A Null Space Algorithm for Constrained Optimization

Ya-xiang Yuan*

State Key Laboratory of Scientific and Engineering Computing,

Institute of Computational Mathematics and Scientific/Engineering Computing,

Chinese Academy of Sciences, P. O. Box 2719, Beijing 100080, China

Email-addresses: yyx@lsec.cc.ac.cn

dedicated to the 65th birthday of Professor H.C. Huang

Abstract

In this paper, we study convergence properties of null step techniques for constrained optimization. In most algorithms that use a null step and range space step, the range space step is normally a quadratic convergent step as it is obtained by Newton's method, but the null step converges much slower as quite often it is computed by quasi-Newton methods. This unbalance suggests us to study a technique that computes the null steps more often than the range space step.

1 Introduction

In this paper we study null step algorithms for the equality constrained optimization problem which has the following form:

$$\min f(x) \tag{1.1}$$

subject to

$$c(x) = 0, \tag{1.2}$$

where $f(x)$ is a real valued function defined in $\Re^n$ and $c(x) = (c_1(x), c_2(x), \ldots, c_m(x))^T$ is a mapping from $\Re^n$ to $\Re^m$.

Consider an iterative algorithm for problem (1.1)-(1.2). At the beginning of the $k-$th iteration, we have a current iterate point x_k, which is an approximate solution. We hope to find a better approximate point x_{k+1}. Write the point that we are searching for in the form $x_k + d$, we would require $c(x_k + d) = 0$. However, generally this leads to a

*This work is partially supported by Chinese NNSF grants 19731010

nonlinear equation which would require an infinite iterative process to solve. Therefore it is reasonable to replace it by a linearized system $c(x_k) + A_k^T d = 0$, where

$$A_k = \nabla c(x_k)^T = (\nabla c_1(x_k), \nabla c_1(x_k), \ldots, \nabla c_m(x_k)) \tag{1.3}$$

is the Jacobian matrix of $c(x)$ at x_k. Many algorithms require the search direction satisfying the linearized constraints. For example, the SQP method which solves the following subproblem:

$$\min g_k^T d + \frac{1}{2} d^T B_k d \tag{1.4}$$

subject to

$$c(x_k) + A_k^T d = 0, \tag{1.5}$$

where $g_k = g(x_k) = \nabla f(x_k)$, and B_k is an approximation to the Hessian of the Lagrangian function. Assume that c_k is in the range space of A_k^T (which is true if A_k has full column rank), then we can see that all the solutions of (1.5) can be written in the form

$$d = -(A_k^T)^+ c_k + (I - (A_k^T)^+ A_k^T) y \tag{1.6}$$

where y is any vector in $\Re^n$. Therefore d can be decomposed into two parts, one is in the range-space and the other in the null space. The range space step, which is often called the vertical step, is

$$v_k = -(A_k^T)^+ c_k. \tag{1.7}$$

The freedom of d lies in the null space. The null step, which is also called the horizontal step, can be expressed as

$$h = Z_k \bar{d}, \tag{1.8}$$

where $\bar{d} \in \Re^{n-r}$, $Z_k \in \Re^{n \times (n-r)}$, r being the rank of A_k and

$$Z_k^T Z_k = I, \qquad A_k^T Z_k = 0. \tag{1.9}$$

We can see that the columns of Z_k form an orthogonal basis of the null space of A_k^T. Using the null step expression, the objective function in the SQP method can be rewritten as

$$\bar{g}_k^T \bar{d} + v_k^T B_k Z_k \bar{d} + \frac{1}{2} \bar{d}^T \bar{B}_k \bar{d}, \tag{1.10}$$

where

$$\bar{g}_k = Z_k^T g_k \tag{1.11}$$

is the reduced gradient, and

$$\bar{B}_k = Z_k^T B_k Z_k \tag{1.12}$$

is the two sided reduced Hessian of the Lagrangian function. This leads to the linear system:

$$\bar{g}_k + Z_k^T B_k v_k + \bar{B}_k \bar{d} = 0. \tag{1.13}$$

The above equation and $d = v_k + Z_k\bar{d}$ give that

$$\begin{aligned} Z_k^T g_k + Z_k^T B_k d &= 0 && (1.14)\\ c_k + A_k^T d &= 0. && (1.15)\end{aligned}$$

The above system can also be easily derived from the KKT condition for the quadratic programming problem (1.4)-(1.5). Actually (1.14) is equivalent to the existence a vector $\lambda \in \Re^m$ such that

$$g_k + B_k d - A_k\lambda = 0 \tag{1.16}$$

holds. Algorithms based on the (1.14)-(1.15) with $Z_k^T B_k$ is replaced by an approximate matrix $\hat{B}_k \in \Re^{(n-r)\times n}$ is so called one-side reduced Hessian method. Under certain conditions, it is shown that the one-side reduced Hessian method is superlinearly convergence (see Nocedal and Overton(1985)). Using $d = Z_k Z_k^T d + Y_k Y_k^T d$, where $Y_k \in \Re^{n\times r}$ satisfies $Y_k^T Y_k = I$ and $Y_k^T Z_k = 0$. We can rewrite (1.14)-(1.15) as

$$\begin{pmatrix} Z_k^T B_k Z_k & Z_k^T B_k Y_k \\ 0 & A_k^T Y_k \end{pmatrix}\begin{pmatrix} Z_k^T d \\ Y_k^T d \end{pmatrix} = -\begin{pmatrix} Z_k^T g_k \\ c_k \end{pmatrix}. \tag{1.17}$$

In his study on the SQP method, Powell(1978) showed that the SQP method is two step q-superlinear convergence provided that $\bar{B}_k = Z_k^T B_k Z_k$ is a good approximation to the two sided reduced Hessian matrix and $Z_k^T B_k Y_k$ is bounded. Replacing $Z_k^T B_k Y_k$ by the zero matrix, We obtained the

$$\begin{aligned} \bar{B}_k Z_k d &= -Z_k^T g_k,\\ A_k^T Y_k Y_k^T d &= -c_k. \end{aligned} \tag{1.18}$$

Denote the solution by d_k, it follows that

$$\begin{aligned} Z_k^T d_k &= -\bar{B}_k^{-1} Z_k^T g_k\\ Y_k Y_k^T d_k &= -(A_k^T)^+ c_k. \end{aligned} \tag{1.19}$$

Thus, remembering that $d = Z_k Z_k^T d + Y_k Y_k^T d$, we have

$$d_k = v_k + h_k, \tag{1.20}$$

where

$$\begin{aligned} h_k &= -Z_k \bar{B}_k^{-1} Z_k^T g_k,\\ v_k &= -(A_k)^+ c_k. \end{aligned} \tag{1.21}$$

Reduced Hessian methods for constrained optimization have been studied by many researchers. Recent works include Gilbert(1991) and Xie and Byrd (1999).

A very good properties of the two sided reduced Hessian method is that we only need to have the two sided reduced Hessian matrix $\bar{B}_k$ which has only $(n-m)\times(n-m)$ elements while the one sided reduced Hessian $Z_k^T B_k$ has $(n-m)\times n$ elements. Thus the step d_k in the two-sided reduced Hessian method is much easy to computed than in the one-side reduced Hessian method, especially when $n-m$ is much less than n. However, an unsatisfactory property of the two sided reduced Hessian method is that its convergence rate is only two step q-superlinearly convergence, namely

$$\lim \frac{\|x_{k+1} - x^*\|}{\|x_{k-1} - x^*\|} = 0, \tag{1.22}$$

while the one sided reduced Hessian method is one step q-superlinearly convergence which says

$$\lim \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0. \tag{1.23}$$

An observation on the two sided reduced Hessian is as follows. In many practical implementations, the two sided reduced Hessian matrix $Z_k^T B_k Z_k$ is replaced by an $(n-m) \times (n-m)$ quasi-Newton matrix. Thus, the method in the null space is a quasi-Newton step while in the range space is a Newton step. As the Newton method is quadratic convergence, while the q-order of convergence of the quasi-Newton method can be very close to one(see Yuan(1984), it is expected that the iterate points approach the solution much faster in the range space than in the null space. To overcome this unbalance, it is intuitive to to take more null space steps than range space steps. The aim of the this paper is to explore this idea.

In the next section, we give an algorithm and shows that the algorithm is locally one step q-superlinearly convergent. In Section 3 an example is given to show the numerical behaviour of the algorithm.

2 The Algorithm and Its Convergence

We consider a method that takes two null space steps after every range space step. At the $k-$th iteration, the vertical step is

$$v_k = -(A_k^T)^+ c_k \tag{2.1}$$

and the null space step is

$$h_k = -Z_k \bar{B}_k^{-1} \bar{g}_k \tag{2.2}$$

We definite the new point

$$\bar{x}_k = x_k + v_k + h_k, \tag{2.3}$$

which is the next iterate point in standard two sided reduced Hessian method. Now we consider to carry out another null step. At the new point $\bar{x}_k$, Let the $A(\bar{x}_k)$ be computed and the projected Thus, the gradient of the objective function $g(\bar{x}_k)$ projected to the null space of the linearized constraints at the point $\bar{x}_k$ is as follow:

$$\tilde{g}_k = \bar{Z}_k \bar{Z}_k^T g(\bar{x}_k). \tag{2.4}$$

Consider the second null space step also have the form $Z_k d$, it is natural to obtain d by solving the following subproblem

$$\min \tilde{g}_k^T Z_k d + \frac{1}{2} d^T \bar{B}_k d, \tag{2.5}$$

which has the solution

$$-\bar{B}_k^{-1} Z_k^T \tilde{g}_k. \tag{2.6}$$

Therefore the second null space step is given by

$$\hat{h}_k = -Z_k \bar{B}_k^{-1} Z_k^T \tilde{g}_k. \tag{2.7}$$

Define $\hat{g}_k = Z_k^T \tilde{g}_k$, it follows that

$$\hat{h}_k = -Z_k \bar{B}_k^{-1} \hat{g}_k, \tag{2.8}$$

which shows that the formula for the second null step is the same as the first null space step except that $\bar{g}_k$ is replaced by $\hat{g}_k$.

Thus we have

$$d_k = v_k + h_k + \hat{h}_k. \tag{2.9}$$

Now we study the local convergence properties of the second null step. Assume that x_k converges to a solution of the optimization problem (1.1)-(1.2) at which the following conditions are satisfied:

Assumption 2.1 *Assume that x^* is a KKT point, namely $c(x^*) = 0$ and there exists $\lambda^* \in \Re^m$ such that*

$$g(x^*) - A(x^*)\lambda^* = 0. \tag{2.10}$$

Assume that $A(x^*)$ has full column rank. Denote the Hessian of Lagrange function by

$$W^* = \nabla^2 f(x^*) - \sum_{i=1}^{m} (\lambda^*)_i \nabla^2 c_i(x^*). \tag{2.11}$$

Let $Z^* \in \Re^{n\times(n-m)}$ be a matrix whose columns form an orthonormal basis of the null space of $A(x^*)$. Assume that $(Z^*)^T W^* Z^*$ is positive definite.

We also assume that the two sided reduced Hessian is a good approximation:

Assumption 2.2 *Assume that*

$$\lim_{k\to\infty} \frac{\|(\bar{B}_k - (Z^*)^T W^* Z^*) d_k\|}{\|d_k\|} = 0. \tag{2.12}$$

First we have the following lemmas.

Lemma 2.3 *If Assumption 2.2 holds, then*

$$Z_k^T W^* h_k + Z_k^T g_k = o(\|h_k\|). \tag{2.13}$$

Proof From the definition of h_k, we have that

$$\bar{B}_k Z_k^T h_k + Z_k^T g_k = 0. \tag{2.14}$$

The above relation and Assumption 2.2 imply that

$$Z_k^T W^* Z_k Z_k^T h_k + Z_k^T g_k = o(\|h_k\|) \tag{2.15}$$

which yields the lemma because $Z_k Z_k^T h_k = h_k$. **QED**

Lemma 2.4 *Assume that $x_k \to x^*$. If Assumptions 2.1 and 2.2 are satisfied, then*

$$Z_k^T(g(\bar{x}_k)) - A(\bar{x}_k)\lambda^*) = Z_k^T W^* v_k + o(\|x_k - x^*\|^2) \tag{2.16}$$

Proof From the assumptions, we have that

$$\begin{aligned} Z_k^T(g(\bar{x}_k)) - A(\bar{x}_k)\lambda^*) &= Z_k^T g_k + Z_k^T W^*(\bar{x}_k - x_k) + o(\|\bar{x}_k - x_k\|) \\ &= Z_k^T g_k + Z_k^T W^*(v_k + h_k) + o(\|\bar{x}_k - x_k\|) \\ &= Z_k^T W^* v_k + o(\|h_k\|) + o(\|\bar{x}_k - x_k\|) \\ &= Z_k^T W^* v_k + o(\|x_k - x^*\|). \end{aligned} \tag{2.17}$$

This shows that the lemma is true. **QED**

Lemma 2.5

$$Z_k^T W^* d_k = -Z_k^T g_k + o(\|x_k - x^*\|). \tag{2.18}$$

Proof Let $\bar{\lambda}_k$ be defined by

$$\bar{\lambda}_k = argmin\|g(\bar{x}_k) - A(\bar{x}_k)\lambda\|_2. \tag{2.19}$$

It is easy to show that

$$\bar{\lambda}_k = \lambda^* + o(1), \tag{2.20}$$

and

$$g(\bar{x}_k) - A(\bar{x}_k)\bar{\lambda}_k = Z(\bar{x}_k)Z(\bar{x}_k)^T g(\bar{x}_k). \tag{2.21}$$

Thus

$$\begin{aligned} Z_k^T W^* \hat{h}_k &= -Z_k^T W^* Z_k \bar{B}_k^{-1} \hat{g}_k = -\hat{g}_k + o(\|\hat{g}_k\|) \\ &= -Z_k \bar{Z}_k \bar{Z}_k^T g(\bar{x}_k) + o(\|\hat{g}_k\|) \\ &= -Z_k^T(g(\bar{x}_k) - A(\bar{x}_k)\bar{\lambda}_k) + o(\|\hat{g}_k\|) \\ &= -Z_k^T(g(\bar{x}_k) - A(\bar{x}_k)\lambda^*) - Z_k^T A(\bar{x}_k)(\lambda^* - \bar{\lambda}_k) + o(\|\hat{g}_k\|) \\ &= -Z_k^T W^* v_k - Z_k^T[A(\bar{x}_k) - A(x_k)](\lambda^* - \bar{\lambda}_k) + o(\|x_k - x^*\|) \\ &= -Z_k^T W^* v_k + o(\|\bar{x}_k - x_k\|) + o(\|x_k - x^*\|) \\ &= -Z_k^T W^* v_k + o(\|x_k - x^*\|). \end{aligned} \tag{2.22}$$

Consequently,

$$
\begin{aligned}
Z_k^T W^* d_k &= Z_k^T W^* h_k + o(\|x_k - x^*\|) \\
&= -Z_k^T g_k + o(\|x_k - x^*\|).
\end{aligned} \tag{2.23}
$$

QED

Now we can establish the local one step q-superlinearly convergence result for our algorithm.

Theorem 2.6 *If Assumptions 2.1 and 2.2 are satisfied, then*

$$\|x_k + d_k - x^*\| = o(\|x_k - x^*\|). \tag{2.24}$$

Proof First we have the relation

$$
\begin{aligned}
Z_k^T W^* d_k &= -Z_k^T g_k + o(\|x_k - x^*\|) \\
&= -Z_k^T (g_k - A(x_k)\lambda^*) + +o(\|x_k - x^*\|) \\
&= -Z_k^T W^* (x_k - x^*) + o(\|x_k - x^*\|),
\end{aligned} \tag{2.25}
$$

which gives that

$$Z_k^T W^* (x_k + d_k - x^*) = o(\|x_k - x^*\|). \tag{2.26}$$

Also we can show that

$$A(x_k) d_k = A(x_k) v_k = -c(x_k) = -A(x_k)(x_k - x^*) + o(\|x_k - x^*\|). \tag{2.27}$$

Thus,

$$A(x_k)(x_k + d_k - x^*) = o(\|x_k - x^*\|). \tag{2.28}$$

Due to our assumptions, we can see that

$$\begin{pmatrix} Z_k^T W^* \\ A(x_k) \end{pmatrix} \to \begin{pmatrix} Z(x^*)^T W^* \\ A(x^*) \end{pmatrix}. \tag{2.29}$$

Because the matrix in the right hand side of the above relation is a non-singular matrix, it can be shown that

$$\|x_k + d_k - x^*\| = o(\|x_k - x^*\|), \tag{2.30}$$

which indicates that the theorem is true. **QED**

3 An Example

We use a simple example to show that our technique ensures locally one step Q-superlinearly convergence. Consider the following 2-dimensional example:

$$
\begin{aligned}
\min f(y,z) &= \frac{1}{2} z^2 - yz + \frac{1}{6(1-z)^3} \Big[-4(z-y)^3 - 6(z-y)^2 (y - z^2) \\
&\quad -12(z-y)(y-z^2)^2 - 17(y-z^2)^3 + 3(1-z)^{-1}(y-z^2)^4 \Big]
\end{aligned} \tag{3.31}
$$

subject to

$$c(y,z) = y + (1-z)^{-2}[(z-y)^2 + (z-y)(y-z^2) + 2(y-z^2)^2] = 0. \tag{3.32}$$

This problem was given by Yuan(1985) to demonstrate the one-fast-one-slow convergence phenomenon of the two sided reduced Hessian method.

The solution of (3.31)-(3.32) is $x^* = (0,0)^T$. Define

$$r_k = \frac{\|x_{k+1} - x^*\|_\infty}{\|x_k - x^*\|_\infty}. \tag{3.33}$$

We apply both the standard two sided reduced Hessian method

$$x_{k+1} = x_k + v_k + h_k, \tag{3.34}$$

and the two sided reduced Hessian method with two null space steps:

$$x_{k+1} = x_k + v_k + h_k + \hat{h}_k \tag{3.35}$$

to problem (3.31)-(3.32). Three different initial points $(0.1, 0.1)^T$, $(0.2, 0.1)^T$ and $(0.0, 0.1)^T$ are used. The values of ratio r_k in all iterations are listed in Table one. Calculations were terminated when $r_k < 10^{-12}$.

	Standard Method			Modified Method		
k	(0.1, 0.1)	(0.2, 0.1)	(0.0, 0.1)	(0.1, 0.1)	(0.2, 0.1)	(0.0, 0.1)
1	1	1.35	0.141	0.144	1.12	0.137
2	0.1	0.394	1.02	2.70×10^{-2}	0.239	2.09×10^{-2}
3	1	0.662	2.62×10^{-2}	6.82×10^{-4}	8.48×10^{-2}	5.45×10^{-4}
4	1.0×10^{-2}	7.28×10^{-2}	0.532	5.21×10^{-7}	8.06×10^{-3}	2.78×10^{-7}
5	1	0.959	2.02×10^{-4}	2.42×10^{-13}	6.81×10^{-5}	8.15×10^{-14}
6	1.0×10^{-4}	4.93×10^{-3}	1		4.46×10^{-9}	
7	1.0	1.0	4.57×10^{-8}		7.07×10^{-18}	
8	1.0×10^{-8}	2.53×10^{-5}	0.894			
9	1.0	0.962	1.67×10^{-15}			
10	1.0×10^{-16}	5.94×10^{-10}	1			
11		1				
12		3.53×10^{-19}				

Table 1. Values of r_k for two methods with different starting points

From the results listed in Table 1, we can see that the standard two sided reduced Hessian method converges to the solution in a one-fast-ond-slow pattern, while the method with two null space steps converges one step q-superlinearly.

Our numerical results in this section and the theoretical analyses in the previous section show that the reduced Hessian method converges one step q-superlinearly if two null spaces are taken in every iteration. Our results provide a remedy for the unbalance (in the null space and the range space) of convergence of the standard reduced Hessian method. Our idea show that it can potentially be more efficiently to take two null space step in every iteration than one null space step. Further works are required to implement a practical algorithm which would requires some globalization techniques such as line searches or trust regions.

References

[1] J.C. Gilbert, "Maintaining the positive definiteness of the matrices in reduced Hessian methods for equality constrained optimization", *Mathematical Programming* 50(1991) 1-28.

[2] J. Nocedal and M.L. Overton, "Projected Hessian update algorithms for nonlinear constrained optimization", *SIAM J. Numer. Anal.* 22(1985) 821-850.

[3] M.J.D. Powell, "A fast algorithm for nonlinearly constrained optimization calculations", in: G.A. Watson, ed., *Numerical Analysis* (Springer-Verlag, Berlin, 1978) pp. 144-157

[4] Y.F. Xie and R.H. Bryd, "Practical update criteria for reduced Hessian SQP: Global analysis", *SIAM J. Optimization* 9(1999) 578-604.

[5] Y. Yuan, "On the least Q-order of convergence of variable metric algorithms", *IMA J. Numerical Analysis* 4(1984) 233-239.

[6] Y. Yuan, "An only 2-step Q-superlinear convergence example for some algorithms that use reduced Hessian approximation", *Mathematical Programming* 32(1985) 224-231.