An interior point technique for nonlinear optimization

José Herskovits

INRIA

UNITÉ DE RECHERCHE
INRIA-ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France
Tél. (1) 39 63 55 11

Rapports de Recherche

1992

ème
anniversaire

N° 1808

Programme 6
Calcul Scientifique, Modélisation et
Logiciels numériques

AN INTERIOR POINT TECHNIQUE FOR NONLINEAR OPTIMIZATION

José HERSKOVITS

Décembre 1992

RR-1808

UNE TECHNIQUE DE POINT INTERIEUR POUR L'OPTIMISATION NON LINEAIRE

José HERSKOVITS(*)
COPPE
Rio de Janeiro Brésil

Résumé

Nous proposons une méthode de minimisation d'une fonction régulière sous contraintes régulières d'égalités et d'inégalités par des algorithmes de points intérieurs. Elle consiste en la résolution itérative, suivant des variables primales et duales, de conditions d'optimalité du premier ordre de Karush-Kuhn-Tucker. Sur ce principe, des algorithmes de différents ordres peuvent être obtenus. Pour commencer, nous considérons le problème sous contraintes d'inégalités et nous présentons un algorithme de base globalement convergent. En particulier, des versions du premier ordre de type quasi-Newton sont établies. Nous considérons ensuite le problème général pour lequel un algorithme de base est proposé. Cette méthode est facile à implémenter car elle n'implique pas la solution de programmes quadratiques, mais seulement de systèmes linéaires d'équations. Plusieurs exemples démontrent la robustesse et l'efficacité de cette méthode.

AN INTERIOR POINT TECHNIQUE FOR NONLINEAR OPTIMIZATION

Abstract

We propose an approach for the minimization of a smooth function under smooth equality and inequality constraints by interior points algorithms. It consists on the iterative solution, in the primal and dual variables, of Karush-Kuhn-Tucker first order optimality conditions. Based on this approach, different order algorithms can be obtained. To introduce the method, in a first stage we consider the inequality constrained problem and present a globally convergent basic algorithm. Particular first order and quasi-Newton versions of the algorithm are also stated. In a second stage, the general problem is consider and a basic algorithm obtained. This method is simple to code, since it does not involve the solution of quadratic programs but merely that of linear systems of equations. Several applications show that it is also strong and efficient.

(*) Professeur invité dans les Projets MENUSIN et MODULEF de l'INRIA-Rocquencourt.

AN INTERIOR POINT TECHNIQUE FOR NONLINEAR OPTIMIZATION

José HERSKOVITS (*)
Mechanical Engineering Program
COPPE / Federal University of Rio de Janeiro
Caixa Postal 68503, 21945 Rio de Janeiro, BRAZIL.

Abstract

We propose an approach for the minimization of a smooth function under smooth equality and inequality constraints by interior points algorithms. It consists on the iterative solution, in the primal and dual variables, of Karush - Kuhn - Tucker first order optimality conditions. Based on this approach, different order algorithms can be obtained. To introduce the method, in a first stage we consider the inequality constrained problem and present a globally convergent basic algorithm. Particular first order and quasi-Newton versions of the algorithm are also stated. In a second stage, the general problem is consider and a basic algorithm is obtained. This method is simple to code, since it does not involve the solution of quadratic programs but merely that of linear systems of equations. Several applications show that it is also strong and efficient.

1. Introduction

This paper is concerned with the solution of the nonlinear constrained optimization problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{submitted to} \quad & g(x) \le 0 \\ \text{and} \quad & h(x) = 0, \end{aligned} \tag{1.1}$$

where $f \in R$, $g \in R^m$ and $h \in R^p$ are smooth functions in R^n.

(*) Before april 93 at INRIA, MODULEF Proj., B.P. 105, 78153 Le Chesnay Cedex,France.

We propose a technique that uses fixed point iterations to solve the nonlinear equations, in the primal and dual variables, given by the equalities included in Karush - Kuhn - Tucker optimality conditions. With the object of ensuring convergence to Karush - Kuhn - Tucker points, this is done in such a way to have the inequalities satisfied at each iteration. Based on the present approach, first order, quasi Newton or Newton algorithms can be obtained.

The algorithms studied here require an initial estimate of x, at the interior of the inequality constraints, and they generate a sequence of points also at the interior of these constraints. When only inequality constraints are considered, the objective is reduced at each iteration. In the general problem, an increase of the objective can be necessary to have the equalities satisfied.

The present method is simple to codify, strong and efficient. It does not involve penalty functions, active set strategies or Quadratic Programming subproblems. It merely requires to solve two linear systems with the same matrix at each iteration and to perform an inaccurate line search. In practical applications, more efficient algorithms can be obtained by taking advantage of the structure of the problem and particularities of the functions in it.

The ideas involved in this approach are first discussed in the framework of the inequality constrained problem and a globally convergent basic algorithm is proposed. Equality constraints are included later.

We say that this algorithm is basic because some of its procedures are very widely defined, allowing several alternatives. The first order algorithm and the quasi Newton one, described in references [3] and [5], can be considered as included in this family. This is also the case of the quasi Newton algorithms presented in [2], [4] and [8] and of an application in solid mechanics presented in [12].

We discuss in the following section the present approach in the case of inequality constrained optimization and present, in Section 3, a basic algorithm for this problem. Global convergence is studied in Section 4 and some particular algorithms are presented in the next section. In section 6 we discuss the inclusion of equality constraints and present a basic algorithm for the general nonlinear smooth optimization problem.

2. The inequality constrained optimization problem

In what follows, we discuss the basic concepts involved in the present approach when applied to the inequality constrained nonlinear programming problem

$$\begin{aligned} &\underset{x}{\text{minimize}} \quad f(x) \\ &\text{submitted to} \quad g(x) \le 0. \end{aligned} \tag{2.1}$$

These ideas will be further extended to problems including equality constraints.

Notation

We call $\lambda \in R^m$ the dual variables vector, $\Omega \equiv \{x \in R^n / \, g(x) \le 0\}$ the feasible set, Ω^0 its interior, $L(x,\lambda) = f(x) + \lambda^t g(x)$ the Lagrangian and $H(x,\lambda) = \nabla^2 f(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 g_i(x)$ its Hessian. $G(x)$ denotes a diagonal matrix such that $G_{ii}(x) = [g_i(x)]$.

Then, Karush - Kuhn - Tucker (KKT) first order optimality conditions are expressed as follows:

$$\nabla f(x) + \nabla g(x)\, \lambda = 0, \tag{2.2}$$

$$G(x)\, \lambda = 0, \tag{2.3}$$

$$g(x) \le 0 \text{ and} \tag{2.4}$$

$$\lambda \ge 0. \tag{2.5}$$

A vector (x^*,λ^*) satisfying KKT conditions will be called a *KKT Pair of the problem* and, a *Stationary Pair* if it only satisfies equalities (2.2) and (2.3). A vector x^* is a *KKT Point* if it exists λ^* such that (x^*,λ^*) constitutes a KKT pair and, we call it, *Stationary Point* if only (2.2) and (2.3) can be satisfied.

The following assumptions about the problem (2.1) are made:

Assumptions

Assumption 2.1. There exists a real number a such that the set $\Omega_a \equiv \{x \in \Omega;\ f(x) \leq a\}$ is a compact and has an interior Ω_a^0.

Assumption 2.2. Each $x \in \Omega_a^0$ satisfies $g(x) < 0$.

Assumption 2.3. The functions f and g are continuously differentiable in Ω_a and their derivatives satisfies Lipschitz condition.

Assumption 2.4. (*Regularity Condition*) For all $x \in \Omega_a$ the vectors of the set $\nabla g_i(x)$; for i such that $g_i(x) = 0$, are linearly independent.

Let us remind some well known some concepts [6], widely used in this paper.

Definitions

Definition 2.1 $d \in R^n$ is a *descent direction* of a real function ϕ at $x \in R^n$ if $d^t \nabla \phi < 0$.

Definition 2.2 $d \in R^n$ is a *feasible direction* of the problem (2.1), at $x \in \Omega$, if for some $\theta > 0$ we have $x + td \in \Omega$ for all $t \in [0,\theta]$.

Definition 2.3 A vector field d(x) defined on Ω is said to be an *uniformly feasible directions field* of the problem (2.1), if there exists $\tau > 0$ such that $x + td(x) \in \Omega$ for all $t \in [0,\tau]$.

It can be proved that d is a feasible direction if $d^t \nabla g_i(x) < 0$ for any i such that $g_i(x) = 0$. Definition 2.3 introduces a condition on the vectorial field d(x), which is strongest than the simple feasibility of any element of d(x). When d(x) constitutes an uniformly feasible directions field, it supports a feasible segment $[x, x + \theta(x)d(x)]$, such that $\theta(x)$ is bounded bellow in Ω by $\tau > 0$.

The algorithms that we obtain, for a given initial interior point, generate a sequence $\{x^k\}$ of interior points with decreasing values of the objective and converging to a KKT point x^* of the problem.

At each iteration it is defined a search direction d, which is a descent direction of the objective and also a feasible direction of Ω. A line search is then performed to ensure that the new point is interior and the objective is lower.

As a consequence of the requirement of feasibility, d must actually constitute an uniformly feasible directions field. Otherwise, the step length may go to zero, forcing convergence to points which are not KKT.

We solve the system of equations (2.2),(2.3) in (x,λ) by means of fixed point iterates. This is done in such a way that (2.4) and (2.5) are verified at each iteration. In this way, we obtain those solutions of (2.2), (2.3) that are Karush-Kuhn-Tucker pairs of the problem.

A Newton's iteration for the solution of (2.2), (2.3) is defined by the following linear system:

$$\begin{bmatrix} B & \nabla g(x) \\ \Lambda \nabla g^t(x) & G(x) \end{bmatrix} \begin{bmatrix} x_0 - x \\ \lambda_0 - \lambda \end{bmatrix} = - \begin{bmatrix} \nabla f(x) + \nabla g(x)\lambda \\ G(x)\lambda \end{bmatrix} \qquad (2.6)$$

where $B = H(x,\lambda)$, Λ is a diagonal matrix with $\Lambda_{ii} = \lambda_i$, (x,λ) is the current point and (x_0,λ_0) is a new estimate. Instead of $H(x,\lambda)$, B can be taken equal to a quasi - Newton estimate or to the identity matrix.

Then, depending on the way that $B \in R^{n \times n}$ symmetric is defined, (3.1) may represent a second order, a quasi - Newton or a first order iteration. However, as a requirement for global convergence, in the present approach B must be taken positive definite.

In what follows, we introduce some modifications on iteration (2.6) in a way to obtain, for a given interior pair (x,λ), a new interior estimate with the objective improved.

With this purpose, we define a direction d_0 in the primal space, as $d_0 = x_0 - x$. Then, (2.6) becomes

$$Bd_0 + \nabla g(x)\lambda_0 = -\nabla f(x) \quad \text{and} \tag{2.7}$$

$$\Lambda \nabla g^t(x) d_0 + G(x)\lambda_0 = 0, \tag{2.8}$$

which now gives a direction d_0 in the primal space and a new estimate of λ. It follows from (2.7), (2.8) that d_0 is zero in a Stationary Point. As it will be proved below, d_0 is a descent direction of f. However, d_0 is not useful as a search direction since it does not always constitute an uniformly feasible directions field. This is due to the fact that as any constraint goes to zero, (2.8) forces d_0 to tend to a direction tangent to the feasible set. In fact, (2.8) is equivalent to

$$\lambda_i \nabla g_i^t(x) d_0 + g_i(x)\lambda_{0i} = 0;\ i=1,m, \tag{2.9}$$

which implies that $\nabla g_i^t(x) d_0 = 0$ for i such that $g_i(x) = 0$.

To avoid this effect, we define the linear system in d and $\bar{\lambda}$

$$Bd + \nabla g(x)\bar{\lambda} = -\nabla f(x)$$

$$\Lambda \nabla g^t(x) d + G(x)\bar{\lambda} = -\rho \Lambda \omega, \tag{2.10}$$

obtained by adding a negative vector in the right side of (2.8), where the scalar factor ρ and $\omega \in R^m$ are positive and $\bar{\lambda}$ is the new estimate of λ. In this case, (2.10) is equivalent to

$$\lambda_i \nabla g_i^t(x) d + g_i(x)\bar{\lambda}_i = -\rho \lambda_i \omega_i;\ i=1,m,$$

and $\nabla g_i^t(x) d = -\rho\omega_i$ for the active constraints. Thus, d is a feasible direction.

We can consider that the inclusion of a negative number in the right hand of (3.4), produces the effect of deflecting d_0 in the sense of the interior of the feasible region, being the deflection of d_0 relative to the i-th constraint proportional to $\rho\omega_i$.

Finally, as the deflection of d_0 is proportional to ρ and d_0 is descent, by establishing upper bounds on ρ it is possible to ensure that d is a descent direction also. Since $d^t\nabla f(x) < 0$, we obtain these bounds by imposing

$$d^t\nabla f(x) \leq \alpha\, d_0^t\nabla f(x), \tag{2.11}$$

which implies $d^t\nabla f(x) < 0$.

Condition (2.11) implies that d is in a circular cone whose axis is $\nabla f(x)$. In general, the derivative of f in the direction of d will be smaller than in the direction of d_0. This is a price that we pay to obtain a feasible descent direction.

Note that d can be obtained by solving

$$Bd_1+\nabla g(x)\lambda_1= 0, \tag{2.12}$$

$$\Lambda\nabla g^t(x)d_1 + G(x)\lambda_1 = -\Lambda\omega \tag{2.13}$$

and computing

$$d = d_0+ \rho\, d_1.$$

In this way, it is easy to establish bounds on ρ to ensure that (2.11) holds.

As a requirement for d(x) to be an uniformly feasible directions field, ρ must be also bounded below.

The ideas pointed above are a basis for the iterative method that we are studying. To determine a new primal point, an inaccurate line search is done in the direction of d, requiring feasibility and a satisfactory decrease of the objective. Different updating rules can be adopted to define a new positive λ.

3. A basic algorithm

In this section we present a basic algorithm for inequality constrained problems that globally converges in the primal space to Karush-Kuhn-Tucker points of the problem. It is basic in the sense that various procedures of the algorithm are very widely defined. This fact, turns possible the implementation of different versions depending of the problem to be solved, the available information about f and g, and the desired rate of convergence.

The algorithm is stated as follows:

ALGORITHM I

Parameters. $\alpha \in (0,1)$, $\eta \in (0,1)$, $\varphi > 0$ and $\nu \in (0,1)$.
Data. $x \in \Omega_a^0$, $\lambda > 0$,$B \in R^{n\times n}$ symmetric and positive definite and $\omega \in R^m$ positive.

Step 1. Computation of a search direction.

(i) Compute (d_0, λ_0) by solving the linear system

$$Bd_0+\nabla g(x)\lambda_0= - \nabla f(x), \tag{3.1}$$

$$\Lambda\nabla g^t(x)d_0 + G(x)\lambda_0 = 0. \tag{3.2}$$

If $d_0= 0$, stop.

(ii) Compute (d_1, λ_1) by solving the linear system

$$Bd_1+\nabla g(x)\lambda_1= 0, \tag{3.3}$$

$$\Lambda\nabla g^t(x)d_1 + G(x)\lambda_1 = - \Lambda\omega. \tag{3.4}$$

(iii) If $d_1^t\nabla f(x) > 0$, set

$$\rho = \inf \{\varphi \|d_0\|^2;\ (\alpha-1)\ d_0^t \nabla f(x)/d_1^t \nabla f(x)\}\ ,\ \text{or} \tag{3.5}$$

$$\rho = \varphi \|d_0\|^2 \ \text{otherwise.} \tag{3.6}$$

(iv) Compute the search direction

$$d = d_0 + \rho\ d_1,\ \text{and} \tag{3.7}$$

$$\bar{\lambda} = \lambda_0 + \rho\ \lambda_1 \tag{3.8}$$

Step 2. Line search.

Compute t, the first number of the sequence

$\{1, \nu, \nu^2, \nu^3, \ldots\}$ satisfying

$$f(x + td) \le f(x) + t\ \eta\ \nabla f^t(x)d \quad \text{and} \tag{3.9}$$

$$g_i(x + td) < 0 \ \text{if}\ \bar{\lambda}_i \ge 0,\ \text{or} \tag{3.10}$$

$$g_i(x + td) \le g_i(x) \ \text{otherwise.} \tag{3.11}$$

Step 3. Updates.

(i) Set

$x := x + td$

and define new values for

$\omega > 0$,

$\lambda > 0$ and

B symmetric and positive definite.

(ii) Go back to *Step 1.* ■

As we are going to show, ρ defined by (3.5) (3.6) is finite in Ω_a and (2.11) holds.

The algorithm includes, in *Step 2*, an inexact line search based on Armijo's procedure for unconstrained optimization [6]. In addition to (3.9), which ensures a reasonable decrease of the function, conditions (3.10) and (3.11) impose to the new primal point to be interior. Moreover, (3.11) prevents saturation of the constraints associated to negative dual variables, as it will be required to prove convergence to Karush - Kuhn - Tucker points.

Different algorithms can be obtained according to the way of updating λ, B and ω. In what follows we introduce some assumptions about λ, B and ω.

Assumptions

Assumption 3.1. There exist positive numbers λ^I, λ^S and β such that $0 \le \lambda \le \lambda^S$ and $\lambda_i \ge \lambda^I$ for any i such that $g_i(x) \ge -\beta$.

Assumption 3.2. There exist positive numbers σ_1 and σ_2 such that $\sigma_1 \|d^2\| \le d^t B\ d \le \sigma_2 \|d^2\|$ for any $d \in R^n$.

Assumption 3.3. There exist positive numbers ω_1 and ω_2 such that $\omega_1 \le \omega \le \omega_2$.

In the next section we prove that any sequence $\{x^k\}$ generated by the algorithm converges to a Karush - Kuhn - Tucker point of the problem for any way of updating λ, B and ω, provided that the previous assumptions are true. Moreover, (x^k, λ_0^k) converges to a Karush - Kuhn - Tucker pair. Depending on the way of updating λ, global convergence in the dual space can also be obtained.

4. Global convergence

The theoretical analysis to be presented now, includes first a proof that the algorithm never fails, then a study of the case when it stops after a finite number of iterations and finally the actual study of convergence.

The algorithm doesn't fails if the solution of the linear systems (3.1), (3.2) and (3.3), (3.4) is unique. This is a consequence of a lemma proved in [8] and stated as follows:

Lemma 4.1. Given any $x \in \Omega_a$, any positive definite symmetric matrix B and any positive λ such that $\lambda_i > 0$ if $g_i(x) = 0$, the matrix

$$M(x,\lambda,B) \equiv \begin{bmatrix} B & \nabla g(x) \\ \Lambda \nabla g^t(x) & G(x) \end{bmatrix}$$

is nonsingular. ■

In consequence, d_0, λ_0, d_1 and λ_1 obtained in *Step 1* of the algorithm are uniquely determined. Moreover, since x is in a compact and λ and B are bounded, it follows that M is bounded out of zero and then, d_0, λ_0, d_1 and λ_1 have an upper bound.

We consider now the special case when $d_0 = 0$ is obtained in the *Step 1* of the algorithm and, in consequence, it stops. Since all the iterates are strictly feasible, it follows from (3.1) and (3.2) that $\nabla f(x) = 0$. Thus, if the algorithm stops, x is a particular Karush - Kuhn - Tucker point. From now on, the case when the algorithm never stops is studied.

Lemma 4.2. The vector d_0, defined in step 1 of the algorithm, satisfies

$$d_0^t \nabla f(x) \leq - d_0^t B\, d_0.$$

Proof. It follows from (3.2) and Assumption 3.1 that $\lambda_i = 0$ implies $\lambda_{0i} = 0$. Then, (3.1) and (3.2) are equivalent to

$$Bd_0 + \nabla g'(x)\lambda_0' = - \nabla f(x), \tag{4.1}$$

$$\Lambda' \nabla g'^t(x) d_0 + G'(x)\lambda_0' = 0, \tag{4.2}$$

where $(\cdot)'$ means that only constraints associated to nonzero λ's are considered. Scalar multiplication of both sides of (4.1) by d_0 yields

$$d_0^t \nabla f(x) = - d_0^t B d_0 - d_0^t \nabla g'(x)\lambda_0',$$

so that, in view of (4.2),

$$d_0^t \nabla f(x) = - d_0^t B\, d_0 + \lambda_0'^t \Lambda'^{-1} G'(x)\lambda_0'.$$

The result of the lemma follows from Assumption 3.2 on B and the fact that $\Lambda'^{-1}G'(x)$ is negative definite. ■

As a consequence of the preceding lemma, we have that d_0 is a descent direction of f.

The updating rule for ρ, stated in *Step 1* of the algorithm, ensures that

$$\rho \le \varphi \|d_0\|^2. \tag{4.3}$$

On the other hand, from Lemma 4.2 and Assumption 3.2, we get

$$- d_0^t \nabla f(x) \ge \sigma_1 \|d_0\|^2$$

and, in view of (3.5),

$$\rho \ge \inf\ [\varphi;\ (1-\alpha)\sigma_1 / d_1^t \nabla f(x)] \|d_0\|^2$$

in the case when $d_1^t \nabla f(x) > 0$. Since d_1 is bounded and (3.6), we deduce that it exists $\varphi_0 > 0$ such that

$$\rho \ge \varphi_0 \|d_0\|^2 \tag{4.4}$$

is always true. Since d_0 is bounded, it follows from (4.3) and (4.4) that ρ is positive and bounded above.

Considering now (4.3) together with (3.7), it can be shown that it exists $\delta > 1$ such that the inequality

$$\|d\| \le \delta \|d_0\| \tag{4.5}$$

is always verified by d_0 and d.

Lemma 4.3. The search direction d satisfies

$$d^t \nabla f(x) \le \alpha\ d_0^t \nabla f(x).$$

Proof. In consequence of (3.7) we have

$$d^t\nabla f(x) = d_0^t\nabla f(x) + \rho\ d_1^t\nabla f(x),$$

that, together with (3.5) and (3.6), yields the result of the lemma. ∎

Since d_0 is a descent direction of f, the last result implies that d also it is.

It follows from the preceding lemma that $d^t\nabla f(x)$ is never zero. This is due to the fact that $d^t\nabla f(x) = 0$ implies $d_0^t\nabla f(x) = 0$ and, from Lemma 4.2, $d_0 = 0$. Thus, the algorithm stops after computing d_0.

Proposition 4.4. Let be a real function Φ in R^n continuously differentiable and such that $\nabla\Phi$ satisfies Lipschitz condition in $\Gamma \in R^n$, that is, there exists a positive constant k such that for all $x,y \in \Gamma$, it is

$$\|\nabla\Phi(y) - \nabla\Phi(x)\| \le k\|y - x\|.$$

Then, the following condition is true:

$$\Phi(y) \le \Phi(x) + (y-x)^t\nabla\Phi(x) + k\|y - x\|^2.$$

Proof. The Mean Value Theorem yields

$$\Phi(y) = \Phi(x) + (y - x)^t\nabla\Phi\ [x + \xi(y - x)]$$

for some $\xi \in (0,1)$. Then,

$$\Phi(y) \le \Phi(x) + (y-x)^t\nabla\Phi(x) + \|y - x\|\|\nabla\Phi[x + \xi(y - x)] - \nabla\Phi(x)\|$$

and the result follows from Lipschitz condition. ∎

Lemma 4.5. There exists $\tau > 0$ such that the conditions (3.9) - (3.11) are verified for any $t \in [0,\tau]$ at any $x \in \Omega_a$ and d computed in *Step 1* of the algorithm.

Proof. Let us consider first condition (3.10), which applies in the case when $\bar{\lambda}_i \ge 0$. It follows from Proposition 4.4 that there exists $k_i > 0$ such that

$$g_i(x + td) \le g_i(x) + t\, d^t \nabla g_i(x) + t^2 k_i \|d\|^2 \tag{4.6}$$

holds for any x and x + td in Ω_a. Therefore, if

$$g_i(x) + t\, d^t \nabla g_i(x) + t^2 k_i \|d\|^2 < 0, \tag{4.7}$$

then (3.10) is true. By (2.10), this condition is equivalent to

$$g_i(x)(1 - t\bar{\lambda}_i / \lambda_i) - \rho\omega_i t + t^2 k_i \|d\|^2 < 0,$$

in the case when $\lambda_i > 0$. The last inequality is verified if

$$1 - t\bar{\lambda}_i / \lambda_i > 0$$

and

$$\rho\omega_i - t k_i \|d\|^2 > 0.$$

Both inequalities hold for any t such that

$$t < \inf\{\lambda_i / \bar{\lambda}_i ;\ \rho\omega_i / k_i \|d\|^2 \}. \tag{4.8}$$

Thus, since $\bar{\lambda}_i$ is bounded, it follows from (4.4), (4.5) and Assumption 3.1 that it exists $\tau_i > 0$ such that (3.10) is satisfied for any $t \in [0, \tau_i]$.

In the case when $\lambda_i = 0$, we have by Assumption 3.1 that $g_i(x) \le \beta$. It follows from (3.2), (3.4) and (3.8) that $\bar{\lambda}_i = 0$. Thus, condition (3.10) is binding in *Step 2* of the algorithm. Since g is continuously differentiable and d is bounded, the previous result is also obtained in consequence of (4.7).

Following the same procedure as before, when $\bar{\lambda}_i < 0$ it can be proved the existence of $\tau_i > 0$ such that (3.11) is satisfied for any $t \in [0, \tau_i]$.

In consequence of Proposition 4.4, there exists $k_f > 0$ such that

$$f(x + td) \le f(x) + t d^t \nabla f(x) + t^2 \|d\|^2 k_f$$

holds for any $x, x + td \in \Omega_a$. Thus, condition (3.9) is verified for any t such that

$$f(x) + td^t\nabla f(x) + t^2\|d\|^2 k_f \leq f(x) + t\,\eta\, d^t\nabla f(x)$$

or, equivalently,

$$t \leq (\eta - 1)d^t\nabla f(x)/k\|d\|^2. \qquad (4.9)$$

It follows from (4.5), Lemmas 4.2 and 4.3 and Assumption 3.2 that (4.9) holds for any $t \leq (1 - \eta)\alpha\sigma_1/\delta^2$. Thus, it exists $\tau > 0$ which satisfies the requirements of the present lemma. ■

In the case when a constraint g_p is linear, the preceding result is valid even taking $\omega_p = 0$, which implies that no deflection relative to g_p is required. This can be shown by taking

$$g_p(x + td) = g_p(x) + t\, d^t\nabla g_p(x)$$

instead of (4.6). Following a procedure similar to that which leads to (4.8), it can be deduced that (3.10) is satisfied for any $t < \lambda_p/\bar{\lambda}_p$ for any $\omega_p \geq 0$. In the case when $\bar{\lambda}_p < 0$, (3.11) holds for any $t > 0$.

Remark that, if B, λ and ω are taken constant, we can consider d as a vector field in Ω. In consequence of the previous lemma, it can be proved that this vector field is an uniformly feasible directions field.

Any sequence $\{x^k\}$ generated by the algorithm is contained in Ω_a. Then, since Ω_a is a compact, $\{x^k\}$ has accumulation points in Ω_a. In what follows, we are first going to prove that this points are stationary points of the problem and then, that there are in fact Karush - Kuhn - Tucker points.

Lemma 4.6. Any accumulation point x^* of the sequence $\{x^k\}$, generated by the algorithm, is a stationary point of the problem. Moreover, $(x^*,\lambda_0(x^*))$ constitutes a stationary pair. We call $\lambda_0^* = \lambda_0(x^*)$.

Proof. Let be a sequence $\{x^k\}_{k\in K}$, where $K \subset N = 1,2,3,..,$ converging to x^*. Since λ, B,ω and ρ are bounded, there exists $K_1 \subset K$ such that

$\{x^k,\lambda^k,B^k,\omega^k,\rho^k\}_{k \in K_1}$ converges to $\{x^*,\lambda^*,B^*,\omega^*,\rho^*\}$. Considering that d continuously depends of x, λ, B, ω and ρ, we deduce that $\{d^k\}_{k \in K_1} \to d^*$, being $d^* = d(x^*,\lambda^*,B^*,\omega^*,\rho^*)$.

Consider now $K_2 \subset K_1$ such that $\{t^k\}_{k \in K_2} \to t^*$. It follows from Lemma 4.5 that $t^* > 0$. Condition (3.9) can be written

$$f(x^{k+1}) \le f(x^k) + \eta t^k d^{kt} \nabla f(x^k),$$

where we assume that $k \in K_2$ and k+1 does not necessarily belong to K_2. Let us call fol(k) the element which follows k in K_2. Then, since fol(k) ≥ k+1, it is

$$f(x^{fol(k)}) \le f(x^k) + \eta t^k d^{kt} \nabla f(x^k)$$

and, taking limits in both sides for $k \to \infty$, we have

$$f(x^*) \le f(x^*) + \eta t^* d^{*t} \nabla f(x^*),$$

thus, $d^{*t}\nabla f(x^*) \ge 0$. Considering now the result of Lemma 4.3 in the limit for $k \to \infty$, we get $d^{*t}\nabla f(x^*) = 0$. It also follows from Lemma 4.3 that $d_0^{*t}\nabla f(x^*) = 0$ and, from Lemma 4.2, that $d_0^* = 0$. Thus, (x^*,λ_0^*) is a Stationary Pair.

■

Next, we are going to prove global convergence to a Karush - Kuhn - Tucker point of the problem. However, this result requires an additional hypothesis about the problem.

Assumption 4.1. The stationary points of the Problem (2.1) are isolated points or they constitute isolated compact sets with the same active constraints.

Theorem 4.7. Any accumulation point x^* of any sequence generated by the algorithm is a Karush - Kuhn - Tucker point of the problem.

Proof. As x^* is a stationary point of the problem, it is only necessary to prove that the Lagrange multipliers $\lambda_0(x^*)$ are non negative.

Consider a constraint g_h such that $g_h(x^*) = 0$. As the method is strictly feasible, $g_h(x^k) < 0$ for all $k \in N$, where $N = 1,2,3,\ldots$ Then, we can define a sequence $\{x^k\}_{k \in K}$, $K \subset N$, converging to x^* and such that $g_h(x^k) > g_h(x^{k-1})$ for any $k \in K$. Note that k-1 may not belong to K. It follows from (3.11) that, for such a sequence, $\bar{\lambda}^{k-1} \geq 0$.

Now, we proceed by contradiction and assume that $\lambda_{0h}(x^*) < 0$. It follows from (4.3) that $\bar{\lambda}_h(x^*) < 0$ also. Then, since Assumption 4.1 holds, it exists a ball $\Gamma(\gamma_h) \equiv \{x / \|x - x^*\| \leq \gamma_h\}$ such that $\bar{\lambda}_h < 0$ for any $x \in \Omega_a^0 \cap \Gamma(\gamma_h)$ and any λ, B, ω, and ρ generated by the algorithm. As $\bar{\lambda}^{k-1} \geq 0$, for any $k \in K$, we have that $\{x^{k-1}\}_{k \in K} \subset \Omega_a^0 - \Gamma(\gamma_h)$, which is a compact. Then, $\{x^{k-1}\}_{k \in K}$ has accumulation points, which are stationary points of the problem. Thus, it exists I such that $\|d^{k-1}\| < \gamma/2$ for any $k > I$, $k \in K$ and, in consequence of the line search, $\|x^k - x^{k-1}\| \leq \|d^{k-1}\| < \gamma/2$. But this result is in contradiction with the fact that the distance between the accumulation points of $\{x^k\}_{k \in K}$ and $\{x^{k-1}\}_{k \in K}$ is greatest than γ. ■

5. Some particular algorithms of the family

Some alternative updating rules for λ, B and ω in the basic algorithm stated above will be discussed in the present section. They lead to particular algorithms with different performances in terms of local or global speed of convergence.

Update of λ. In the case of the dual variables λ, the following updating rule can be stated in *Step 3*:

Set, for i = 1,m,

$$\lambda_i := \sup\ [\lambda_{0i};\ \varepsilon\ \|d_0\|^2\ ]. \tag{5.1}$$

If $g_i(x) \geq -\beta$ and $\lambda_i < \lambda^I$, set $\lambda_i = \lambda^I$. ■

The parameters ε, β and λ^I are taken positive. In this rule, λ_i is a second order perturbation of λ_{0i}, given by Newton's iteration (2.6). If β and λ^I are taken small enough, after a finite number of iterations, λ_i for the active constraints becomes equal to λ_{0i}.

If Assumptions 3.2 and 3.3 are verified, λ defined above satisfies Assumption 3.1. In effect, since λ_0 and d_0 are bounded, λ_i has a positive upper bound λ^S.

Far from the solution it is convenient to have a search direction not pointing to the constraints but slipping on their boundary. In this way the steps will be longer and, consequently, the efficiency of the algorithm improved.

This effect can be obtained by increasing the dual variable as the corresponding constraint goes to zero since, as it follows from (2.9), $\nabla g_i^t(x)d_0$ becomes smaller as λ_i grows. The following rule satisfies this requirement in a very simple way:

Set, for i = 1,m,

$$\lambda_i := -g_i^{-1}(x). \qquad (5.2)$$

If $\lambda_i > \lambda^S$, set $\lambda_i = \lambda^S$. ■

We got this updating rule based on Dikin's algorithm for Linear Programming, presented in [1] and also studied in [11]. In the case of linear programs, d_0 obtained with this rule is the same that the search direction given by Dikin's algorithm. In effect, since $H(x,\lambda) = 0$, it follows from (2.7), (2.8) and (5.2) that

$$\nabla g(x)\lambda_0 = -\nabla f(x) \quad \text{and} \qquad (5.3)$$

$$-G(x)^{-1}\nabla g^t(x)d_0 + G(x)\lambda_0 = 0. \qquad (5.4)$$

Substitution of λ_0 from (5.4) in (5.3) gives the linear system in d_0

$$[\nabla g(x)G(x)^{-2}\nabla g^t(x)]d_0 = -\nabla f(x),$$

that shows the equivalency with Dikin's algorithm, [11].

Update of B. In *ALGORITHM I,* B is a symmetric and positive definite $R^{n\times n}$ matrix and has to satisfy Assumption 3.2. It can be taken B equal to the identity, to an estimate of $H(x,\lambda)$ or, in the case when $H(x,\lambda)$ is positive definite, to $H(x,\lambda)$ itself.

A quasi Newton estimate, following the same procedures as in Successive Quadratic Programming (SQP) algorithms, can be constructed. We use the BFGS formula with Powell's modification to ensure positive definiteness [9], although it is not clear whether Assumption 3.2 will then be always satisfied.

Update of ω. As it was pointed before, the deflection of d_0 relative to the i-th constraint is proportional to ω_i. Let us call, for $x \in \Omega$ and $d \in R^n$ given, $\theta_i = \max \{ t; g_i(x + td) \le 0 \}$. It should be convenient to chose ω in a way to have similar resulting values for all θ_i; $i = 1, m$. As it follows from (4.8) in Lemma 4.5, this requires the use of second order information about the constraints. When this information is not available, some kind of estimate can be done.

Asymptotic convergence. A study about about the local speed of convergence of these algorithms is not presented in this paper, but only some comments about this matter are made. In the case when (5.1) and BFGS rules are used for λ and B, the convergence is two-steps superlinear, provided a unitary step length is obtained after a finite number of iterations. This result can be obtained in a similar way as in Theorem 4.6 in [8], by showing that the search directions of the present method and of the SQP algorithm locally differ by a term that goes to zero faster than $\|d_0\|$ and then, extending to the present algorithm the proof of two-steps superlinearly convergence of the SQP method, in [9]. However, to satisfy the requirement about the step length, d must be such that Definition 2.2 holds for some $\theta > 1$. It is clear that θ increases whenever ρ grows but, unfortunately, in some problems the upper bound on ρ may be not large enough to allow the step length to be unitary. This effect, which is similar to Maratos' effect [10], in theory can produce rates of convergence lower than superlinear.

In a quasi - Newton algorithm described in [2], Maratos' effect is avoided by looking at each iteration for a decrease of an estimate of the Lagrangian, instead of the objective function. However, global convergence is not strongly proved, since the estimate Lagrange multipliers changes at each iteration and oscillations between several accumulation points are possible.

An algorithm that also solves optimality conditions by means of fixed point iterates was presented in [8]. The authors obtained global and local superlinear convergence by applying a technique presented by Mayne and Polak [7] in a different context. First, as in the present approach, a feasible descent direction d is obtained. Then, the constraints are computed at $(x + d)$ and an approximate projection, $\tilde{x}$, of $(x + d)$ on the active constraints is found. Finally, a new primal variable is determined by doing a search along a parabola which is tangent to d and contains $\tilde{x}$. In this search, a decrease of the objective and the feasibility of the new iterate are required.

6. Including Equality Constraints

We consider in this section the general nonlinear programming problem (1.1). A technique will be proposed to extend the domain of application of the present approach to problems including equality constraints. The simplest way probably consists on defining a suitable penalty function of the equalities and then, minimizing that function submitted only to the inequality constraints. Unfortunately, this approach brings up numerical problems encountered in penalty methods.

In what follows, we discuss our approach to solve Karush - Kuhn - Tucker conditions of problem (1.1) and present an algorithm based on this approach. This algorithm requires an initial point at the interior of Ω, defined in

Section 2, not necessarily verifying the equality constraints. It generates a sequence $\{x^k\}$ of interior points ($\{x^k\} \in \Omega^0$) which converges to a Karush - Kuhn - Tucker point x^* of the problem. In general, the equalities are only active at the limit. Then, to have the equalities satisfied, an increase of the objective function can be required.

Karush - Kuhn - Tucker first order optimality conditions of problem (1.1) can be expressed as follows:

$$\nabla f(x) + \nabla g(x)\,\lambda + \nabla h(x)\,\mu = 0, \tag{6.1}$$

$$G(x)\,\lambda = 0, \tag{6.2}$$

$$h(x) = 0, \tag{6.3}$$

$$g(x) \le 0 \text{ and} \tag{6.4}$$

$$\lambda \ge 0. \tag{6.5}$$

where $\mu \in R^p$ is the dual variables vector corresponding to the equality constraints. Now, the Lagrangian is $L(x,\lambda,\mu) = f(x) + \lambda^t g(x) + \mu^t h(x)$, and its second derivative becomes $H(x,\lambda,\mu) = \nabla^2 f(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 g_i(x) + \sum_{i=1}^{p} \mu_i \nabla^2 h_i(x)$.

A Newton's iteration for the solution of (6.1) to (6.3) is defined by the following system:

$$\begin{bmatrix} B & \nabla g(x) & \nabla h(x) \\ \Lambda \nabla g^t(x) & G(x) & 0 \\ \nabla h^t(x) & 0 & 0 \end{bmatrix} \begin{bmatrix} x_0 - x \\ \lambda_0 - \lambda \\ \mu_0 - \mu \end{bmatrix} = - \begin{bmatrix} \nabla f(x) + \nabla g(x)\lambda + \nabla h(x)\mu \\ G(x)\lambda \\ h(x) \end{bmatrix} \tag{6.6}$$

being $B = H(x,\lambda,\mu)$, (x,λ,μ) the current point and (x_0,λ_0,μ_0) a new estimate. Again, B can be taken equal to a quasi - Newton estimate of $H(x,\lambda,\mu)$ or to the identity matrix.

If we define $d_0 = x_0 - x$, then (6.6) becomes

$$Bd_0 + \nabla g(x)\lambda_0 + \nabla h(x)\mu_0 = -\nabla f(x), \tag{6.7}$$

$$\Lambda\nabla g^t(x)d_0 + G(x)\lambda_0 = 0 \text{ and} \tag{6.8}$$

$$\nabla h^t(x)d_0 = -h(x), \tag{6.9}$$

which is independent of the current value of μ. As in Section 2, in this case we can also deduce that d_0 is not useful as a search direction, since it does not always constitute an uniformly feasible directions field.

Let us consider now the auxiliary function

$$\Phi_c(x) = f(x) + \sum_{i=1}^{p} c_i |h_i(x)|, \tag{6.10}$$

where c_i are positive constants. It can be shown that, if c_i are large enough, then $\Phi_c(x)$ is an Exact Penalty Function of the equality constraints, [6]. In other words, it exists a finite c such that the the minimum of Φ_c submitted only to the inequality constraints occurs at the solution of the problem (1.1). Then, the use of Φ_c as a penalty function is numerically very advantageous, since it does not require penalty parameters going to infinite. On the other side, Φ_c has not derivatives at points where there are active equality constraints. Then, to minimize Φ_c, non smooth optimization techniques are required.

Suppose now that $x \in \Omega$ and c is such that

$$sg(h_i)(c_i + \mu_{0i}) < 0; \; i = 1,2,..,p, \tag{6.11}$$

where $sg(\cdot) = (\cdot)/|(\cdot)|$. Then, d_0 given by (6.7) - (6.9) is a descent direction of Φ_c. This assert follows from

$$d_0^t\nabla\Phi_c(x) \le -d_0^tB\, d_0,$$

which is proved in a similar way as the result of *Lemma 4.2.* In this case, a feasible direction with respect to the inequality constraints and, at the same time, descent direction of Φ_c can be obtained by adding a negative vector in the right side of (6.8). This one can be taken as the search direction of an algorithm for the solution of the problem (1.1), together with a line search ensuring a decrease of Φ_c and feasibility of the inequality constraints. Even if this approach is satisfactory, it has the disadvantage of involving a non differentiable function in the line search.

We propose here an algorithm that uses Φ_c in the line search, but avoiding the points where this one is nonsmooth. Let be $\Delta \equiv \{x \in \Omega / \ h(x) \leq 0\}$ and $\Delta^0 \equiv \{x \in \Omega^0 / \ h(x) \leq 0\}$. For a given initial point in Δ^0, this algorithm generates a sequence $\{x^k\}$ in Δ^0 with decreasing values of Φ_c. With this purpose, the system (6.8) - (6.10) is modified in a way to obtain a feasible direction of Δ and, at the same time, descent direction of Φ_c.

The algorithm is stated as follows:

ALGORITHM II

Parameters. $\alpha \in (0,1)$, $\eta \in (0,1)$, $\varphi > 0$ and $\nu \in (0,1)$.
Data. $x \in \Delta^0$, $\lambda > 0$, $B \in R^{n\times n}$ symmetric and positive definite and $\omega^i \in R^m$, $\omega^e \in R^p$ and $c \in R^p$ positive.

Step 1. Computation of a search direction.

(i) Compute (d_0, λ_0, μ_0) by solving the linear system

$$Bd_0 + \nabla g(x)\lambda_0 + \nabla h(x)\mu_0 = -\nabla f(x),$$

$$\Lambda \nabla g^t(x) d_0 + G(x)\lambda_0 = 0,$$

$$\nabla h^t(x) d_0 = -h \qquad (6.12)$$

If $d_0 = 0$, stop.

(ii) Compute (d_1, λ_1, μ_1) by solving the linear system

$$Bd_1 + \nabla g(x)\lambda_1 + \nabla h(x)\mu_1 = 0,$$

$$\Lambda \nabla g^t(x) d_1 + G(x)\lambda_1 = -\Lambda\omega^i, \qquad (6.13)$$

$$\nabla h^t(x) d_1 = -\omega^i. \qquad (6.14)$$

(iii) If $c_i < -1.2\ \mu_{0i}$, then set $c_i = -2\ \mu_{0i}$; $i = 1,..,p$.

(iv) If $d_1^t \nabla \phi_c(x) > 0$, set

$$\rho = \inf\ \{\varphi \| d_0 \|^2;\ (\alpha-1)\ d_0^t \nabla \phi_c(x)/d_1^t \nabla \phi_c(x)\},\ \text{or}$$

$$\rho = \varphi \| d_0 \|^2\ \ \text{otherwise.}$$

(v) Compute the search direction

$$d = d_0 + \rho\ d_1,\ \text{and}$$

$$\bar{\lambda} = \lambda_0 + \rho\ \lambda_1$$

Step 2. Line search.

Compute t, the first number of the sequence

$\{1,\nu,\nu^2,\nu^3,...\}$ satisfying

$$\phi_c(x + td) \le \phi_c(x) + t\ \eta\ d_0^t \nabla \phi_c(x),$$

$$h(x + td) \le 0\ \ \text{and}$$

$$g_i(x + td) < 0\ \text{if}\ \bar{\lambda}_i \ge 0,\ \text{or}$$

$$g_i(x + td) \le g_i(x)\ \text{otherwise.}$$

Step 3. Updates.

(i) Set

$$x := x + td$$

and define new values for

$\omega^i > 0$, $\omega^e > 0$,

$\lambda > 0$ and

B symmetric and positive definite.

(ii) Go back to *Step 1.* ■

In consequence of (6.13) and (6.14) d_1, obtained in *Step 1*, is a descent direction of the active equality and inequality constraints. Thus, d_1 points to the interior of Δ. Updating of c ensures that d_0 is a descent direction of the resulting ϕ_c and, updating of ρ, that d is a descent direction also.

In the case when linear equality constraints are included, it is easy to find an initial point verifying them. Then, it follows from (6.12) that d_0 is on those constrains. Taking $\omega_i^e = 0$ for any i corresponding to the linear equality constraints, we have that they are always active. In consequence, when all the equalities are linear, no penalty function is required and a decrease of the objective is obtained at each iteration.

Global convergence of this algorithm can be proved using similar techniques as in Section 4. We are not going to develop this prove since it does not involve new interesting ideas.

7. References

[1] Dikin, I. I., "About the Convergence of an Iterative Procedure" (in Russian), Soviet Mathematics Doklady, Vol. 8, pp. 674 -675, 1967.

[2] Herskovits, J. "A Two-Stage Feasible Directions Algorithm Including Variable Metric Techniques for Nonlinear Optimization", Research Report n° 118, INRIA, BP 105, 78153 Le Chesnay CEDEX, France, 1982.

[3] Herskovits, J. "A Two-Stage Feasible Directions Algorithm for Nonlinear Constrained Optimization", Mathematical Programming, Vol. 36, pp. 19-38, 1986.

[4] Herskovits, J and Asquier, "Quasi - Newton Interior Points Algorithms for Nonlinear Constrained Optimization", in DGOR - Proceedings of Operations Research 1990, Vienna, August 1990.

[5] Herskovits, J. and Coelho, C.A.B. "An Interior Point Algorithm for Structural Optimization Problems", in Computer Aided Optimum Design of Structures: Recent Advances, edited by C.A. Brevia and S.Hernandez, Computational Mechanics Publications, Springer - Verlag, June 1989.

[6] Luenberger D.G., "Linear and Nonlinear Programming", 2nd. Edition, Addison - Wesley , 1984.

[7] Mayne D.Q. and Polak E., "Feasible Directions Algorithms for Optimization Problems With Equality and Inequality Constraints", Mathematical Programming 11, pp. 67 - 80, 1976.

[8] Panier, E.R.,Tits A.L. and Herskovits J. "A QP - Free, Globally Convergent, Locally Superlinearly Convergent Algorithm for Inequality Constrained Optimization", SIAM Journal of Control and Optimization, Vol 26, pp 788-810, 1988.

[9] Powell, M.J.D. "The Convergence of Variable Metric Methods for Nonlinearly Constrained Optimization Calculations", in Nonlinear Programming 3, edited by O.L. Mangasarian, R.R. Meyer and S.M. Robinson,

Academic Press, London, 1978.

[10] Powell M. J. D., " Variable Metric Methods for Constrained Optimization", in Mathematical Programming - The State of the Art, Edited by A. Bachem, M. Grotschet and B. Korte, Springer - Verlag, Berlin, 1983.

[11] Vanderbei, R. J.and Lagarios, J. C., "I. I. Dikin's Convergence Result for the Affine Scaling Algorithm", Technical Repport, AT&T Bell Laboratories, Murray Hill, NJ,1988.

[12] Zouain, N., Borges, L.A., Herskovits, J. and Feijóo, R. "An Iterative Algorithm for Limit Analysis with Nonlinear Yield Functions", Research Repport 034/90, Laboratorio Nacional de Computacion Cientifica, Rua Lauro Muller 455, 22290 Rio de Janeiro, Brazil, 1990.(to be published in Solids and Structures).

ISSN 0249-6399