A Trust-Region Method Based on a Smoothing Penalty Function for Constrained Optimization Problems [1]

Bo Zhang, Zhensheng Yu and Jieqiong Guo

College of Science, University of Shanghai for Science and Technology, Shanghai, 200093, P.R.China

Abstract: By using a smoothing penalty function, we present a trust region method for equality constrained optimization problems, the penalty function is an approximation of the l_1 function penalty function and overcomes the nonsmoothness of the l_1 function. We design a sequence unconstrained optimization method which use the trust region method as the inner algorithm and obtain the global convergence of the proposed method.

Keywords: Equality constrained optimization, Trust-region method, Smoothing penalty function, Global convergence

1. Introduction

In this paper, we consider the equality constrained optimization problem:

$$\begin{aligned} \min \quad & f(x), \\ s.t. \quad & c(x) = 0, \end{aligned} \tag{1}$$

where $f(x) : R^n \to R$, $c(x) = (c_1(x), c_2(x), ..., c_m(x))^T, c_i(x) : R^n \to R^m, (i = 1, 2, ..., m)$, $(m \leq n)$ are assumed to be twice continuously differentiable.

Trust region method is one of the most well-known method for solving problem (1). Due to its strong convergence and robustness, trust region methods have been proved to be efficient for solving problem (1), and there are many research on trust region methods available for solving such problem, see, for examples, [1, 5, 7, 8, 10].

To obtain the next iteration point in trust region methods, one often use a penalty function as the merit function, the following l_1 penalty function is commonly used as the merit function.

$$\Psi(x, \alpha) = f(x) + \alpha \sum_{i=1}^{m} |c_i(x)|, \tag{2}$$

where $\alpha > 0$ is a penalty parameter.

The nonsmooth of function $\Psi(x, \alpha)$ often bring us some difficulty, for example, the Matatos effect will occur. This motivates the use of a smoothing penalty function.

[1] This work was supported by Innovation Program of Shanghai Municipal Education Commission(No.10YZ99) and Shanghai Leading Discipline Project (No. S30501)

Recently, [9] introduced a smoothing penalty function for nonlinear optimization, which is defined as follows:

$$\Phi(x,\alpha,\mu) = f(x) + \alpha \sum_{i=1}^{m} \frac{1}{\mu}(ln2 + ln(1 + cosh(\mu c_i(x)))) \tag{3}$$

where $\mu > 0$ is a smoothing parameter.

This function can be seen as an approximation of the function (2), the numerical examples in [9] showed that the sequence unconstrained penalty method based on this smoothing function behaviors well.

In this paper, we aim to use the smoothing penalty function (3) as the merit function and combine it with the trust region method to solve the problem (1). Under some reasonable assumptions, we establish the global convergence of the proposed algorithm.

This paper is organized as follows: In Section 2, we discussion some properties of the smoothing penalty function. In Section 3, propose our algorithm and establish its global convergence. The conclusion is given in Section 4.

2. Smoothing Penalty Function

In this section, we discuss some properties of the smoothing penalty function (3). For convenience, we set

$$\phi(x,\mu) = \sum_{i=1}^{m} \frac{1}{\mu}(ln2 + ln(1 + cosh(\mu c_i(x)))),$$

so

$$\Phi(x,\alpha,\mu) = f(x) + \alpha\phi(x,\mu).$$

In what follows, we discuss some properties of the function (3). The first property gives differentiability, first order and second order expressions.

Proposition 2.1 For any stationary $\mu \in R^+$, if $f(x)$, $c_i(x), (i = 1, 2, \cdots, m)$ are k times continuously differentiable, then $\Phi(x,\alpha,\mu)$ is continuously differentiable too. If $f(x)$, $c_i(x), (i = 1, 2, \cdots, m)$ is twice continuously differentiable, then

$$\nabla_x \Phi(x,\alpha,\mu) = \nabla_x f(x) + \alpha \sum_{i=1}^{m} \frac{\sinh(\mu c_i(x))}{1 + \cosh(\mu c_i(x))} \nabla_x c_i(x)$$

and

$$\nabla^2_{xx}\Phi(x,\alpha,\mu) = \nabla^2_{xx} f(x) + \alpha\mu \sum_{i=1}^{m} \frac{1}{1+\cosh(\mu c_i(x))} \nabla_x c_i(x) \nabla_x c_i(x)^T$$

$$+\alpha \sum_{i=1}^{m} \frac{\sinh(\mu c_i(x))}{1+\cosh(\mu c_i(x))} \nabla^2_{xx} c_i(x)$$

Similar to lemma2.2.1 in literature[12], we can get the following properties:

Proposition 2.2 If $f(x)$,$c_i(x), (i = 1, 2, \cdots, m)$ is convex, then $\Phi(x,\alpha,\mu)$ is convex too.

Proposition 2.3 $\Phi(x,\alpha,\mu) > \Psi(x,\alpha)$,$x \in R^n$.

Proposition 2.4 $\sup\limits_{x\in R^n} (\Phi(x,\alpha,\mu) - \Psi(x,\alpha)) \le \frac{m\alpha}{\mu} \ln 4$.

Proposition 2.5 $\lim\limits_{\mu\to\infty} \Phi(x,\alpha,\mu) = \Psi(x,\alpha)$.

Proposition 2.6 If $\mu_1 < \mu_2$, then $\Phi(x,\alpha,\mu_1) > \Phi(x,\alpha,\mu_2)$, $x \in R^n$.

3. Algorithm and Global convergence

In this section, we give the smoothing trust region method and the proof of global convergence. The algorithm consists of three parts: the first part builds the algorithm framework; the second part presents trust region inner iteration; the last part gives the main algorithm.

We first give the algorithm framework as follows:
Step 1 Given $x_0 \in R^n$, $\alpha_0 > 0, \mu_0 > 0, \varepsilon > 0, k = 0$; Further a fixed parameter $\eta_1 \gg 1, \eta_2 \gg 1, \eta_3 \gg 1$.
Step 2 If $\sum_{i=1}^{m} c_i(x(\alpha_k,\mu_k)) > \varepsilon$, turn to step3;
else return x_k.
Step 3 With fixed $\alpha = \alpha_k, \mu = \mu_k$, using the initial value x_j to compute

$$\min \Phi(x,\alpha,\mu) \tag{4}$$

and get the solution as x_{k+1}
Step 4 Update α_k or μ_k depending on the infeasibility of x_{k+1}.
If $\sum_{i=1}^{m} c_i(x(\alpha_k,\mu_k)) \le \eta_1/\mu_k$,then update μ_k:

$$\mu_{k+1} = \eta_2\mu_k, \alpha_{k+1} = \alpha_k.$$

Else update α_k:

$$\alpha_{k+1} = \eta_3\alpha_k, \mu_{k+1} = \mu_k, x_{k+1} = x_k.$$

Step5 $k = k + 1$, turn to step 2.

Now, let's consider the trust region method for solving the problem (4). The corresponding trust region subproblem for (4) is as follows:

$$\begin{aligned} \min \quad & m_k(d) = g_k^T d + \frac{1}{2} d^T B_k d, \\ s.t. \quad & \|d\| \le \Delta_k, \end{aligned} \tag{5}$$

where $g_k = \nabla_x \Phi(x,\alpha,\mu)$, $B_k = \nabla^2_{xx}\Phi(x,\alpha,\mu)$. Δ_k is the trust region radius.

The method for solving the subproblem was described as followings:

Algorithm 1(Newton-CG-Steihaug)[10]
Step0 Given $\varepsilon > 0, d_0 = 0, r_0 = g_k, p_0 = -r_0, j = 0$;
Step1 If $\|r_0\| < \varepsilon$, return $d = d_0$;
Step2 If $p_j^T B_k p_j \le 0$, then find τ such that
$d = d_j + \tau p_j$ minimizes $m_k(d)$ and satisfies $\|d\| = \Delta_k$;
Step3 Set $\alpha_j = r_j^T r_j/(p_j^T B_k p_j), d_{j+1} = d_j + \alpha_j p_j$;
Step4 If $\|d_{j+1}\| \ge \Delta_k$,
find $\tau \ge 0$ such that $d = d_j + \tau p_j$, and $\|d\| = \Delta_k$, return d;
Step5 Set $r_{j+1} = r_j + \alpha_j B_k p_j$;
Step6 If $r_{j+1} < \varepsilon\|r_0\|$, return $d = d_{j+1}$;

Step7 Set $\beta_{j+1} = r_{j+1}^T rj + 1/(r_j^T r_j), p_{j+1} = r_{j+1} + \beta_{j+1} p_j, j = j+1$,turn to step2.

The trust-region Newton-CG method has a number of attractive computational and theoretical properties. First, it is global convergent. Its first step along the direction $-\nabla\Phi(x_k, \alpha_k, \mu_k)$ identifies the Cauchy point for the subproblem and any subsequent CG iterates only serve to improve the model value. Second, it requires no matrix factorizations without worrying about fill-in during a direct factorization.

Main algorithm

Step0 Given a appropriate sequence $\alpha_0 > 0, \varepsilon > 0, 0 < \delta < 1, \lambda \geq 1,$
$\gamma \geq \lambda, \eta_1 \gg 1, \eta_2 \gg 1, \eta_3 \gg 1, \mu_0 > 0, k = 0;$
Step1 If $\sum\limits_{i=1}^{m} c_i(x(\alpha_k, \mu_k)) > \varepsilon$, turn to Step2;
else return x_k;
Step2 $i = 0$, set $\tilde{x}_0 = x_k$;
Step3 If $\|\nabla\Phi(\tilde{x_i}, \alpha_k, \mu)\| \leq \varepsilon$, turn to Step9;
Step4 Obtain d_i by solving $\min \Phi(x, \alpha, \mu)$ by Algorithm 1;
Step5 Compute $A_{red,i} = \Phi(\tilde{x}_i, \alpha_k, \mu_k) - \Phi(\tilde{x}_i + d_i.\alpha_k, \mu_k)$,
$P_{red,i} = m_k(0) - m_k(d_i)$,
$\rho_i = A_{red,i}/P_{red,i}$;
Step6 If $\rho_i \geq \delta$, then set $\Delta_i = \max\{\Delta_i, \gamma\|d_i\|\}$
else $\Delta_{i+1} \in [\|d_i\|, \lambda\|d_i\|]$;
Step7 If $\rho_i \geq \varepsilon$, then $\tilde{x}_{i+1} = \tilde{x}_i + d_i$,
else $\tilde{x}_{i+1} = \tilde{x}_i$;
Step8 $i = i + 1$, turn to step3;
Step9 $x_k = \tilde{x}_{i+1}$;
Step10 Update α_k or μ_k depending on the infeasibility of x_{k+1}.
If $\sum\limits_{i=1}^{m} c_i(x(\alpha_k, \mu_k)) \leq \eta_1/\mu_k$, then update μ_k:

$$\mu_{k+1} = \eta_2\mu_k, \alpha_{k+1} = \alpha_k;$$

Else update α_k:

$$\alpha_{k+1} = \eta_3\alpha_k, \mu_{k+1} = \mu_k, x_{k+1} = x_k;$$

Step11 $k = k + 1$, turn to Step1.

After limited iterations, we can get a $x(\alpha_k, \mu_k)$ meets the $\sum\limits_{i=1}^{m} c_i(x(\alpha_k, \mu_k)) < \varepsilon$, so $x(\alpha_k, \mu_k)$ is the optimal solution of problem (3). Then we can get the solution of original problem.

We denote the Lagrangian of (1) is given by

$$L(x, \lambda) = f(x) + c(x)^T\lambda$$

where $\lambda \in R^m$ is a lagrange multiplier.
We recall the second order conditions for (1) and the well-known result [15].

Theorem 1. *Let* $(x^\star, \lambda^\star)$ *fulfill*

$$\begin{aligned} \nabla_x L(x^\star, \lambda^\star) = 0, \\ c(x^\star) = 0, \\ d^T \nabla^2_{xx} L(x^\star, \lambda^\star) d > 0 \end{aligned} \tag{6}$$

where $d \in \{y : \nabla c(x^\star)^T y = 0\}$. *Then* $x^\star$ *is a strict local minimizer of* (1).

In what follows, we introduce two theorems to prove the global convergence of the algorithm.

The following preliminary results can be found in [9].

Theorem 2. *Let* $(x^\star, \lambda^\star, \mu^\star)$ *satisfy the second order conditions for a minimizer of problem* (1). *Then for* $\alpha > \bar{\alpha}$ *with*

$$\bar{\alpha} = \| (\mu^\star, \lambda^{\star)} \|_\infty$$

$x^\star$ *is a strict unconstrained local minimizer of* $\Psi(x, \alpha)$.

The result shows that Φ indeed is an exact penalty function for (1). A proof can be found in [13, 14].

Theorem 3. *Let* $(x^\star, \alpha^\star)$ *satisfy the second order conditions for a minimizer of problem* (1). *Then for* $\mu \to \infty$ *there exists a minimizer* $x(\mu)$ *of* $\Phi(x, \alpha, \mu)$ *and* $x(\mu) \to x^\star$ *as* $\mu \to \infty$.

This result shows that we can get a local minimizer $x(\mu)$ of problem (1) as increasing the smoothing factor.

Theorem 4. *Assume that* $\varepsilon > 0$ *in algorithm meets the conditions* $\phi(x, \mu) \leq \varepsilon$, *then the algorithm stops in finite steps.*
Proof. *Suppose this theorem is false, then there exists a sequence* $\{\alpha_k\}$, *and* $\lim_{k\to\infty} \alpha_k = \infty$,

$$\sum_{i=1}^{m} c_i(x(\alpha_k, \mu_k)) > \varepsilon \tag{7}$$

For every k *holds.*
On the other side, there exists a vector $x^\star \in R^n$,

$$s.t. \quad \sum_{i=1}^{m} c_i(x^\star) < \varepsilon \tag{8}$$

$$f(x^\star) + \alpha_k \sum_{i=1}^{m} c_i(x^\star) \geq f(x(\alpha_k, \mu_k)) + \alpha_k \sum_{i=1}^{m} c_i(x(\alpha_k, \mu_k))$$

$$\geq f(x(\alpha_o, \mu_0)) + \alpha_k \sum_{i=1}^{m} c_i(x(\alpha_k, \mu_k))$$

which means

$$\alpha_k \sum_{i=1}^{m} c_i(x^\star) - \alpha_k \sum_{i=1}^{m} c_i(x(\alpha_k, \mu_k)) \geq f(x(\alpha_o, \mu_0)) - f(x^\star)$$

Both sides divided by α_k,

$$\sum_{i=1}^{m} c_i(x^\star) - \sum_{i=1}^{m} c_i(x(\alpha_k, \mu_k)) \geq \frac{1}{\alpha_k}(f(x(\alpha_o, \mu_0)) - f(x^\star)) \to 0$$

So $\sum_{i=1}^{m} c_i(x^\star) - \sum_{i=1}^{m} c_i(x(\alpha_k, \mu_k)) \geq 0$.

But, from (7) and (8), we know $\sum_{i=1}^{m} c_i(x^\star) - \sum_{i=1}^{m} c_i(x(\alpha_k, \mu_k)) < 0$,
Therefore, the hypothesis is not established. The contradiction shows that the algorithm stops in finite steps.

Theorem 5. *If the algorithm stops in finite steps at* $x(\alpha_k, \mu_k)$, *then* $x(\alpha_k, \mu_k)$ *is the local minimum point of problem* $\min \Phi(x, \alpha, \mu)$.
Proof If algorithm stops in finite steps at $x(\alpha_k, \mu_k)$, *then* $x(\alpha_k, \mu_k)$*is the local minimum point of* $\Phi(x, \alpha_k, \mu_k)$.
Suppose this theorem is false, $x(\alpha_k, \mu_k)$*is the local minimum point of* $\Phi(x, \alpha_k, \mu_k)$, *satisfying:*

$$\phi(x(\alpha_k, \mu_k), \mu_k) = 0$$

Then there exist a sequence $\{x_k\}$, $x_k \to x(\alpha_k, \mu_k)$, $x_k \neq x(\alpha_k, \mu_k)$, *and* $f(x_k) < f(x(\alpha_k, \mu_k))$.
Since $\phi(x, \mu)$ *is continuous, so*

$$\phi(x_k, \mu) = 0$$

$$\Phi(x_k, \alpha_k, \mu_k) < \Phi(x(\alpha_k, \mu_k), \mu_k)$$

This contradicts to $x(\alpha_k, \mu_k)$ *is the local minimum point of* $\Phi(x, \alpha_k, \mu_k)$. *So* $x(\alpha_k, \mu_k)$ *is the local minimum point of problem* $\min \Phi(x, \alpha, \mu)$.

4. Conclusion

We have developed a penalty trust-region method to solve equality constrained problems. With this method, we can transfer the equality constrained problem into a simple, smoothing, unconstrained problem, and then use trust-region method to deal with it. The accuracy of the algorithm is also controlled by the smooth parameter and penalty parameter.

References

[1] M. J. D. Powell, On the global convergence of trust region algorithms for unconstrained optimization, Math.Prog., 1984, 29: 297-303.
[2] A. Vardi, A trust region algorithm for equality constrained minimization: convergence properties and implementation, SIAM J.Numer.Anal., 1985, 22: 575-591.
[3] G. D. Pillo and L. Grippo, An exact penalty function method with global convergence properties for nonlinear programming problem, Math. Prog, 1986, 36: 1-18.

[4] M. J. D. Powell, Y. X. Yuan, A trust region algorithms for equality constrained optimization, Math. Programming. 1991, 49: 189-211.

[5] Y. X. Yuan, On the convergence of trust region method, Math. Number. Sinica., 1994,16: 334-336.

[6] Janesch, S.M.H., Santos,L.T., Exact penalty methods with constrained subproblems. Investigacion Operativa, 1997.

[7] J.L. Zhang, Trust Region Algorithm for Nonlinear Optimization.Ph.D thesis, Institute of Applied Mathematics, Chinese Academy of Science,2001. Andrew.R, ConnTrust Region Methods, CGT Productions, 1999.

[8] Y.H.Dai, A new family of trust region algorithms for unconstrained optimization, Journal of computational mathematics, 2003,Vol.21,No.2: 221-228.

[9] M.Herty, A.Klar, A.K.Singh, P.Spellucci, Smoothed Penalty Algorithms for Optimization for Nonlinear Models. Comput Optim Appl, 2007, 37:157-176.

[10] Y.Feng, H.W.Liu, S.S.Zhou, S.Y.Liu, A Smoothing Trust-region New-CG Method for minimax problem. Applied Mathematics and Computation, 2008, 199:581-589.

[11] X.F.YU, D.G.Pu, S.H.Gui, A New Penalty Trust-Region Method for Constrained Optimization Problems. Journal of Shang Hai Second Polytechnic University, 2009, 26:280-285.

[12] Yi Qin, Smoothing function and its application in the problems of optimization and complementarity, Master thesis, University of Shanghai for Science and technology, 2009.

[13] J.V.Burke, An exact penalization viewpoint of constrained optimization.SIAM J. Control Optim.1991,29:968-971.

[14] S.-P.Han, O.L.Mangasarian, O.l., Exact penalty functions in nonlinear programming. Math.Program. 1979, 17:251-256.

[15] J.F.Bonnans, J.C.Gilbert, C.Lemarechal, C.A.Sagastizabal, Numerical Optimization. Springer, Berlin,1997.