Optimal algorithm for large sparse equality constrained problems with uniformly bounded spectrum

Z. Dostál

FEI VŠB-Technical University Ostrava
Tř 17. listopadu, CZ-70833 Ostrava, Czech Republic

1 Introduction

An important ingredient in development of effective methods for the solution of very large problems is identification of the algorithms that can solve some special cases with the optimal (i.e. asymptotically linear) complexity. For example, the interior point methods were applied successfully to the solution of very large problems of nonlinear optimization with many constraints and tens of thousands decision variables observing that the Hessian matrix with a special pattern of the distribution of the nonzero elements may be decomposed with nearly linear complexity [7].

In this paper, we review our recent algorithm [2] for the solution of the convex equality constrained quadratic programming problems. This algorithm can be implemented in such a way that, for the class of problems with the spectrum of the Hessian matrix in a given positive interval, it can find approximate solutions at the cost proportional to that of the matrix-vector multiplication. If applied to the class of problems with the sufficiently sparse Hessian and the constraint matrix, then the algorithm is optimal.

We have tested our algorithms by solving the class of well conditioned problems of varying dimensions with the quadratic form q_t defined for $t \in \{2, 3, \ldots\}$ by the symmetric Toeplitz matrix $A = A_t$ of the order $2 * t^2$ determined by the nonzero entries $a_{11} = 12, a_{12} = a_{1,t} = -1$ and by the vectors $b = b_t$ defined by the entries $b_i = -1, i = 1, \ldots, 2 * t$. Using the Gershgorin theorem, it is easy to see that the eigenvalues λ_i of any A_t satisfy $8 \le \lambda_i \le 16$. The equality constraints were defined by the matrix $C = C_t$ with t rows comprising $2 * t^2$ entries which are zeros except $c_{i,t^2-i+1} = 1$ and $c_{i,t^2+i} = -1, i = 1, \ldots, t$.

In the whole paper, $q(x) = \frac{1}{2}x^T Ax - b^T x$ will always denote a strictly convex quadratic function defined on R^n, with the Hessian matrix $\nabla^2 q = A \in R^{n \times n}$ symmetric positive definite and $x, b \in R^n$. The eigenvalues of A will be denoted $\lambda_i(A)$,

$$\lambda_{\min}(A) = \lambda_1(A) \le \ldots \le \lambda_n(A) = \lambda_{\max}(A) = \|A\|.$$

The Euclidean norm and the $A-$energy norm of x will be denoted by $\|x\|$ and $\|x\|_A$, respectively. Thus $\|x\|^2 = x^\top x$ and $\|x\|_A^2 = x^\top Ax$. Analogous notation will be used for the induced matrix norm.

2 Equality constrained problems

We shall start with the problem of finding the minimizer of the quadratic function $q(x)$ subject to the linear equality constraints, that is

$$\text{minimize} \;\; q(x) \;\; \text{subject to} \;\; x \in \Omega_E \tag{2.1}$$

with $\Omega_E = \{x \in R^n : Cx = d\}, C \in R^{m\times n}$, and $d \in R^m$. We require neither that C is a full row rank matrix nor $m \leq n$, but we shall assume that d belongs to the range of C to guarantee that Ω_E is not empty. Our development is based on the augmented Lagrangian method [1] which reduces (2.1) to a sequence of the problems of the form

$$\text{minimize} \;\; L(x, \mu^k, \rho_k) \;\; \text{subject to} \;\; x \in R^p \tag{2.2}$$

where

$$L(x, \mu^k, \rho_k) = q(x) + (\mu^k)^T(Cx - d) + \frac{\rho_k}{2}\|Cx - d\|^2 \tag{2.3}$$

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

$$g(x, \mu, \rho) \;\;=\;\; \nabla_x L(x, \mu, \rho) \;\;=\;\; Ax - b + C^T\mu + \rho C^T(Cx - d). \tag{2.4}$$

Our algorithm with the adaptive precision control reads as follows.

Algorithm 2.1. (Semi-monotonic augmented Lagrangians for equality constraints (SMALE))
Given $\eta > 0,\; \beta > 1,\; M > 0,\; \rho_0 > 0,$ and $\mu^1 \in R^m$, set $k = 0$.
Step 1. {Inner iteration with adaptive precision control.}
Find x^k such that

$$\|g(x^k, \mu^k, \rho_k)\| \leq \min\{M\|Cx^k - d\|, \eta\}. \tag{2.5}$$

Step 2. {Update μ.}

$$\mu^{k+1} = \mu^k + \rho_k(Cx^k - d). \tag{2.6}$$

Step 3. {Update ρ provided the increase of the Lagrangian is not sufficient.}
If $k > 0$ and

$$L(x^k, \mu^k, \rho_k) < L(x^{k-1}, \mu^{k-1}, \rho_{k-1}) + \frac{\rho_k}{2}\|Cx^k - d\|^2 \tag{2.7}$$

then

$$\rho_{k+1} = \beta\rho_k, \tag{2.8}$$

else

$$\rho_{k+1} = \rho_k. \tag{2.9}$$

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 the conjugate gradient method [1]. Algorithm 2.1 differs from those considered by Hager [8] and Dostál, Friedlander and Santos [5] by the condition on the update of the penalization parameter in Step 3.

Algorithm 2.1 has been proved to be correctly defined and to enjoy a kind of optimal convergence of the feasibility error [2]. To present our optimality result related to the conjugate gradient implementation of Step 1, let $\mathcal{T}$ denote any set of indices and assume that for any $t \in \mathcal{T}$ there is defined a problem

$$\text{minimize} \;\; q_t(x) \;\; \text{s.t.} \;\; x \in \Omega_E^t \tag{2.10}$$

with $\Omega_E^t = \{x \in R^{n_t} : C_t x = 0\}$, $q_t(x) = \frac{1}{2}x^T A_t x - b_t^T x$, $A_t \in R^{n_t \times n_t}$ symmetric positive definite, $C_t \in R^{m_t \times n_t}$, and $b_t, x \in R^{n_t}$.

Theorem 2.2. *Let* $\{x_t^k\}, \{\mu_t^k\}$ *and* $\{\rho_{t,k}\}$ *be generated by Algorithm 2.1 for (2.10) with* $\|b_t\| \geq \eta_t > 0,\ \beta > 1,\ M > 0,\ \rho_{t,0} = \rho_0 > 0,\ \ \mu_t^0 = 0$. *Let* $0 < a_{\min} < a_{\max}$ *and* $0 < c_{\min} < c_{\max}$ *be given constants. Let Step 1 be implemented by the conjugate gradient method which generates the iterates* $x_t^{k,0}, x_t^{k,1}, \ldots, x_t^{k,l} = x_t^k$ *for the solution of (2.10) starting from* $x_t^{k,0} = x_t^{k-1}$ *with* $x_t^{-1} = 0$*, where* $l = l_{k_t}$ *is the first index satisfying*

$$\|g(x_t^{k,l}, \mu_t^k, \rho_k)\| \leq M\|C_t x_t^{k,l}\| \tag{2.11}$$

or

$$\|g(x_t^{k,l}, \mu_t^k, \rho_k)\| \leq \epsilon\|b_t\| \min\{1, M\}. \tag{2.12}$$

Let the class of problems (2.10) satisfies

$$a_{\min} \leq \lambda_{\min}(A_t) \leq \lambda_{\max}(A_t) = \|A_t\| \leq a_{\max} \ \text{ and } \ c_{\min} \leq \sigma_{\min}(C_t) \leq \|C_t\| \leq c_{\max}, \tag{2.13}$$

where $\sigma_{\min}(C_t)$ *denote the least nonzero singular value of* C_t. *Then the following statements hold:*
(i) Algorithm 2.1 generates an approximate solution $x_t^{k_t}$ *of any problem (2.10) which satisfies*

$$\|x^k - \overline{x}\| \leq \epsilon\|b_t\| \tag{2.14}$$

at $O(1)$ *matrix-vector multiplications by the Hessian of the augmented Lagrangian* L_t *for (2.10).*
(ii) The images of the Lagrange multipliers $C^T\mu^k$ *are bounded and converge to* $C^T\overline{\mu}$*, where* $\overline{\mu}$ *denotes any vector of Lagrange multipliers of the solution.*
Proof: See [4].

We have implemented Algorithm 2.1 in Matlab and solved a class of problems of the varying dimension defined in the introduction. We solved the problem with $\eta_t = \|b_t\|, \beta = 10, \rho = 200, M = 1$ and $\mu_0 = 0$ using the stopping criterium $\|g_t(x, \mu, \rho)\| \leq 10^{-5}\|b_t\|$ and $\|C_t x\| \leq 10^{-5}\|b_t\|$. The results are in Table 2.1.

Table 1. Performance of the SMALE

Equality constrains =bandwidth	Dimension n	cg iterations	Outer iterations
10	200	25	4
50	5000	22	4
100	20000	18	3
250	125000	18	3
500	500000	17	3

We conclude that we can observe optimality in practice for well conditioned problems. More numerical experiments and theoretical results may be found in [2, 4].

3 Conclusions

Theoretical results concerning optimality of the recently proposed algorithms for equality constrained quadratic programming were presented and illustrated by numerical experiments. An interesting feature of the presented algorithm is its capability to find the approximate solution of the class of problems with the uniformly bounded spectrum of the Hessian matrix at $O(1)$ matrix-vector multiplications. No assumptions concerning regularity of solution are used and the results are valid even for linearly dependent constraints. The results may be generalized to bound and equality constrained problems [3, 6].

Acknowledgement: This research has been supported by grant 1ET400300415.

References

[1] D. P. Bertsekas, *Nonlinear Optimization*, Athena Scientific, Belmont (1999).

[2] Z. Dostál, Semi-monotonic inexact augmented Lagrangians for quadratic programming with equality constraints, to appear in *Optimization Methods & Software.*

[3] Z. Dostál, Inexact semi-monotonic Augmented Lagrangians with optimal feasibility convergence for quadratic programming with simple bounds and equality constraints, to appear in *SIAM Journal for Numerical Analysis.*

[4] Z. Dostál, An optimal algorithm for a class of equality constrained quadratic programming problems with bounded spectrum, submitted.

[5] Z. Dostál, A. Friedlander and S. A. Santos (1999). Augmented Lagrangians with adaptive precision control for quadratic programming with equality constraints. *Computational Optimization and Applications,* 14, 37-53.

[6] Z. Dostál, A. Friedlander and S. A. Santos, Augmented Lagrangians with Adaptive Precision Control for Quadratic Programming with Simple Bounds and Equality constraints, *SIAM Journal on Optimization* 13,4(2003)1120-1140.

[7] J. Gondzio and R. Sarkissian, Parallel interior point solver for structured linear programs, *Mathematical Programming* 96 (2003) No 3, 561-584.

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