# Combining Trust Region and Affine Scaling for Linearly Constrained Nonconvex Minimization

THOMAS F. COLEMAN AND YUYING LI [1]
COMPUTER SCIENCE DEPARTMENT, CORNELL UNIVERSITY

**Abstract.** An interior point method is proposed for a general nonlinear (nonconvex) minimization with linear inequality constraints. This method is a combination of the trust region idea for nonlinearity and affine scaling technique for constraints. Using this method, the original objective function is monotonically decreased.

In the proposed approach, a Newton step is derived directly from the complementarity conditions. A trust region subproblem is formed which yields an approximate Newton step as its solution asymptotically. The objective function of the trust region subproblem is the quadratic approximation to the original objective function plus an augmented quadratic convex term. Similar to an augmented Lagrangian function, this augmentation adds positive curvature in the range space of the constraint normals.

The global convergence is achieved by possibly using trust regions with different shapes. A reflection technique, which accelerates convergence, is described. Explicit sufficient decrease conditions are proposed.

Computational results of a two-dimensional trust region implementation are reported for large-scale problems. Preliminary experiments suggest that this method can be effective; a relatively small number of function evaluations are required for some medium and large test problems.

**Key words.** trust region, Dikin-affine scaling, an interior point method, Newton, reflection

**1. Introduction.** Interior point methods have proven to be an efficient approach for solving large scale linear and convex programming problems: see [24] for a comprehensive bibliography. An appealing property of these methods is that a small number of iterations is typically required to obtain an accurate solution for a large problem. This property, if achievable, is even more attractive for a *nonlinear* programming problem since function evaluation can be a dominant and costly computation.

Interior point methods share a common characteristic: they avoid approaching the boundary prematurely. The majority of interior point methods can be interpreted as following the central path to optimality, e.g., [21, 31]. The exception is the classical affine scaling algorithm [15, 33]. An affine scaling method uses a diagonal scaling technique to compute a damped step which ultimately leads to convergence to a solution. Despite the absence of a polynomial convergence property, an affine scaling method is the only type of interior point method which monotonically decreases the original objective function; it stands out for its simplicity and typically good computational performance, e.g., [1, 5, 33].

There has been great interest in generalizing interior point methods to nonlinear (nonconvex) programming problems, e.g., [16, 36, 9, 12, 14, 35, 25, 19, 4]. However, this has proven to be a difficult and challenging task.

For a nonconvex problem, a minimization method is typically able to compute

[1] Research partially supported by the Applied Mathematical Sciences Research Program (KC-04-02) of the Office of Energy Research of the U.S. Department of Energy under grant DE-FG02-90ER25013.A000, and in part by NSF through grant DMS-9505155, ONR through grant N00014-96-1-0050, and by the Cornell Theory Center which receives major funding from the National Science Foundation and IBM Corporation, with additional support from New York State and members of its Corporate Research Institute.

This article is submitted to the proceeding of the International Nonlinear Programming Conference, Beijing, September 2-5, 1996.

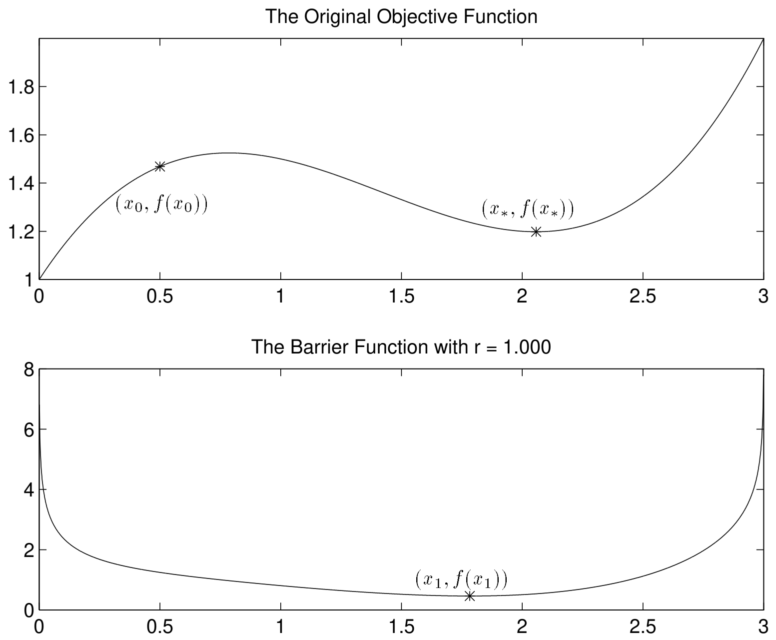


FIG. 1.1. *Loss of the Information from the Starting Point:* $\min_{0 \le x \le 3} f(x)$

a local minimizer at best. This does not prevent its usefulness since an initial point is often provided in applications and the local minimizer in the neighborhood of this initial point can be sufficient. The majority of interior point methods are based on the view that it is worthwhile to sacrifice decrease of the original objective function in order to gain centrality, e.g., [21, 31]. This view may not be reasonable for nonconvex problems; achieving centrality can cause loss of information provided by the initial point and possibly cause convergence to a local minimizer with a higher objective function value than that of the initial point. FIG. 1.1 illustrates this effect: $x_1$ is the minimizer of the barrier function for the original bound constrained problem with the barrier parameter $r = 1$. Using a barrier function method with the initial $x_0$ and $r = 1$, the iterates will converge to the local minimizer $x_*$. Using an algorithm which monotonically decreases the objective function with the same initial $x_0$, the iterates will converge to the local minimizer $x = 0$.

Monotonicity offers a natural connection to the original problem. Monotonicity and the simplicity of an affine scaling method makes it particularly appealing and suitable for constrained nonconvex minimization. Based on this philosophy, Newton type affine scaling algorithms have been considered for various structured linear and nonlinear programming problems [8, 7, 26, 9, 27, 25, 2, 14]. These methods represent a generalization of the affine scaling methods to piecewise linear and nonlinear minimization. Such algorithms have fast local convergence and typically solve a large problem in a small number of iterations. For minimization with simple bounds, Newton steps are derived from the first order necessary conditions directly and globalization is successfully achieved using a trust region approach [9]. For $l_p$ problems and minimization of the sums of Euclidean distances, the resulting algorithms become a natural enhancement of the classical IRLS and Weiszfeld algorithms [26, 25].

We consider the problem of minimizing a general nonlinear (nonconvex) function subject to linear inequality constraints,

$$\begin{array}{lrl} & & \min_{x \in \Re^n} f(x) \\ (1.1) & \text{subject to} & Ax \geq b, \end{array}$$

where $A^T = [a_1, \cdots, a_m] \in \Re^{n \times m}$. Let $\mathcal{F} \stackrel{\text{def}}{=} \{x : Ax \geq b\}$. The strictly feasible region $int(\mathcal{F}) \stackrel{\text{def}}{=} \{x : Ax > b\}$ is assumed to be non-empty and a strictly feasible initial point $x_0 \in int(\mathcal{F})$ is given. Moreover, $f(x)$ is twice continuously differentiable in $\mathcal{F}$.

In this article, a trust region and affine scaling interior point method (TRAM) is proposed for a general nonlinear minimization problem (1.1) with linear inequality constraints. Specifically, TRAM exhibits the following characteristics:

- the iterates $\{x_k\}$ are in the interior of the feasible region and the dual multiplier approximations are only feasible in the limit;
- the Newton steps are derived from the complementarity conditions;
- the original objective function $f(x)$ is decreased monotonically;
- the trust region subproblems yield approximate Newton steps asymptotically; these trust region subproblems can be solved using existing trust region techniques;
- a reflection technique is used to accelerate convergence.

The presentation of the paper is organized as follows. In §2, an affine scaling Newton process is firstly derived from the complementarity conditions of the problem (1.1). Section §3 includes globalization of this Newton process within a trust region context. A trust region subproblem which yields approximate Newton steps asymptotically is motivated in §3.1. The shapes of the trust regions necessary for convergence are analyzed in §3.2. The proposed TRAM algorithm is outlined §3.3. A two-dimensional trust region generalization is included in §3.4. Explicit decrease conditions for convergence to a local minimizer are described in §3.6. A useful reflection technique is described in §3.5. Preliminary computational results are presented in §4. Concluding remarks and possible future research are included in §5.

In a separate paper [11], convergence properties of the proposed method are presented.

**2. An Affine Scaling Newton Approach.** The first order optimality conditions of a constrained minimization problem consist of the complementarity conditions, primal feasibility, and dual feasibility. The following is a classical characterization of a local minimizer of the linearly constrained problem (1.1), e.g., see Fletcher [18].

**First Order Necessary Conditions**

*If $x_*$ is a local minimizer of (1.1), then there exist Lagrangian multipliers $\lambda_*$ such that $x_*, \lambda_*$ satisfy:*

$$\nabla f_* - A^T \lambda_* = 0, \tag{2.1}$$

$$Ax_* - b \geq 0, \tag{2.2}$$

$$\lambda_* \geq 0, \tag{2.3}$$

$$(\lambda_*)_i (a_i^T x_* - b_i) = 0, \qquad 1 \leq i \leq m. \tag{2.4}$$

**Second Order Sufficiency Conditions**

*If at $x_*$ there exist $\lambda_*$ which satisfy (2.1)-(2.4), $d^T\nabla^2 f_* d > 0$ for all $d \neq 0$, $a_i^T d \geq 0$ for all $a_i^T x_* - b_i = 0$, and $a_i^T d = 0$ for all $(\lambda_*)_i > 0$, then $x_*$ is a strict local minimizer of (1.1).*

Condition (2.2) is referred to as primal feasibility, (2.3) is dual feasibility. Conditions (2.1) and (2.4) are called the complementarity conditions and can be equivalently expressed as an $(m+n)$-by-$(m+n)$ nonlinear system of equations

$$\text{diag}(Ax-b)\lambda = 0 \qquad \text{and} \qquad A^T\lambda - \nabla f = 0. \tag{2.5}$$

If there is no index $i$ such that $(\lambda_*)_i = a_i^T x_* - b_i = 0$, then *strict complementarity* is said to hold.

Complementarity has played a central role in the successful primal-dual interior point methods for linear and convex programming problems. The primal dual interior point method [23], proposed by Kojima, Mizuno and Yoshise in 1987, maintains both the primal-dual strict feasibility and can be considered as a damped Newton process on the perturbed complementarity conditions. For (1.1), this perturbed complementarity conditions correspond to

$$\text{diag}(Ax-b)\lambda = \rho e \qquad \text{and} \qquad A^T\lambda - \nabla f(x) = 0.$$

The parameter $\rho > 0$ is decreased to zero as a solution is approached.

Fiacco and McCormick [17] first considered this perturbed KKT condition for an inequality constrained nonlinear programming problem. If $f(x)$ is convex, the solution to this nonlinear system, as the parameter $\rho > 0$ varies, defines a central path in the strictly feasible region. For a convex programming problem, primal and dual feasibility together with complementarity sufficiently characterize a solution.

Extension of a primal and dual method to a general nonconvex problem has proven to be difficult and challenging [16, 14, 19, 4]. The first order necessary conditions is no longer sufficient in characterizing a local minimizer for a *nonconvex* minimization problem. For example, a 2-norm penalty function has been used to measure satisfaction of the first order necessary conditions for a nonconvex problem in [16]. This method does not necessarily converge to a local minimizer of the original minimization problem since the second order necessary conditions may fail to hold.

We believe that monotonic decrease of the original objective function, if achievable, is an important property of a minimization algorithm for nonconvex problems. This property enables a direct connection to the original minimization problem. It can make use of information such as an available starting point. We demonstrate that a Newton process based on the complementarity conditions (2.5) can be used in combination with the trust region idea to achieve convergence by monotonically decreasing the original objective function.

Let $F(x,\lambda) = 0$ denote the complementarity conditions (2.5), i.e.,

$$F(x,\lambda) \stackrel{\text{def}}{=} \begin{bmatrix} \text{diag}(Ax-b)\lambda \\ \nabla f(x) - A^T\lambda \end{bmatrix}.$$

The Newton step for $F(x,\lambda) = 0$ at $(x_k, \lambda_k)$ is

$$\begin{bmatrix} \text{diag}(\lambda_k)A & \text{diag}(Ax_k - b) \\ \nabla^2 f_k & -A^T \end{bmatrix} \begin{bmatrix} d_k^N \\ d_k^\lambda \end{bmatrix} = -\begin{bmatrix} \text{diag}(Ax_k-b)\lambda_k \\ \nabla f_k - A^T\lambda_k \end{bmatrix}. \tag{2.6}$$

If $x_k$ is strictly feasible, then $\mathrm{diag}(Ax_k - b)$ is positive definite. An alternative way of computing the Newton step (2.6) is to solve a $n$-by-$n$ symmetric linear system,

$$(A^T(\mathrm{diag}(\lambda_k)\mathrm{diag}(Ax_k - b)^{-1})A + \nabla^2 f_k)d_k^N = -\nabla f_k, \tag{2.7}$$

with the dual multipliers $\lambda_{k+1}^N = \lambda_k^N + d_k^\lambda$ updated as

$$\lambda_{k+1}^N = -\mathrm{diag}(\lambda_k)\mathrm{diag}(Ax_k - b)^{-1}Ad_k^N.$$

The Newton system (2.6) has appealing properties in a neighborhood of a local minimizer $x_*$ satisfying the second order sufficiency conditions. This is described in Theorem 2.1. Its proof is straightforward and can be found in [11]. Let $(v; w)$ denote the column vector in $\Re^{n+m}$ formed from the vectors $v \in \Re^n$ and $w \in \Re^m$.

THEOREM 2.1. *Assume that $f(x) : \mathcal{F} \to \Re$ is twice continuously differentiable and the second order sufficiency conditions of (1.1) are satisfied at $(x_*; \lambda_*)$. Assume further that strict complementarity holds at $x_*$ and $\{a_i : (\lambda_*)_i > 0\}$ are linearly independent. Then*

*(a) The Jacobian matrix $\nabla F(x_*, \lambda_*)$ is nonsingular;*

*(b) The symmetric matrix $(A^T(diag(\lambda)diag(Ax - b)^{-1})A + \nabla^2 f(x))$ is positive definite when $x \in int(\mathcal{F})$ and $(x; \lambda)$ is sufficiently close to $(x_*; \lambda_*)$.*

**3. A Trust Region Globalization.** An immediate consequence of Theorem 2.1 is that, if $x_k \in int(\mathcal{F})$ is in a sufficiently small neighborhood of a solution, the Newton step $d_k^N$ of (2.5) is descent for the objective function $f(x)$. This suggests that, using this Newton process, local convergence can be achieved from the interior of the feasible region with a monotonic decrease of the objective function.

Maintaining strict feasibility can be done with a simple backtracking technique. This technique can be used to avoid approaching a boundary prematurely, which is important for the success of an interior point method.

Assume that $x_k + d_k$ brings a sufficient decrease of the objective function but possibly violates strict feasibility. Then a damped step $\alpha_k d_k$ can be taken where

$$\alpha_k \stackrel{\text{def}}{=} \theta_k \alpha_k^*, \quad 0 < \theta_0 \le \theta_k < 1, \quad \alpha_k^* \stackrel{\text{def}}{=} \min(1, \beta_k) \tag{3.1}$$

and

$$\beta_k \stackrel{\text{def}}{=} \min\{-\frac{a_i^T x_k - b_i}{a_i^T d_k}, \quad -\frac{a_i^T x_k - b_i}{a_i^T d_k} > 0\}.$$

The parameter $\theta_k$ determines the damped step $\alpha_k d_k$. Its choice is important for the convergence properties of an affine scaling method for linear programming [28, 32].

It is also possible to maintain strict dual feasibility, $\lambda > 0$. However, since global convergence is achieved by monotonic decrease of $f(x)$, and maintaining dual feasibility can lead to a smaller stepsize, we allow violation of dual feasibility. Hence only primal strict feasibility, $Ax > b$, is maintained in the proposed method; however, the dual multipliers play an important role in determining the next step, particularly asymptotically.

The local Newton process needs to be globalized. Next we show that a Newton step can be globalized using a trust region approach. The main trust region subproblem is motivated in §3.1. A trust region subproblem, which is useful occasionally for quickly departing from a nearly binding constraint, is described in §3.2. The model algorithm is summarized in §3.3. A two dimensional trust region subproblem is considered in §3.4. A reflection line search is proposed in §3.5 to accelerate convergence.

**3.1. A Trust Region Subproblem.** Assume that $x_k$ is strictly feasible. Let $\psi_k(d)$ denote the quadratic approximation for $f(x)$ at $x_k$, i.e.,

$$\psi_k(d) \stackrel{\text{def}}{=} \nabla f_k^T d + \frac{1}{2} d^T \nabla^2 f_k d. \tag{3.2}$$

Further, assume that the quadratic approximation $\psi_k(d)$ is a "good" approximation to $f(x_k + d) - f(x_k)$ within the trust region $\|d\|_2 \leq \Delta_k$. The goal is to produce a step $s_k$, $x_k + s_k \in int(\mathcal{F})$, such that the quadratic approximation $\psi_k(d)$ (and hence $f(x)$) achieves a good decrease. The Newton step (2.6) locally produces a good decrease for $f(x)$ but it may not be descent globally. Furthermore, a damped step of the trust region solution $\min_{\|d\|_2 \leq \Delta_k} \psi_k(d)$ may fail to produce a sufficient decrease due to feasibility restriction. We want to formulate a trust region subproblem with the following properties:

- it is closely connected to the Newton step $d_k^N$ for fast local convergence;
- it produces a strictly feasible step which yields a sufficient decrease.

Let $C_k$ and $D_k$ denote diagonal matrices:

$$C_k \stackrel{\text{def}}{=} \text{diag}(|\lambda_k|), \qquad D_k \stackrel{\text{def}}{=} \text{diag}(Ax_k - b).$$

Replacing $\text{diag}(\lambda_k)$ by $C_k$ in (2.6), the *modified* Newton step $d_k$,

$$\begin{bmatrix} C_k A & D_k \\ \nabla^2 f_k & -A^T \end{bmatrix} \begin{bmatrix} d_k \\ \lambda_{k+1} \end{bmatrix} = - \begin{bmatrix} 0 \\ \nabla f_k \end{bmatrix}, \tag{3.3}$$

approximates the Newton step sufficiently accurately, hence retains fast convergence. Moreover, this modified Newton step $d_k$ is a minimizer of the augmented quadratic $\psi_k(d) + \frac{1}{2} d^T A^T D_k^{-1} C_k A d$, which can be considered a quadratic convex regularization of the constrained problem (1.1) at $x_k$. The quadratic approximation $\psi_k(d)$ is sufficiently decreased if its convex regularization $\psi_k(d) + \frac{1}{2} d^T A^T D_k^{-1} C_k A d$ is sufficiently decreased. Using the augmented quadratic as the objective function, a trust region subproblem consistent with the approximate Newton step $d_k$ is

$$\begin{aligned} &\min_{d \in \Re^n} \psi_k(d) + \frac{1}{2} d^T A^T D_k^{-1} C_k A d \\ &\text{subject to} \quad \|(d; D_k^{-\frac{1}{2}} A d)\|_2 \leq \Delta_k. \end{aligned} \tag{3.4}$$

The affine scaling $D_k^{-\frac{1}{2}}$ controls the shape of the trust region so that a posterior damping for strict feasibility can retain a large portion of the trust region solution. Letting $\hat{d} \stackrel{\text{def}}{=} D_k^{-\frac{1}{2}} A d$, the trust region subproblem (3.4) is equivalent to

$$\begin{aligned} &\min_{d \in \Re^n, \hat{d} \in \Re^m} \psi_k(d) + \frac{1}{2} \hat{d}^T C_k \hat{d} \\ &\text{subject to} \quad A d - D_k^{\frac{1}{2}} \hat{d} = 0 \\ &\qquad\qquad\quad \|(d; \hat{d})\|_2 \leq \Delta_k. \end{aligned} \tag{3.5}$$

Since (3.4) and (3.5) are equivalent, we subsequently use these two formulations interchangeably depending on the context. Since problem (3.5) is a 2-norm trust region subproblem with consistent equality constraints, existing techniques for computing a solution can be applied to (3.5).

The augmented term $\frac{1}{2}d^T A^T D_k^{-1} C_k A d$ in the objective function of the subproblem (3.4) serves a similar purpose to the augmented Lagrangian function for a constrained minimization problem [20]: it adds positive curvature in the space spanned by the constraint normals. In addition, the curvature augmentation is such that the minimizer of this augmented quadratic $\psi_k(d) + \frac{1}{2}d^T A^T D_k^{-1} C_k A d$ lies in the null space of the binding constraint normals asymptotically. Together with the affine scaling $D_k^{-\frac{1}{2}}$ in the trust region bound, the augmentation implies that a damped step of the solution $p_k$ of the subproblem yields a sufficient decrease, if $p_k$ yields a sufficient decrease.

Consider the Newton step $(p_k^N; \hat{p}_k^N)$ of the trust region subproblem (3.5), i.e.,

$$\hat{H}_k(p_k^N; \hat{p}_k^N) = -\begin{bmatrix} \nabla f_k \\ 0 \end{bmatrix} + \begin{bmatrix} A^T \\ -D_k^{\frac{1}{2}} \end{bmatrix} \lambda_{k+1}^p, \tag{3.6}$$

where $\hat{H}_k \stackrel{\text{def}}{=} \hat{H}(x_k \lambda_k)$ and

$$\hat{H}(x, \lambda) \stackrel{\text{def}}{=} \begin{bmatrix} \nabla^2 f(x) & 0 \\ 0 & C \end{bmatrix}, \qquad C = \operatorname{diag}(|\lambda|).$$

The matrix $\hat{H}_k$ is the Hessian of the augmented quadratic in (3.5).

A Newton step $p_k^N$, with respect to (2.5), is the approximate Newton step $d_k$ (3.3) since

$$\begin{bmatrix} C_k A & D_k \\ \nabla^2 f_k & -A^T \end{bmatrix} \begin{bmatrix} p_k^N \\ \lambda_{k+1}^p \end{bmatrix} = -\begin{bmatrix} 0 \\ \nabla f_k \end{bmatrix}. \tag{3.7}$$

Theorem 3.1 and 3.2 indicate that the trust region subproblem is consistent with the original nonlinear minimization (1.1) with respect to the second order optimality. Proofs for these theorems can be found in [11].

THEOREM 3.1. *Assume that $[A, D_*]$ has full row rank at $x_* \in \mathcal{F}$ and the complementarity conditions are satisfied at $(x_*; \lambda_*)$. Let the columns of $Z_*$ denote an orthonormal basis for the null space of $[A, D_*]$. If $Z_*^T \hat{H}_* Z_*$ is positive semidefinite, then $d^T \nabla^2 f_* d \geq 0$ for any $d$ satisfying $a_i^T d = 0$ for all $i$ with $a_i^T x_* - b_i = 0$.*

Assume that the pair $(x_*; \lambda_*)$ satisfies the first order necessary conditions with strict complementarity, then the *second* order necessary conditions are satisfied at $x_*$ if $Z_*^T \hat{H}_* Z_*$ is positive semidefinite. Theorem 3.2 indicates that quadratic convergence ensues if a damped Newton step $p_k^N$ is taken locally.

THEOREM 3.2. *Assume $0 < \mu < 1$ and $f(x) : \mathcal{F} \to \Re$ is twice continuously differentiable. Assume that $\{x_k\}$ converges to $x_*$, a point satisfying the second order sufficiency conditions with strict complementarity, and $\{a_i : a_i^T x_* - b_i = 0\}$ are linearly independent. Then, for sufficiently large $k$, $Z_k^T \hat{H}_k Z_k$ is positive definite and*

$$f(x_k + s_k) - f(x_k) < \mu \psi_k(s_k), \qquad \text{where } s_k = \theta_k \alpha_k^* p_k^N \text{ and } \theta_0 \leq \theta_k < 1,$$

*and $\alpha_k^*$ is defined in (3.1). In addition, if $x_{k+1} = x_k + s_k$ for sufficiently large $k$ and $|\theta_k - 1| = O(\|(x_k;\lambda_k) - (x_*;\lambda_*)\|_2)$, then $\{(x_k;\lambda_k^p)\}$ converges quadratically to $(x_*;\lambda_*)$.*

**3.2. Moving Away from a Nearly Binding Constraint.** The trust region subproblem (3.4) addresses complementarity, second order necessary requirement and fast local convergence. The only condition yet to be considered is dual feasibility. The trust region subproblem (3.4) is not effective near a non-optimal point satisfying all the necessary optimality conditions except dual feasibility, since both the solution of the trust region subproblem (3.4) and the Newton step $p_k^N$ are zero at such a point. This situation can occur in a strictly linear problem; however, the nonlinearity of a problem (1.1) can either alleviate or exacerbate this problem. On the one hand, the iterates may approach a boundary more slowly due to nonlinearity of $f(x)$. On the other hand, once close to a boundary, the nonlinearity of $f(x)$ may make it harder to move away.

Starting from a relatively centered initial point, the above situation may be rare or may not occur at all. However, when encountered, it is necessary to move away from some nearly binding hyperplane $a_i^T x - b_i = 0$. This situation can be identified by a measurement for satisfaction of complementarity conditions and dual feasibility.

Let $j_0$ identify a hyperplane from which departure is desired, e.g.,

$$(\lambda_k)_{j_0} \stackrel{\text{def}}{=} \min\{(\lambda_k)_i \;:\; |a_i^T x - b_i| < \sigma \text{ and } (\lambda_k)_i < 0\}, \tag{3.8}$$

where $\sigma > 0$ is a small constant. Let $\tilde{D}_k$ be the modified affine scaling matrix of $D_k$:

$$(\tilde{D}_k)_{ii} \stackrel{\text{def}}{=} \begin{cases} (D_k)_{ii} & \text{if } i \neq j_0 \\ 1 & \text{otherwise.} \end{cases} \tag{3.9}$$

When there is no $i$ with $|a_i^T x_k - b_i| < \sigma$ and $(\lambda_k)_i < 0$, it is assumed that $j_0 = 0$ and thus $D_k = \tilde{D}_k$. The trust region $\|(d;\tilde{D}_k^{-\frac{1}{2}}Ad)\|_2 \le \Delta_k$ is elongated along the normal $a_{j_0}$ of the hyperplane identified, see FIG. 3.1. The following subproblem is more appropriate near a non-optimal complementarity point:

$$\begin{aligned} &\min_{d\in\Re^n} \psi_k(d) + \frac{1}{2} d^T A^T \tilde{D}_k^{-1} C_k A d \\ &\text{subject to} \quad \|(d;\tilde{D}_k^{-\frac{1}{2}}Ad)\|_2 \le \Delta_k. \end{aligned} \tag{3.10}$$

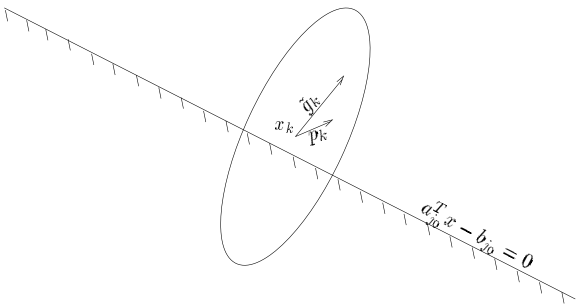


FIG. 3.1. *Leaving the Hyperplane $a_{j_0}^T x - b_{j_0} = 0$*

**3.3. The Proposed Trust Region Algorithm.** Using the trust region subproblems derived in §3.1 and §3.2, the general framework of the proposed trust region and affine scaling interior point method is described in FIG. 3.2. Assume that a strictly feasible initial point $x_0 \in int(\mathcal{F})$ is given. The proposed algorithm generates a sequence $\{x_k\}$ in the interior of the feasible region $\mathcal{F}$ which converge to a solution.

Consider the $k$-th iteration and let $x_k \in int(\mathcal{F})$. Assume further that $\psi_k(d)$ approximates $f(x_k+d)-f(x_k)$ sufficiently accurately; otherwise, the usual trust region size adjustment mechanism can be used to ensure this. The goal of the computation at the $k$-th iteration is to produce a step $s_k$ such that $x_k + s_k \in int(\mathcal{F})$ and $\psi_k(s_k)$ is sufficiently decreased. This task is accomplished in **Step 2**. A step $s_k$ is computed via solving a trust region subproblem which minimizes a quadratic objective function subject to a trust region bound constraint in the 2-norm measure,

$$\begin{aligned} &\min_{d\in\Re^n} \psi_k(d) + \frac{1}{2} d^T A^T S_k^{-1} C_k A d \\ &\text{subject to } \|(d; S_k^{-\frac{1}{2}} d)\|_2 \le \Delta_k, \end{aligned} \tag{3.11}$$

where $S_k$ equals either $D_k^{\frac{1}{2}}$ or $\tilde{D}_k^{\frac{1}{2}}$. The presence of the affine scaling $S_k$ is crucial: it allows for a sufficiently large step. The affine scaling matrix $S_k$ typically equals $D_k$ equals $\mathrm{diag}(Ax_k - b)$. Occasionally $S_k$ equals $\tilde{D}_k$ to facilitate staying away from the boundary. Asymptotically, however, $S_k$ equals $D_k$ and the trust region solution is an approximate Newton step for the complementarity conditions (2.5). Therefore fast local convergence can be achieved.

The affine scaling matrix $S_k \in \{D_k, \tilde{D}_k\}$ can be specified to satisfy the sufficient decrease conditions for optimality, see §3.6. Next we illustrate how a two-dimensional subspace trust region approach can replace the full space trust region subproblem (3.11).

**3.4. A Two-dimensional Trust Region Algorithm.** Computing a solution to the full-space trust region subproblem (3.11) can be too costly for a large scale problem (1.1). Similar to a subspace approach for unconstrained minimization [3], an appropriate small-dimensional trust region subproblem can be used to approximate the full space trust region subproblem, e.g.,

$$\begin{aligned} &\min_{d\in\Re^n, \hat{d}\in\Re^m} \quad \psi_k(d) + \frac{1}{2}\hat{d}^T C_k \hat{d} \\ &\text{subject to} \quad Ad - S_k^{\frac{1}{2}}\hat{d} = 0, \quad (d;\hat{d}) \in \mathcal{S}_k \\ &\qquad\qquad\quad \|(d;\hat{d})\|_2 \le \Delta_k. \end{aligned} \tag{3.12}$$

Here $\mathcal{S}_k$ denotes a small-dimensional subspace in $\Re^{n+m}$, e.g., $|\mathcal{S}_k| = 2$. Performance of a subspace algorithm depends on the choice of the subspace $\mathcal{S}_k$. In order to set up $\mathcal{S}_k$ effectively, we analyze a few important approximate solutions to the trust region subproblem.

Let us first consider the typical choice of $S_k = D_k$. If the second order information is absent, i.e., $\nabla^2 f_k = 0$ and $C_k = 0$ in (3.5), the solution of the trust region solution (3.5) is in the projected gradient direction $(g_k; \hat{g}_k)$,

$$\begin{aligned} g_k &\stackrel{\text{def}}{=} -(\nabla f_k - A^T\lambda_k), \\ \hat{g}_k &\stackrel{\text{def}}{=} D_k^{-\frac{1}{2}} A g_k, \end{aligned} \tag{3.13}$$

**TRAM** Let $0 < \mu < \eta < 1$ and $x_0 \in int(\mathcal{F})$.

**Step 1** Evaluate $f_k$, $\nabla f_k$ and $B_k \approx \nabla^2 f_k$; compute an approximate Lagrangian multiplier $\lambda_k$ and let $C_k = \mathrm{diag}(|\lambda_k|)$.

**Step 2** Compute a step $s_k$, $x_k + s_k \in int(\mathcal{F})$, based on the trust region subproblem

$$\min_{d \in \Re^n} \psi_k(d) + \frac{1}{2} d^T A^T S_k^{-1} C_k A d$$
$$\text{subject to } \|[d; S_k^{-\frac{1}{2}} d]\|_2 \leq \Delta_k.$$

**Step 3** Compute

$$\rho_k = \frac{f(x_k + s_k) - f(x_k)}{\psi(s_k)}.$$

**Step 4** If $\rho_k > \mu$ then set $x_{k+1} = x_k + s_k$. Otherwise set $x_{k+1} = x_k$. Update $\Delta_k$ as specified.

**Updating Trust Region Size $\Delta_k$**
Let $0 < \gamma_1 < 1 < \gamma_2$ be given.
1. If $\rho_k \leq \mu$ then set $\Delta_{k+1} \in (0, \gamma_1 \Delta_k]$.
2. If $\rho_k \in (\mu, \eta)$ then set $\Delta_{k+1} \in [\gamma_1 \Delta_k, \Delta_k]$.
3. If $\rho_k \geq \eta$ then set $\Delta_{k+1} \in [\Delta_k, \gamma_2 \Delta_k]$.

FIG. 3.2. *A Trust Region and Affine Scaling Interior Point Method*

where $\lambda_k$ is the least squares solution to

$$\begin{bmatrix} A^T \\ -D_k^l \end{bmatrix} \lambda \stackrel{\text{LS}}{=} \begin{bmatrix} \nabla f_k \\ 0 \end{bmatrix}, \tag{3.14}$$

with $l = \frac{1}{2}$; more generally, $l \geq \frac{1}{2}$ can be used. Moving along the projected gradient direction $(g_k; \hat{g}_k)$ is an attempt to satisfy the complementarity conditions (2.5) via a least squares approach.

Asymptotically, the Newton step $p_k^N$ is a solution to the trust region subproblem. However, globally $p_k^N$ may not always be a good approximate solution, e.g., $p_k^N$ may not exist or it may be a poor approximate solution when the Hessian of the augmented quadratic is indefinite.

In order to yield a good decrease for $\psi_k(d) + \frac{1}{2} d^T A^T D_k^{-1} C_k A d$ when its Hessian is indefinite, a negative curvature direction is important. Since a Newton step $p_k^N$ is a Newton step for the augmented quadratic objective, i.e.,

$$(A^T (C_k D_k^{-1}) A + \nabla^2 f_k) p_k^N = -\nabla f_k, \tag{3.15}$$

negative curvature can be detected, then computed in the process of computing the desired approximate Newton step. A negative curvature direction of (3.15) can be computed by either a partial Cholesky factorization [20] or using a conjugate gradient process with an incomplete Cholesky factorization as a preconditioner.

An effective subspace $\mathcal{S}_k$ can be formed from the first and second order approximate solutions to the trust region subproblem (3.5). Let $d_k^c$ denote the solution of (3.15) if its coefficient matrix is positive definite and the computed negative curvature direction of (3.15) otherwise. We propose to choose $S_k$ to be spanned by the projected gradient $g_k$ and the second order direction $d_k^c$. The projected gradient is especially important when a problem is nearly degenerate. The addition of the Newton or the negative curvature directions are important for fast global and local convergence. The two dimensional subspace trust region subproblem is

$$
\begin{aligned}
&\min_{d\in\Re^n,\hat{d}\in\Re^m} && \psi_k(d)+\frac{1}{2}d^TA^TD_k^{-1}C_kAd \\
&\text{subject to} && d\in\operatorname{span}\{g_k,d_k^c\} \\
& && \|(d;D_k^{-\frac{1}{2}}Ad)\|_2\le\Delta_k.
\end{aligned}
\tag{3.16}
$$

Since $[g_k,\hat{g}_k]$ and $[d_k^c,\hat{d}_k^c]$ both satisfy $Ad-D_k^{\frac{1}{2}}\hat{d}=0$, the 2-dimensional trust region subproblem (3.16) can be solved by first computing an orthonormal basis for $\mathcal{S}_k$ in $O(n+m)$ flops and then solving a trust region problem in $\Re^2$ with a constant cost. Computation of $g_k$ and $d_k^c$ is the main cost for a 2-dimensional subspace trust region algorithm.

With respect to the choice $S_k=\tilde{D}_k$, a Newton direction or a negative curvature direction can be computed from

$$
(A^T(C_k\tilde{D}_k^{-1})A+\nabla^2 f_k)\tilde{d}_k=-\nabla f_k. \tag{3.17}
$$

A first order approximate solution for (3.10) is a projected gradient direction $\tilde{g}_k$,

$$
\tilde{g}_k \stackrel{\text{def}}{=} -(\nabla f_k - A^T\tilde{\lambda}_k), \tag{3.18}
$$

where $\tilde{\lambda}_k$ solves

$$
\begin{bmatrix} A^T \\ -\tilde{D}_k^l \end{bmatrix}\lambda \stackrel{\text{LS}}{=} \begin{bmatrix} \nabla f_k \\ 0 \end{bmatrix}, \qquad l\ge\frac{1}{2}. \tag{3.19}
$$

The projected direction $\tilde{g}_k$ is descent for $\psi_k(d)$ and leads to departure from the hyperplane $a_{j_0}^Tx-b_{j_0}=0$.

Let $\tilde{d}_k^c$ denote the Newton step (3.17) when it exists and a negative curvature direction of (3.17) otherwise. Using the projected gradient direction $\tilde{g}_k$ and the second-order direction $\tilde{d}_k^c$, the two-dimensional trust region subproblem is

$$
\begin{aligned}
& && \min_{d\in\Re^n}\psi_k(d)+\frac{1}{2}d^TA^TS_k^{-1}C_kAd \\
&\text{subject to} && d\in\operatorname{span}\{\tilde{g}_k,\tilde{d}_k^c\} \\
& && \|(d;S_k^{-\frac{1}{2}}Ad)\|_2\le\Delta_k, \qquad S_k=\tilde{D}_k.
\end{aligned}
$$

Compared to the unsymmetric linear system (3.3), the $n$-by-$n$ symmetric Newton system (3.15) is preferable since it leads to a negative curvature direction when the

Hessian of $\psi_k(d) + \frac{1}{2}d^T A^T D_k^{-1} C_k A d$ is indefinite. The disadvantage of using the symmetric $n$-by-$n$ linear system (3.15) rather than the unsymmetric $(m+n)$-by-$(m+n)$ linear system (3.3) is the increasing ill-condition of (3.15) as a solution is approached. The stability of the related linear system for linear programming has been studied [37, 34]. Similar issues are yet to be investigated for the linear system (3.15) for a nonlinear programming problem (1.1).

**3.5. A Reflection Line Search.** The effectiveness of an interior point method depends, in part, on the ability to avoid getting close to the boundary prematurely. A trust region algorithm using the subproblem (3.4) is similar to a Dikin affine scaling algorithm in that no explicit effort is made to stay central when computing a descent direction. For a Dikin affine scaling method, it is desirable that an initial point $x_0$ is approximately centered and damping is used to avoid getting close to a boundary prematurely.

Explicit effort can be made to facilitate staying away from boundary and achieve further decrease. A special line search has been used in [8] for linear $l_1$ problems. For a nonlinear minimization with bound constraints, a reflection line search and its effectiveness is illustrated in [9, 10, 2]. We now examine a similar reflection technique for the problem (1.1) with linear inequality constraints.

Assume that $x_k$ is the current strictly feasible point and $d_k$ is an approximate solution to the trust region subproblem (3.11). If $S_k = D_k$ and $x_k$ is close to a hyperplane $a_i^T x = b_i$, then a solution $p_k$ to the trust region subproblem (3.11) will be nearly tangential to the hyperplane. However, the trust region may be outside of the feasible region, see FIG. 3.3. To maintain strict feasibility, it is possible to simply take a damped step by backtracking. Let us consider the reflection direction $p_k^R$ of $p_k$ against the hyperplane $a_i^T x - b_i = 0$, i.e.,

$$p_k^R = p_k - \frac{2a_i^T p_k}{a_i^T a_i} a_i.$$

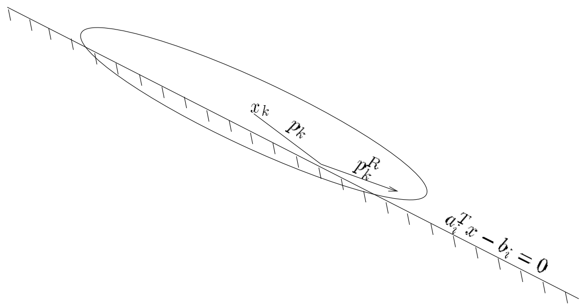


FIG. 3.3. *Trust Region and Reflection*

Use of the reflection direction $p_k^R$ has several benefits:

1. $p_k^R$ points away from the hyperplane $a_i^T x = b_i$;
2. moving along $p_k^R$ is likely to yield continual decrease of the objective $\psi_k(d)$;
3. $p_k^R$ can be computed with little cost.

This reflection process can be repeated for each $p_k^R$. The effectiveness of the reflection search is demonstrated with computational examples, see Table 4.7 in §4.

**3.6. Sufficient Decrease Conditions.** Consider the current strictly feasible point $x_k$. A step $s_k$, with $\|s_k\|_2 \leq \Delta_k$ and $x_k + s_k \in int(\mathcal{F})$, needs to be computed to generate a sufficient decrease of the quadratic approximation $\psi_k(d)$. Typically, the trust region subproblem (3.11) with $S_k$ equal to $D_k$ leads to such a step. However, occasionally, the subproblem with $S_k$ equal to $\tilde{D}_k$, forcing departure from a constraint, is more appropriate. To decide on the scaling $S_k$, we examine sufficient decrease conditions for optimality.

The sufficient decrease conditions emerge naturally in the two dimensional subspace trust region context. From inspection of (3.13) it follows that, the complementarity conditions are satisfied at $x_k$ if and only if $(g_k; \hat{g}_k) = 0$, where

$$\begin{aligned} g_k &= -(\nabla f_k - A^T \lambda_k), \\ \hat{g}_k &= -D_k^{\frac{3}{2}} \lambda_k, \end{aligned}$$

and $\lambda_k$ solves

$$\begin{bmatrix} A^T \\ -D_k \end{bmatrix} \lambda \stackrel{\text{LS}}{=} \begin{bmatrix} \nabla f_k \\ 0 \end{bmatrix}.$$

Let $\theta_k \in [\theta_0, 1)$, $0 < \theta_0 < 1$, be a damping parameter for strict feasibility, as in (3.1). Let $g_k^* \stackrel{\text{def}}{=} \alpha_k g_k$, $\alpha_k = \theta_k \alpha_k^*$, denote a damped minimizer of the augmented quadratic objective along $g_k$ within the feasible trust region, i.e., $\alpha_k^*$ solves

$$\min_{0 \leq \alpha \leq \Delta_k / \|(g_k; D_k^{-\frac{1}{2}} A g_k)\|_2} \{\psi_k(\alpha g_k) + \frac{\alpha^2}{2} g_k^T A^T D_k^{-1} C_k A g_k, \ x_k + \alpha g_k \in \mathcal{F}\}. \tag{3.20}$$

This suggests that a sufficient decrease of $\psi_k(d)$ measured against $\psi_k(g_k^*)$ leads to satisfaction of complementarity. This is expressed in condition (AS.1) in FIG. 3.4.

The projected gradient $\tilde{g}_k$ attempts to satisfy (3.18),

$$\begin{bmatrix} A^T \\ -\tilde{D}_k \end{bmatrix} \tilde{\lambda} = \begin{bmatrix} \nabla f_k \\ 0 \end{bmatrix}.$$

Assume that strict complementarity is satisfied at $x_k$. Then dual feasibility is satisfied when $\tilde{g}_k = 0$ and $\tilde{D}_k \tilde{\lambda}_k = 0$. The damped minimizer $\tilde{g}_k^* \stackrel{\text{def}}{=} \alpha_k \tilde{g}_k$, $\alpha_k = \theta_k \alpha_k^*$, is along the projected gradient $\tilde{g}_k$ where $\alpha_k^*$ solves

$$\min_{0 \leq \alpha \leq \Delta_k / \|(\tilde{g}_k; \tilde{D}_k^{-\frac{1}{2}} A \tilde{g}_k)\|_2} \{\psi_k(\alpha \tilde{g}_k) + \frac{\alpha^2}{2} \tilde{g}_k^T A^T \tilde{D}_k^{-1} C_k A \tilde{g}_k, \quad x_k + \alpha \tilde{g}_k \in \mathcal{F}\}. \tag{3.21}$$

This leads to condition (AS.2) in FIG. 3.4 for dual feasibility.

Finally condition (AS.3) is sufficient for second order optimality and fast local convergence when $p_k^*$ is defined as a damped minimizer, e.g.,

$$p_k^* \stackrel{\text{def}}{=} \alpha_k p_k, \quad \alpha_k = \theta_k \alpha_k^*, \quad \alpha_k = \min(1, \beta_k), \tag{3.22}$$

and $\beta_k$ is the stepsize to the boundary along $p_k$ and $(p_k; \hat{p}_k)$ is a solution to the trust region subproblem (3.5).

**Sufficient Decrease Conditions**. Let $0 < \beta_{cs}, \beta_{df}, \beta_q < 1$ and $\beta_s > 0$. Let $g_k$ be defined by (3.13), $\tilde{g}_k$ be defined by (3.18) and $p_k$ be a solution to (3.5). Let $g_k^*$, $\tilde{g}_k^*$ and $p_k^*$ be as defined as in (3.20), (3.21) and (3.22) respectively. Let $\|s_k\|_2 \leq \beta_s \Delta_k$. The sufficient decrease conditions are:

**(AS.1)** $\psi_k(s_k) < \beta_{cs}(\psi_k(g_k^*) + \frac{1}{2} g_k^{*T} A^T D_k^{-1} C_k A g_k^*)$;

**(AS.2)** $\psi_k(s_k) < \beta_{df}(\psi_k(\tilde{g}_k^*) + \frac{1}{2} \tilde{g}_k^{*T} A^T \tilde{D}_k^{-1} C_k A \tilde{g}_k^*)$;

**(AS.3)** $(\psi_k(s_k) + \frac{1}{2} s_k^T A^T D_k^{-1} C_k A s_k) < \beta_q(\psi_k(p_k^*) + \frac{1}{2} p_k^{*T} A^T D_k^{-1} C_k A p_k^*)$.

FIG. 3.4. *Sufficient Decrease Conditions for the First and Second Order Optimality*

Conditions (AS.1), (AS.2) and (AS.3) are closely related to the sufficient decrease conditions proposed for the bound constrained minimization problem [9] and the nonlinear $l_1$ problem [25]. In [11], these conditions will be rigorously analyzed to establish convergence properties for the linearly constrained minimization (1.1). Essentially, under reasonable assumptions, if (AS.1) is satisfied at each iteration, then any limit point of $\{x_k\}$ will satisfy the complementarity conditions. If, in addition, (AS.2) is satisfied *asymptotically*, then any limit point with strict complementarity also satisfies the first order necessary conditions. Furthermore, if (AS.3) is satisfied *asymptotically*, then the second order necessary condition is satisfied at a limit point with strict complementarity. Finally, with the additional assumption the second order sufficiency conditions at a limit point, convergence is locally quadratic.

Clearly, a step $s_k$ can be determined to satisfy (AS.1), (AS.2) and (AS.3) at every iteration if $g_k$, $\tilde{g}_k$ and $p_k$ are computed. However since (AS.2) and (AS.3) only need to be satisfied *asymptotically*, less costly alternatives exist. In particular, the reflection search discussed in §3.5 can be combined with a dogleg line search to satisfy these conditions.

**4. Computational Experience.** To illustrate its potential, preliminary computational experience is reported for our TRAM implementation. A trust region subproblem is solved via a two dimensional approximation as described in §3.4. The projected gradient $g_k$ is computed using a sparse least squares solver, e.g., [30]. A reflection technique described in §3.5 is incorporated. The details of the implemented two-dimensional subspace algorithm are described in FIG. 4.1.

A large scale nonlinear minimization test problem (1.1) is generated in a similar fashion to the test problems for a nonlinear minimization subject to simple bounds [9, 13]. Nonlinear objective functions $f(x)$ are chosen from two classes: the nonlinear test problem collection for unconstrained minimization [29], and the molecule minimization problem. In the latter, the objective function $f(x)$ has the following formulation [22, 6],

$$f(x) = \sum_{(i,j) \in \mathcal{S}} (\|x_i - x_j\|_2^2 - d_{ij}^2)^2,$$

where $x_i$ denotes the position of the atom and $d_{ij}$ is the known distance between a pair of atoms $(i, j)$.

The inequality constraints $Ax \geq b$ consist of lower bounds, upper bounds and inequality constraints obtained from a random sparse matrix $C^T = [c_1, c_2, \cdots, c_{m_0}] \in$

| Number of Function Evaluations | | | | |
|---|---|---|---|---|
| $(m,n)$ | $(350,100)$ | $(875,250)$ | $(1750,500)$ | $(3500,1000)$ |
| GENROSE | 14 | 13 | 20 | 14 |
| GENSING | 21 | 34 | 43 | 43 |
| CHAINSING | 17 | 22 | 35 | 26 |
| DEGENSING | 23 | 28 | 25 | 31 |
| GENWOOD | 13 | 17 | 16 | 20 |
| CHAINWOOD | 16 | 12 | 15 | 16 |
| BROYDEN1A | 12 | 13 | 23 | 26 |
| BROYDEN1B | 12 | 14 | 18 | 28 |
| BROYDEN2A | 16 | 22 | 27 | 22 |
| BROYDEN2B | 17 | 21 | 17 | 22 |
| TOINTBROY | 19 | 14 | 43 | 34 |
| CRAGGLEVY | 19 | 21 | 28 | 29 |
| AUGMLAGN | 77 | 78 | 52 | 141 |
| BROWN3 | 12 | 15 | 18 | 17 |
| BVP | 53 | 6 | 11 | 5 |
| VAR | 16 | 13 | 15 | 17 |

TABLE 4.1
*Number of Function Evaluations*

$\Re^{n\times m_0}$ with the 5-point difference pattern. Solving an unconstrained minimization problem first, the constraints $Ax \geq b$ are formulated in a fashion similar to the bound constraints in the test problems used in [9, 13].

For the results reported subsequently, computation is terminated when

$$\text{either} \quad \theta_k \leq \text{tol} \quad \text{or} \quad \frac{f(x_k)-f(x_{k+1})}{\max(1,|f(x_k)|)} \leq \text{tol},$$

where tol equals $10^{-8}$.

Table 4.1 lists the number of function evaluations required for each testing problem using standard nonlinear test functions. The number of iterations required is typically small, e.g., less than 20. In addition, computation of the projected gradient $\tilde{g}_k$ is infrequent for the generated test problems in Table 4.1; $\tilde{g}_k$ is computed 42 times for the 64 tests.

Table 4.2 tabulates the statistics of the computed solutions. The numbers under the column *opt* are the optimality measurement $\theta_k$ at termination and the values under *deg* are the strict complementarity measurement $\min(Ax_k - b + |\lambda_k|)$ at termination. It is evident that there are many constraints active (typically far more than the number of iterations required to solve a problem). A substantial portion of the active constraints at the computed solution are not simple bounds. Fairly good accuracy is obtained, even in the presence of near degeneracy. Note that column "deg" measures only one type of degeneracy. Ill conditioning can also come from near singularity of the projected Hessian (which seems to occur for problems GENSING, CHAINGENSING and DEGENSING). In Table 4.2, for problems GENWOOD, CHAINWOOD and BROYDEN2A, the gradients of the active constraints are structurely rank deficient,

| Statistics at Termination | | | | |
|---|---|---|---|---|
| problems | #fevals | activities | opt | deg |
| GENROSE | 14 | 504 | $10^{-10}$ | $10^{-3}$ |
| GENSING | 43 | 646 | $10^{-4}$ | $10^{-3}$ |
| CHAINSING | 26 | 677 | $10^{-3}$ | $10^{-3}$ |
| DEGENSING | 31 | 671 | $10^{-7}$ | $10^{-5}$ |
| GENWOOD | 20 | 798 | $10^{-5}$ | $10^{-5*}$ |
| CHAINWOOD | 16 | 807 | $10^{-6}$ | $10^{-3*}$ |
| BROYDEN1A | 26 | 693 | $10^{-9}$ | $10^{-4}$ |
| BROYDEN1B | 28 | 695 | $10^{-9}$ | $10^{-4}$ |
| BROYDEN2A | 22 | 483 | $10^{-7}$ | $10^{-4*}$ |
| BROYDEN2B | 22 | 511 | $10^{-8}$ | $10^{-4}$ |
| TOINTBROY | 34 | 709 | $10^{-7}$ | $10^{-4}$ |
| CRAGGLEVY | 29 | 377 | $10^{-8}$ | $10^{-4}$ |
| AUGMLAGN | 141 | 442 | $10^{-8}$ | $10^{-3}$ |
| BROWN3 | 17 | 686 | $10^{-8}$ | $10^{-4}$ |
| BVP | 5 | 1 | $10^{-5}$ | $10^{-7}$ |
| VAR | 17 | 646 | $10^{-9}$ | $10^{-4}$ |

TABLE 4.2
*Characteristics of the Computed Solutions* $(m, n) = (3500, 1000)$

| Number of Function Evaluations | | | | |
|---|---|---|---|---|
| (250,100) | (500,200) | (750,300) | (1000,400) | (1250,500) |
| 48 | 54 | 64 | 65 | 65 |
| 41 | 53 | 42 | 53 | 53 |
| 36 | 46 | 67 | 97 | 97 |
| 36 | 42 | 44 | 65 | 65 |
| 31 | 40 | 69 | 65 | 65 |
| 32 | 66 | 67 | 73 | 73 |
| 30 | 50 | 76 | 95 | 95 |
| 44 | 55 | 55 | 65 | 65 |
| 33 | 47 | 65 | 52 | 52 |
| 46 | 39 | 56 | 63 | 63 |
| 37.7 | 49.2 | 60.5 | 69.3 | 82.7 |

TABLE 4.3
*Number of Function Evaluations for Molecule Problems with Constraints*

e.g., GENWOOD and BROYDEN2A are deficient by 1, CHAINWOOD by 2.

Tables 4.3-4.5 illustrate performance of our 2-dimensional TRAM implementation when the objective function is a molecule minimization. Here the objective function is a quartic; this class of problems is difficult to solve due to the presence of negative curvature. Again $\tilde{g}_k$ is computed infrequently: 6 times for the 50 problem instances in Table 4.4. As indicated by Table 4.3 and 4.4, the number of iterations grows with

| Number of Function Evaluations | | | | |
|---|---|---|---|---|
| (350,100) | (700,200) | (1050,300) | (1400,400) | (1750,500) |
| 36 | 46 | 72 | 64 | 70 |
| 49 | 31 | 34 | 56 | 101 |
| 28 | 31 | 113 | 77 | 67 |
| 30 | 46 | 52 | 127 | 63 |
| 35 | 24 | 72 | 94 | 62 |
| 19 | 28 | 49 | 35 | 69 |
| 20 | 36 | 55 | 46 | 41 |
| 27 | 35 | 37 | 87 | 52 |
| 26 | 28 | 38 | 46 | 65 |
| 38 | 40 | 55 | 49 | 49 |
| 30.7 | 34.5 | 57.7 | 68.1 | 63.9 |

TABLE 4.4
*Number of Function Evaluations for Molecule Problems with Constraints*

| #fevals | activities | opt | deg |
|---|---|---|---|
| 64 | 208 | $10^{-6}$ | $10^{-5}$ |
| 56 | 79 | $10^{-6}$ | $10^{-6}$ |
| 77 | 171 | $10^{-9}$ | $10^{-5}$ |
| 127 | 103 | $10^{-9}$ | $10^{-4}$ |
| 94 | 122 | $10^{-7}$ | $10^{-5}$ |
| 35 | 260 | $10^{-9}$ | $10^{-4}$ |
| 46 | 262 | $10^{-7}$ | $10^{-6}$ |
| 87 | 219 | $10^{-9}$ | $10^{-5}$ |
| 46 | 165 | $10^{-6}$ | $10^{-5}$ |
| 49 | 155 | $10^{-8}$ | $10^{-5}$ |

TABLE 4.5
*Characteristics of the Computed Solutions for Molecule Problems* $(m, n) = (1750, 500)$

the problem size, although the number of function evaluations are still significantly less than the number of active constraints at the solution, see Table 4.5. The growth of the number of iterations, as the problem dimension is increased, is probably due to the severe nonlinear and nonconvex nature of the molecule minimization problem rather than the technique of handling constraints. This is certainly supported by the computational results in Table 4.6 in which each molecule testing problem is unconstrained. We observe a similar growth in the number of function evaluations as the problem size is increased. This suggests that, for these problems, the approximation of the full space trust region subproblem by a two-dimensional trust region subproblem becomes increasingly less accurate as the problem size is increased.

In Table 4.7 the effectiveness of reflection is illustrated: the first column under each problem size entry list the number of function evaluations with a reflection and the second column without a reflection. The maximum number of reflections allowed

| Number of Function Evaluations | | | | |
|---|---|---|---|---|
| 100 | 200 | 300 | 400 | 500 |
| 55 | 94 | 108 | 152 | 119 |
| 39 | 72 | 103 | 132 | 138 |
| 46 | 128 | 75 | 124 | 138 |
| 53 | 84 | 98 | 114 | 101 |
| 57 | 84 | 120 | 99 | 167 |
| 48 | 78 | 112 | 127 | 187 |
| 71 | 76 | 98 | 138 | 116 |
| 41 | 95 | 89 | 120 | 124 |
| 66 | 64 | 100 | 206 | 171 |
| 66 | 103 | 90 | 125 | 120 |

TABLE 4.6
*Number of Function Evaluations for Unconstrained Molecule Problems*

| Number of Function Evaluations | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| (250,100) | | (500,200) | | (750,300) | | (1000,400) | | (1250,500) | |
| 46 | 51 | 83 | 85 | 65 | 94 | 86 | 108 | 77 | 124 |
| 33 | 58 | 58 | 58 | 53 | 87 | 80 | 155 | 120 | 196 |
| 33 | 44 | 69 | 71 | 73 | 177 | 97 | 229 | 81 | 122 |
| 42 | 70 | 39 | 63 | 42 | 99 | 83 | 101 | 69 | 81 |
| 44 | 84 | 40 | 66 | 109 | 121 | 86 | 91 | 81 | 139 |
| 28 | 40 | 77 | 101 | 90 | 117 | 75 | 134 | 130 | 141 |
| 26 | 32 | 51 | 59 | 78 | 134 | 76 | 148 | 83 | 161 |
| 37 | 46 | 82 | 127 | 93 | 161 | 77 | 93 | 94 | 130 |
| 41 | 58 | 48 | 56 | 51 | 71 | 75 | 258 | 71 | 138 |
| 30 | 52 | 30 | 62 | 54 | 95 | 69 | 92 | 79 | 106 |
| 36 | 53.5 | 57.7 | 74.8 | 70.8 | 115.6 | 80.4 | 140.9 | 88.5 | 133.8 |

TABLE 4.7
*With and Without Reflection for the Constrained Random Molecule Problems*

is 20 at each iteration and the cost of reflection is insignificant. It is evident that the reflection technique significantly improved the efficiency of the algorithm.

**5. Concluding Remarks.** The main objective of this article is to motivate a trust region and affine scaling interior point method (TRAM) for general nonlinear (nonconvex) minimization with linear inequality constraints.

Based on the belief that a monotonic decrease of the original objective function is important for general nonlinear minimization, a Newton step based on the nonlinear system of equations expressing the complementarity conditions is considered. A trust region subproblem is formed consistent with the Newton step. This trust region subproblem yields an approximate Newton step for the complementarity conditions asymptotically. The quadratic objective function of the trust region subproblem is the quadratic Taylor approximation $\psi_k(d)$ to the original objective function $f(x)$ plus a convex quadratic term determined by affine scaling and the Lagrangian multiplier approximation. Affine scaling in both the quadratic objective function and the trust region bound constraint allows the fast convergence of the iterates from the strictly feasible region. Since the effectiveness of an interior point method depends, in parts, on avoiding approaching the boundary of the feasible region prematurely, a reflection line search is proposed to facilitate this.

The proposed algorithm TRAM is related to a Dikin-affine scaling algorithm. However, our typical affine scaling is $D_k^{-\frac{1}{2}}$ rather than $D_k^{-1}$ and there is an augmented term in the objective function contributes additional scaling. In addition, the proposed algorithm is a Newton-type algorithm for a nonlinear minimization and has fast local convergence.

Explicit decrease conditions are proposed for complementarity, dual feasibility and second order optimality. Global and local convergence analysis of the TRAM algorithm is presented in [11].

A two-dimensional subspace trust region framework is analyzed and implemented for large scale problems. Within this subspace framework, the main computation of each iteration can be done using (partial) Cholesky factorization and least squares solve. Preliminary computational results suggest that the proposed method can be effective.

**A 2-dimensional TRAM Algorithm**. Assume that $x_0 \in \Re^n$ with $Ax_0 > b$ is given. Let $\mu = 0.05$ and $\eta = 0.75$.

**Step 1** Evaluate $f_k$, $\nabla f_k$, $\nabla^2 f_k$; Compute the projected gradient $g_k$, Lagrangian multiplier $\lambda_k$ and $C_k$; Compute a complementarity measure $\theta_k^{\mathrm{c}} = \|D_k\lambda_k\|_\infty$, dual feasibility measure $\theta_k^{\mathrm{d}} = \min_{(\lambda_k)_i<0}(|(\lambda_k)_i|)$ and a first order optimality measure $\theta_k = (\theta_k^{\mathrm{c}} + \theta_k^{\mathrm{d}})/(1+\theta_k^{\mathrm{c}}+\theta_k^{\mathrm{d}})$;

**Step 2** **if** $\theta_k^{\mathrm{c}} < 10^{-2}\min(1, \theta_k^{\mathrm{d}})$, compute $\tilde{g}_k$; **end**;

**if** $\theta_k^{\mathrm{c}} > 10^{-2}\min(1,\theta_k^{\mathrm{d}})$ **or** $\psi_k(g_k^*) + \frac{1}{2}g_k^{*T}A^TD_k^{-1}C_kAg_k^* < \psi_k(\tilde{g}_k^*) + \frac{1}{2}\tilde{g}_k^{*T}A^T\tilde{D}_k^{-1}C_kA\tilde{g}_k^*$

$$S_k = D_k;$$

**else**

$$S_k = \tilde{D}_k;\ g_k^* = \tilde{g}_k^*;\ g_k = \tilde{g}_k;$$

**end**

Compute $d_k^c$ by (3.15) and let $p_k$ solve a 2-dimensional suproblem

$$\begin{aligned} &\min_{d\in\Re^n} \psi_k(d) + \frac{1}{2}d^TA^TS_k^{-1}C_kAd \\ \text{subject to}\quad & d \in \mathrm{span}\{g_k, d_k^c\} \\ & \|[d; S_k^{-\frac{1}{2}}Ad]\|_2 \le \Delta_k. \end{aligned}$$

**if** $(\psi_k(p_k^*) + \frac{1}{2}p_k^{*T}A^TD_k^{-1}C_kAp_k^*) \le 0.99(\psi_k(g_k^*) + \frac{1}{2}g_k^{*T}A^TD_k^{-1}C_kAg_k^*)$

$s_k = p_k^*$;

**else**

let $p_k^R$ be the damped minimizer of $\psi_k(d)$ along the reflection path of $p_k$;
$s_k = \mathrm{argmin}\{\psi_k(d):\ d = \alpha g_k^* + (1-\alpha)p_k^R,\ 0 \le \alpha \le 1\}$;

**end**

**Step 3** Compute

$$\rho_k = \frac{f(x_k+s_k) - f(x_k)}{\psi_k(s_k)};$$

**Step 4** If $\rho_k > \mu$ then set $x_{k+1} = x_k + s_k$. Otherwise set $x_{k+1} = x_k$. Update $\Delta_k$ as specified in FIG. 3.2.

FIG. 4.1. *A 2-dimensional Trust Region Affine Scaling Method for Minimization Subject to Linear Inequality Constraints*

## REFERENCES


[1] N. K. ADLER, M. RESENDE, AND G. VEIGA, *An implementation of karmarkar's algorithm for linear programming*, Mathematical Programming, 44 (1989).

[2] M. A. BRANCH, T. F. COLEMAN, AND Y. LI, *A subspace, interior and conjugate gradient method for large-scale bound-constrained minimization*, Tech. Report TR95-1525, Computer Science Department, Cornell University, 1995.

[3] R. BYRD, R. SCHNABEL, AND G. SHULTZ, *Parallel quasi-Newton methods for unconstrained optimization*, tech. report, Department of Computer Science, University of Colorado, 1988.

[4] R. H. BYRD, J. C. GILBERT, AND J. NOCEDAL, *A trust region method based on interior point techniques for nonlinear programming*, Tech. Report OTC96/02, Optimization Technology Center, Northwestern University, 1996.

[5] Y.-C. CHEN, D. J. HOUCK, J.-M. LIU, M. MEKETON, R. J. V. L. SLUTSMAN, AND P. WANG, *The AR&T KORBX System*, AT&T Tecnnical Journal, (1989).

[6] T. F. COLEMAN, *Large-scale numerical optimization: Introduction and overview*, in Encyclopedia of Computer Science and Technology, A. Kent and J. G. Williams, eds., Marcel Dekker, INC., 1993, pp. 167–195.

[7] T. F. COLEMAN AND Y. LI, *A global and quadratically-convergent method for linear $l_\infty$ problems*, SIAM Journal on Scientific and Statistical Computing, 29 (1992), pp. 1166–1186.

[8] ———, *A globally and quadratically convergent affine scaling method for linear $l_1$ problems*, Mathematical Programming, 56, Series A (1992), pp. 189–222.

[9] ———, *An interior, trust region approach for nonlinear minimization subject to bounds*, SIAM Journal on Optimization, 6 (1996).

[10] ———, *A reflective Newton method for minimizing a quadratic function subject to bounds on the variables*, SIAM Journal on Optimization, 6 (1996).

[11] ———, *A trust region and affine scaling interior point method for nonconvex minimization with linear inequality constraints*, Tech. Report TR97-1642, Computer Science Department, Cornell University, 1997. Submitted to Math. Prog.

[12] T. F. COLEMAN AND J. LIU, *An interior Newton method for quadratic programming*, Tech. Report TR93-1388, Computer Science Department, Cornell University, 1993.

[13] A. R. CONN, N. I. M. GOULD, AND P. L. TOINT, *Testing a class of methods for solving minimization problems with simple bounds on the variables*, Mathematics of Computation, 50 (1988), pp. 399–430.

[14] J. E. DENNIS, M. H. JR., AND L. N. VICENTE, *Trust-region interior-point sqp algorithms for a class of nonlinear programming problems*, Tech. Report TR94-45, Department of Computational and Applied Mathematics, Rice University, 1994.

[15] I. DIKIN, *Iterative solution of problems of linear and quadratic programming*, Doklady Akademiia Nauk SSSR, 174 (1967), pp. 747–748.

[16] A. S. EL-BAKRY, R. A. TAPIA, T. TSUCHIYA, AND T. ZHANG, *On the formulation and theory of the Newton interior-point method for nonlinear programming*, (1992).

[17] A. V. FIACCO AND G. P. MCCORMICK, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley and Sons, 1968.

[18] R. FLETCHER, *Practical Methods of Optimization: Volume 2, Constrained Optimization*, John Wiley and Sons, 1981.

[19] A. FORSGREN AND P. E. GILL, *Primal-dual interior methods for nonconvex nonlinear programming*, Tech. Report NA 96-3, Department of Mathematics, University of California, San Diego, 1996.

[20] P. E. GILL, W. MURRAY, AND M. H. WRIGHT, *Practical Optimization*, Academic Press, 1981.

[21] C. C. GONZAGA, *Path following methods for linear programming*, SIAM Review, 34 (1992), pp. 167–227.

[22] B. A. HENDRICKSON, *The molecule problem: Determining conformation from pairwise distances*, Cornell University Ph.D thesis, Computer Science, 1991.

[23] M. KOJIMA, S. MIZUNO, AND A. YOSHISE, *A primal-dual interior point method for linear programming*, in Progress in Mathematical Programming Interior-point and Related Methods, N. Megiddo, ed., Springer-Verlag, 1989.

[24] R. KRANICH, *Interior point methods for mathematical programming: A bibliograph*, tech. report, Available through NETLIB: send e-mail to netlib@research.att.com containing the message "send intbib.bib from bib.

[25] Y. LI, *A Newton acceleration of the Weiszfeld algorithm for minimizing the sum of Euclidean*

*distances*, Computational Optimization and Applications, to appear.

[26] Y. Li, *A globally convergent method for $l_p$ problems*, SIAM Journal on Optimization, (1993), pp. 609–629.

[27] Y. Li, *A trust region and affine scaling method for nonlinearly constrained minimization*, Tech. Report TR 94-1463, Computer Science Department, Cornell University, 1994, (*Submitted to SIAM J. Optimization*).

[28] R. Monteriro, T. Tsuchiya, and Y. Wang, *A simplified global convergence proof of the affine scaling algorithm*, Annals of Operation Research, (1993).

[29] J. J. Moré, B. S. Garbow, and K. E. Hillstrom, *Testing unconstrained optimization software*, ACM Transactions on Mathematical Software, 7 (1981), pp. 17–41.

[30] C. SUN, *Dealing with dense rows in the solution of sparse linear least squares*, Tech. Report CTC95TR227, Cornell Theory Center, Cornell University, 1995.

[31] M. Todd, *Potential-reduction methods in mathematical programming*, Tech. Report TR1112, School of Operations Research and Industrial Engineering, Cornell University, 1995.

[32] T. Tsuchiya and R. Monteriro, *Superlinear convergence of the affine scaling algorithm*, Mathematical Programming, (1996).

[33] R. J. Vanderbei, M. S. Meketon, and B. A. Freedman, *A modification of Karmarkar's linear programming algorithm*, Algorithmica, 1 (1986), pp. 395–407.

[34] S. Vavasis, *Stable numerical algorithms for equilibrium systems*, Tech. Report 92-1280, Computer Science Dept., Cornell University, 1992.

[35] M. H. Wright, *Some properties of hte hessian of the logarithmic barrier function*, Mathematical Programming, 67 (1994), pp. 265–295.

[36] S. Wright, *An interior point algorithm for linearly constrained optimization*, SIAM J. Optimization, 2 (1992), pp. 450–473.

[37] ———, *Modified cholesky factorizations in interior-point algorithms for linear programming*, Tech. Report MCS-P600-0596, Mathematics and Computer Science Division, Argonne National Laboratory, 1996.