WIS-CS-72-166
COMPUTER SCIENCES DEPARTMENT
The University of Wisconsin
1210 West Dayton Street
Madison, Wisconsin 53706

A QUADRATICALLY CONVERGENT LAGRANGIAN ALGORITHM FOR NONLINEAR CONSTRAINTS

by

J. B. Rosen and J. L. Kreuser

Technical Report #166

November 1972

Received November 17, 1972

This research was supported in part by NSF Grants GJ362, GJ32552, and GJ35292.

ABSTRACT

An algorithm for the nonlinearly constrained optimization problem is presented. The algorithm consists of a sequence of major iterations generated by linearizing each nonlinear constraint about the current point, and adding to the objective function a linear penalty for each nonlinear constraint. The resulting function is essentially the Lagrangian. A Kantorovich-type theorem is given, showing quadratic convergence in terms of major iterations. This theorem insures quadratic convergence if the starting point (or any subsequent point) satisfies a condition which can be tested using computable bounds on the objective and constraint functions.

A QUADRATICALLY CONVERGENT LAGRANGIAN ALGORITHM FOR NONLINEAR CONSTRAINTS

1. INTRODUCTION

Among the most appealing methods for nonlinear constraint problems, from a theoretical standpoint, are various versions of the cutting plane and penalty function methods, (see e.g., [McCormick, 71]). However, many of the penalty function methods become ill-conditioned due, for example, to the penalty becoming infinitely large. The cutting plane methods become ill-conditioned due to the linear dependence of the linearizations in the constraints. These difficulties are largely overcome by the algorithm presented earlier [Rosen, Kreuser, 71], and discussed in more detail in this paper. The penalty terms remain finite and the nonlinearities in the constraints are put into the objective function, [Beale, 67].

The algorithm is motivated by the existence of efficient computational algorithms for convex linearly constrained problems [Fletcher, 71], and reduces the original problem to a sequence of such problems (major iterations). A major iteration is generated by linearizing each nonlinear constraint about the current (infeasible) point, and adding to the objective function a linear external penalty for each violated nonlinear constraint. The resulting function is essentially the Lagrangian corresponding to these violated constraints.

A similar penalty is used by Kelley and Speyer (1970) in their improvement of an earlier method [Rosen, 61].

The idea of putting the nonlinearities in the objective was discussed by Beale (67) and a certain implementation was considered by Wilson (63).

Robinson (72) has developed quadratic convergence for a similar algorithm. A comparison of the two algorithms is discussed in a later section. The assumption here on the Hessian of the Lagrangian is stronger than in Robinson's paper, however here we obtain explicit estimates for the quadratic convergence neighborhood in terms of information at the point x^0 . The estimates here could be used as a stopping criteria with an explicit bound given on the distance from the final point obtained by the algorithm to the optimum point.

For the algorithm presented below a Kantorovich type theorem is given showing quadratic convergence in terms of major iterations. Computational results verify this quadratic convergence even when some of the assumptions of the theorem are not satisfied.

2. THE ALGORITHM AND ASSUMPTIONS

2.1 The Problem

The problem may be stated as

$$\text{Min } \phi_0(x)$$

subject to

$$\phi_j(x) \leq 0$$

$$x \in E^n$$

$\phi_j : E^n \to E^1$ are convex and differentiable, $j=0,1,\ldots,m$ and an optimal solution x^* is assumed to exist.

2.2 Definitions

For any fixed x^k we define the linearization of $\phi_j(x)$ about x^k as

$$h_j^k(x) \equiv \phi_j(x^k) + \nabla\phi_j'(x^k)(x-x^k) \;, \quad j=1,2,\ldots,m$$

where $\nabla\phi$ denotes the gradient (column) vector and $\nabla\phi'$

is its transpose. Also define the set of indices

$$I(x) \equiv \{j \mid \phi_j(x) \geq 0\} \text{ , and}$$

the corresponding Jacobian matrix

$$J = J(x) = [\nabla\phi_j(x)]_{j \in I(x)}$$

Finally, we denote by $J^{\dagger}(x)$ the generalized inverse of $J(x)$, where

$$J^{\dagger} = (J'J)^{-1}J' \text{ ,}$$

provided $J'J$ is non-singular. Also let

$$\phi'(x) = (\phi_1(x), \phi_2(x), \ldots, \phi_m(x))$$

Given

$$\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$$

Define

$$\psi(x,\lambda) \equiv \phi_0(x) + \sum_{j=1}^{m} \lambda_j \phi_j(x)$$

then

$$\nabla\psi(x,\lambda) = \nabla_x\psi(x,\lambda) = \nabla\phi_0(x) + \sum_{j=1}^{m} \lambda_j \nabla\phi_j(x)$$

and

$$\nabla^2\psi(x,\lambda) = \nabla_x^2\psi(x,\lambda) = \nabla^2\phi_0(x) + \sum_{j=1}^{m}\lambda_j\nabla^2\phi(x)$$

To simplify notation we let

$$H_j(x) = \nabla^2\phi_j(x) \qquad j=0,1,\ldots,m$$

$$H(x) = \nabla^2\psi(x,\lambda)$$

with the dependence on λ implied.

Given x^k and λ^k define $\tilde{x}^k$ to be that point for which

$$\nabla\psi(\tilde{x}^k,\lambda^k) = 0$$

Note that the existence of $\tilde{x}^k$ will follow from the strict convexity in x of $\psi(x,\lambda)$.

Define $\hat{x}^k$ as

$$\hat{x}^k - x^k = -J(x^k)(J'(x^k)J(x^k))^{-1}\phi(x^k)$$

Note that

$$J'(x^k)(\hat{x}^k - x^k) = -\phi(x^k)$$

so that $\hat{x}^k$ satisfies the linearized constraints. Furthermore $\hat{x}^k$ will exist since we assume that the columns of $J(x^k)$ (or gradients of constraints) are linearly independent.

2.3 The Algorithm

Step 0: Start with an arbitrary point x^0 .

Step 1: Given x^k compute λ^k by

$$\lambda^k = -J^{\dagger}(x^k)\nabla\phi_0(x^k)$$

Step 2: Compute x^{k+1} as optimum point of

$$\min_x \ \phi_0(x) + \sum_{\lambda_j^k>0} \lambda_j^k \phi_j(x)$$

subject to

$$h_j^k(x) \leq 0 \ , \qquad j=1,2,\ldots,m$$

Step 3: Are Kuhn-Tucker conditions for the problem 2.1 satisfied

No: $x^{k+1} \rightarrow x^k$, go to step 1.

Yes: $x^k = x^{k+1}$ STOP.

Remark 2.3.1

It is assumed that the problem has at least one nonlinear constraint. If there are any linear constraints they are included with the linearized constraints in the subproblem only and do not appear in the Lagrangian function. The point x^0 defines a set of m nonlinear constraints which are violated, $\phi_i(x^0) > 0$, $i=1,2,\ldots,m$. It is assumed that $1 \leq m \leq n$, and that these m constraints are the active nonlinear constraints at x^* . By active we mean that $\phi_i(x^*) = 0$, $\lambda_i^* > 0$ $i=1,2,\ldots,m$, where the λ_i^* are the corresponding set of optimal dual multipliers.

Remark 2.3.2

The convexity of the problem may allow a global convergence proof to be given or at least a global convergence criteria is available. If x^* and λ^* are the optimal pair, then because of the convexity of the constraint set we have

$$\psi(x^k,\lambda^{k-1}) \leq \psi(x^*,\lambda^*) = \phi_0(x^*)$$

so that if

$$\psi(x^k,\lambda^{k-1}) \geq \psi(x^{k-1},\lambda^{k-2}) + \sigma(\| x^k - x^{k-1} \|)$$

($\sigma(\cdot)$ is a forcing function) then we have a global convergence proof. This has been observed computationally. The idea is further explored in [Kreuser, 73].

Remark 2.3.3

We desire a least squares estimate of λ^k so that we might compute λ^k in general as

$$\min_{\lambda \geq 0} \| \nabla\phi_0(x^k) + J(x^k)\lambda \|_2^2$$

The point x^* is assumed to satisfy the strict complementary slackness assumption with linear independence in the gradients of the active constraint set. Therefore in the limit $(x^k \to x^*)$, which is our main concern here, the procedures are identical.

2.4 Assumptions

The following are assumed to hold on the set

$$\Omega_0(\rho) \equiv \{x \mid \| x - x^0 \| \leq \rho\}$$

for some $\rho > 0$.

2.4.1 $$\|(J'(x)J(x))^{-1}\| \leq \beta^2$$

where

$$J(x) = [\nabla\phi_1(x), \nabla\phi_2(x), \ldots, \nabla\phi_m(x)]$$

2.4.2 $$\|H_j(x)\| \leq \beta_H \qquad j=0,1,\ldots,m$$

2.4.3 $$\text{cond}(H_j(x)) \leq C_H \qquad j=1,2,\ldots,m$$

(cond(·) is the condition number of the matrix, i.e., the ratio of the maximum to the minimum eigenvalue) and $H_0(x)$ is positive semi-definite

2.4.4 $$\|\nabla\phi_i(x)\| \leq D \qquad i=0,1,\ldots,m$$

2.4.5 $$1 \leq \frac{D}{\|\nabla\phi_0(x)\|} \leq \sigma$$

Let

$$M_1 = \frac{\beta_H}{\sigma C_H}$$

$$M_2 = 2m^{1/2}\beta D\beta_H$$

$$R = M_2/M_1 = 2m^{1/2}\beta D\sigma C_H$$

$$\omega = \frac{5\beta\beta_H}{\sigma\ C_H} R^{5/2}$$

$$\theta(x^k) \equiv \beta R^{1/2} \| \phi(x^k) \| + \frac{2\sigma C_H}{\beta_H} R^{1/2} \| \nabla\psi(x^k,\lambda^k) \|$$

Note that $\theta(x)$ takes on the character of an optimality function [Mangasarian, 72].

3. CONVERGENCE THEOREM

Given the problem, 2.1, definitions, 2.2, algorithm, 2.3, and assumptions, 2.4, as previously stated, assume x^0 satisfies

3.0.1 $$\frac{\beta\beta_H}{2\sigma C_H} \| \phi(x^0) \| + \| \nabla\psi(x^0,\lambda^0) \| \leq \frac{1}{20\beta R^3}$$

$$(\text{or} \quad \theta(x^0) \leq \frac{1}{2\omega})$$

with $\rho \geq \frac{1}{\omega}$

Then the sequences $\{x^k\}$, $\{\tilde{x}^k\}$, and $\{\hat{x}^k\}$ exist and satisfy the following:

3.0.2 $$\| x^{k+1} - x^k \| \leq \theta(x^k) \leq \omega \| x^k - x^{k-1} \|^2$$

3.0.3 $$x^k \in \Omega_0 (\tfrac{1}{\omega}) \quad \text{and} \quad x^k \to x^* \in \Omega_0 (\tfrac{1}{\omega})$$

$$\text{with} \quad \| \phi(x^k) \| \to 0$$

3.0.4 $$\lambda^k \to \lambda^* \quad \text{with} \quad \|\nabla\psi^k\| \to 0$$

3.0.5 $$\| x^{k+1} - x^* \| \leq \frac{1}{\omega} \left(\frac{1}{2}\right)^{2^k}$$

Proof: The proof follows by establishing the following five lemmas:

Lemma 3.1: $R = \frac{M_2}{M_1}$ is an upper bound on the condition number of $\nabla_x^2 \psi(x, \lambda^k)$ for $x \in \Omega_0(\rho)$ and

$$\lambda^k = -(J'(x^k)J(x^k))^{-1} J'(x^k)\nabla\phi_0(x^k)$$

$$\text{with} \quad x^k \in \{x^k\}$$

Proof: We have

$$\nabla\phi_0(\tilde{x}^k) + J(\tilde{x}^k)\lambda^k = 0$$

so

$$\| \nabla\phi_0(\tilde{x}^k) \|^2 = \lambda'^k J'(\tilde{x}^k)J(\tilde{x}^k)\lambda^k \geq \frac{\| \lambda^k \|^2}{\beta^2}$$

or

$$\| \lambda^k \| \leq \beta \| \nabla\phi_0 \| \leq \beta D$$

equivalently

$$\| \lambda^k \|_1 \leq m^{1/2}\beta D$$

Also

$$\| \nabla\phi_0 \| = \| J\lambda^k \| \leq \Sigma |\lambda_i^k| \, \|\nabla\phi_i\|$$

$$\leq D \, \Sigma|\lambda_i^k| = D \, \| \lambda^k \|_1$$

so by 2.4.5

$$\frac{1}{\sigma} \leq \frac{\| \nabla\phi_0 \|}{D} \leq \| \lambda^k \|_1 \leq m^{1/2}\beta D$$

Now consider

$$\nabla^2\psi = H_0 + \sum_{i=1}^{m} \lambda_i^k H_i \qquad , \lambda_i^k \geq 0$$

Then for any y

$$M_1 \| y \|^2 = \frac{\beta_H}{\sigma C_H} \| y \|^2 \leq y' \nabla^2 \psi y \leq$$

$$\leq (1+m^{1/2}\beta D)\beta_H \| y \|^2$$

$$\leq 2m^{1/2}\beta D\beta_H \| y \|^2 = M_2 \| y \|^2$$

∎

Lemma 3.2:

$$\| x^{k+1} - x^k \| \le R^{1/2} \| \hat{x}^k - x^k \|$$

$$+ (1+R^{1/2}) \| \tilde{x}^k - x^k \|$$

Proof: We have

$$\| x^{k+1} - x^k \| \le \| x^{k+1} - \tilde{x}^k \| + \| x^k - \tilde{x}^k \|$$

Now: since $\hat{x}^k$ is feasible

$$\psi(x^{k+1}, \lambda^k) \le \psi(\hat{x}^k, \lambda^k)$$

Also

$$\psi(x^{k+1}, \lambda^k) = \psi(\tilde{x}^k, \lambda^k) + 1/2(x^{k+1} - \tilde{x}^k)^2 H(x')$$

and

$$\psi(\hat{x}^k, \lambda^k) = \psi(\tilde{x}^k, \lambda^k) + 1/2(\hat{x}^k - \tilde{x}^k)^2 H(x'')$$

therefore we have

$$M_1 \| x^{k+1} - \tilde{x}^k \|^2 \le M_2 \| \hat{x}^k - \tilde{x}^k \|^2$$

this gives

$$\| x^{k+1} - x^k \| \le R^{1/2} \| \hat{x}^k - \tilde{x}^k \|$$

$$+ \| x^k - \tilde{x}^k \|$$

and so

$$\| x^{k+1} - x^k \| \le R^{1/2} \| \hat{x}^k - x^k \|$$

$$+ (1+R^{1/2}) \| x^k - \tilde{x}^k \|$$

Lemma 3.3:

$$\| \tilde{x}^k - x^k \| \le \frac{1}{M_1} \| \nabla\psi(x^k,\lambda^k) \|$$

Proof: We have

$$\psi(x^k,\lambda^k) = \psi(\tilde{x}^k,\lambda^k) + 1/2(x^k - \tilde{x}^k)^2 H(x')$$

and

$$\psi(\tilde{x}^k,\lambda^k) = \psi(x^k,\lambda^k) + \nabla\psi'(x^k,\lambda^k)(\tilde{x}^k - x^k)$$

$$+ 1/2(\tilde{x}^k - x^k)^2 H(x'')$$

adding these we have

$$1/2(x^k - \tilde{x}^k)^2 H(x') + 1/2(\tilde{x}^k - x^k)^2 H(x'')$$

$$= -\nabla\psi'(x^k,\lambda^k)(\tilde{x}^k - x^k)$$

so

$$M_1 \| \tilde{x}^k - x^k \|^2 \le \| \nabla\psi(x^k,\lambda^k) \| \| \tilde{x}^k - x^k \|$$

If $\tilde{x}^k = x^k$ then the lemma holds trivially. Otherwise $\tilde{x}^k \neq x^k$ and

$$\| \tilde{x}^k - x^k \| \leq \frac{1}{M_1} \| \nabla\psi(x^k, \lambda^k) \|$$

∎

Lemma 3.4:

$$\| \hat{x}^k - x^k \| \leq 1/2 M_2 \beta m^{1/2} \| x^k - x^{k-1} \|^2$$

Proof:

$$\| \hat{x}^k - x^k \|^2 = \phi'(x^k)(J'(x^k)J(x^k))^{-1}\phi(x^k)$$

$$\leq \beta^2 \| \phi(x^k) \|^2$$

since

$$h_j^{k-1}(x^k) \leq 0 \qquad j=1,2,\ldots,m$$

we have

$$\phi_j(x^k) = h_j^{k-1}(x^k) + 1/2(x^k - x^{k-1})^2 H(x')$$

so

$$\| \phi(x^k) \|^2 \leq m(\frac{M_2}{2})^2 \| x^k - x^{k-1} \|^4$$

and

$$\| \hat{x}^k - x^k \| \leq 1/2M_2\beta m^{1/2} \| x^k - x^{k-1} \|^2$$

<u>Lemma 3.5:</u>

$$\| \nabla\psi(x^k,\lambda^k) \| \leq 2\beta M_2^2 \| x^k - x^{k-1} \|^2$$

Proof:

Let

$$J^{\dagger}(x^k) = (J'(x^k)J(x^k))^{-1}J'(x^k)$$

From Kuhn-Tucker conditions at x^k, we have

3.5.1 $$\nabla\psi(x^k,\lambda^{k-1}) = -J(x^{k-1})\mu^k$$

for some $\mu^k \geq 0$ so

$$\nabla\phi_0(x^k) + J(x^k)\lambda^{k-1} = -J(x^{k-1})\mu^k$$

and

$$-J^{\dagger}(x^k)\nabla\phi_0(x^k) - \lambda^{k-1} = J^{\dagger}(x^k)J(x^{k-1})\mu^k$$

and so

$$\lambda^k = \lambda^{k-1} + J^{\dagger}(x^k)J(x^{k-1})\mu^k$$

Now

$$\nabla\psi(x^k,\lambda^k) = \nabla\psi(x^k,\lambda^{k-1}) + J(x^k)(\lambda^k - \lambda^{k-1})$$

$$= -J(x^{k-1})\mu^k + J(x^k)J^{\dagger}(x^k)J(x^{k-1})\mu^k$$

$$= - [I - J(x^k)J^{\dagger}(x^k)]J(x^{k-1})\mu^k$$

$$= -P(x^k)J(x^{k-1})\mu^k$$

where $P(x^k) \equiv [I - J(x^k)J^{\dagger}(x^k)]$ is a projection operator.

Now since

$$P(x^k)J(x^k) = 0$$

and

$$||P(x^k)|| \leq 1$$

we have

3.5.2 $$|| \nabla\psi(x^k,\lambda^k)|| \leq ||J(x^k) - J(x^{k-1})|| \, ||\mu^k||$$

$$\leq M_2 ||\mu^k|| \, ||x^k - x^{k-1}||$$

We also have from 3.5.1

$$|| \nabla\psi(x^k,\lambda^{k-1})||^2 = \mu^{k'}J'(x^{k-1})J(x^{k-1})\mu^k$$

$$\geq \frac{1}{\beta^2} ||\mu^k||^2$$

so

3.5.3 $$\| \mu^k \| \le \beta \, \| \nabla\psi(x^k, \lambda^{k-1}) \|$$

We also have

$$\nabla\psi(x^{k-1}, \lambda^{k-1} + \mu^k) = \nabla\psi(x^{k-1}, \lambda^{k-1}) + J(x^{k-1})\mu^k$$

$$= \nabla\psi(x^{k-1}, \lambda^{k-1}) - \nabla\psi(x^k, \lambda^{k-1})$$

from 3.5.1.

Then since

$$\min_{\lambda} \| \nabla\psi(x^{k-1}, \lambda) \|$$

is attained for $\lambda = \lambda^{k-1}$ we have

3.5.4 $$\begin{aligned} \| \nabla\psi(x^{k-1}, \lambda^{k-1}) \| &\le \| \nabla\psi(x^{k-1}, \lambda^{k-1} + \mu^k) \| \\ &\le \| \nabla\psi(x^{k-1}, \lambda^{k-1}) - \nabla\psi(x^k, \lambda^{k-1}) \| \\ &\le M_2 \| x^k - x^{k-1} \| \end{aligned}$$

also

3.5.5 $$\begin{aligned} \| \nabla\psi(x^k, \lambda^{k-1}) \| &\le \| \nabla\psi(x^{k-1}, \lambda^{k-1} \| \\ &\quad + \| \nabla\psi(x^k, \lambda^{k-1}) - \nabla\psi(x^{k-1}, \lambda^{k-1}) \| \\ &\le 2M_2 \| x^k - x^{k-1} \| \end{aligned}$$

from 3.5.4.

Then combining 3.5.2, 3.5.3, and 3.5.5 we have

$$\| \nabla\psi(x^k,\lambda^k) \| \leq 2\beta M_2^2 \| x^k - x^{k-1} \|^2$$

∎

Combining the results of Lemmas 3.1-3.5 we obtain

$$\| x^{k+1} - x^k \| \leq \theta(x^k) \leq \omega \| x^k - x^{k-1} \|^2$$

Since

$$\theta(x^0) \leq \frac{1}{2\omega}$$

we have

$$x^k \in \Omega_0(\tfrac{1}{\omega}) \quad \text{and} \quad x^k \to x^* \in \Omega_0(\tfrac{1}{\omega})$$

Similarly $\lambda^k \to \lambda^*$ with $\| \nabla\psi^k \| \to 0$

The estimate

$$\| x^{k+1} - x^* \| \leq \frac{1}{\omega} (\tfrac{1}{2})^{2^k}$$

follows directly from the above as, for example in [Kantorovich and Akilov, 64].

∎

4. NUMERICAL EXAMPLE AND COMPUTATIONAL RESULTS

We first give a simple numerical example in 3 variables with 2 quadratic constraints and a linear objective function. The constants needed to determine the domain of convergence are easily obtained for this example. The problem is

4.1 $$\min_x \left\{ c'x \;\middle|\; \begin{array}{l} \phi_1(x) \leq 0 \\ \phi_2(x) \leq 0 \end{array} \right\}$$

where

$$c = \quad (-.65, -.5, -.7)$$

$$\phi_1(x) = .15x_1^2 + .2x_2^2 + .1x_3^2 - .45$$

$$\phi_2(x) = .25x_1^2 + .15x_2^2 + .3x_3^2 - .7$$

This has the optimal solution

$$x_1^* = x_2^* = x_3^* = 1.0$$

with

$$\lambda_1^* = .5 \qquad \lambda_2^* = 1.0$$

The following values of the bounds defined in section 2.4 are readily computed for x close to x^*.

We obtain

$$\beta \simeq 4.3$$

$$\beta_H \simeq .60$$

$$C_H \simeq 2$$

$$D \simeq 1.1$$

$$\sigma = 1$$

and

$$M_1 = \frac{\beta_H}{\sigma C_H} = .3$$

$$M_2 = 2m^{1/2}\beta D\beta_H = 7.81$$

$$R = 26.0$$

$$\omega = \frac{5\beta\beta_H}{\sigma C_H} R^{5/2} \simeq 22{,}237.$$

Thus quadratic convergence is guaranteed if x^0 is chosen so that (see 3.0.1).

4.2 $$.63 \,\|\phi(x^0)\| + \|\nabla\psi(x^0,\lambda^0)\| \leq .66 \times 10^{-6}$$

Furthermore, the minimum point x^* lies within a specified neighborhood of x^0, given by

4.3 $$\| x^{*}-x^{0} \| \leq \frac{1}{\omega} \approx .45 \times 10^{-4}$$

The above problem was run and the results are tabulated below.

k	x_1^k	x_2^k	x_3^k
1	.400000000000+001	.300000000000+001	.200000000000+001
2	.140814126335+001	.209147185050+001	.212616067179+001
3	.141540253018+001	.117819855510+001	.115405773821+001
4	.100344693378+001	.104418228872+001	.105338714695+001
5	.100177824399+001	.100017776652+001	.100032113343+001
6	.999999796560+000	.100000080687+001	.100000114609+001
7	.999999998328+000	.999999998674+000	.100000000516+001

k	λ_1^k	λ_2^k
1	.357401079274-001	.391755067482+000
2	.275816335414+000	.511728502746+000
3	.370833146850+000	.804316163533+000
4	.497025418582+000	.959390616610+000
5	.499789601640+000	.999061042181+000
6	.500000219458+000	.999999111523+000
7	.499999999953+000	.999999997883+000

k	$c'x^k$	$\psi(x^k,\lambda^{k-1})$	$\psi(x^k,\lambda^k)$
1	-.550000000000+001	-.550000000000+001	-.305991140607+001
2	-.344934022092+001	-.269906777978+001	-.220022306410+001
3	-.231695134025+001	-.203577323012+001	-.189138667240+001
4	-.191170265582+001	-.186181790886+001	-.185055577832+001
5	-.185146953675+001	-.185004705302+001	-.185000013909+001
6	-.185000107495+001	-.185000000423+001	-.185000000335+001
7	-.185000000335+001	-.185000000335+001	-.185000000335+001

k	$\|\| \phi(x^k) \|\|$	$\|\| \nabla\psi(x^k,\lambda^k) \|\|$
1	.717251699384+001	.297532849921+000
2	.215592360871+001	.193892243334+000
3	.485033297270+001	.104205265276+000
4	.567733292677-001	.234713633264-001
5	.131846727134-002	.780431236867-003
6	.960995496161-006	.637321935539-006
7	.565608340368-012	.142970029057-008

The problem was started with $x^0 = (4, 3, 2)$ which does not satisfy the quadratic convergence criterion. However, the sequence obtained does converge and if we take the initial point to be x^6 the test 4.2 is satisfied and we do get quadratic convergence from that point on. It is seen that x^6 also satisfies 4.3.

The algorithm described here has been tested computationally on a variety of nonlinear constraint test problems. A code [Kreuser, 71] based on Goldfarb's synthesis of Rosen's linearly constrained gradient projection and Davidon-Fletcher-Powell unconstrained minimization was used to solve the subproblem at each major iteration.

The largest problem solved thus far consisted of 30 variables and 30 quadratic constraints of which 20 were active at the optimum point.

Typical computational behavior is illustrated by the following tabulated results. The problem consisted of a linear objective function in 15 variables having 15 ellipsoidal constraints, of which 10 were active at the optimum. An interior starting point $(x^0 = 0)$ was chosen and the corresponding multipliers were zero $(\lambda^0 = 0)$.

k	CONSTRS. VIOLATED	$\|\| \phi(x^k) \|\|$	$\|\| x^k - x^* \|\|$	$\|\| \lambda^k - \lambda^* \|\|$	$\phi_0^k - \phi_0^*$	CUMULATIVE GRADIENT EVALS.	STD. TIME
0	0	.000	.387+1	.344+1	+.920+4	1	.000
1	15	.477+5	.155+2	.344+1	-.368+5	17	.028
2	15	.114+5	.620+1	.275+1	-.147+5	29	.131
3	15	.243+4	.191+1	.212+1	-.453+4	41	.219
4	15	.329+3	.314+0	.114+1	-.747+3	54	.308
5	10	.953+1	.118-1	.259+0	-.281+2	73	.388
6	10	.145-1	.181-4	.104-1	-.427-1	90	.455
7	0	.338-7	.232-8	.161-4	-.996-7	107	.523

(Note that this is the same problem for which results are presented in [Rosen, Kreuser, 71]. The results given here were obtained with a double precision version of the program, .477 + 5 means $.477 \times 10^5$).

The linearization about an interior point for convex constraints gives x^1 as an exterior point (in fact, all 15 constraints are violated). Bounds were placed on the variables so that the violation would not be too great (it still was $.477 \times 10^5$). The sequence of points still converged from the exterior of the domain as shown. The optimal function value is 9203, so that the relative error in ϕ_0^7 is approximately 1.1×10^{-11}. These calculations were done on a UNIVAC 1108, for which we have time (secs.) = 27 x (std. time). A gradient evaluation consists of computing the gradient of the Lagrangian function.

It should be noted that the sufficient conditions for quadratic convergence are certainly not satisfied by x^0 in this example. We do nevertheless get convergence which appears to be at least quadratic for $k = 5, 6,$ and 7. The behavior of other test problems was similar and details will be given in [Kreuser, 73].

5. EXTENSION AND COMPARISON

The algorithm presented here can be modified rather easily to handle the general nonlinear programming problem. Consider the general problem

$$\text{Min} \quad \phi_0(x)$$

subject to

$$\phi_j(x) \leq 0 \qquad j=1,2,\ldots,m$$

$$\phi_j(x) = 0 \qquad j=m+1,m+2,\ldots,\ell$$

The extended algorithm is then given by

$$\text{Pick } \varepsilon > 0$$

Given

$$x^k$$

Define

$$I(x^k) \equiv \{j \mid 1 \leq j \leq m \text{ and } \phi_j(x^k) \geq -\varepsilon\}$$

and corresponding Jacobian

$$J = J(x) = [\nabla\phi_j(x)]_{j \varepsilon I(x) \cup \{j \mid j=m+1,m+2,\ldots,\ell\}}$$

(See also [Kreuser, 73] for alternate definitions.)

and

$$J^{\dagger} = (J'J)^{-1}J'$$

Step 0: Start with an arbitrary point x^0 .

Step 1: Given x^k compute λ^k by

$$\lambda^k = -J^{\dagger}(x^k)\nabla\phi_0(x^k)$$

Step 2: Compute x^{k+1} as optimum point of

$$\operatorname*{Min}_x \ \phi_0(x) + \sum_{\substack{\lambda_j^k > 0 \\ 1 \leq j \leq m}} \lambda_j^k \ \phi_j(x) + \sum_{j=m+1}^{\ell} \lambda_j^k \ \phi_j(x)$$

subject to the complementary condition

$$\begin{cases} h_j^k(x) = 0 & \text{if} \quad \lambda_j^k > 0 \quad 1 \le j \le m \\ h_j^k(x) \le 0 & \text{otherwise} \end{cases}$$

and $h_j^k(x) = 0$ for $j=m+1,m+2,\ldots,\ell$

This algorithm has been tested successfully using some of Colville's problems [Colville, 68] where the convexity condition is not satisfied. It is shown in [Kreuser, 73] that under certain conditions on the Hessian of the Lagrangian that the same quadratic convergence is obtained.

The algorithm proposed by Robinson [Robinson, 72] will now be compared with the extended algorithm given above.

Robinson's algorithm for the subproblem is (given x^k and λ^k)

$$\min_x \; \phi_0(x) + \sum_j \lambda_j^k[\phi_j(x) - h_j^k(x)]$$

subject to

$$h_j^k(x) \le 0 \qquad j=1,2,\ldots,m$$

As we approach the optimum point x^* the constraints which are active $(\phi_j(x^*) = 0,\ \lambda_j^* > 0)$ at the optimum also satisfy the condition that the linearizations are active at points x^k close to x^* . So the subproblem is equivalent to

$$\min_x \phi_0(x) + \sum_j \lambda_j^k[\phi_j(x) - h_j^k(x)]$$

subject to the complementary condition

$$h_j^k(x) = 0 \qquad \text{if} \qquad \lambda_j^k > 0$$

$$h_j^k(x) \leq 0 \; , \quad \text{otherwise}$$

Since $\lambda_j^k \; h_j^k(x) = 0$, $1 \leq j \leq m$, it follows that the subproblem becomes

$$\min_x \phi_0(x) + \sum_{\lambda_j^k > 0} \lambda_j^k \; \phi_j(x)$$

subject to the complementary condition.

Thus, once x^k is close enough to x^* , the two objective functions are the same, provided the λ_j^k are the same.

We assume the same values of λ_j^k are used and we wish to compare the values of λ_j^{k+1} obtained.

Consider multipliers μ_j^{k+1} such that

$$\nabla\phi_0(x^{k+1}) + \Sigma\ \lambda_j^k \nabla\phi_j(x^{k+1}) = -\ \Sigma\ \mu_j^{k+1}\nabla\phi_j(x^k)$$

and multipliers ω_j^{k+1} such that

$$\nabla\phi_0(x^{k+1}) + \Sigma\ \lambda_j^k[\nabla\phi_j(x^{k+1}) - \nabla\phi_j(x^k)]$$

$$= -\Sigma\ \omega_j^{k+1}\nabla\phi_j(x^k)$$

In Robinson's algorithm the new λ_j 's are given by

$$\lambda_j^{k+1} = \omega_j^{k+1}$$

Equating the two expressions above gives

$$-\ \Sigma\mu_j^{k+1}\nabla\phi_j(x^k) = -\ \Sigma(\omega_j^{k+1} - \lambda_j^k)\nabla\phi_j(x^k)$$

so that if the $\nabla\phi_j$ are linearly independent we have

$$\mu_j^{k+1} + \lambda_j^k = \omega_j^{k+1}$$

then with

$$\lambda_j^{k+1} = \mu_j^{k+1} + \lambda_j^k$$

the algorithms would be identical in the limit.

The relationship between Newton's algorithm, the Rosen-Kreuser algorithm presented here, and the Robinson algorithm are considered in detail in [Kreuser, 73]. Considered also are various alternative techniques for computing the multipliers λ^k .

REFERENCES

[1] Beale, E. M. L., Numerical Methods, in "Nonlinear Programming," Ed. G. Abadie, North-Holland Publishing Company, Amsterdam, 1967.

[2] Colville, A. R., A Comparative Study on Nonlinear Programming Codes , IBM New York Scientific Center Report No. 320-2949, June, 1968.

[3] Fletcher, R., Minimization Subject to Linear Constraints, Conference on Numerical Methods for Nonlinear Optimization, Dundee, Scotland, June, 1971, Published in "Numerical Methods for Nonlinear Optimization," Ed. F. A. Lootsma, Academic Press, New York, 1972.

[4] Kantorovich, L. V. and Akilov, G. P., "Functional Analysis in Normed Spaces," Macmillan, New York, 1964.

[5] Kelley, H. J., and Speyer, J. L., Accelerated Gradient Projection, in "Proceedings of Colloquium on Optimization." Lectures on Mathematics 132. Springer-Verlag, 1970.

[6] Kreuser, J., Gradient Projection Code Manual, University of Wisconsin Computing Center, 1971.

[7] Kreuser, J., "Convergence Rates for Nonlinear Constraint Lagrangian Methods," Ph.D. Thesis, Computer Sciences Department, University of Wisconsin, 1973.

[8] Mangasarian, O. L., Dual, Feasible Direction Algorithms, Technical Summary Report No. 1173, Mathematics Research Center, University of Wisconsin, Madison, February, 1972.

[9] McCormick, Garth P., Penalty Function versus Non-penalty Function Methods for Constrained Nonlinear Programming Problems, Technical Paper RAC-TP-418, Research Analysis Corporation, June, 1971.

[10] Ortega, J. M. and Rheinboldt, W. C., "Iterative Solution of Nonlinear Equations in Several Variables," Academic Press, New York, 1970.

[11] Robinson, Stephen M., A Quadratically-Convergent Algorithm for General Nonlinear Programming Problems, Technical Summary Report No. 1215, Mathematics Research Center, University of Wisconsin, Madison, May, 1972.

[12] Rosen, J. B., The Gradient Projection Method for Nonlinear Programming. Part II, Nonlinear Constraints, J. Soc. Ind. Appl. Math., 9, 514-553.

[13] Rosen, J. B. and Kreuser, J., A Gradient Projection Algorithm for Nonlinear Constraints, Conference on Numerical Methods for Nonlinear Optimization, Dundee, Scotland, June, 1971, Published in "Numerical Methods for Nonlinear Optimization," Ed. F. A. Lootsma, Academic Press, New York, 1972, pp. 297-300.

[14] Wilson, R. B., "A Simplicial Algorithm for Concave Programming," Ph.D. Thesis, Graduate School of Business Administration, Harvard University, 1963.

BIBLIOGRAPHIC DATA SHEET	1. Report No. WIS-CS-166-72	2.	3. Recipient's Accession No.
4. Title and Subtitle A QUADRATICALLY CONVERGENT LAGRANGIAN ALGORITHM FOR NONLINEAR CONSTRAINTS			5. Report Date November, 1972
			6.
7. Author(s) J. B. Rosen and J. L. Kreuser			8. Performing Organization Rept. No.
9. Performing Organization Name and Address Computer Sciences Department The University of Wisconsin 1210 West Dayton Street Madison, Wisconsin 53706			10. Project/Task/Work Unit No.
			11. Contract/Grant No. GJ-362, GJ-32552, GJ-35292
12. Sponsoring Organization Name and Address National Science Foundation Washington, D. C.			13. Type of Report & Period Covered
			14.

15. Supplementary Notes

16. Abstracts

An algorithm for the nonlinearly constrained optimization problem is presented. The algorithm consists of a sequence of major iterations generated by linearizing each nonlinear constraint about the current point, and adding to the objective function a linear penalty for each nonlinear constraint. The resulting function is essentially the Lagrangian. A Kantorovich-type theorem is given, showing quadratic convergence in terms of major iterations. This theorem insures quadratic convergence if the starting point (or any subsequent point) satisfies a condition which can be tested using computable bounds on the objective and constraint functions.

17. Key Words and Document Analysis. 17a. Descriptors

Nonlinear Programming
Optimization
Nonlinearly Constrained Problems
Quadratic Convergence

17b. Identifiers/Open-Ended Terms

17c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
Available to Public	UNCLASSIFIED	33
	20. Security Class (This Page) UNCLASSIFIED	22. Price

FORM NTIS-35 (10-70) USCOMM-DC 40329-P71