Constrained Optimization

Sara Rosato
Politecnico di Torino
s317547

Florentin-Cristian Udrea
Politecnico di Torino
s319029

***Abstract*—In this review we're going to examine and compare the results obtained using three numerical methods for the equality constrained quadratic programming problem: GMRES, Schur complement approach and Null Space Method.**

I. Introduction

A quadratic program is an optimization problem with a quadratic objective function and linear constraints.
The general problem (QP) can be written as the following:

$$\begin{array}{ll} min_x & q(x) = \frac{1}{2}x^TQx + c^Tx \\ s.t. & Ax = b \end{array}$$

where

- Q $\in \mathbb{R}^{n,n}$ is a symmetric positive semidefinite matrix
- c, x are vectors $\in \mathbb{R}^n$
- A $\in \mathbb{R}^{m,n}$, b $\in \mathbb{R}^m$
- m is the number of constraints

Quadratic programs can usually be solved in a finite number of computations, but the computational cost and time depend on the characteristics of the objective function and the number of constraints.
Applying the KKT conditions to (QP), at a minimum x^* we obtain the system:

$$\begin{array}{c} Qx^* + A^T\lambda^* = -c \\ Ax^* = b \end{array} \tag{1}$$

Introducing the matrix

$$K = \begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \tag{2}$$

and the vectors

$$w^* = \begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} \tag{3}$$

$$d = \begin{bmatrix} -c \\ b \end{bmatrix} \tag{4}$$

we can rewrite (1) as

$$K \cdot w^* = d \tag{5}$$

Solving (5) we can find a candidate minimum point x^* with Lagrangian multiplier λ^*.
If A is full rank and we consider a matrix Z whose columns form a basis of Ker(A), the reduced-Hessian matrix Z^TQZ is positive definite. On this assumptions, K is non-singular and it implies that there is a unique solution (x^*, λ^*) for (5).

II. Problem

The quadratic objective function is:

$$min_{x\in\mathbb{R}^n} \quad \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n-1} x_i \cdot x_{i+1} + \sum_{i=1}^{n} x_i \tag{6}$$

with equality constraints:

$$\begin{gathered} x_1 + x_{1+K} + x_{1+2K} + ... = 1 \\ x_2 + x_{2+K} + x_{2+2K} + ... = 1 \\ \vdots \\ x_K + x_{2K} + x_{3K} + ... = 1 \end{gathered} \tag{7}$$

We've solved problem (6) obtaining the results reported in the successive sections, testing our code with parameters $n = 10^4$, $n = 10^5$ and $K = 100$, $K = 500$. For this latter value of n, the matrix Q turns out too big, so we transformed it into a sparse matrix. The reason being that sparse matrices are easier to store, since the majority of their non-diagonal elements are zeros.
By inspecting the problem description (7) and the corresponding A matrix, we can observe that it is full rank. We also found that matrix Q is symmetric positive definite.
Given the considerations made in section I, we conclude that the matrix K is non-singular (thus invertible) and its intertia is (n, m, 0).
In order to have a useful metric to compare all the methods, we've used the equations of (1), computing the norm of the residuals, yield the results:

$$KKT\_gradL\_norm = ||Qx^* + A^T\lambda^* + c|| \tag{8}$$

$$KKT\_eq\_norm = ||Ax^* - b|| \tag{9}$$

In particular equation (8) shows how close we are to the solution and equation (9) show how well we respected the constraints. In both cases the closer to 0 the norms are, the higher the accuracy is.

III. Full system factorization

This method consists in the decomposition of the matrix K as $K = LDL^T$, where L is the lower triangular and D diagonal.
However, we can't use full system factorization in our problem since the minimum n+m we have is of the order $\sim 10^4$ and this could cause a fill-in phenomenon. We have therefore decided to proceed with other methods.

IV. GMRES

In order to solve the linear system (5) without decomposition, we have used the iterative solver GMRES. At first we used the parameters $tol = 10^{-6}$, $maxit = 200$, on the problem described by $n = 10^4$ and $K = 100$.
In this way GMRES converged at iteration 115 with elapsed time $\sim$ 10 seconds and having KKT_gradL_norm = $9.9037 \cdot 10^{-5}$, KKT_eq_norm = $2.9262 \cdot 10^{-6}$.
The 1-dimension vector w^* has, as expected, $n + K$ rows, since it gives both x^* (first n rows) and λ^*

$$w^* = \begin{bmatrix}0.0004 & 0.0008 & 0.0011 & ... - 1 & -1 & -1\end{bmatrix}^T$$

One drawback of the GMRES method is its linearly increasing cost per iterations, which may become prohibitively large in practical applications. We have tried with a bigger number of iterations and a lower tolerance, respectively 2000 and 10^{-8}, to see if we could reach better results. They are, in fact, slightly more precise, but the computational time is significantly higher ($\sim$ 123 seconds). We decided that the increment in precision was not worth the extra computational time needed to perform the method, since they are comparable. Thus the following results were taken having as parameters the ones initially used.

TABLE I
GMRES RESULTS: ELAPSED TIME (S)

	$n = 10^4$	$n = 10^5$
K=100	10.1752	0.2469
K=500	18.2654	0.9434

TABLE II
GMRES RESULTS: KKT_GRADL_NORM

	$n = 10^4$	$n = 10^5$
K=100	$9.9037 \cdot 10^{-5}$	$2.6841 \cdot 10^{-4}$
K=500	$2.2275 \cdot 10^{-4}$	$2.9531 \cdot 10^{-4}$

TABLE III
GMRES RESULTS: KKT_EQ_NORM

	$n = 10^4$	$n = 10^5$
K=100	$2.9264 \cdot 10^{-6}$	$1.4123 \cdot 10^{-4}$
K=500	$1.9133 \cdot 10^{-5}$	$4.6047 \cdot 10^{-5}$

TABLE IV
GMRES RESULTS: F(X_STAR)

	$n = 10^4$	$n = 10^5$
K=100	100.0000	100.0000
K=500	500.0000	500.0000

V. SCHUR COMPLEMENT APPROACH

By the considerations made on matrix Q in section II we understand that Schur complement approach is feasible since it only asks for Q to be symmetric positive semi-definite.
Schur complement $\hat{Q} = AQ^{-1}A^T$ is symmetric positive definite under these conditions and belongs to $\mathbb{R}^{K,K}$.
To find x^*, at first we have computed λ^* solving the linear system

$$\hat{Q}\lambda^* = -b - AQ^{-1}c \tag{10}$$

obtaining

$$\lambda^* = \begin{bmatrix}-1.0000 - 1.0000 - 1.0000 - 1.0000...\end{bmatrix}^T$$

with dimension $\mathbb{R}^{K,1}$.
We've used this result in

$$x^* = Q^{-1}(-c - A'\lambda^*) \tag{11}$$

leading to the final solution $x^* \in \mathbb{R}^n$.
Since Q is rappresented in different ways for n=10^4 and n=10^5 (as specified in section II) we used different methods for computing $\hat{Q}$: respectively directly computing the inverse of Q and using the backslash operator.

We can say that Schur complement method is convenient here since matrix Q is easy to invert and K is relatively small.

In the tables (V), (VI), (VII) and (VIII) are reported all the results obtained with Schur Complement method.

TABLE V
SCHUR RESULTS: ELAPSED TIME (S)

	$n = 10^4$	$n = 10^5$
K=100	35.2247	0.1900
K=500	33.9837	1.2846

TABLE VI
SCHUR RESULTS: KKT_GRADL_NORM

	$n = 10^4$	$n = 10^5$
K=100	$7.9029 \cdot 10^{-15}$	0
K=500	$4.1992 \cdot 10^{-14}$	0

TABLE VII
SCHUR RESULTS: KKT_EQ_NORM

	$n = 10^4$	$n = 10^5$
K=100	$5.6613 \cdot 10^{-6}$	0.2707
K=500	$2.4407 \cdot 10^{-6}$	0.1303

TABLE VIII
SCHUR RESULTS: F(X_STAR)

	$n = 10^4$	$n = 10^5$
K=100	100.0000	102.7070
K=500	500.0000	502.9141

VI. Null space method

We can use null space method even if Q is non singular, differently from Schur complement method.
Null space method is based on the assumptions that a particular solution $\hat{x}$ of the linear system

$$Ax = b \tag{12}$$

is known and the null space matrix $Z \in \mathbb{R}^{n,n-m}$, full rank, is given such that $AZ = 0$.
In order to find the general solution of (12), we used an arbitrary vector $v \in \mathbb{R}^{n-m}$, defining x as

$$x = Zv + \hat{x} \tag{13}$$

Replacing this equivalence in the first equation of (1) and multiplying by Z^T, we obtain the linear system

$$Z^T QZv = -Z^T(c + Q\hat{x}) \tag{14}$$

that can be solved with a Cholesky factorization of the reduced-Hessian matrix $Z^T QZ$ to determine v and, as a consequence of (13), the vector x^*.
To quantify the Lagrangain multiplier, we have solved the system

$$AA^T\lambda^* = -Ac - AQx^* \tag{15}$$

Results obtained with Null spaced method and elapsed time are reported in tables (IX), (X), (XI), (XII)

TABLE IX
NULL SPACE RESULTS: ELAPSED TIME (S)

	$n = 10^4$	$n = 10^5$
K=100	58.0505	139.8493
K=500	55.2803	5.7247

TABLE X
NULL SPACE RESULTS: KKT_GRADL_NORM

	$n = 10^4$	$n = 10^5$
K=100	$9.0386 \cdot 10^{-15}$	$1.6002 \cdot 10^{-13}$
K=500	$7.1856 \cdot 10^{-15}$	$1.4359 \cdot 10^{-13}$

TABLE XI
NULL SPACE RESULTS: KKT_EQ_NORM

	$n = 10^4$	$n = 10^5$
K=100	$2.8152 \cdot 10^{-15}$	$8.5140 \cdot 10^{-15}$
K=500	$2.9708 \cdot 10^{-15}$	$7.8748 \cdot 10^{-15}$

TABLE XII
NULL SPACE RESULTS: F(X_STAR)

	$n = 10^4$	$n = 10^5$
K=100	100.0000	100.0000
K=500	500.0000	500.0000

As we can see from tab. (IX), null space method is efficient when $n - m$ is small, otherwise the null-space matrix Z is too expensive to compute.

VII. Conclusions

It is difficult to determine ”apriori” which method is better to solve QP, since each one has its strengths and weaknesses which are based on the specific problem, as explained in the previous sections.

In general GMRES requires a lot of iterations, thus a lot of computational time if high precision is required, so it’s preferable to use the other two methods reported.

If both Schur complement approach and Null-Space method are feasible, in general Schur complement approach is the best choice as it solves smaller linear systems which is computationally more efficient. Also, a risk of the Null-Space method is that it may define a Z matrix which could make the system (14) ill conditioned.

For our problems we’ve noticed, as shown in the table (VIII), that Schur Complement Approach is not completely precise when n is large (e.g. 10^5). We also can observe in table VII that for large n the KKT equality constraints (9) are not fully respected. This issue is probably given by the fact that the matrix Q is sparse (as previously explained). The Schur Complement Method, though, is clearly faster than the other methods considered.

On that note, depending on the problem that we are trying to solve, we should ponder a trade-off between accuracy and computational time.

On the other hand, having a relatively smaller problem ($n = 10^4$) both methods have satisfactory KKT equality conditions (Null-Space gives most accurate results) and converge to the same solution. The Schur Complement Approach is faster, thus (all else being equal) better. This last consideration can be observed in fig 1.

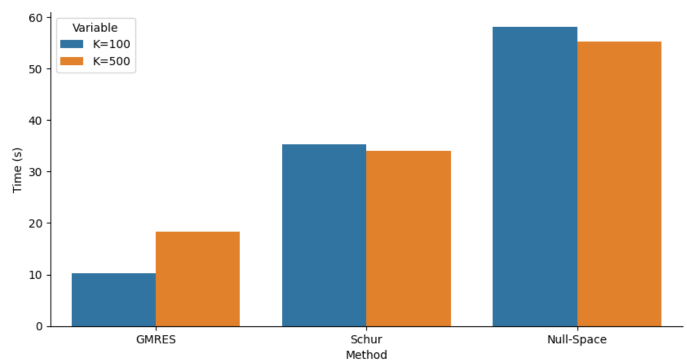

Fig. 1. Time comparison between considered methods having $n = 10^4$

References

[1] ”Springer series in operations research and financial engineering” by Thomas V. Mikosch, Sidney I. Resnick, Stephen M. Robinson, 2006