A Nonmonotone Wedge Trust Region Method with Self-correcting Geometry for Derivative-free Optimization

Weili ZHENG, Qinghua ZHOU*
College of Mathematics and Information Science, Hebei University, Baoding, China

Abstract: In this paper, we consider a nonmonotone wedge trust region method with self-correcting geometry for derivative-free optimization. After establishing the whole algorithm, we investigate its qualities and prove the global convergence of the new algorithm under some mild conditions.

Keywords: Wedge trust region; Nonmonotone technique; Self-correcting geometry; Derivative-free optimization

1. Introduction

In this paper, we limit our discussion to the unconstrained optimization problem

$$\min f(x), \quad x \in R^n, \tag{1}$$

where the objective function $f(x)$ is a smooth function of several variables and its derivatives are unavailable or unreliable. This class of problems exists widely in practice and is called as the derivative-free optimization problems. For example, the function evaluations are the outcomes of a so-called black box or the simulation results of computer program. All the situations above reduce to an impossible mission to compute the relative derivatives of the function. The Hooke-Jeeves method [1] in 1961 and the Nelder-Mead method [2] in 1968 are the early literatures of pattern search methods which can deal with this situation. In 1969 and 1973, Winfield [3] put forward the first model-based method for derivative-free optimization. From then on, many researchers investigated such problem and contributed rich results on derivative-free methods. In 2002, Marazzi and Nocedal [11] proposed the genius idea of constructing linear or quadratic model within the wedge trust region. By introducing a "taboo region", their method succeeded figure out the problem in persisting the geometry of the interpolation sets.

Considering the effectiveness of nonmonotone strategy when coping with the problems which have strong non-quadratic qualities, we proposed a hybrid algorithm which combines nonmonotone strategy with wedge trust region methods in this paper.

Totally, this paper is organized as follows. In section 2 we first introduce some preliminaries about interpolation schemes, then design the new self-correcting geometry strategy and propose our nonmonotone wedge trust region algorithm with self-correcting geometry. In section 3, we prove the global convergence of our algorithm and some conclusions are made in section 4.

2. A Non-monotone Wedge Trust Region Method with Self-correcting Geometry for Derivative-free Optimization

2.1. Interpolation models

There are several different methods to construct the interpolation model, such as Lagrange interpolation [4, 5], Newton interpolation, and radial basis function interpolation [6].

Let us consider p_n^d, the space of polynomials of degree $\le d$ in R^n, and let $p_1 \triangleq p+1$ be the dimension of the space. One knows that for $d=1, p_1=n+1$ and for $d=2, p_1=(n+1)(n+2)/2$.

A basis $\Phi=\{\phi_0(x),\phi_1(x),\cdots,\phi_p(x)\}$ of p_n^d is a set of p_1 polynomials of degree $\le d$ that span p_n^d. For any such basis Φ, any polynomial $m(x)\in p_n^d$ can be written as $m(x)=\sum_{j=0}^{p}\alpha_j\phi_j(x)$, where $\alpha_0,\cdots,\alpha_p$ are real coefficients.

We say that the polynomial $m(x)$ interpolates the function $f(x)$ at a given point if $m(y)=f(y)$.

We give a set $Y=\{y^0,y^1,\cdots,y^p\}\subset R^n$ of the interpolation points. Let $m(x)$ denote a polynomial of degree d in R^n that interpolates a given function $f(x)$ at the points in Y. The $\alpha_0,\alpha_1,\cdots,\alpha_p$ can then be determined by solving the linear system

$M(\Phi,Y)\alpha_\Phi=f(Y)$, where

$$M(\Phi,Y)=\begin{pmatrix}\phi_0(y^0) & \cdots & \phi_p(y^0)\\ \vdots & \ddots & \vdots \\ \phi_o(y^p) & \cdots & \phi_p(y^p)\end{pmatrix}$$

$$\alpha_\Phi=\begin{pmatrix}\alpha_0\\ \alpha_1\\ \vdots\\ \alpha_p\end{pmatrix} \text{ and } f(Y)=\begin{pmatrix}f(y^0)\\ f(y^1)\\ \vdots\\ f(y^p)\end{pmatrix}.$$

Obviously, the necessary and sufficient conditions which can make the above system to have a unique solution are that the matrix $M(\Phi,Y)$ has to be nonsingular.

Definition 2.1. [7] The $Y=\{y^0,y^1,\cdots,y^p\}\subset R^n$ is poised for polynomial interpolation in R^n if the corresponding matrix $M(\Phi,Y)$ is nonsingular for some basis Φ in p_n^d.

Definition 2.2. [7, 8] Given a set of interpolation $Y=\{y^0,y^1,\cdots,y^p\}\subset R^n$, a basis of $p_1=p+1$ polynomials $l_j(x)(j=0,\cdots,p)\in p_n^d$ is called a basis of Lagrange polynomials if

$$l_j(y^i)=\delta_{ij}=\begin{cases}1, i=j,\\ 0, i\neq j.\end{cases}$$

If Y is poised, then Lagrange polynomials exist and are unique. Moreover, they have a lot of useful properties. Particularly, we are interested in the crucial fact that, if $m(x)$ interpolates $f(x)$ at the points in Y, then for all x,

$$m(x)=\sum_{j=0}^{p} f(y^j)l_j(x). \tag{2}$$

It can also be shown that $\sum_{j=0}^{p} l_j(x)=1, \forall x\in R^n$.

Definition 2.3. [7] Let $\Lambda>0$ and a set $B\subset R^n$. Let $\bar{\phi}$ be the natural of monomials of p_n^d. A poised set $Y=\{y^0,y^1,\cdots,y^p\}$ is said to be Λ-poised in B if and only if for the basis of Lagrange polynomials associated with Y. One has that $\Lambda\geq \max_{0\leq i\leq p}\max_{x\in B}|l_i(x)|$.

Lemma 2.1. Give a $B(x,\Delta)\overset{\Delta}{=}\{y\in R^n \,\|\, \|y-x\|\leq\Delta\}$, a poised interpolation $Y\subset B(x,\Delta)$, and its associated basis of Lagrange polynomials $\{l_j(y)\}_{j=0}^{p}$, there exist constants $k_{ef}>0$ and $k_{eg}>0$ such that, for any interpolating polynomial $m(x)$ of degree one or higher of the form (2) and any given point $Y\subset B(x,\Delta)$, we have [9]

$$\|f(y)-m(y)\|\leq k_{ef}\sum_{j=0}^{p}\left\|y^i-y\right\|^2|l_j(y)| \text{ and}$$

$$\|\nabla f(y)-\nabla m(y)\|\leq k_{eg}\Lambda\Delta, \Lambda=\max_{j=0,\cdots,p}\max_{x\in B(x,\Delta)}|l_j(x)|.$$

Lemma 2.2. [9] Given a closed bounded domain B, any initial interpolation set $Y\subset B$, and a constant $\Lambda>1$. Consider the procedure: find $j\in\{0,\cdots,p\}$ and a point $x\in B$ such that $|l(x)|\geq\Lambda$, and replace y^j by x to obtain a new set Y. Then this procedure terminates after a finite number of iterations with a model which is Λ-poised in B.

2.2. Nonmonotone wedge trust region methods

In general, the model is quadratic in trust region framework, written as

$$m_k(x_k+s)=f(x_k)+g_k^T s+\frac{1}{2}s^T G_k s, \quad s.t.\|s\|\leq\Delta_k, \tag{3}$$

where $\nabla m_k(x_k+s)=G_k s+q_k, \nabla m_k(x_k)=g_k$, $\nabla m_k^2(x_k)=G_k$, and Δ_k is the trust-region radius at k-th iteration. G_k is a symmetric approximation to $\nabla^2 f(x_k)$. In a derivative-free case, $G_k\neq\nabla^2 f(x_k)$, $g_k\neq\nabla f(x_k)$. In addition, $m_k(y)$ satisfies the follow interpolation condition $m_k(y)=f(y), \forall y\in Y_k$.

The wedge trust region method was deliberately discussed by Marazzi [10] and introduced the main idea in Marazzi & Nocedal [11]. The wedge constraint is added to the trust region subproblem, and we have

$$\min_s m_k(x_k+s) \tag{4}$$

$$s.t.\|s\|\leq\Delta_k \tag{5}$$

$$s\notin W_k, \tag{6}$$

Where W_k is a set which contains the "taboo region" area, and the purpose is to avoid the new point falling into it. As for solving the wedge trust region subproblem (4)-(6), we usually first solve the standard trust region sub-problem without the wedge constraint and get a solution s_k^e at the k-th iteration. If s_k^e satisfies the wedge constraint, we set $s_k=s_k^e$ as the trail step. Otherwise, the wedge constraint is active. By rotating s_k, we find a vector satisfying the wedge constraint. Then we set the trail point $x_k^+=x_k+s_k$.

Recently, nonmonotone techniques are widely used in the trust region methods. Due to the high efficiency of nonmonotone techniques, many researchers are interested in working on the nonmonotone techniques for solving optimization problems. With the continuous development of nonmonotone technique, Li and Deng [12] introduced another nonmonotone method, and they define

$$\hat{r}_k=\frac{D_k-f(x_k^+)}{m_k(x_k)-m_k(x_k^+)} \quad (7)$$

$$D_k=\begin{cases} f(x_k) & k=1 \\ h_k D_{k-1}+(1-h_k)f(x_k) & k? 2 \end{cases} \quad (8)$$

And the $\eta_k=\frac{f(x_k)}{f(x_k)-D_{k-1}}$. This paper applied the D_k to the trust region with self-correcting geometry method.

Algorithm2.1. A new Nonmonotone Self-correcting Geometry Process

Step 0. Initialization

The current iterate x_k, the current interpolation set Y_k, the current trust region radius $\Delta_k>0$, the ratio $\hat{r}_k$, the switch value Δ_c for updating radius and the trial point x_k^+ are given, and the $\beta>0,0<\beta_2<1<\beta_1$ and $\eta\in[0,1)$ are also given.

Step 1. Successful Iteration

If $\hat{r}_k\ge\eta$, then define $x_{k+1}=x_k^+$, $\Delta_{k+1}=\beta_1\Delta_k$, and $Y_{k+1}=Y_k\cup\{x_k^+\}\backslash\{y^\tau\}$, $y^\tau\in\arg\max_{y\in Y_k}\|y-x_k^+\|$.

Step 2. Replace the Comparatively bad Interpolation Points

If $\hat{r}_k<\eta$, the $A_k\overset{\Delta}{=}\{y^j\in Y_k\,\|\,\|y^j-x_k\|>\beta\Delta_k\}$ is nonempty and the $B_k\overset{\Delta}{=}\{y^j\in Y_k\,\|\,|l_j(x_k^+)|>\Lambda\}$ is nonempty, then $x_{k+1}=x_k$ and $Y_{k+1}=Y_k\cup\{x_k^+\}\backslash\{y^\tau\}$,

Where τ is an index of any point in $A_k\cup B_k$, for instance, the $y^\tau\in\arg\max_{y^j\in A_k\cup B_k}\|y^j-x_k\|^2|l_j(x_k^+)|$. if $\Delta_k\ge\Delta_c$, set $\Delta_{k+1}=\beta_2\Delta_k$; else set $\Delta_{k+1}=\Delta_k$.

Step 3. Replace the Distant Interpolation Points

If $\hat{r}_k<\eta$, $A_k\ne\phi$ and $B_k=\phi$, then set $x_{k+1}=x_k$ and $Y_{k+1}=Y_k\cup\{x_k^+\}\backslash\{y^\tau\}$, where τ is an index of any point in A_k, for instance, $y^\tau\in\arg\max_{y^j\in A_k}\|y^j-x_k\|$. If $\Delta_k\ge\Delta_c$, set $\Delta_{k+1}=\beta_2\Delta_k$; else set $\Delta_{k+1}=\Delta_k$.

Step 4. Replace the Badly-poised Interpolation Points

If $\hat{r}_k<\eta$, $A_k=\phi$ and $B_k\ne\phi$, then set $x_{k+1}=x_k$ and $Y_{k+1}=Y_k\cup\{x_k^+\}\backslash\{y^\tau\}$, where τ is an index of any point in B_k, for instance, $y^\tau\in\arg\max_{y^j\in B_k}\|l_j(x_k^+)\|$. If $\Delta_k\ge\Delta_c$, set $\Delta_{k+1}=\beta_2\Delta_k$; else set $\Delta_{k+1}=\Delta_k$.

Step 5. Unchange the Interpolation Set

If $\hat{r}_k<\eta$, $A_k\cup B_k=\phi$, then set $x_{k+1}=x_k$ and $\Delta_{k+1}=\beta_2\Delta_k$.

Now we state our new algorithm as follows.

Algorithm 2.2. A New Nonmonotone Wedge Trust Region Method with Self-correcting Geometry

Step 0. Initialization

Given an initial point x_0, an radius $\Delta_0\in(0,+\infty)$, an initial tolerance ε_0, an initial interpolation set Y_0 such that Y_0 is poised and the initial interpolation model function $m_0(x)$ corresponding to Y_0. Constants $\eta\in[0,1),\mu\in(0,1),\theta>0,0<\beta_2<1<\beta_1$ and $\Lambda>1,k=0$. $v_0\ne x_k$.

Step 1. Criticality Test

Step1a: Set $i=0$ and $\hat{m}_i=m_k$.

Step1b: If $\|\nabla\hat{m}_i(x_k)\|<\varepsilon_i$, set $\varepsilon_{i+1}=\mu\|\nabla\hat{m}_i(x_k)\|$, compute a Λ-poised model $\hat{m}_{i+1}$ in $B(x_k,\varepsilon_{i+1})$, set $i:=i+1$, and return to step 1b.

Step 1c: Set $\hat{m}_i=m_k$, $\Delta_{k+1}=\theta\|\nabla m_k(x_k)\|$, and set $\upsilon_i=x_k$ if a new model has been computed.

Step 2. Compute the Replaced Point

Choose the point that is the farthest from the current iterate as the replaced point y^τ, i.e.,

$$y^\tau=\arg\max_{y\in Y_k}\|y-x_k\|.$$

Step 3. Solve the Wedge Trust Region Subproblem

Solve the wedge trust region subproblem (4)-(6) for getting s_k, and set the trial point $x_k^+\overset{\Delta}{=}x_k+s_k$.

Step 4. Update the Iteration and Interpolation Set

Compute $\rho_k=\frac{D_k-f(x_k^+)}{m_k(x_k)-m_k(x_k^+)}$ Use algorithm 2.1 to get Y_{k+1}, x_{k+1} and Δ_{k+1}.

Step 5. Update the Model and Lagrange Polynomials

If $Y_{k+1}\ne Y_k$, recompute the interpolation model $m_{k+1}(x)$ using the Lagrange polynomial for every associated with Y_{k+1}. Set $k=k+1$, go to Step 1.

3. Global Convergence

In this section, we discuss the global convergence of the new algorithm, and the convergence analysis is based on the convergence theory of basic trust region methods in [14], and it is an extension of the convergence analysis of algorithm WEDGE [11]. We first give some assumptions [13].

(A1) the objective function f is continuously differentiable in an open set Ω containing all iterates generated by the algorithm and its gradient ∇f is Lipschitz continuous in Ω with constant L.

(A2) there exists a constant k_{low} such that $f(x) \geq k_{low}$ for every $x \in \Omega$.

(A3) there exists a constant $k_H \geq L$ such that $G_k \leq k_H$ for every $k \geq 0$.

Lemma 3.1. (see Lemma 6.1.3 in [16] and [13]**)** At the k-th iteration, the solution of the wedge trust region subproblem satisfies the fraction of Cauchy decrease condition:

$$m_k(x_k) - m_k(x_k^+) \geq k_c \|g_k\| \min\left\{\Delta_k, \frac{\|g_k\|}{G_k}\right\}, \tag{9}$$

where $k \in (0,1)$ is a constant.

Lemma 3.2. (see [13]) Suppose that assumptions A1 and A3 hold, that the model is quadratic for all k sufficiently large and that there is a finite number of successful iterations. Then

$$\lim_{k\to\infty} \inf \|g_k\| = 0. \tag{10}$$

Lemma 3.3. (see Lemma 5.2 in [9]) Suppose that assumptions A1 and A3 hold and that m_k is a quadratic model. Then, for any constant $\Lambda > 1$, if the k-th iteration is unsuccessful, $A_k = \phi$ and

$$\Delta_k \leq \min\left[\frac{1}{k_H}, \frac{(1-\eta)k_c}{2k_{ef}(\beta+1)^2(p\Lambda+1)}\right]\|g_k\| \square k_\Lambda \|g_k\|,$$ then $B_k \neq \phi$.

Lemma 3.4. Suppose that assumptions A1 and A3 hold. Suppose also that, for some $k_o \geq 0$ and all $k \geq k_o$, the model is quadratic and

$$\|g_k\| \geq k_g \tag{11}$$

for some $k_g > 0$. The there exists a constant $k_\Delta > 0$ such that, for all $k \geq k_0$,

$$\Delta_k \geq k_\Delta. \tag{12}$$

Proof. Assume that, for some $k \geq 0$,

$$\Delta_k < \min(k_\Lambda k_g, \mu k_g, \Delta_c) \tag{13}$$

If the k-th iteration is successful, i.e., $\hat{r}_k \geq \eta$, then $\Delta_{k+1} \geq \Delta_k$. Otherwise, $\hat{r}_k < \eta$, then there are three cases that may occur.

The first case is when $A_k \neq \phi$ and $B_k \neq \phi$, step2 of algorithm 2.1 is executed. Observe that (13) ensure $\Delta_k < \Delta_c$, therefore $\Delta_{k+1} = \Delta_k$.

The second case is when $A_k \neq \phi$ and $B_k = \phi$. If $i > 0$, then (11) and (13) ensure that

$$\Delta_k < \mu \|g_{k_i}\| = \varepsilon_i, \tag{14}$$

Where k_i is the index of the last iteration before k where a new Λ-poised model has been recomputed in the criticality test. Therefore, step 3 of Algorithm 2.1 is executed. Together with $\Delta_k < \Delta_c$, we have $\Delta_{k+1} = \Delta_k$.

The third case is conducted when $A_k = \phi$. Under the condition (13), Lemma 3.3 can infer the $B_k \neq \phi$. Since (13) and (14) hold in this case, step 4 of Algorithm 2.1 is executed and $\Delta_{k+1} = \Delta_k$.

As the consequence, the trust region radius can be decreased only if $\Delta_k \geq \min(k_\Lambda k_g, \mu k_g, \Delta_c)$, and algorithm 2.1 implies (12) with

$$k_\Delta = \min\left[\Delta_0, \gamma_1 \min(k_\Lambda k_g, \mu k_g, \Delta_c)\right].$$

Lemma 3.5. Suppose that assumptions A1-A3 hold, the model is quadratic for all k sufficiently large and the number of successful iterations is infinite. Then the (10) also holds.

Proof. Assume that the lemma is not true, i.e., there exists some $k_g > 0$ such that (11) holds for all sufficiently large k. then we have from Lemma 3.4 that (12) holds for all k, including all successful iterations with k large enough. However from (9), we have that

$$\begin{aligned} D_k - f(x_k^+) &\geq \eta(m_k(x_k) - m_k(x_k^+)) \\ &\geq k_c \|g_k\| \min\left\{\Delta_k, \frac{\|g_k\|}{G_k}\right\} \\ &\geq k_c k_g \min\left\{\frac{k_g}{k_H}, k_\Delta\right\} \\ &\square\, k_d > 0 \end{aligned}.$$

Since, by assumption, there are infinitely successful iterations. We obtain that

$$\lim_{k\to\infty} f(x_k) = f(x_0) - \sum_{i=1, i\in S}^{\infty} k_d = -\infty,$$

which contradicts assumption A2, where S is a set of successful iterations. Therefore the lemma holds.

Lemma 3.6. [14, 9] Suppose that assumptions A1 and A3 hold, then

$$\left|f(x_k^+) - m_k(x_k^+)\right| \leq \|\nabla f(x_k) - g_k\| \Delta_k + k_H \Delta_k^2 \tag{15}$$

Lemma 3.7. Suppose that A1 and A3 hold, that $g_k \neq 0$, that

$$\|\nabla f(x_k) - g_k\| \le \frac{1}{2} k_c (1-\eta)\|g_k\|, \quad (16)$$

and that

$$\Delta_k \le \frac{k_c}{2k_H}(1-\eta)\|g_k\|. \quad (17)$$

Then iteration k is successful.

Proof (see theorem8.4.3, p. 286, in Conn, Gould, and Toint [15]) Observe first that A3, (17) and (9) imply that

$$m_k(x_k) - m_k(x_k^+) \ge k_c \|g_k\| \min\left[\frac{\|g_k\|}{k_H}, \Delta_k\right].$$

$$= k_c \|g_k\| \Delta_k$$

Hence, successively using (7), this last inequality, (15) (16) and (17), we obtain that

$$\hat{\eta}_k = \frac{D_k - f(x_k^+)}{m_k(x_k) - m_k(x_k^+)} \ge \frac{f(x_k) - f(x_k^+)}{m_k(x_k) - m_k(x_k^+)}$$

$$\Rightarrow \left|\hat{\eta}_k - 1\right| \le \left|\frac{f(x_k^+) - m_k(x_k^+)}{m_k(x_k) - m_k(x_k^+)}\right|$$

$$\le \left|\frac{\|\nabla f(x_k) - g_k\|}{k_c\|g_k\|}\right| + \left|\frac{k_H \Delta_k}{k_c\|g_k\|}\right| \le 1 - \eta$$

Thus we have that $\hat{\eta}_k \ge \eta$, and iteration k is successful.

Theorem 3.8. Suppose that assumptions A1-A3 hold and that the model is quadratic for all k sufficiently large, then $\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0$.

4. Conclusions

In this paper, we consider a nonmonotone wedge trust region method with self-correcting geometry for derivative-free optimization. We analyzed the properties of the new algorithm and proved the global convergence theory under some mild conditions. In the near future, we would like to design and check the effectiveness of different nonmonotone strategy for unconstrained optimization problems.

5. Acknowledgment

The work is supported by the Natural Science Foundation of Hebei Province (Grant No. A2014201003, F2014201 100).

References

[1] R. House and T. A. Jeeyes. "Direct search solution of numerical and statistical problem, Journal of the Association for Computing Machinery," vol. 8, 1961, pp. 212-219.

[2] J.A. Nelder and R. Mead, "A simplex method for function minimization," the Computer Journal, vol. 7, 1965, pp. 308-313.

[3] D. Winfield, "Function minimization by interpolation in a data table," J. Inst. Math. Appl, vol. 12, 1973, pp. 339-347.

[4] P. G. Ciarlet and P. A. Raviart, "General Lagrange and Hermite interpolation in R^n with applications to finite element methods," Archive for Rational Mechanics and Analysis, vol. 46, 1972, pp. 178-199.

[5] M. J. D. Powell, "On the Lagrange function of quadratic models that are defined by interpolation," Optimization Methods, and Software, vol. 16, 2001, pp. 289-309.

[6] R. Oeuvray, Trust-Region Methods Based on Radial Basis Function with Application to Biomedical Imaging, PhD thesis, Institut de Mathematiques, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland, 2005.

[7] A. R. Conn, K. Scheinberg and L. N. Vicente, Introduction to Derivative-Free Optimization, MPS-SIAM, Series on Optimization, SIAM, Philadephia, PA, USA, 2008.

[8] W. Sun, Q. K. Du and J. R. Chen Computational Methods, Science Press, Beijing, 2007.

[9] K. Scheinberg and Ph. L. TOINT, "Self-correcting geometry in model-based algorithms for derivative-free unconstrained optimization," SIAM Journal on Optimization, vol. 20, 2010, pp. 3512-3532.

[10] M. Marazzi, "Nonlinear Optimization with and without Derivatives," PhD thesis, Departmant of Industrial Engineering an Management Science, Northwestern University, Illinois, USA, 2001.

[11] M. Marazzi and J. Nocedal, "Wedge trust region methods for derivative free optimization," Mathematical Programming, vol. 91, 2002, pp. 289-305.

[12] Z. Li and N. Yang. "A New Class of Nonmonotone Trust Region Methods and its Convergence," Acta Mathematicae Applicatae Sinica, vol. 3, 1999, pp. 457-465.

[13] L. Zhang, W. Sun, Raimundo J. B. de Sampaio and J. Yuan, "A Wedge Trust Region Method with Self-correcting Geometry for Derivative-free Optimization," Numerical Algebra, Control and Optimization, vol. 5(2), 2015, pp. 169-184.

[14] A. R. Conn, N. I. M. Gould and Ph. L. Toint, Trust-Region Methods, SIAM, 2000.

[15] A.R. Conn, K. SCHEINBERG, and PH. L. Yoint, Trust-Region Method, MOS-SIAM. Sre, Optim. 1. SIAM, Philadephia, 2000.

[16] W. Sun and Y. Yuan. Optimization Theory and Methods, Nonlinear Programming, Springer Optimization and Its Applications, Volume 1. Springer, New York, 2006.