1(y) = u_2(y) \quad \text{for } x \in [0,1],$$

where

$$\begin{array}{l} \Omega_1 = (0,1) \times (0,1), \qquad \Omega_2 = (1,2) \times (0,1), \\ P(x,y) = -1 \quad \text{for} \quad (x,y) \in (0,1) \times (0.5,1), \\ P(x,y) = 0 \quad \text{for} \ \ (x,y) \in (0,1) \times (0,0.5), \\ P(x,y) = 2 \quad \text{for} \ \ (x,y) \in (1,2) \times (0.0.5), \\ P(x,y) = 0 \quad \text{for} \ \ (x,y) \in (1,2) \times (0.5,1). \end{array}$$

The discretization scheme consists of a regular grid of 21×21 nodes for each subdomain Ω_i. The initial approximation used was the null vector. The problem was solved with all possible combinations of $M \in \{0.1, 10, 1000\}$ and $\rho_0 \in$ { 1E3, 1E5, 1E7 }. The stopping criteria used were the relative precision $\|\nabla L\|/\|P\|$ <1E−5 and the feasibility tolerance $\|Bx\| \leq$ 1E−8. For the other parameters, the choices $\alpha = .1$, $\beta = 10$ and $\eta = 1$ were made. In just one case, namely for ρ_0 =1E3 and M =1E3, the penalty parameter was updated. The number of outer iterations and the total number of conjugate gradient iterations used in the solution of the auxiliary problems in Step 1 are in Table 1.

Table 1. No preconditioning

ρ_0	Outer iterations			CG iterations		
	$M = 0.1$	$M = 10$	$M = 1000$	$M = 0.1$	$M = 10$	$M = 1000$
1E3	3	4	8	300	400	342
1E5	3	3	4	221	214	303
1E7	2	2	4	251	223	220

The computational results suggest that the use of large penalty parameters in the algorithm presented is efficient for this type of problems. These results are in agreement with the theory which predicts that the number of outer iterations is small for large penalty parameters (Theorem 3.3). Also in [1, 15] results relating the spectrum of the penalized matrix and the performance of the conjugate gradient method for problems with the number of constraints m much smaller than the dimension p explain the fact that the number of inner iterations in Step 1 is not sensitive to the growth of the values of the penalty parameters, as observed here. It is not hard to see that the spectrum of the matrix $B + \rho D^T D$ is distributed in two disjoints intervals, the first $p - m$ eigenvalues are smaller than the largest eigenvalue of B and that the last m eigenvalues are at least ρ times the respective largest m eigenvalues of $D^T D$. This suggests a way of preconditioning the augmented matrices in case the conjugate gradient method is used in the inner loop, namely to precondition only the matrix B and then to orthonormalize the constraints. The results in Table 2 show that the number of the conjugate gradient iterations can be considerably reduced in this way.

Table 2. SSOR preconditioning of B with orthogonalization of constraints

ρ_0	Outer iterations			CG iterations		
	$M = 0.1$	$M = 10$	$M = 1000$	$M = 0.1$	$M = 10$	$M = 1000$
1E3	3	4	7	130	106	116
1E5	2	3	4	82	82	107
1E7	2	2	13	122	93	110

For comparison, we carried out the computations also with straightforward preconditioning of the augmented matrix, but the number of conjugate gradient iterations nearly doubled as compared with Table 1. Our explanation is that the preconditioned matrix has no gap in the spectrum so that the distribution of the spectrum is not favorable for the conjugate gradient method. An alternative strategy for preconditioning the augmented matrices has been presented in [10, 11].

5 Conclusions

We have proved global convergence results for an augmented Lagrangian method that uses adaptive precision control in the solution of the auxiliary problems for quadratic programming problems with equality constraints. The precision is controlled by the feasibility of the current iteration. We show that the rate of convergence does not depend on the accuracy of the solution of the auxiliary problems and that the penalty parameter ρ is bounded.

The numerical experiments suggest that the algorithm may be used for efficient solution of large sparse problems using initial large values for the penalty parameters and a strategy of preconditioning of the augmented matrices that takes advantage of the properties of the spectrum of these matrices when the number of constraints is relatively small.

The results presented here are relevant for the solution of quadratic programming problems with equality constraints and simple bounds. An extension to these more general problems is under consideration by the authors.

References

[1] O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, Cambridge (1994).

[2] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, London (1982).

[3] A. R. Conn, N.I.M. Gould, Ph. L. Toint, A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds, *SIAM Journal on Numerical Analysis* **28** (1988), 545-572.

[4] Z. Dostál, Domain decomposition and PDE with periodic boundary conditions, *Domain Decomposition Methods fot Partial Differential Equations*, eds. D. E. Keyes, T. F. Chan, G. Meurant, J. S. Scroggs, and R. G. Voigt, SIAM, Philadelphia (1992) 258-261.

[5] Z. Dostál, A. Friedlander, S. A. Santos, Analysis of semicoercive contact problems using symmetric BEM and augmented Lagrangians, *Eng. Appl. Bound. El.* (1996), in print.

[6] Z. Dostál, A. Friedlander, S. A. Santos, Augmented Lagrangians with Adaptive Precision Control for Quadratic Programming with Simple Bounds and Equality constraints, RP 74/96, IMECC-UNICAMP, Campinas 1996.

[7] R. Glowinski, P. Le Tallec, *Augmented Lagrangians and Operator Splitting Methods*, SIAM, Philadelphia (1989).

[8] G. H. Golub, C. F. Van Loan, *Matrix Computations*, 2ed., John Hopkins University Press, Baltimore (1989).

[9] W. W. Hager, Approximations to the multiplier method, *SIAM Journal on Numerical Analysis* **22** (1985 16-46.

[10] W. W. Hager, Dual techniques for constraint optimization, *JOTA* **55** (1987) 37-71.

[11] W. W. Hager, Analysis and implementation of a dual algorithm for constraint optimization, *JOTA* **79** (1993) 37-71.

[12] W. W. Hager, Updating the inverse of a matrix, *SIAM Review* **31** (1989), 221-239.

[13] M. R. Hestenes, Multiplier and gradient methods, *JOTA* **4** (1969) 303-320.

[14] P. Le Tallec, T. Sassi, Domain decomposition for non-matching grids: Augmented Lagrangians approach, *Mathematics of Computation* **64** (1995) 1367-1396.

[15] D. G. Luenberger, *Linear and Nonlinear Programming*, 2nd edition, Addison-Wesley (1984).

[16] M. J. D. Powell, A method for nonlinear constraints in minimization problems, in *Optimization*, ed. R. Fletcher, Academic Press, New York (1969) 283-298.

[17] R. T. Rockafellar, The multiplier method of Hestenes and Powell applied to convex programming, *JOTA* **12** (1973) 555-562.

[18] R. T. Rockafellar, A dual approach to solving nonlinear programming problems by unconstrained optimization, *Mathematical Programming* **5** (1973) 354-373.

[19] R. T. Rockafellar, Augmented Lagrange multiplier functions and duality in nonconvex programming, *SIAM Journal on Control* **12** (1974) 268-285.

[20] J. C. Simo, T. A. Laursen, An augmented Lagrangian treatment in contact problems involving friction, in *Comp.Structures* **42** (1992) 97-116.