On the Development of a Trust Region Interior-Point Method for Large Scale Nonlinear Programs*

Miguel Argáez
University of Texas-El Paso
Dept. of Mathematical Sciences
El Paso, Texas 79968
mar@math.utep.edu

Leticia Velázquez
University of Texas-El Paso
Dept. of Mathematical Sciences
El Paso, Texas 79968
leti@math.utep.edu

Cristina Villalobos
University of Texas-Pan American
Dept. of Mathematics
Edinburg, Texas 78539
mcvilla@panam.edu

ABSTRACT

The focus of this paper is to present a new methodology for solving general nonlinear programs. We propose the use of interior-point methodology, trust-region globalization strategies, and conjugate gradient method to find a solution to large scale problems.

Categories and Subject Descriptors

G.1.6 [**Numerical Analysis**]: Optimization—*nonlinear programming*

General Terms

Algorithms, Theory

Keywords

Interior-point method, Newton's method, trust region method, and conjugate gradient

1. INTRODUCTION

We study the general nonlinear program in the form

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0 \\ & x \geq 0, \end{array} \tag{1}$$

where $h(x) = (h_1(x), \ldots, h_m(x))^T$ and $f, h_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$ $(m \leq n)$ are twice continuously differentiable functions. The Lagrangian function associated with problem (1) is

$$\ell(x, y, z) = f(x) + h(x)^T y - x^T z,$$

where $y \in \mathbb{R}^m$ and $z \geq 0 \in \mathbb{R}^n$ are the Lagrange multipliers associated with the equality and inequality constraints, respectively.
For $\mu > 0$, the perturbed Karush-Kuhn-Tucker (KKT) conditions for problem (1) are

$$F_\mu(x, y, z) \equiv \begin{pmatrix} \nabla f(x) + \nabla h(x) y - z \\ h(x) \\ XZe - \mu e \end{pmatrix} = 0, \tag{2}$$

$$(x, z) > 0,$$

where $X = \operatorname{diag}(x)$, $Z = \operatorname{diag}(z)$, and $e = (1, \ldots, 1)^T \in \mathbb{R}^n$. The Newton's method applied to (2) leads to the following nonsymmetric and indefinite linear system

$$\begin{pmatrix} \nabla_x^2 \ell & \nabla h(x) & -I \\ \nabla h(x)^T & 0 & 0 \\ Z & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = - \begin{pmatrix} \nabla_x \ell \\ h(x) \\ e_c \end{pmatrix} \tag{3}$$

where $e_c = XZe - \mu e$, $\nabla_x \ell = \nabla f(x) + \nabla h(x) y - z$, and $\nabla_x^2 \ell = \nabla^2 f(x) + \sum_{i=1}^m \nabla^2 h_i(x) y_i$.
The interest in large scale applications has motivated the use of inexact Newton steps for solving (3). Therefore we propose to decouple this system in such a way that we can take advantage of the structure of the problem, and allow implementations for solving large scale problems.

2. QUADRATIC SUBPROBLEM

We derive a quadratic subproblem associated with the perturbed KKT conditions which forms the central framework for solving problem (1). Once Δx is known, from the third block of equations of (3), we have

$$\Delta z = -X^{-1}(e_c + Z \Delta x).$$

Now, substituting Δz into system (3) we obtain a smaller system of equations, known as the saddle point problem,

$$\begin{pmatrix} Q & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} c \\ h(x) \end{pmatrix} \tag{4}$$

where $Q = \nabla_x^2 \ell + X^{-1} Z$, $A = \nabla h(x)^T$, and $c = \nabla_x \ell + X^{-1} e_c$. One advantage of this formulation is that we have improved the chances of Q being positive definite on the nullspace of A, especially far away from a solution of the problem, even though $\nabla_x^2 \ell$ may not be. This occurs because the term

*The first two authors were supported in part by the U.S. Department of the Army Grant DAAD19-01-1-0741, and the third author by DAAD19-02-1-0008.

TAPIA'03, October 15–18, 2003, Atlanta, Georgia, USA.

$X^{-1}Z$ is a positive diagonal matrix. It is well known that if Q is positive definite and A is full rank, then Δx obtained from (4) is the unique global minimizer of the following quadratic subproblem:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\Delta x^T Q \Delta x + c^T \Delta x \\ \text{subject to} & A\Delta x + h(x) = 0 \end{array} \tag{5}$$

with Q, A, and c defined as in (4).
However, it is possible that Q may not be positive definite on the nullspace of A and/or A is not full rank, and therefore it is not possible to find an Δx solution of (4). Usually this is the case far away of an optimal solution. So, the purpose is to present a global strategy that yields to find an approximate solution, and that the iterates coincides with the Newton step associated to the problem close to an optimal solution.

3. TRUST REGION SUBPROBLEM

In line with this purpose, we introduce the trust region subproblem as a globalization strategy. In our case, we want to improve even further the chances that the matrix Q be positive definite on the nullspace of A and to allow the use of the conjugate gradient algorithm for obtaining an approximate solution of Δx. Therefore we propose to obtain a solution or an approximate solution of (4) using a trust region globalization strategy. The subproblem is given by

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\Delta x^T Q \Delta x + c^T \Delta x \\ \text{subject to} & A\Delta x + h(x) = 0 \\ & \|\Delta x\| \leq \Delta, \end{array} \tag{6}$$

where $\Delta > 0$ is the trust region radius.

Observation 1. *A solution or an approximate solution Δx of (6) can be expressed as a direct sum of one element in the row space of A, $\Delta x_p \in \mathcal{R}(A^T)$, and the other one in the nullspace of A, $\Delta x_h \in \mathcal{N}(A)$, because A is a linear operator from $\mathbb{R}^n$ to $\mathbb{R}^m$. That is*

$$\Delta x = \Delta x_p + \Delta x_h, \tag{7}$$

where Δx_p and Δx_h are perpendicular. We call Δx_p and Δx_h the particular and homogeneous solutions of (6), respectively. (See [4, 3])

To find these elements, we propose solving the following two subproblems.

3.1 Particular Solution

The particular solution Δx_p is given by the following linear least squares subproblem

$$\begin{array}{ll} \text{minimize} & \|A\Delta x_p + h(x)\| \\ \text{subject to} & \|\Delta x_p\| \leq \tau\Delta \end{array} \tag{8}$$

where $\Delta x_p \in \mathcal{R}(A^T)$, and $\tau \in (0, 1)$.

3.2 Homogeneous Solution

Once Δx_p is known and assuming $A\Delta x_p + h(x) = 0$, then from (6) and (7) we obtain, after some algebraic operations, the following subproblem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\Delta x_h^T Q \Delta x_h + (Q\Delta x_p + c)^T \Delta x_h \\ \text{subject to} & A\Delta x_h = 0 \\ & |\Delta x_h\| \leq \sqrt{\Delta^2 - \|\Delta x_p\|^2}. \end{array} \tag{9}$$

Now, let $\Delta x_h = Pw$ for some $w \in \mathbb{R}^n$, where P is the projection onto $\mathcal{N}(A)$. Then making the substitution of Δx_h in (9), we obtain the following unconstrained trust region subproblem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}w^T(PQP)w + (Q\Delta x_p + c)^T Pw \\ \text{subject to} & \|Pw\| \leq \sqrt{\Delta^2 - \|\Delta x_p\|^2}. \end{array} \tag{10}$$

We propose solving (10) using the conjugate gradient method and Steihaug's termination test.
After obtaining Δx_p from (8) and w from (10), we define an inexact Newton direction $(\Delta x, \Delta z)$ by

$$\Delta x = \Delta x_p + Pw \quad \text{and} \quad \Delta z = -X^{-1}(e_c + Z\Delta x). \tag{11}$$

4. GLOBALIZATION STRATEGY

We follow the same globalization philosophy presented by [1, 2] which consists in treating the variable y as a parameter. Then, only the variables x and z are taken into account in our globalization strategy.
The direction $(\Delta x, \Delta z)$ given by (11) is accepted if it allows a sufficient decrease for the merit function introduced in [1, 2],

$$M_\mu(x, z; y, \rho) = \ell(x, y, z) + \rho\,(\frac{1}{2}\|h(x)\|^2 + x^T z - \mu \sum_{i=1}^{n} \ln(x_i z_i))$$

for ρ sufficiently large.
The fundamental idea of our globalization strategy is to apply a trust region and inexact Newton's method to the perturbed KKT conditions (2) for a fixed μ until we arrive to a specified neighborhood that measures nearness of a central region previously defined. We consider the notion of quasicentral path, introduced in [1, 2], as a central region to follow for obtaining an optimal solution of problem (1). The quasicentral path, parameterized by $\mu > 0$, is defined as t he collection of points $(x, z) \in \mathbb{R}^{n+n}$ satisfying

$$\begin{pmatrix} h(x) \\ XZe - \mu e \end{pmatrix} = 0, \qquad (x, z) > 0.$$

For a fixed $\mu > 0$, we say that a point $(x, z) > 0 \in \mathbb{R}^{n+n}$ is close enough to the quasicentral path if it satisfies the following inequality:

$$\|h(x)\|^2 + \|W(XZe - \mu e)\|^2 \leq \gamma\mu$$

where $W = (XZ)^{-1/2}$ and $\gamma \in (0, 1)$.

5. CONCLUSIONS

We have described a new methodology for solving large scale nonlinear programs using interior-point methods combined with a trust region strategy and conjugate gradient method. Our future work involves an implementation of this methodology on a set of large scale problems.

6. ACKNOWLEDGMENTS

The authors thank the referees for their valuable comments.

7. REFERENCES

[1] M. Argáez. *Exact and Inexact Newton Linesearch Interior-Point Algorithms for Nonlinear Programming Problems*. PhD thesis, Rice University, Houston, TX, 1997.

[2] M. Argáez and R. A. Tapia. On the global convergence of a modified augmented lagrangian linesearch interior-point newton method for nonlinear programming. *J. Optim. Theor. Appl.*, 114(1):1–25, August 2002.

[3] M. Lalee, J. Nocedal, and T. Platenga. On the implementation of an algorithm for large-scale equality constrained optimization. *SIAM J. Optim.*, 8(3):682–706, 1998.

[4] E. O. Omojokun. *Trust Region Algorithms for Optimization with Nonlinear Equality and Inequality Constraints*. PhD thesis, University of Colorado, Boulder, CO, 1989.