A Pattern Search Algorithm for Nonlinearly Constrained Optimization Problems

Jean Medard Tchoukouegno Ngnotchouye (medard@aims.ac.za)
African Institute for Mathematical Sciences (AIMS)

Supervised by Prof Montaz Ali
University of Witwatersrand

June 5, 2007

Abstract

In this essay, we discuss a pattern search algorithm for non-linearly constrained global optimization problems. We consider a general nonlinear problem with explicit linear constraints and we use an augmented Lagrangian approach to embed the nonlinear constraints in an auxiliary objective function, the augmented Lagrangian, and an equivalent of the resulting linearly-contrained problem is solved in the inner iteration of our algorithm using the Generating Set Search algorithm of Kolda, Lewis and Torczon. The updates of Lagrange multipliers estimates and the penalty parameters are treated in a derivative-free setting in the outer iteration. We also study the convergence of our algorithm based on a similar problem in the gradient-based case.

Key words: Direct search methods, Generating Set Search, augmented Lagrangian, constrained optimization problems, global convergence.

Contents

Notations

Symbol		Definition
$\mathbb{R}$	$\equiv$	The set of real numbers,
$\mathbb{R}^n$	$\equiv$	The n^{th} dimensional set of real numbers,
GSS	$\equiv$	Generating Set Search,
$f(x)$	$\equiv$	Objective function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$,
$c(x)$	$\equiv$	Function of nonlinear constraints $c : \mathbb{R}^n \longrightarrow \mathbb{R}^m$,
$\mathcal{E}$	$\equiv$	Set of indices of equality constraints
$\mathcal{I}$	$\equiv$	Set of indices of inequality constraints
$n_{\mathcal{E}}$	$\equiv$	Number of equality constraints, $n_{\mathcal{E}} = \lvert\mathcal{E}\rvert$,
$n_{\mathcal{I}}$	$\equiv$	Number of inequality constraints, $n_{\mathcal{I}} = \lvert\mathcal{I}\rvert$,
$c_{\mathcal{E}}(x)$	$\equiv$	Function of nonlinear equality constraints $c_{\mathcal{E}} : \mathbb{R}^n \longrightarrow \mathbb{R}^{n_{\mathcal{E}}}$,
$J_{\mathcal{E}}(x)$	$\equiv$	Jacobian matrix of $c_{\mathcal{E}}$,
$c_{\mathcal{I}}$	$\equiv$	Function of nonlinear inequality constraints $c_{\mathcal{I}} : \mathbb{R}^n \longrightarrow \mathbb{R}^{n_{\mathcal{I}}}$,
$J_{\mathcal{I}}(x)$	$\equiv$	Jacobian matrix of $c_{\mathcal{I}}$,
∇f or $\nabla_x f$	$\equiv$	Gradient of the function f,
I or $I_{n\times n}$	$\equiv$	Identity matrix of order n,
$\mathcal{U}$	$\equiv$	Set of unsuccessful iterations,
$\mathcal{S}$	$\equiv$	Set of successful iterations,
K°	$\equiv$	Polar cone of a cone K,
$\|.\|$	$\equiv$	Euclidian norm in a multidimensional set of real numbers,
u^T	$\equiv$	transpose of the vector or matrix u.

1. Introduction

We are interested throughout this essay in solving the problem of determining the minimum of a given function under nonlinear constraints and explicit linear constraints; that is, in solving the problem

$$\begin{aligned} \text{Minimize} \quad & f(x) \\ \text{Subject to:} \quad & c_i(x) = 0 \quad i \in \mathcal{E}, && (1.1) \\ & c_i(x) \geq 0 \quad i \in \mathcal{I}, && (1.2) \\ \text{and} \quad & Ax \geq b, && (1.3) \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R},\ c : \mathbb{R}^n \to \mathbb{R}^{n_\mathcal{E}+n_\mathcal{I}}$, are functions and $\mathcal{E},\ \mathcal{I}$ are sets of indices; $n_\mathcal{E} = |\mathcal{E}|$ and $n_\mathcal{I} = |\mathcal{I}|,\ A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$.

There are many approaches to solve this problem. One can use Lagrangian method (see for example [1], page 71) and solve the first order optimality conditions for a KKT point. One can also use an augmented Lagrangian approach by adding to the Lagrangian a penalty term (See [1] or [2]). Another approach is to use interior-exterior method by simultaneously using a penalty and a barrier term [3]. We will neither discuss the latter nor the first approach in this essay .

Following Kolda, Lewis and Torczon [4] and Conn & Al [5] who considered this problem only with nonlinear equalities, we discuss an augmented Lagrangian direct search algorithm to minimize $f(x)$ under the constraints (1.1), (1.2) and (1.3) with the linear constraints kept explicit. After introducing slack variables, we use an augmented Lagrangian in the form

$$L(x, z, \lambda, \nu; \mu) = f(x) + \lambda^T c_\mathcal{E}(x) + \nu^T (c_\mathcal{I}(x) - z^s) + \frac{1}{2\mu}\|c_\mathcal{E}(x)\|^2 + \frac{1}{2\mu}\|c_\mathcal{I}(x) - z^s\|^2, \quad (1.4)$$

where $\lambda,\ \nu$ are vectors of Lagrange multipliers, μ is the penalty parameter and $z^s = (z_i^2)_{i\in\mathcal{I}}$ is the vector of slack variables. The problem now reduces to a sequence of minimizations of the augmented Lagrangian (1.4) under linear constraints (1.3). It happens that the dimension of the problem deteriorates the performance of direct search methods [6]. So for the variable x kept fixed, we explicitly carry out the minimization of the augmented Lagrangian (1.4) with respect to z, and come out with an equivalent problem with the same size as the original Lagrangian minimization problem.

More precisely, we solve at the inner iteration of our algorithm the linearly constrained problem consisting of minimizing the resulting augmented Lagrangian with the constraints (1.3) by using the linearly-constrained Generating Set Search(GSS) method of Kolda & Al ([6], [4]) up to a given tolerance. At the outer iteration, if nonlinear feasibility is reached, we update the Lagrange multipliers, otherwise, we update the penalty parameters. The tolerance parameter for the inner iteration is also updated at the outer iteration. A derivative-free stopping condition is given like in [4]. As it so happens, we can still apply convergence results of Conn & Al [5] under small changes of the constants appearing.

This essay is structured as follows, the definitions of an optimization problem, optimality conditions, global and local convergence and other prerequisites are addressed in the second chapter.

In the third chapter, we present the linearly-constrained GSS algorithm of Kolda, Lewis and Torczon and discuss the proof of convergence. In the fourth chapter which is the corner-stone of this essay, we discuss the augmented Lagrangian GSS method for nonlinear optimization with explicit linear constraints. We close the essay with a conclusion and future research directions.

2. Optimization: Optimality Conditions, Measure of Stationarity

In this chapter, we give the general definition of an optimization problem, we introduce the notions of stationary points, optimality conditions and augmented Lagrangian method. Theses notions are important for the understanding of the rest of the essay. More details can be found in [2], [1] or any other book on nonlinear optimization.

2.1 Optimization Problem

A general optimization problem consists of solving

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{Subject to} & x \in S, \end{array} \tag{2.1}$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is called *objective function* and S the *feasible region*. The specification of the optimization problem (2.1) depends on special properties of f and S.

- When $S = \mathbb{R}^n$, (2.1) is called *unconstrained optimization problem*.
- When f is linear and S is a set of linear inequalities or equalities[1], (2.1) is called a *linear optimization problem*.
- When $S = \{x \in \mathbb{R}^n \,|\, l \leq x \leq u\}$ where l, $u \in \mathbb{R}^n$, (2.1) is called *bounded optimization problem*.
- When $S = \{x \in \mathbb{R}^n \,|\, l \leq c(x) \leq u\}$ where $c : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is a function and l, $u \in \mathbb{R}^m$, (2.1) is called a general *non linear optimization problem*. We can also have in this case a set S with only one side inequalities (allows l_i to be $-\infty$ or u_i to be $+\infty$) or with equalities ($l_i = u_i$) where the subscript denote the component of a given vector..

When the functions defining the problem (2.1) are smooth, the problem is said to be smooth; otherwise, the problem is non smooth. Also, when S is convex and f is convex, (2.1) is a convex optimization problem. To solve (2.1), we need conditions that will enable us to detect optimality or more generally stationarity. When the problem is smooth, classical optimality conditions are based on the gradients of the functions f and c. When the problem is non smooth, we can seek weak optimality conditions based on the general Clarke differential of the objective function. In the constrained case, optimality is based on the so called Lagrangian function under suitable "constraints qualification". We address these problems in the next sections.

[1] Example: $S = \{x \in \mathbb{R}^n \,|\, Ax \geq b\}$ where $A \in \mathbb{R}^{m \times n}$ is a matrix and $b \in \mathbb{R}^m$

2.2 Optimality Conditions

2.2.1 Unconstrained Optimization Problems

Here, we consider the problem

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{Subject to} & x \in \mathbb{R}^n. \end{array} \tag{2.2}$$

The proposition that follows say that if optimality is reached at a point $x_* \in \mathbb{R}^n$, then the gradient of the function f at x_* vanish. Conversely, if the gradient of f at x_* vanishes and the Hessian matrix of f at x_* is positive definite[2], then x is a local minimum of f. We start with the definition of local and global minimum.

Definition 2.1. *A point $x_* \in \mathbb{R}^n$ is a local solution of (2.2) or a stationary point of f if there exists a scalar $\delta > 0$ such that*

$$f(x_*) \leq f(x) \quad \text{for all } x \in B(x_*, \delta) = \{x \in \mathbb{R}^n, \|x - x_*\| \leq \delta\}.$$

A point $x_* \in \mathbb{R}^n$ is said to be a global solution of (2.1) if $f(x_*) \leq f(x)$, for any $x \in \mathbb{R}^n$.

We now state a well known result on necessary and sufficient optimality conditions for unconstrained optimization problems.

Proposition 2.1. (i) *Suppose f is continuously differentiable. Let $x_* \in \mathbb{R}^n$ such that x_* is a local solution of (2.2). Then $\nabla f(x_*) = 0$.*

(ii) *Conversely, suppose f is twice continuously differentiable, we assume that x_* is such that $\nabla f(x_*) = 0$ and the Hessian matrix $\nabla^2 f(x_*)$ is positive definite. Then x_* is a local solution of (2.2).*

2.2.2 Constrained Optimization Problems

Here we consider problem (2.1) in the form

$$\begin{array}{lll} \text{Minimize} & f(x) & \\ x \in \Omega \subset \mathbb{R}^n & & \\ \text{Subject to} & c_i(x) \geq 0 & i \in \mathcal{I} \\ & c_j(x) = 0 & j \in \mathcal{E}, \end{array} \tag{2.3}$$

where Ω is an open set containing a feasible region, $|\mathcal{E}| = n_{\mathcal{E}}$ and $|\mathcal{I}| = n_{\mathcal{I}}$. The feasible set S of this problem is the intersection of Ω and the explicit (in general) nonlinear constraints in (2.3):

$$S = \Omega \cap \{x \,|\, c_i(x) \geq 0,\ c_j(x) = 0,\ i \in \mathcal{I},\ j \in \mathcal{E}\}.$$

In this case, the definition of stationary point is different.

[2] A matrix A is said to be positive definite if for all $d \in \mathbb{R}^n,\ d \neq 0$, we have $d^T A d > 0$.

Definition 2.2. *A point $x_* \in S$ is a stationary point of (2.3) if for every feasible direction $d \in S$, we have $\nabla f(x_*)^T d \geq 0$.*

The most important result in this section is the existence of Lagrange multipliers, which derive from Farcas lemma.

Theorem 2.1. *[2](Farcas lemma) Let $a_0, a_1, \ldots, a_q$ be a set of vectors of $\mathbb{R}^n$. Define*

$$Z = \left\{ z \,|\, z^T a_i \geq 0, \text{ for } i = 1, \ldots, q \right\}. \tag{2.4}$$

If for every $z \in Z,\ z^T a_0 \geq 0$, then there exists a set of non negative scalars $t_1, \ldots, t_q$, such that

$$a_0 = \sum_{i=1}^{q} a_i t_i.$$

This theorem says that if Z is at one side of the hyperplane represented by $z^T a_0 = 0$, then a_0 is in the cone generated by $a_j,\ j = 1, \ldots, q$.
There are some simple generalizations of this theorem that we state in the corollary that follows:

Corollary 2.1. *Let $a_0, a_1, \ldots, a_q, a_{q+1}, \ldots, r$ be a set of vectors of $\mathbb{R}^n$. Define*

$$Z = \left\{ z \,|\, z^T a_i \geq 0, \text{ for } i = 1, \ldots, q;\ z^T a_i = 0, \text{ for } i = q+1, \ldots, r \right\},$$
$$S = \left\{ s \,|\, s^T a_i > 0, \text{ for } i = 1, \ldots, q;\ s^T a_i \geq 0, \text{ for } i = q+1, \ldots, r \right\}.$$

(a) *If for every $z \in Z,\ z^T a_0 \geq 0$, then there exists a set of scalars $t_1, \ldots, t_r$ with $t_i \geq 0$ for $i = 1, \ldots, q$, such that*

$$a_0 = \sum_{i=1}^{r} a_i t_i.$$

(b) *If S is not empty, and if for every $s \in S,\ s^T a_0 \geq 0$, then there exists scalars $t_1, \ldots, t_r$ with $t_i \geq 0$ for $i = 1, \ldots, q$, such that*

$$a_0 = \sum_{i=1}^{r} a_i t_i.$$

Now we state and prove a lemma which is directly linked to the problem we are solving namely (2.3).

Lemma 2.1. *[2] Assume that the functions $f,\ c_i,\ i \in \mathcal{I} \cup \mathcal{E}$ in (2.3) are continuously differentiable in the open set Ω and let x_* be a local minimizer for that problem. We define the set of binding constraints at x_* to be the set of indices $\mathcal{I}_* = \{ i \,|\, c_i(x_*) = 0,\ i \in \mathcal{I} \}$. Let*

$$\bar{S}_0 = \left\{ s \,|\, s^T \nabla c_i(x_*) > 0,\ i \in \mathcal{I}_* \text{ and } s^T \nabla c_j(x_*) = 0,\ j \in \mathcal{E} \right\}.$$

If $\nabla c_j(x_*),\ j \in \mathcal{E}$ are linearly independent, then for every $s \in \bar{S}_0,\ s^T \nabla f(x_*) \geq 0$.

Proof. We prove the case where $n_{\mathcal{E}} < n$ and $\bar{S}_0 \neq \emptyset$. The other cases are obvious. Let s be any nonzero element of $\bar{S}_0$. We assume without loss of generality that $\|s\| = 1$. Denote by $c_{\mathcal{E}}$ the function whose components are $c_j,\ j \in \mathcal{E}$ and $J_{\mathcal{E}}(x_*)$ its Jacobian matrix at x_*. As $\nabla c_j(x_*),\ j \in \mathcal{E}$ are linearly independent, $J_{\mathcal{E}}(x_*)$ has rank $n_{\mathcal{E}}$. Its null space then has rank $n - n_{\mathcal{E}}$ and there is an orthonormal set of vector containing s that is a basis for that null space We now consider the system of equations where θ is a scalar

$$\gamma_s(x, \theta) = \begin{pmatrix} c_{\mathcal{E}}(x) \\ B^T x - B^T x_* \\ s^T x - s^T x_* - \theta \end{pmatrix} = 0,$$

and where B and s form an orthogonal basis for the null space of $J_{\mathcal{E}}(x_*)$. The Jacobian matrix of this system is

$$\frac{d\gamma_s(x_*, 0)}{dx} = \begin{pmatrix} J_{\mathcal{E}}(x_*)^T \\ B^T \\ s^T \end{pmatrix},$$

which is non singular with inverse $\left[J_{\mathcal{E}}(x_*) \left[J_{\mathcal{E}}(x_*)^T J_{\mathcal{E}}(x_*) \right]^{-1} \ B \ s \right]$. The implicit function theorem implies the existence of a continuous function $x_s(\theta)$ near $s = 0$ with $x_s(0) = x_*$ and $x_s'(0) = s$ (this is true since $D_\theta \gamma_s(x, 0)^T = [0, \ldots, -1]$). For small θ, $x_s(\theta)$ can be thought of as a feasible arc emanating from x_* (the equality constraints are satisfied and for small θ, the constraints were $c_i(x_*) > 0$ are strictly satisfied; for those constraints with indices in $\mathcal{I}_*$, $\frac{dc_i}{d\theta}(x_s(0)) = \nabla c_i(x_*)^T x_s'(0) = \nabla c_i(x_*)^T s > 0$ (because $s \in \bar{S}_0$) and hence are strictly satisfied along the arc). Since x_* is an assumed local minimizer, $\frac{df}{d\theta}(x_s(0)) \geq 0$; but this means that $\nabla f(x_*)^T s \geq 0$ which was to be proved. ∎

The theorem that follows is an important step towards the proof of existence of Lagrange multipliers.

Theorem 2.2. *[2] Assume $f,\ c_i,\ x_*,\ \mathcal{I}_*$ are as in Lemma 2.1. Then there exists values $(w_1, \ldots, w_{n_{\mathcal{E}}}),\ u_0$ and $u_i,\ i \in \mathcal{I}_*$ not all equal to zero such that*

$$u_0 \geq 0 \quad \textit{and} \quad u_i \geq 0 \quad i \in \mathcal{I}_*, \tag{2.5}$$

and

$$\nabla f(x_*) u_0 - \sum_{i \in \mathcal{I}_*} \nabla c_i(x_*) u_i + \sum_{j=1}^{n_{\mathcal{E}}} \nabla c_j(x_*) w_j = 0. \tag{2.6}$$

Proof. There are three cases to consider.
(a) The $\nabla c_j(x_*)$'s are linearly dependent. We take $u_0 = 0$ and $u_i = 0,\ i \in \mathcal{I}_*$ and the w_j's to be some non zero values such that $\sum_{j=1}^{n_{\mathcal{E}}} \nabla c_j(x_*) w_j = 0$ (linear dependence of the $\nabla c_j(x_*)$).
(b) The $\nabla c_j(x_*)$'s are linearly independent and the Jacobian matrix $J_{\mathcal{E}}(x_*)$ has rank n (i.e., $n_{\mathcal{E}} = n$). Then we take $u_0 = 1,\ u_i = 0,\ i \in \mathcal{I}_*$ and $w = J_{\mathcal{E}}(x_*)^{-1} \nabla f(x_*)$.
(c) The matrix $J_{\mathcal{E}}(x_*)$ has rank $n_{\mathcal{E}} < n$. We assume for simplicity that $\mathcal{I}_*$ is in the form $\mathcal{I}_* = \{1, \ldots, p\}$. We define

$$\bar{S}_k = \left\{ s \,|\, s^T \nabla c_l(x_*) > 0,\ l = k+1, \ldots, p \text{ and } s^T \nabla c_j(x_*) = 0,\ j \in \mathcal{E} \right\}.$$

Note that $\bar{S}_p$ is not empty, because $n_{\mathcal{E}} < n$ ($0 \in \bar{S}_p$). It is straightforward that $\bar{S}_k$ being empty implies that $s^T \nabla c_{k+1}(x_*) \leq 0$ for all $s \in \bar{S}_{k+1}$. Let K be the smallest integer such that $\bar{S}_k \neq \emptyset$. If $K = 0$, then from Lemma 2.1, $\nabla f(x_*)^T s \geq 0$ for all $s \in \bar{S}_0$. Corollary 2.1 assures us of the existence of values $(w_1, \ldots, w_{n_{\mathcal{E}}})$ and $u_i,\ i \in \mathcal{I}_*$ such that (2.6) holds with $u_0 = 1$. If $K > 0$, the same corollary implies the existence of values $(w_1, \ldots, w_{n_{\mathcal{E}}})$ and $u_i,\ i = K+1, \ldots, p$ such that (2.6) holds with $u_i = 0,\ i = 0, \ldots, K-1$ and $u_K = 1$. ■

This theorem has the drawback that it can allow u_0 to be zero (like in the cases we saw in the proof of the theorem). In such a case, the contribution of the objective function to optimality disappears. To avoid that, additional assumptions are needed and are related to the geometry of the constraints.

Theorem 2.3. *[2] (Karush, Kuhn and Tucker)*
Assume $f,\ c_i,\ x_*,\ \mathcal{I}_*$ are as in Lemma 2.1. Let

$$\bar{Z} = \left\{ s \mid s^T \nabla c_i(x_*) > 0,\ i \in \mathcal{I}_* \text{ and } s^T \nabla c_j(x_*) = 0,\ j \in \mathcal{E} \right\}.$$

Assume that a first order constraints qualification of the following form holds at x_* :
(CQ) Every $z \in \bar{Z}$ is the tangent of a differentiable arc emanating from x_* that is contained in the region of feasible points of (2.3).
Then there exists $u_i,\ i \in \mathcal{I}_*,\ w_j,\ j \in \mathcal{E}$ such that $u_i \geq 0,\ i \in \mathcal{I}_*$ and

$$\nabla f(x_*) - \sum_{i \in \mathcal{I}_*} \nabla c_i(x_*) u_i + \sum_{j=1}^{n_{\mathcal{E}}} \nabla c_j(x_*) w_j = 0.$$

The quantities $u_i,\ i \in \mathcal{I}_*,\ w_j,\ j \in \mathcal{E}$ and $u_i \geq 0,\ i \in \mathcal{I}_*$ are called **Lagrange multipliers** *of the problem (2.3).*

Proof. Let $z \in \bar{Z}$ and $a_z(\theta)$ be an arc tangent to z emanating from x_* and contained in the feasible region. As x_* is a local minimum of f, we have

$$\frac{df(a_z(\theta))}{d\theta} = \nabla f(x_*)^T \frac{da_z(\theta)}{d\theta} = \nabla f(x_*)^T z \geq 0,$$

and the result derive from Farcas lemma. ■

There are other constraints qualifications that ensure the result of the Theorem 2.3. The ones mostly used in nonlinear programming literature are:

- **Mangasarian Fromovitz Constraints Qualification (MFCQ)**
  The vectors $\nabla c_j(x_*),\ j \in \mathcal{E}$ are linearly independent and the set

  $$\bar{S} = \left\{ s \mid s^T \nabla c_i(x_*) > 0,\ i \in \mathcal{I}_* \text{ and } s^T \nabla c_j(x_*) = 0,\ j \in \mathcal{E} \right\}$$

  is not empty.

- **Linear Independence Constraints Qualification (LICQ)**
  The vectors $\nabla c_j(x_*),\ j \in \mathcal{E}$ and $\nabla c_i(x_*),\ i \in \mathcal{I}_*$ are linearly independent.

When any of these constraints qualification is satisfied at a feasible point x_*, then x_* is said to be a *regular* point.

We now summarize the optimality conditions for a nonlinear problem of the form (2.3). We define the Lagrangian as

$$L(x, \lambda, \nu) = f(x) + \sum_{i \in \mathcal{I}} c_i(x)\nu_i + \sum_{j \in \mathcal{E}} c_j(x)\lambda_j,$$

and Theorem 2.3 can be restated as the following theorem.

Theorem 2.4. *(First order necessary optimality conditions)*
Assume x_* is regular and is a local minimizer of (2.3), then there exists Lagrange multiplier vectors $\bar{\lambda},\ \bar{\nu}$ such that

$$\begin{aligned}
&\nabla_x L(x_*, \bar{u}) = \nabla_x f(x_*) + \sum_{i \in \mathcal{I}} \nabla_x c_i(x_*)\bar{\nu}_i + \sum_{j \in \mathcal{E}} \nabla_x c_j(x_*)\bar{\lambda}_j = 0 && \textit{(Dual objective),}\\
&\bar{\nu}_i \leq 0,\ i \in \mathcal{I} && \textit{(Dual feasibility),}\\
&c_i(x_*) \geq 0,\ i \in \mathcal{I} && \textit{(Primal feasibility),}\\
&c_j(x_*) = 0,\ j \in \mathcal{E} && \textit{(Primal feasibility),}\\
&\bar{\nu}_i c_i = 0,\ i \in \mathcal{I} && \textit{(Complementary slackness).}
\end{aligned}$$

The triple $(x_*, \bar{\lambda}, \bar{\nu})$ is said to be a **Kurush-Kuhn-Tucker (KKT) point** *or a first order stationary point for problem (2.3).*

Note that the complementary slackness assures us that when a constraint $c_i,\ i \in \mathcal{I}$ is not binding (holds as an equality) at the point x_*, then the associated Lagrange multiplier $\bar{\nu}_i$ vanishes.

2.3 Global Convergence and Local Convergence

In nonlinear optimization, the notion of global convergence of an algorithm refers to first order convergence from an arbitrary initial point. By order, we refer to the order of the associated derivatives. In contrast, local convergence means convergence when the initial point is close enough to a minimizer. In addition, local convergence is normally associated with a rate of convergence.

In this essay, we will be mostly concern by first order global convergence of direct search methods to a stationary point of a given optimization problem. It is important to note that global convergence does not mean convergence to a global minimizer in the sense of Definition 2.1.

In general it is impossible to prove second order convergence results (i.e. convergence to a minimizer) for first order methods like the line-search methods [6]. In practice however, these algorithms reliably find local minimizers. The global analysis of such methods establishes first order convergence result. Surprisingly, similar first order convergence results holds for a particular class of direct search methods which are zeroth order methods.

2.4 Geometry of Cones and Measure of Stationarity

A set K is said to be a cone if for all $x \in K$, and for all scalar $\alpha > 0,\ \alpha x \in K$. This means that a cone is a set which is closed under non negative scalar multiplication. The polar cone K° of a cone K is defined by

$$K^\circ = \left\{ x \,|\, x^T y \leq 0, \text{ for all } y \in K \right\}.$$

Roughly speaking it is the set of vectors making an angle of at least 90° with all vectors in K. Given two polar cones K and K°, any vector v in $\mathbb{R}^n$ can be written uniquely as

$$v = v_K + v_{K^\circ}, \tag{2.7}$$

with $v_K^T v_{K^\circ} = 0$, $v_K = P_K(v) \in K$, $v_{K^\circ} = P_{K^\circ}(v) \in K^\circ$ where P_K is the projection operator onto K. The relation in equation (2.7) is called the polar decomposition[3] of v. A cone K is finitely generated if there exists a finite set of vectors $\mathcal{G}$ such that $\mathcal{G}$ positively spans K[4].
Given a set Ω, every point $x \in \Omega$ has two important cones associated to it: the cone of tangents and the cone of normals. A vector v is tangent to Ω at x if for every sequence x_i converging to x, for every sequence α_i decreasing to zero, there exists a sequence v_i converging to v such that $x_i + \alpha_i v_i$ remains in Ω. Roughly speaking, a vector v is tangent to Ω at x if one can proceed from x along v and still remain in Ω at least for a small distance. A classical result is that the set of all vectors tangent to Ω at x is a convex cone. This cone is called the (Clarke) tangent cone to Ω at x and it is denoted by $\mathcal{T}_\Omega(x)$. Its polar is called the normal cone to Ω at x and is denoted by $\mathcal{N}_\Omega(x)$. Roughly speaking it is the set of vectors pointing "outward" from Ω at x. It is important to note that when x is in the interior of Ω, then $\mathcal{N}_\Omega(x) = \{0\}$ and $\mathcal{T}_\Omega(x) = \mathbb{R}^n$. Two examples of polar and normal cones to Ω at x are depicted in Figure 2.1.
We now consider Problem (2.3) in the following form:

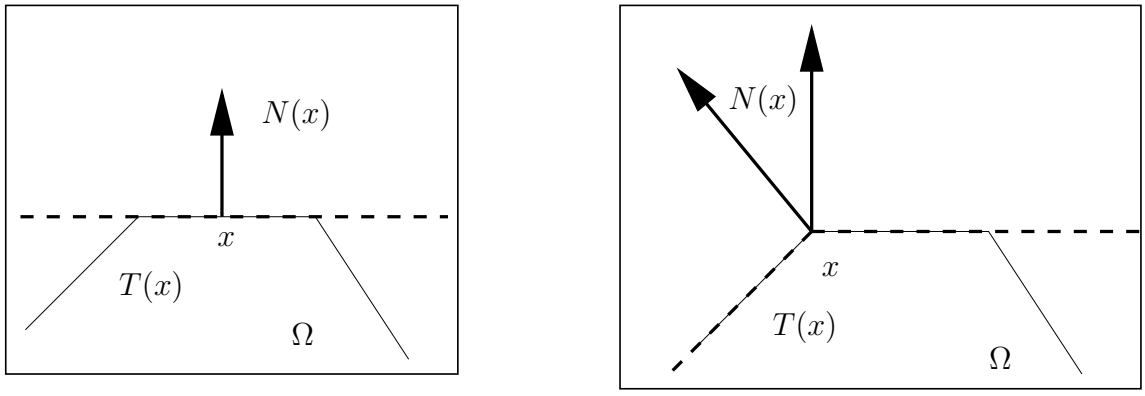


Figure 2.1: Examples of tangent and normal cones. $T(x)$ is the region below the dashed lines. In the first figure, $N(x)$ is the half line from x in the direction of the upward pointing arrow. In the second picture, $N(x)$ is the cone generated by the two arrows.

[3] It is also called the Moré decomposition of v.

[4] A set of vector $\mathcal{G}$ positively spans a set K if every element in K can be written as a linear combination of the vectors in $\mathcal{G}$ with non negative coefficients.

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{Subject to} & c(x) \leq 0, \end{array} \tag{2.8}$$

where here $c : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is a differentiable function. We still denote $\mathcal{I}_*$ the set of binding constraints at a given point x_*.
If a suitable constraints qualification holds at x_* (LICQ for example), then the normal cone to the feasible set $\Omega = \{x \in \mathbb{R}^n \,|\, c(x) \leq 0\}$ at x_* is generated by the set of vectors $\{\nabla c_i(x_*), \quad i \in \mathcal{I}_*\}$ that is

$$\mathcal{N}_\Omega(x_*) = \left\{ v \,|\, v = \sum_{i \in \mathcal{I}_*} u_i \nabla c_i(x_*),\ u_i \geq 0 \quad \text{for all } i \in \mathcal{I}_* \right\}.$$

It is obvious that if x_* is a local minimizer of (2.8), Theorem 2.3 reads

$$-\nabla f(x_*) \in \mathcal{N}_\Omega(x_*). \tag{2.9}$$

In this case, (x_*, u_i) or x_* is called a KKT point for (2.8) and the u_i's the associated Lagrange multipliers. Note that if x_* is in the interior of $\mathcal{N}_\Omega(x_*)$, then $\mathcal{N}_\Omega(x) = \{0\}$ and (2.9) becomes $\nabla f(x_*) = 0$: the necessary condition for the unconstrained case.

Looking at the definition of the polar decomposition of a vector of $\mathbb{R}^n$ given in (2.7), we see that x_* is a KKT point for (2.8) is equivalent to

$$P_{\mathcal{T}_\Omega(x_*)}(-\nabla f(x_*)) = 0.$$

The necessary condition (2.9) has the drawback that it does not serve as a continuous measure of constrained stationarity because the normal cone does not depend continuously on x at the boundary of the feasible region[5]. For this reason, other continuous measures of stationarity need to be defined. The two we highlight in this section are widely used in the literature on nonlinear programming. The first measure is the following. For $x \in \Omega$, let

$$\chi(x) = \begin{array}{ll} \text{Maximize} & -\nabla f(x)^T \omega, \\ \text{Subject to} & x + \omega \in \Omega, \\ & \|\omega\| \leq 1. \end{array} \tag{2.10}$$

Loosely speaking, $\chi(x)$ captures the degree to which the direction of steepest descent is outward pointing with respect to the portion of the feasible region near x. The geometry of this measure of stationarity is depicted in Figure 2.2.

The result that follows gives some properties of $\chi(x)$.

Proposition 2.2. *Assume Ω is convex. The function $\chi(x)$ has the following properties:*

1. *χ is continuous,*
2. *$\chi(x) \geq 0$,*

[5] In fact, for the case of a convex polyhedra, if two points x and y are closed and x is on the boundary of Ω and y is interior to Ω, $\mathcal{N}_\Omega(x)$ is a half line or a cone and $\mathcal{N}_\Omega(y) = \{0\}$ which are not "closed".

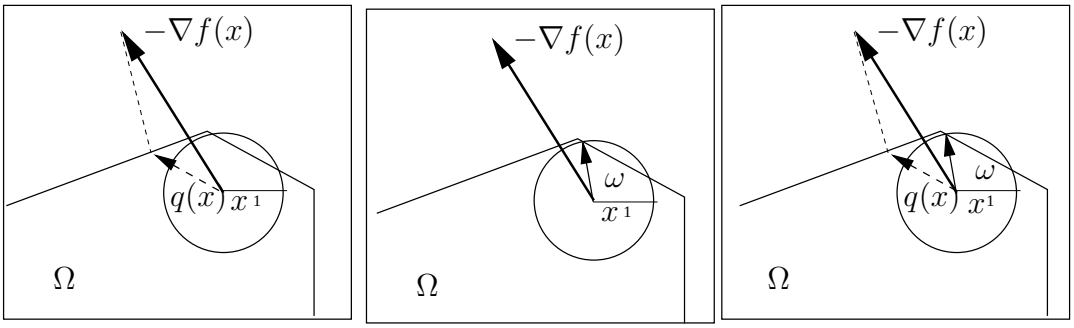


Figure 2.2: The geometry of $\chi(x)$ in the left picture and $q(x)$ in the middle picture. The ω that gives the measure $\chi(x)$ and $q(x)$ are superimposed on the right picture for comparison

3. $\chi(x) = 0$ *if and only if* x *is a KTT point for (2.8).*

This proposition also holds for a set Ω that has a geometrically "nice" boundary vis-à-vis the above mentioned constraints qualifications.
If $\chi(x) = 0$, then there is no direction making an angle less than 90° with the direction of the steepest descent along which one can move for any distance an remains feasible. Therefore, showing that $\chi(x_k) \to 0$ as $k \to \infty$ constitutes a global first order convergence result.

The second measure of stationarity we will address in this essay involves projection onto Ω under the assumptions that Ω is convex[6]. We define

$$q(x) = P_\Omega(x - \nabla f(x)) - x. \tag{2.11}$$

The geometry of $q(x)$ is depicted in Figure 2.2. A point x_* is a KKT point for (2.8) if and only if $q(x_*) = 0$. Thus a global first order convergence result can also be obtained by showing that $\|q(x_k)\| \to 0$ as $k \to \infty$. Like χ, q reflects the extend to which the direction of steepest descent is outward pointing with respect to the feasible region. However, unlike $\chi(x)$, $\|q(x)\|$ depends on the global geometry of the feasible region, since the projection $P_\Omega(x - \nabla f(x))$ is not necessarily within a predetermined distance from x.

2.5 Augmented Lagrangian Penalty Function for Nonlinear Programming

This method is part of a large class of methods called penalty methods. They consist of reintroducing the constraints of a nonlinear optimization problem into an auxiliary objective function and solving the resulting unconstrained problem[7]. This is done in such a way that when the constraints are violated, an additional amount is added to the objective function and this amount is controlled by a parameter called *penalty parameter*.

[6] This is the case for linearly-constrained problems.
[7] Or with simple constraints as bound on variables or linear constraints.

For simplicity and the purpose of this essay, we consider a nonlinear program with only equality constraints

$$\begin{array}{ll} \text{Minimize} & f(x), \\ x \in \Omega \subset \mathbb{R}^n & \\ \text{Subject to} & c(x) = 0, \end{array} \tag{2.12}$$

where $c : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ and $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and Ω is a subset of $\mathbb{R}^n$ containing some simple constraints. Let $\lambda = (\lambda_1, \ldots, \lambda_m)$ be the vector of Lagrange multipliers. The augmented Lagrangian method leads to the auxiliary function:

$$L(x, \lambda; \mu) = f(x) + \sum_{j=1}^{m} c_j(x)\lambda_j + \frac{1}{2\mu}\sum_{j=1}^{m}(c_j(x))^2,$$

where μ is called the penalty parameter and the problem consists now of solving a sequence of problems of the form

$$\begin{array}{ll} \text{Minimize} & L(x, \lambda; \mu) \\ x \in \Omega \subset \mathbb{R}^n, & \end{array} \tag{2.13}$$

with fixed value of Lagrange multiplier and penalty parameter. The following result ensures the convergence of the method towards the solutions of (2.12).

Theorem 2.5. *[1] We assume $\Omega \neq \emptyset$ is a closed set and that $f(x)$ and $c(x)$ are continuous function on Ω. For $k = 0, 1, \ldots,$ let x_k be the global solution of the problem*

$$\begin{array}{ll} \text{Minimize} & L(x, \lambda_k; \mu_k) \\ \text{Subject to } x \in \Omega, & \end{array} \tag{2.14}$$

where $\{\lambda_k\}$ is bounded and $0 < \mu_{k+1} < \mu_k$ for all k, $\mu_k \to 0$. Then every limit point of the sequence $\{x_k\}$ is a global solution of (2.12).

3. A pattern Search Algorithm for Linearly-Constrained Optimization Problems

We investigate in this chapter a pattern search method for linearly-constrained optimization problems called the Generating Set Search (GSS) method. The implementation of this method, that we will not discuss here, is done in [7], [8], [9], [10]. This method will be used in the next chapter to solve a more general nonlinear problem.

3.1 A Simple and Illustrative Example: Compass Search Method

The Compass search method is an example of coordinate search methods. This is a class of methods where the search directions are the coordinates vectors of $\mathbb{R}^n$, e_i, $i = 1, \ldots, n$. The idea in two dimensions is the following. At the current iterate x_k with the step-length parameter Δ_k, the method samples the objective function f along the feasible trial points $x_k + \Delta_k e_1$ (West), $x_k + \Delta_k e_2$ (North), $x_k - \Delta_k e_1$ (East), and $x_k - \Delta_k e_2$ (South). If a feasible improvement is found at iteration k in one direction, the iteration is deemed successful ($k \in \mathcal{S}$, the set of successful iterations) and we move in that direction, for example $x_{k+1} = x_k + \Delta_k e_2$. If no feasible improvement is found, the iteration is deemed unsuccessful ($k \in \mathcal{U}$, the set of unsuccessful iterations); the iterate remains unchanged ($x_{k+1} = x_k$) and the step-length parameter is halved. We illustrate an instance of compass search in Figure 3.1 where we applied it to the two dimensional function $f(x, y) = x^2 + y^2$ with three linear constraints. The level set (contours) of the function f are in the background, the feasible region is the inside of the triangle and the search directions at any iterate form a cross.

3.2 An Overview of Generating Set Search (GSS) Methods

The problem we consider in this section is the following.

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{Subject to:} & Ax \geq b, \end{array} \tag{3.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is the objective function, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. We denote throughout this chapter the set of constraints or the feasible set by

$$\Omega = \{x \,|\, Ax \geq b\}.$$

We start by an outline of Generating Set Search methods for constrained optimization problems. Let $x_0 \in \Omega$ be an initial iterate and Δ_0 be the initial choice of step-length parameter with

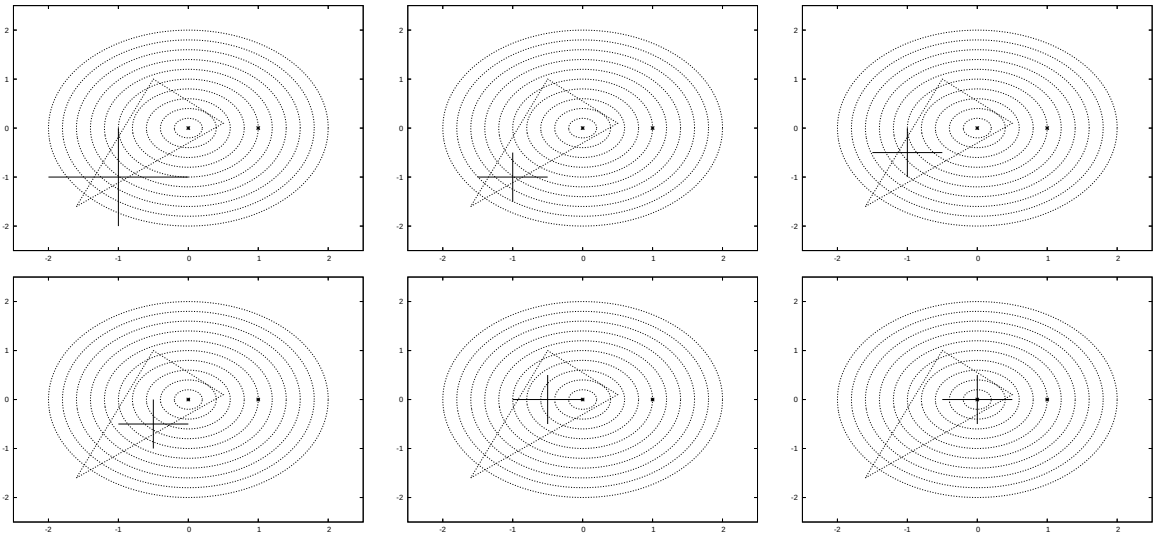


Figure 3.1: From the left to the right: At the initial configuration, all the trial points are infeasible. At $k = 0$, we contract the set of search directions, $k \in \mathcal{U}$. At $k = 1$, there is feasible improvement in the North direction, so we move North $k \in \mathcal{S}$; at $k = 2$ move West, $k \in \mathcal{S}$; at $k = 3$ move North, $k \in \mathcal{S}$; at $k = 3$ move North, $k \in \mathcal{S}$; at $k = 4$ move West, $k \in \mathcal{S}$.

$\Delta_0 > \Delta_{tol} > 0$, where Δ_{tol} serves as a measure of termination. The search proceeds for iteration $k = 0,\ 1,\ 2 \ldots$ until $\Delta_k < \Delta_{tol}$. At each iteration, x_k is the best point find thus far.
The first step at each iteration is to select a set of search directions. At points near the boundary, these search directions should capture the geometry of the feasible region. The number of search directions here is denoted by p_k and the set of search directions by

$$\mathcal{D}_k = \{d_k^{(1)}, \ldots, d_k^{(p_k)}\}.$$

The second step is to construct feasible trial points of the form

$$x_k + \tilde{\Delta}_k^{(i)} d_k^{(i)} \quad i \in \{1, \ldots, p_k\},$$

with $\tilde{\Delta}_k^{(i)} \in [0,\ \Delta_k]$ which is chosen to ensure feasibility. These trial points are points where the objective function may be evaluated in the search for a new best point (with the lowest value of the objective function) to replace x_k. The third step is to determine whether the iteration is successful or unsuccessful and correspondingly update the iterate and the step-length parameter. We now study each step in detail.

3.3 Search Directions

The directions to be used in a derivative-free algorithm should conform to the geometry of the feasible region for the near boundary iterates and be such that the evaluation of the objective function along them is sufficiently indicative of the local behavior of the objective function. These directions should have the property that, performing samplings of the objective function along them possibly with small and smaller step-size, it is either possible

1. to realize that the current iterate is a good approximation of a stationary point of the objective function or

2. to find a specific direction along which the objective function decreases.

So the set of search directions should have the property that the local behavior of the objective function along them provide sufficient information to overcome the lack of gradient and to ensure desirable convergence properties of the algorithm. For linearly-constrained optimization, in order to capture the geometry of the feasible region, the set of search directions at near boundary iterates should contain the generators of the ε-tangent and ε-normal cone, that we address in the next section.

3.3.1 ε-tangent and ε-normal Cones

We studied tangent and normal cones to a set at a point in a general setting in section 2.4. Here, we discuss the particular case of convex polyhedral, that is sets defined by a linear inequalities.

Let a_i^T be the i^{th} row of the constraints matrix A. We define the set of points where the i^{th} constraint is binding by

$$A_i = \{y \,|\, a_i^T y = b_i\}.$$

The set of indices for the binding constraints at $x \in \Omega$ is

$$I(x) = \{i \,|\, x \in A_i\}.$$

The normal cone at x, denoted $N(x)$ is the cone generated by the set $\{-a_i \,|\, i \in I(x)\} \cup \{0\}$. The presence of $\{0\}$ means that $N(x) = \{0\}$ if there are not binding constraints at x. The tangent cone at x, denoted $T(x)$ is the polar of the normal cone. While implementing linearly-constrained GSS methods, one is interested in the faces of the polyhedra defining the constraints that are near to the iterates. This suggests the definitions of the so-called ε-tangent and ε-normal cones. Given $x \in \Omega$, the indices of the ε-binding constraints are given by

$$I(x, \varepsilon) = \{i \,|\, \mathrm{dist(x,\, A_i)} \leq \varepsilon\},$$

where $\mathrm{dist(x,\, A_i)}$ is the distance from x to the hyperplane A_i. The vectors $-a_i$ for $i \in I(x, \varepsilon)$ are the outward pointing normals to the faces of the boundary of Ω within a distance ε of x. We define the ε-normal cone denoted $N(x, \varepsilon)$ to be the cone generated by $\{-a_i \,|\, i \in I(x, \varepsilon)\} \cup \{0\}$. The presence of $\{0\}$ here means that $N(x) = \{0\}$ if $I(x, \varepsilon) = \emptyset$. The corresponding polar cone is the ε-tangent cone and it is denoted by $T(x, \varepsilon)$. As the set A_i is closed, when $\varepsilon = 0$, we recover the aforementioned tangent and normal cones at x, $T(x)$ and $N(x)$. An illustration of ε-tangent and ε-normal cones is given in Figure 3.2.

The set $x + T(x, \varepsilon)$ approximates the feasible region near [1] x. Note that if $I(x, \varepsilon) = \emptyset$, then $N(x, \varepsilon) = \{0\}$ so that $T(x, \varepsilon) = \mathbb{R}^n$. This means that if the boundaries are at a distance more than ε to an iterate x_k, then the problem looks unconstrained in the neighborhood of x_k.

[1] with respect to ε.

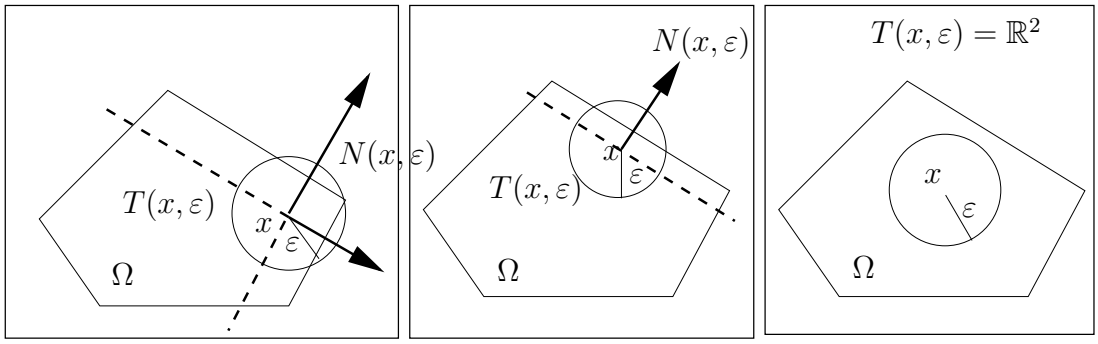


Figure 3.2: Examples of ε-tangent and normal cones. For the figure at the left there are two ε-binding constraints at x. For the middle figure, only one constraint is ε-binding at x. For the figure at the right none of the constraints is binding at x.

Another important observation is that one can proceed from x along any direction in $T(x, \varepsilon)$ for a distance of at least ε and remain inside the feasible region. This is formalized in the proposition that follows.

Proposition 3.1. *If $x \in \Omega$ and $v \in T(x, \varepsilon)$ satisfies $\|v\| \leq \varepsilon$, then $x + v \in \Omega$.*

Proof. Let $x \in \Omega$ and $v \in T(x, \varepsilon)$ such that $\|v\| \leq \varepsilon$. As $T(x, \varepsilon) = (N(x, \varepsilon))^\circ$, we have that $-a_i^T v \leq 0 \ \forall i \in I(x, \varepsilon)$. So for $i \in I(x, \varepsilon)$, $a_i^T(x+v) = a_i^T x + a_i^T v \geq a_i^T x \geq b_i$ because $x \in \Omega$. For $i \notin I(x, \varepsilon)$, the face A_i where the i^{th} constraint is not binding is at distance of at least ε of x. So $x + v \in \Omega$. ■

We prove in the next proposition that the number of distinct ε-normal cones is finite for a given ε.

Proposition 3.2. *Let $x \in \Omega$ and $\varepsilon > 0$. There are at most 2^m distinct sets $I(x, \varepsilon)$. Consequently, there are at most 2^m distinct cones $N(x, \varepsilon)$ and at most 2^m distinct cones $T(x, \varepsilon)$.*

Proof. Each $I(x, \varepsilon)$ is a subset of $\{1, \ldots, m\}$ of which are exactly 2^m possible subsets including the empty set, where m is the number of rows of the constraints matrix A. As there is a one to one relationship between $I(x, \varepsilon)$ and $N(x, \varepsilon)$, the remainder of the proof follows. ■

3.3.2 Conditions on the Search Directions for GSS

For linearly-constrained GSS, the set of search directions $\mathcal{D}_k$ can be computed at each iteration. $\mathcal{D}_k$ is partitioned into two subsets that play different roles in the analysis.

$$\mathcal{D}_k = \mathcal{G}_k \cup \mathcal{H}_k.$$

The set $\mathcal{G}_k$ is called the set of *core directions*. It is required to generate the ε-tangent cone $T(x_k, \varepsilon_k)$. This requirement is what lead to the name *generating set search*. This set plays an important role in the analysis.

The possibly empty set $\mathcal{H}_k$ accommodates any remaining direction in $\mathcal{D}_k$ and is meant in general to accelerate the progress of the search, but it has little effect on the analysis.

The fact that the generators of $T(x_k, \varepsilon_k)$ be contained in $\mathcal{G}_k$ is very important for the feasibility of the iterates and for the proof of convergence. As a practical matter, the set of search direction should also include the generators of $N(x_k, \varepsilon_k)$. This implies that the set of search directions positively spans $\mathbb{R}^n$. Also, we need the following condition to be satisfied [11].

Condition 1. There exists a constant $\kappa_{min} > 0$ independent of k such that for every k for which $T(x_k, \varepsilon_k) \neq \{0\}$, the set $\mathcal{G}_k$ generates $T(x_k, \varepsilon_k)$ and satisfies $\kappa(\mathcal{G}_k) \geq \kappa_{min}$.

$\kappa(\mathcal{G}_k)$ is defined as

$$\kappa(\mathcal{G}_k) = \min_{\substack{v \in \mathbb{R}^n \\ v_K \neq 0}} \max_{d \in \mathcal{G}_k} \frac{v^T d}{\|v_K\| \|d\|}, \tag{3.2}$$

where K is a cone generated by $\mathcal{G}_k$ and v_K the projection of v onto K. $\kappa(\mathcal{G}_k)$ is called the *cosine measure* of the set of direction $\mathcal{G}_k$. A simple result easy to prove is that if $\mathcal{G} \neq \emptyset$, then $\kappa(\mathcal{G}) > 0$ for any set of directions $\mathcal{G}$.

The need for condition 1 arises when there is a linearity (a subvector space) present in $T(x_k, \varepsilon_k)$ as in the middle picture in Figure 3.2. It prevents choosing core directions nearly orthogonal to the direction of the steepest descent.

Although there are only finitely many ε-tangent cones, the set of possible generators for each cone is not necessarily unique. To enforce Condition 1, we should for example, have a memory, which prevents us from choosing a set of core directions that has been chosen before. As the number of $\mathcal{G}_k$ is finite, by Proposition 3.2, Condition 1 is satisfied; take κ_{min} to be the minimum of all the $\kappa(\mathcal{G}_k)$. Another condition easy to enforce and necessary for the proofs of convergence of GSS methods is the following.

Condition 2. Whenever $T(x, \varepsilon) \neq \{0\}$, the search directions in $\mathcal{D}_k$ are uniformly bounded above and below away from zero, that is there exist $\beta_{min}, \ \beta_{max}$ satisfying $\beta_{max} > \beta_{min} > 0$ such that $\beta_{min} \leq \|d\| \leq \beta_{max}, \ $ for all $d \in \mathcal{D}_k$.

3.4 Choice of Step-length Control Parameter and Globalization

The step-length parameter Δ_k determine the length of the step along the search directions at each iteration. For the unconstrained case a full step is taken at each iteration so that the trials points are in the form

$$\{x_k + \Delta_k d_k^{(i)} \mid i = 1, \ldots, r_k\},$$

where r_k is the number of search directions at iteration k.
For the constrained case, there are many techniques to determine the step-length parameters.

The general form for the trial points at each iteration is

$$\{x_k + \widetilde{\Delta}_k^{(i)} d_k^{(i)} \mid i = 1, \ldots, p_k\},$$

where $\widetilde{\Delta}_k^{(i)} \in [0, \Delta_k]$. The simplest formula for choosing $\widetilde{\Delta}_k^{(i)}$ might be the following:

$$\widetilde{\Delta}_k^{(i)} = \begin{cases} \Delta_k & \text{if } x_k + \Delta_k d_k^{(i)} \in \Omega \\ 0 & \text{otherwise.} \end{cases}$$

This means that infeasible trial points are rejected. Another technique is to stop at the boundary, that is to take the maximum value of $\widetilde{\Delta}_k^{(i)}$ such that $x_k + \Delta_k d_k^{(i)} \in \Omega$. More generally, for GSS methods, when a trial point is feasible, we take the full step in the sense of the following condition that need to be satisfied.

Condition 3. If $x_k + \widetilde{\Delta}_k^{(i)} d_k^{(i)} \in \Omega$, take $\widetilde{\Delta}_k^{(i)} = \Delta_k$.

The conditions on the choice of the step length-control parameters and their updates depend on the criterion of acceptance of a step. The GSS algorithms are greedy, that is, at each iteration k, x_k is always the best feasible point discovered thus far; i.e., $f(x_k) \leq f(x_j)$ for $j \leq k$.
For an iteration k to be deemed successful, it is required that

$$x_k + \widetilde{\Delta}_k d_k \in \Omega \text{ and } f(x_k + \widetilde{\Delta}_k d_k) < f(x_k) - \rho(\Delta_k) \text{ for some } d_k \in \mathcal{D}_k \text{ and } \widetilde{\Delta}_k \in [0, \Delta_k]. \quad (3.3)$$

The function ρ is called the *forcing function* and must satisfy the following conditions.

Condition 4. (General requirements for the forcing function)

1. The function ρ is a non-negative continuous function;
2. The function ρ is $o(t)$ as $t \downarrow 0; i.e., \frac{\rho(t)}{t} \to 0$ as $t \downarrow 0$;
3. The function ρ is nondecreasing; i.e, $\rho(t_1) \leq \rho(t_2)$ if $t_1 \leq t_2$.

Examples of forcing function are $\rho(t) = 0, \quad \rho(t) = 5t^3$ and $\rho(t) = \frac{t^2}{1+t}$, with $t \geq 0$.
In the case of successful iterations (one that satisfies (3.3)), the next iterate is defined by

$$x_{k+1} = x_k + \widetilde{\Delta}_k d_k \text{ for } k \in \mathcal{S},$$

where $\mathcal{S}$ is the set of successful iteration. In addition, Δ_k is update according to

$$\Delta_k = \phi_k \Delta_k, \quad \phi_k \geq 1 \text{ for } k \in \mathcal{S},$$

where ϕ_k is called the expansion parameter.

For the k^{th} iteration to be deemed unsuccessful, we must have

$$x_k + \Delta_k d \notin \Omega \text{ or } f(x_k + \Delta_k d) \geq f(x_k) - \rho(\Delta_k) \text{ for every } d \in \mathcal{G}_k.$$

In this case, the best point is unchanged

$$x_{k+1} = x_k \text{ for } k \in \mathcal{U},$$

where $\mathcal{U}$ is the set of unsuccessful iterations. In addition, the step-length parameter is reduced

$$\Delta_{k+1} = \theta_k \Delta_k, \quad \theta_k \in (0, 1) \text{ for } k \in \mathcal{U}.$$

The parameter θ_k is called the contraction parameter.

Another way to choose the expansion parameter ϕ_k for successful iteration is discussed by Lucidi and Siandrone in algorithm 2 in [12]. They used line-search techniques to expand the step-length control parameter along "good" directions. More generally, the choice of the step-length control parameter depend on the globalization strategy. This is a way to enforce global convergence of the algorithm. The idea is to choose the updates of the step-length parameters in order to enforce the condition $\liminf_{k\to\infty} \Delta_k = 0$ or a stronger condition $\lim_{k\to\infty} \Delta_k = 0$; at least for k in a subsequence of the sequence of iterations.
We now discuss globalization via sufficient decrease condition.

3.5 Globalization via Sufficient Decrease Condition

The notion of sufficient decrease is well known is gradient based optimization. Enforcing sufficient decrease ties the choice of step-length parameter to the expected decrease, as estimated by the initial rate of decrease $-\nabla f(x_k)^T d_k$. In the context of derivative-free methods for which we are concerned, the underlying assumption is that the value of $\nabla f(x_k)$ is not available. Therefore, we can not use the type of sufficient decrease conditions often used in gradient-based methods. However, as in [6] and [13], we can use alternatives that use step-length control parameter Δ_k rather than $\nabla f(x_k)$. In the context of GSS methods for linearly-constrained optimization, a sufficient decrease globalization strategy requires the following conditions to be satisfied for the forcing function ρ and the choice of the contraction parameter θ_k.

Condition 5.
(a) The forcing function for sufficient decrease: The forcing function ρ is such that $\rho(t) > 0$ for $t > 0$.
(b) Contracting Δ_k for sufficient decrease: A constant $\theta_{\max} < 1$ exists such that $\theta_k \leq \theta_{\max}$ for all k.

We recall that θ_k is the contraction parameter in the update of the step-length parameter Δ_k. Enforcing Condition 5(a), which ensure that indeed there is sufficient decrease as opposed to simple decrease ($\rho(t) = 0$ for all $t \geq 0$), is easy. One can choose $\rho(t)$ to be for example $\rho(t) = 10^{-5}t^2$. The upper bound on the contraction factor θ_k is to make sure that a predictable fraction of reduction on Δ_k is made at the conclusion of an unsuccessful iteration. This is to avoid very small reduction of Δ_k that can prevent the method to terminate. A fixed contraction parameter (for example $\theta_k = \frac{1}{3},\ \forall k$) satisfies obviously condition 5(b).

Algorithm 3.5.1. Linearly constrained GSS using a sufficient decrease globalization strategy
Initialization.
Let $x_0 \in \Omega$ be the initial guess.
Let $\Delta_{\text{tol}} > 0$ be the tolerance used to test for convergence.
Let $\Delta_0 > \Delta_{\text{tol}}$ be the initial value of the step-length control parameter.
Let $\varepsilon_{\max} > \beta_{\max}\Delta_{\text{tol}}$ be the maximum distance used to identify nearby constraints ($\varepsilon_{\max} = +\infty$ is permissible).
Let ρ be the forcing function satisfying condition 4 and 5(a).

Algorithm. For each iteration $k = 0, 1, \ldots$

Step 1. Let $\varepsilon_k = \min\{\varepsilon_{\max}, \beta_{\max}\Delta_k\}$. Choose a set of search directions $\mathcal{D}_k = \mathcal{G}_k \bigcup \mathcal{H}_k$ satisfying condition 1 and 2.

Step 2. If there exist $d_k \in \mathcal{D}_k$ and a corresponding $\widetilde{\Delta}_k \in [0, \Delta_k]$ satisfying condition 3 such that $x_k + \widetilde{\Delta}_k d_k \in \Omega$ and $f(x_k + \widetilde{\Delta}_k d_k) < f(x_k) - \rho(\Delta_k)$,
then:
- set $x_{k+1} = x_k + \widetilde{\Delta}_k d_k$.
-set $\Delta_{k+1} = \phi_k \Delta_k$ for any choice of $\phi_k \geq 1$.

Step 3. Otherwise for every $d \in \mathcal{G}_k$, either $x_k + \widetilde{\Delta}_k d \notin \Omega$ or $f(x_k + \widetilde{\Delta}_k d_k) \geq f(x_k) - \rho(\Delta_k)$.
In this case:
-set $x_{k+1} = x_k$ (no change),
-set $\Delta_{k+1} = \theta_k \Delta_k$ for some choice of $\theta_k \in (0, 1)$ satisfying condition 5(b).
If $\Delta_{k+1} < \Delta_{\text{tol}}$, then terminate.

Figure 3.3: The GSS algorithm for linearly-constrained optimization problem of Kolda, Lewis and Torczon.

The GSS algorithm using the sufficient decrease globalization strategy is presented in Figure 3.3. Note that when using sufficient decrease condition, the update of the step-length control parameter $\Delta_k^{(i)}$ within an iterate is such that if $x_k + \Delta_k^{(i)} d_k \notin \Omega$, one can simply find the step to the nearest constraints from x_k along $d_k^{(i)}$. So $\Delta_k^{(i)}$ is maximal such that $x_k + \Delta_k^{(i)} d_k \in \Omega$.
We now state a result which assures that when the forcing function satisfies Condition 4, $\liminf_{k\to\infty} \Delta_k = 0$; the proof can be found in [6].

Proposition 3.3. *Let f be bounded below. Suppose that $\rho(t)$ is as specified in Condition 4, then GSS algorithm given in Figure 3.3 produce iterates satisfying*

$$\liminf_{k\to\infty} \Delta_k = 0.$$

This result is very important for the proof of convergence of GSS methods. In the next section, we will prove that when k goes to infinity, there is a big-O relationship between the measure of

stationarity $\chi(x_k)$ at x_k and Δ_k. The immediate consequence will be the convergence of the GSS algorithm to a stationary point of the linearly-constrained optimization problem (3.1).

3.6 Convergence Analysis

In this section, we prove the convergence of the algorithm presented in the previous section.

The first result relates the norm of the projection of the direction of the steepest descent onto the ε_k tangent cone at the iterate x_k to the step-length control parameter.

Theorem 3.1. *[14] Suppose that the gradient of f is Lipschitz continuous with constant M on Ω. Consider the linearly constrained algorithms given in Figure 3.3, which satisfy Condition 1, 2 and 3. If $k \in \mathcal{U}$ and ε_k satisfies $\varepsilon_k = \beta_{\max}\Delta_k$, then*

$$\|P_{T(x_k,\varepsilon_k)}(-\nabla f(x_k))\| \leq \frac{1}{\kappa_{\min}}\left(M\Delta_k\beta_{\max} + \frac{\rho(\Delta_k)}{\Delta_k\beta_{\min}}\right). \tag{3.4}$$

Here $\kappa_{\min}$ is from Condition 1 and $\beta_{\min}$ and $\beta_{\max}$ are from Condition 2.

From Theorem 3.1, we can deduce a relationship between the measure of stationarity $\chi(x_k)$ of the iterate x_k and the step-length control parameter. Before we proceed, we need to recall some results on the geometry of cones and polyhedra and set some more notations. We define the quantity

$$\nu_{\min} = \min\left\{\kappa(\mathcal{A}) : \mathcal{A} = \underset{i\in I(x,\varepsilon)}{\cup}\{-a_i\}, \quad x \in \Omega,\ \varepsilon \geq 0,\ I(x,\varepsilon) \neq \emptyset\right\}, \tag{3.5}$$

where κ is given in (3.2). We have that $\nu_{\min} > 0$ because $\kappa(\mathcal{G}) > 0$, for all $\mathcal{G} \neq \emptyset$.
Let us prove this elementary result concerning the projection of a point onto an hyperplane of $\mathbb{R}^n$ which will be helpful to prove results on the convergence of GSS algorithm.

Lemma 3.1. *Let $a \in \mathbb{R}^n, \quad b \in \mathbb{R}$ with $a \neq 0$ and consider the hyperplane $\mathcal{L} = \{x \,|\, a^T x = b\}$. Let $x_0 \in \mathbb{R}^n$. The distance of x_0 to the hyperplane $\mathcal{L}$ which is in fact the projection of x_0 onto $\mathcal{L}$ is given by*

$$\mathrm{dist}(\mathrm{x}_0, \mathcal{L}) = \inf_{\mathrm{x}\in\mathcal{L}} \|\mathrm{x} - \mathrm{x}_0\| = \frac{|\mathrm{b} - \mathrm{a}^{\mathrm{T}}\mathrm{x}_0|}{\|\mathrm{a}\|}.$$

Proof. For simplicity, we will solve the problem with the norm squared and we will take the square root of the solution. So the lagrangian associated to this new problem is given by

$$L(x, \lambda) = \|x - x_0\|^2 - \lambda(a^T x - b),$$

where λ is the Lagrange multiplier. We have $\nabla_x L(x,\lambda) = x - x_0 - \lambda a$ and then the optimality conditions reads

$$x - x_0 - \lambda a = 0 \tag{3.6}$$
$$a^T x - b = 0. \tag{3.7}$$

Multiplying to the left (3.6) by a^T and using (3.7), we find $b - a^T x_0 = \lambda \|a\|^2$ so that $\lambda = \frac{b-a^T x_0}{\|a\|^2}$. Inserting this into (3.6) and solving for x, we get $x = x_0 + \frac{b-a^T x_0}{\|a\|^2} a$. Substituting into the objective function squared , we get the optimal value of the new problem $[\mathrm{dist}(\mathrm{x}_0, \mathcal{L})]^2 = \|\frac{b-a^T x_0}{\|a\|^2} a\|^2 = \frac{|b-a^T x_0|^2}{\|a\|^2}$ and taking the square root, we find the desired result. ■

We recall that given a convex cone K and any vector v, there is a unique closest point of K to v, the projection of v onto K, which we will denote v_K. Thus $v_{N(x,\varepsilon)}$ is the projection of v onto the ε-normal cone at x, $N(x,\varepsilon)$ while $v_{T(x,\varepsilon)}$ is the projection of v onto the ε-tangent cone at x, $T(x,\varepsilon)$.

We now prove that if one can move from x to $x+v$ and remain feasible, then v is pointing inward with respect to the constraints near x.

Proposition 3.4. *If $x \in \Omega$ and $x + v \in \Omega$, then for any $\varepsilon \geq 0$, $\quad \|v_{N(x,\varepsilon)}\| \leq \frac{\varepsilon}{\nu_{\min}}$ where $\nu_{\min}$ is the constant in (3.5).*

Proof. Let $N = N(x,\varepsilon)$. If $v_N = 0$, then the result is obvious. So we only consider the case $v_N \neq 0$. We know that N is generated by the outward pointing normals to the binding constraints within a distance ε of x. This means that the set $\mathcal{A} = \{-a_i \,|\, i \in I(x,\varepsilon)\}$ generates N. By lemma 3.1 applied to the hyperplane $A_i = \{y \,|\, a_i^T y = b_i\}$ and x, we have

$$\frac{a_i^T x - b_i}{\|a_i\|} \leq \varepsilon \quad \text{for all } i \in I(x,\varepsilon).$$

At the same time, $x + v \in \Omega$, so we have that

$$a_i^T x + a_i^T v \geq b_i \quad \text{for all } i.$$

The last two relations then lead to

$$-a_i^T v \leq a_i^T x - b_i \leq \varepsilon \|a_i\| \quad \text{for all } i.$$

Since N is generated by $\mathcal{A} \subset -A = [-a_1, \ldots, -a_n]$ and $v_N \neq 0$, by the definition of $\nu_{\min}$ and κ, we have

$$\nu_{\min} \|v_N\| \leq \kappa \left(\bigcup_{i \in I(x,\varepsilon)} \{-a_i\} \right) \|v_N\| \leq \max_{i \in I(x,\varepsilon)} \frac{-a_i^T v}{\|a_i\|} \leq \max_{i \in I(x,\varepsilon)} \frac{\varepsilon \|a_i\|}{\|a_i\|} = \varepsilon$$

and the proof is completed. ■

Now for $x \in \Omega$ and $v \in \mathbb{R}^n$, consider the following generalization of the measure of stationarity $\chi(x)$ defined in Chapter 2

$$\hat{\chi}(x,v) = \max_{\substack{x+\omega \in \Omega \\ \|\omega\| \leq 1.}} \omega^T v, \tag{3.8}$$

It is easy to see that $\chi(x) = \hat{\chi}(x, -\nabla f(x))$. This generalization help to understand that the result that follow is a geometrical fact about polyhedra. It relates $\hat{\chi}(x,v)$ with the projection of v onto the ε-tangent and ε-normal cones at x.

Proposition 3.5. *If $x \in \Omega$, then for all $\varepsilon \geq 0$,*

$$\hat{\chi}(x, v) \leq \|v_{T(x,\varepsilon)}\| + \frac{\varepsilon}{\nu_{\min}} \|v_{N(x,\varepsilon)}\|,$$

where $\nu_{\min}$ is given at (3.5).

Proof. We set $N = N(x, \varepsilon)$ and $T = T(x, \varepsilon)$. Using the polar decomposition of v, $v = v_N + v_T$, we obtain

$$\hat{\chi}(x, v) = \max_{\substack{x+\omega \in \Omega \\ \|\omega\| \leq 1}} \omega^T v \leq \max_{\substack{x+\omega \in \Omega \\ \|\omega\| \leq 1}} \omega^T v_T + \max_{\substack{x+\omega \in \Omega \\ \|\omega\| \leq 1}} \omega^T v_N.$$

The first term in the right hand side of the preceding inequality satisfies

$$\max_{\substack{x+\omega \in \Omega \\ \|\omega\| \leq 1}} \omega^T v_T \leq \|v_T\|,$$

by the Cauchy-Schwartz inequality. Meanwhile, using the polar decomposition of ω, we have that

$$\omega^T v_N = (\omega_N + \omega_T)^T v_N \leq \omega_N^T v_N,$$

since $\omega_T^T v_N \leq 0$. Thus,

$$\max_{\substack{x+\omega \in \Omega \\ \|\omega\| \leq 1}} \omega^T v_N \leq \max_{\substack{x+\omega \in \Omega \\ \|\omega\| \leq 1}} \omega_N^T v_N \leq \max_{\substack{x+\omega \in \Omega \\ \|\omega\| \leq 1}} \|\omega_N\| \|v_N\|.$$

However, since $x + \omega \in \Omega$, Proposition 3.4 assures us that $\|\omega_N\| \leq \frac{\varepsilon}{\nu_{\min}}$. Therefore ,

$$\hat{\chi}(x, v) \leq \|v_T\| + \frac{\varepsilon}{\nu_{\min}} \|v_{N(x,\varepsilon)}\|.$$

■

We are now ready to state and prove the key result for the convergence of linearly constrained GSS methods.

Theorem 3.2. *Suppose that the gradient of f is Lipschitz continuous with constant M on Ω. Suppose also that the set $\mathcal{F} = \{x \in \Omega \,|\, f(x) \leq f(x_0)\}$ is bounded. Consider the linearly constrained GSS algorithm given in figure 3.3, which satisfy condition 1, 2 and 3. If $k \in \mathcal{U}$ and ε_k satisfies $\varepsilon_k = \beta_{\max} \Delta_k$, then*

(i) *There exists $\gamma > 0$ such that for all $x \in \mathcal{F}$,*

$$\|\nabla f(x)\| < \gamma. \tag{3.9}$$

(ii)

$$\chi(x_k) \leq \left(\frac{M}{\kappa_{\min}} + \frac{\gamma}{\nu_{\min}} \right) \Delta_k \beta_{\max} + \frac{1}{\kappa_{\min} \beta_{\min}} \frac{\rho(\Delta_k)}{\Delta_k}. \tag{3.10}$$

Here $\nu_{\min}$ is from (3.5) and the other constants are like in theorem 3.1.

Proof. (i) As $\mathcal{F}$ is bounded, there exist a constant $\alpha > 0$ such that $\|x\| \leq \alpha$ for all $x \in \mathcal{F}$. Now let $x \in \mathcal{F}$; as $x_0 \in \mathcal{F}$, we use the fact that the gradient of f is Lipschitz continuous with constant M on Ω to obtain

$$\|\nabla f(x) - \nabla f(x_0)\| \leq M\|x - x_0\| \leq M(\|x\| + \|x_0\|) \leq 2M\alpha.$$

But now

$$\|\nabla f(x)\| \leq \|\nabla f(x) - \nabla f(x_0)\| + \|\nabla f(x_0)\| \leq 2M\alpha + \|\nabla f(x_0)\|.$$

So part (i) of the theorem is proven with $\gamma = 2M\alpha + \|\nabla f(x_0)\|$.

(ii) Since $\varepsilon = \Delta_k \beta_{\max} \geq 0$, Proposition 3.5 applied with $x = x_k$ and $v = -\nabla f(x_k)$ gives

$$\chi(x_k) \leq \| \left[-\nabla f(x_k) \right]_{T(x_k, \varepsilon_k)} \| + \frac{\Delta_k \beta_{\max}}{\nu_{\min}} \| \left[-\nabla f(x_k) \right]_{N(x_k, \varepsilon_k)} \|.$$

By Theorem 3.1 , we have $\| \left[-\nabla f(x_k) \right]_{T(x_k, \varepsilon_k)} \| \leq \frac{1}{\kappa_{\min}} \left(M \Delta_k \beta_{\max} + \frac{\rho(\Delta_k)}{\Delta_k \beta_{\min}} \right)$. Because the projection onto a convex set is contractive, we also have

$$\| \left[-\nabla f(x_k) \right]_{N(x_k, \varepsilon_k)} \| \leq \|\nabla f(x_k)\|.$$

Combining all this with part (i) we get

$$\chi(x_k) \leq \frac{1}{\kappa_{\min}} \left(M \Delta_k \beta_{\max} + \frac{\rho(\Delta_k)}{\Delta_k \beta_{\min}} \right) + \frac{\Delta_k \beta_{\max}}{\nu_{\min}} \gamma.$$

Rearranging the terms of this last expression yields the desired bound on $\chi(x_k)$. ■

A direct consequence of this result is the following global convergence result of the linearly-constrained Generating Set Search methods.

Corollary 3.1. *Suppose that the gradient of f is Lipschitz continuous with constant M on Ω. Suppose also that the set $\mathcal{F} = \{x \in \Omega \,|\, f(x) \leq f(x_0)\}$ is bounded. Consider the linearly-constrained GSS algorithm given in figure 3.3, which satisfy Condition 1, 2, 3, 4, 5. Then we have*

$$\liminf_{k \to \infty} \ \chi(x_k) = 0$$

Proof. This follows from Theorem 3.2 and Proposition 3.3. ■

4. An augmented Lagrangian GSS Algorithm for General Nonlinear Problems with Explicit Linear Constraints

In this chapter, we study an extention of the GSS methods for optimization problem with combination of general nonlinear constraints and explicit linear constraints. This work follows the paper of Kolda, Lewis and Torczon [4] where they considered nonlinear equality constraints. Here we consider problems with nonlinear equality and inequality constraints and we aim to design a GSS algorithm to solve it. We introduce slack variables to transform the inequalities into equalities. The minimization of the resulting problem can be carry out explicitly with respect to the slack variable for the original variable fixed. This leads to an equivalent problem with the same dimension as the original problem. We give an algorithm to solve this equivalent problem with accompanying proofs of convergence.

4.1 Problem, Formulations and First Results

We recall the problem at hand in this work

$$\begin{aligned} \text{Minimize} \quad & f(x) && (4.1)\\ \text{Subject to:} \quad & c_{\mathcal{E}}(x) = 0, && (4.2)\\ & c_{\mathcal{I}}(x) \geq 0, && (4.3)\\ \text{and} \quad & Ax \geq b, && (4.4) \end{aligned}$$

where the functions $f,\ c_{\mathcal{E}},\ c_{\mathcal{I}},\ A$ and b are as in the introduction. We denote, as in the previous chapter, the set of linear constraints by

$$\Omega = \{x \,|\, Ax \geq b\}.$$

We introduce the vector of non negative slack variables $z^s = (z_i^2)_{i\in\mathcal{I}}$ (the upperscript s stand for squared) so that the inequalities constraints (4.3) become

$$c_i(x) - z_i^2 = 0, \quad \text{for } i \in \mathcal{I}$$

and the problem (4.1)-(4.4) becomes

$$\begin{array}{lll} \text{Minimize} & f(x) & \\ \text{Subject to:} & c_j(x) = 0 & j \in \mathcal{E} \\ & c_i(x) - z_i^2 = 0 & i \in \mathcal{I} \\ \text{and} & x \in \Omega. & \end{array} \tag{4.5}$$

Clearly x_* solves (4.1)-(4.4) if and only if (x_*, z_*) solves (4.5) with $z_{*,i} = \sqrt{c_i(x_*)}$. The augmented Lagrangian associated to (4.5) is given by

$$L(x, z; \lambda, \nu; \mu) = f(x) + \lambda^T c_{\mathcal{E}}(x) + \nu^T (c_{\mathcal{I}}(x) - z^s) + \frac{1}{2\mu}\|c_{\mathcal{E}}(x)\|^2 + \frac{1}{2\mu}\|c_{\mathcal{I}}(x) - z^s\|^2, \quad (4.6)$$

where $\lambda \in \mathbb{R}^{n_\mathcal{E}}, \quad \nu \in \mathbb{R}^{n_\mathcal{I}}$ are vectors of Lagrange multipliers and μ is the penalty parameter. We recall that solving (4.1)-(4.4) is equivalent to a sequence of minimization of the augmented Lagrangian (4.6) with respect to (x, z), the sequence being relative to a decreasing sequence of penalty parameters and a bounded sequence of Lagrange multipliers (See Section 2.5).

As noted in the introduction to this chapter, we can carry out the minimization of the augmented Lagrangian given in (4.6) with respect to z explicitly with x fixed. This corresponds to solving the problem

$$\begin{aligned} \min_z L(x, z, \lambda, \nu; \mu) &= f(x) + \lambda^T c_{\mathcal{E}}(x) + \tfrac{1}{2\mu}\|c_{\mathcal{E}}(x)\|^2 \\ &\quad + \sum_{i\in\mathcal{I}} \min_{r_i \geq 0} \left\{ \nu_i (c_i(x) - r_i) + \tfrac{1}{2\mu}(c_i(x) - r_i)^2 \right\}. \end{aligned} \quad (4.7)$$

This problem is solved if we can solve, for the variable r_i, the problem

$$\min_{r_i \geq 0} \left\{ \nu_i (c_i(x) - r_i) + \frac{1}{2\mu}(c_i(x) - r_i)^2 \right\}. \quad (4.8)$$

The objective function of the minimization problem in (4.8) is a quadratic function of r_i. Its unconstrained (global) minimum is reached when its derivative with respect to r_i vanishes, that is when

$$-\nu_i - \frac{1}{\mu}(c_i(x) - r_i) = 0,$$

and this gives

$$\hat{r_i} = c_i(x) + \mu\nu_i.$$

Two cases can occurred; either $\hat{r_i} \geq 0$ and then $\hat{r_i}$ solves (4.8) or the solution of (4.8) is $r_{i*} = 0$. So the solution of (4.8) is

$$r_{i*} = \max\{0, c_i(x) + \mu\nu_i\}.$$

It follows that $c_i(x) - r_{i*} = \min\{c_i(x), -\mu\nu_i\}$ and if we write

$$c_i^-(x, \nu; \mu) = \min\{c_i(x), -\mu\nu_i\} \quad (4.9)$$

and denote by $c_\mathcal{I}^-(x, \nu; \mu)$ the function whose components are the $c_i^-(x, \nu; \mu)'$s, the optimal value of (4.7) is then given by

$$L(x, z_*; \lambda, \nu, \mu) = f(x) + \lambda^T c_{\mathcal{E}}(x) + \frac{1}{2\mu}\|c_{\mathcal{E}}(x)\|^2 + \sum_{i\in\mathcal{I}} \left(\nu_i c_i^-(x) + \frac{1}{2\mu}(c_i^-(x))^2 \right).$$

By rearranging the terms in the summation and completing the square, this expression is equal to

$$L(x, \lambda, \nu, \mu) = f(x) + \lambda^T c_{\mathcal{E}}(x) + \frac{1}{2\mu}\|c_{\mathcal{E}}(x)\|^2 + \frac{\mu}{2}\sum_{i\in\mathcal{I}} \left(\left[\min\{\nu_i + \frac{1}{\mu}c_i(x), 0\} \right]^2 - \nu_i^2 \right). \quad (4.10)$$

We have then proven the following result.

Proposition 4.1. (i) *The problem*

$$\begin{array}{ll} \textit{Minimize} & L(x,z,\lambda,\nu;\mu), \\ \textit{Subject to:} & (x,z) \in \Omega \times \mathbb{R}^{n_{\mathcal{I}}}, \end{array} \tag{4.11}$$

is equivalent to the problem

$$\begin{array}{ll} \textit{Minimize} & L(x,\lambda,\nu;\mu), \\ \textit{Subject to:} & x \in \Omega. \end{array} \tag{4.12}$$

(ii) $x_* = x_*(\lambda,\nu,\mu)$ *is solution of (4.12) if and only if* $(x_*, z_*) = (x_*(\lambda,\nu;\mu), z_*(\lambda,\nu;\mu))$ *is solution to (4.11) and*

$$z_{*i}^2 = \max\{0, c_i(x_*) + \mu\nu_i\}, \quad i \in \mathcal{I}.$$

In the remainder of this section, we will design an *augmented Lagrangian Generating Set Search algorithm* to solve (4.1)-(4.4). A particular feature of our algorithm will be that in its "inner iteration", we solve problem (4.12) with the linearly-constrained GSS method studied in Chapter 3. At the present stage, there are two important things to care about, one theoretical and the other computational. Theoretically, we need, for the purpose of the analysis of convergence to know how hypotheses on data $(f(c),\ c(x))$ infer on this new problem. From a computational point of view, we need to suitably define the updates of Lagrange multipliers and penalty parameters.

The following result from [1] address the theoretical problem.

Proposition 4.2. (a) *If* f *and* c *are continuous on a subset* S *of* $\mathbb{R}^n$, *then* $L(.,\lambda,\nu;\mu)$ *is continuous on* S *for each* $\lambda,\ \nu$ *and* $\mu > 0$.

(b) *If* f *and* c *are* C^1 *on an open subset* S *of* $\mathbb{R}^n$, *then* $L(.,\lambda,\nu;\mu)$ *is* C^1 *on* S *for each* $\lambda,\ \nu$ *and* $\mu > 0$.

(c) *If* f, c *are* C^2 *on an open subset* S *of* $\mathbb{R}^n$, *then* $L(.,\lambda,\nu;\mu)$ *is* C^2 *on the set*

$$\hat{S}_{\nu,\mu} = S \cap \{x \,|\, c_i(x) \neq -\nu_i\mu,\ \forall i \in \mathcal{I}\},$$

for each $\lambda,\ \nu$ *and* $\mu > 0$.

For the computational problem, if $x_k(\lambda_k,\nu_k;\mu_k)$ is obtained by a minimization of $L(.,\lambda_k,\nu_k;\mu_k)$, we will consider the first-order Lagrange multiplier iteration given by Bertsekas [1] which fits in the context of derivative-free methods we are interested in. That is

$$\lambda_{k+1} = \lambda_k + \frac{1}{\mu_k} c_{\mathcal{E}}(x_k(\lambda_k,\nu_k;\mu_k)), \tag{4.13}$$

$$\nu_{k+1} = \nu_k + \frac{1}{\mu_k} c_{\mathcal{I}}^{-}\left[x_k(\lambda_k,\nu_k;\mu_k), \nu_k;\mu_k\right]. \tag{4.14}$$

However, in view of the definition of $c_{\mathcal{I}}^{-}(x_k, \nu_k; \mu_k)$ given in (4.9), we can also write (4.14) as

$$\nu_{k+1,i} = \nu_{k,i} + \frac{1}{\mu_k} \min\{c_i(x_k(\lambda_k, \nu_k; \mu_k)), -\mu_k \nu_{k,i}\}$$

and finally

$$\nu_{k+1,i} = \min\{\nu_{k,i} + \frac{1}{\mu_k} c_i(x_k(\lambda_k, \nu_k; \mu_k)), 0\}. \tag{4.15}$$

We will choose as update of the penalty parameter a sequence satisfying $\mu_{k+1} \leq \mu_k, \quad \mu_k \to 0$.

4.2 The GSS Augmented Lagrangian

For computational reasons, the constraints $c_{\mathcal{E}}(x)$ and $c_{\mathcal{I}}(x)$ are assumed to be partitioned into $q_{\mathcal{E}} + q_{\mathcal{I}}$ disjoint subsets $\{\mathcal{Q}_j\}_{j=1}^{q_{\mathcal{E}}}$ and $\{\mathcal{Q}_l\}_{l=q_{\mathcal{E}}+1}^{q_{\mathcal{I}}}$ such that $\cup_{j=1}^{q_{\mathcal{E}}} \mathcal{Q}_j = \mathcal{E}$, and $\cup_{l=q_{\mathcal{E}}+1}^{q_{\mathcal{I}}} \mathcal{Q}_l = \mathcal{I}$. This partition enables the algorithm to place, at each iteration, greather emphasize on achieving feasibility for subsets of constraints that are, for any particular reason, proportionally more violated than the others. The Lagrangian given in (4.10) now takes the form

$$\begin{aligned} L(x, \lambda, \nu, \mu) \;\; = \;\; & f(x) + \sum_{j=1}^{q_{\mathcal{E}}} \sum_{i \in \mathcal{Q}_j} \left[\lambda_i c_i(x) + \tfrac{1}{2\mu_j} c_i(x)^2 \right] \\ & + \sum_{l=q_{\mathcal{E}}+1}^{q_{\mathcal{I}}} \tfrac{\mu_l}{2} \sum_{i \in \mathcal{Q}_l} \left[\left[\min\{\nu_i + \tfrac{1}{\mu_l} c_i(x), 0\} \right]^2 - \nu_i^2 \right]. \end{aligned} \tag{4.16}$$

The vector $\mu = (\mu_j, \mu_l; j = 1, \ldots, q_{\mathcal{E}}, l = q_{\mathcal{E}} + 1, \ldots, q_{\mathcal{I}})$ contains the penalty parameters associated with each partition of $c(x)$.

4.3 The Inner Iteration Subproblem

At the k^{th} outer iteration of the augmented Lagrangian method, an approximate solution of the following problem is required:

$$\begin{array}{ll} \text{Minimize} & L(x, \lambda_k, \nu_k; \mu_k) \\ \text{Subject to:} & x \in \Omega, \end{array} \tag{4.17}$$

where $\Omega = \{x \,|\, Ax \geq b\}$ and $L(x, \lambda_k, \nu_k; \mu_k)$ is given by (4.16) and the vectors λ_k ν_k and μ_k are updated at each outer iteration.

The stopping criterion for the approximate solution of (4.17) originally depends on the gradient of the augmented Lagrangian. Although the gradient of the augmented Lagrangian is very useful for the analysis of the convergence, we will neither use it nor approximate it in our algorithm.

In view of (4.16), the gradient of the augmented Lagrangian in given by

$$\begin{aligned}\nabla_x L(x,\lambda,\nu,\mu) &= \nabla_x f(x) + \sum_{j=1}^{q_{\mathcal{E}}} \sum_{i\in\mathcal{Q}_j} \left[\lambda_i \nabla_x c_i(x) + \tfrac{1}{\mu_j} c_i(x) \nabla_x c_i(x)\right] \\ &\quad + \sum_{l=q_{\mathcal{E}}+1}^{q_{\mathcal{I}}} \mu_l \sum_{i\in\mathcal{Q}_l} \min\{\nu_i + \tfrac{1}{\mu_l} c_i(x), 0\} \min\{\tfrac{1}{\mu_l}\nabla_x c_i(x), 0\},\end{aligned}$$

where the minimum applied to a vector is to understand componentwise. But for $l = 1, \ldots, q_{\mathcal{I}}$ and $i \in \mathcal{Q}_l$, the ℓth component of the term in the last summation is given by

$$\begin{aligned}&\left[\min\{\nu_i + \tfrac{1}{\mu_l} c_i(x),\, 0\} \min\{\tfrac{1}{\mu_l}\nabla_x c_i(x),\, 0\}\right]_\ell \\ &= \min\{\nu_i + \tfrac{1}{\mu_l} c_i(x), 0\} \min\{\tfrac{1}{\mu_l}\tfrac{\partial c_i(x)}{\partial x_\ell}, 0\} \\ &= \begin{cases} \tfrac{\nu_i}{\mu_l}\tfrac{\partial c_i(x)}{\partial x_\ell} + \tfrac{1}{\mu_l^2} c_i(x)\tfrac{\partial c_i(x)}{\partial x_\ell} \text{ if } \nu_i + \tfrac{1}{\mu_l} c_i(x) \le 0 \text{ and } \tfrac{1}{\mu_l}\tfrac{\partial c_i(x)}{\partial x_\ell} \le 0 \\ 0 \text{ otherwise} \end{cases} \\ &= \tfrac{1}{\mu_l} \widetilde{P}_+ \left(\nu_i \tfrac{\partial c_i(x)}{\partial x_\ell} + \tfrac{1}{\mu_l} c_i(x) \tfrac{\partial c_i(x)}{\partial x_\ell}\right).\end{aligned}$$

In the last equality, we used the notation

$$\widetilde{P}_+ \left(\nu_i \frac{\partial c_i(x)}{\partial x_l} + \frac{1}{\mu_l} c_i(x) \frac{\partial c_i(x)}{\partial x_\ell}\right) = \begin{cases} \nu_i \tfrac{\partial c_i(x)}{\partial x_\ell} + \tfrac{1}{\mu_l} c_i(x) \tfrac{\partial c_i(x)}{\partial x_\ell} \text{ if } \nu_i + \tfrac{1}{\mu_l} c_i(x) \le 0 \text{ and } \tfrac{\partial c_i(x)}{\partial x_\ell} \le 0 \\ 0 \text{ otherwise.} \end{cases}$$

So that we finally obtain

$$\begin{aligned}\nabla_x L(x,\lambda,\nu,\mu) &= \nabla_x f(x) + \sum_{j=1}^{q_{\mathcal{E}}} \sum_{i\in\mathcal{Q}_j} \left[\lambda_i \nabla_x c_i(x) + \tfrac{1}{\mu_j} c_i(x) \nabla_x c_i(x)\right] \\ &\quad + \sum_{l=q_{\mathcal{E}}+1}^{q_{\mathcal{I}}} \sum_{i\in\mathcal{Q}_l} \widetilde{P}_+ \left(\nu_i \nabla_x c_i(x) + \tfrac{1}{\mu_l} c_i(x) \nabla_x c_i(x)\right).\end{aligned} \tag{4.18}$$

For the convergence of the algorithm, we will require as in [4] that the following condition be satisfied by the functions f and c.

Condition 6. The functions $f(x)$ and $c(x)$ are twice continuously differentiable for all $x \in \Omega$.

We state the following result which is a direct consequence of Condition 6.

Proposition 4.3. *Assume that Condition 6 is satisfied. Let M be the Lipschitz constant of $\nabla f(x)$, $J_{\mathcal{E}}(x)$, $J_I(x)$, $J_{\mathcal{E}}(x)^T c_{\mathcal{E}}(c)$ and $J_{\mathcal{I}}(x)^T c_{\mathcal{I}}(x)$ on Ω. We have that*

$$\|\nabla_x L(x,\lambda,\nu,\mu) - \nabla_x L(y,\lambda,\nu,\mu)\| \le M \left(1 + \|\lambda\| + \|\nu\| + \sum_{j=1}^{q_{\mathcal{E}}+q_{\mathcal{I}}} \frac{1}{\mu_j}\right) \|x - y\|. \tag{4.19}$$

So $\nabla_x L(x, \lambda, \nu, \mu)$ is Lipschitz continuous with Lipschitz constant

$$M_k = M\left(1 + \|\lambda\| + \|\nu\| + \sum_{j=1}^{q_{\mathcal{E}}+q_{\mathcal{I}}} \frac{1}{\mu_j}\right). \tag{4.20}$$

Proof. See Appendix A. ■

4.4 The derivative-free Stopping Criterion and the Full Algorithm

We recall that in the inner iteration, the linearly-constrained GSS algorithm of Kolda, Lewis and Torczon in used to solve the subproblem (4.17). Specifically, we use algorithm 3.3 with the forcing function

$$\rho(\Delta) = \hat{\alpha}\Delta^2, \tag{4.21}$$

where $\hat{\alpha}$ is a small parameter (example $\hat{\alpha} = 10^{-3}$) and the contraction and expansion parameters

$$\theta_{k,i} = \frac{1}{2} \text{ and } \phi_{k,i} = 1 \text{ for all iteration } i \text{ of the GSS algorithm.}$$

The asymptotic behavior of the measure of stationarity of the iterates produced in the inner iteration is regulated by the step-length control parameter $\Delta_{k,i}$. We update at each outer iteration a tolerance parameter, δ_k, so that the stopping criterion in the inner iteration is to find an iterate $u \in \mathcal{U}$ (the subsequence of unsuccessful linearly-constrained GSS iterations) for which

$$\Delta_{k,u} \leq \delta_k, \tag{4.22}$$

where $\delta_k \to 0$.

The algorithm we propose here follows the idea of Kolda & Al [4] and Conn & Al. [5]. The latter authors studied the problem assuming availability of derivative information. They made the following assumptions for the purpose of their convergence analysis. We need similar assumptions for the consistency of our algorithm with their progenitors.

Condition 7. The set Ω is not empty.

Condition 8. The iterates $\{x_k\}$ produced by the GSS augmented Lagrangian algorithm lie within a closed, bounded set $\mathcal{B}$.

Condition 8 implies that the the augmented Lagrangian $L(x, \lambda_k, \nu_k; \mu_k)$ is bounded below and looking at Proposition 4.2, the conditions of Proposition 3.3 are satisfied. Thus the stopping criterion given by (4.22) will eventually be satisfied and thus the inner iteration terminates.

The test of convergence for the outer iteration need to also be derivative-free unlike the one given in [5]. For some δ_* (the stationarity tolerance) and η_* (non-linear equalities feasibility tolerance), we stop at the outer iteration if

$$\delta_k \leq \delta_* \text{ and } \|c_{\mathcal{E}}(x_k)\| \leq \eta_* \text{ and } c_{\mathcal{I}}(x_k) \geq 0. \tag{4.23}$$

Algorithm 4.4.1. A generating set search augmented Lagrangian algorithm

Step 0. **[Initialization]**. A partition of the set $\{1, \ldots, q_{\mathcal{E}} + q_{\mathcal{I}}\}$ into $q_{\mathcal{E}} + q_{\mathcal{I}}$ disjoints subsets $\{\mathcal{Q}\}_{j=1}^{q_{\mathcal{E}}}$ and $\{\mathcal{Q}\}_{j=1}^{q_{\mathcal{I}}}$ is given, as well as initial vector of Lagrange multiplier estimates λ_0 and ν_0 and positive penalty parameters μ_0 such that $\mu_{0,j} < 1$, $\mu_{0,l} < 1$, $j = 1, \ldots, q_{\mathcal{E}}$, $l = q_{\mathcal{E}} + 1, \ldots, q_{\mathcal{I}}$.
Set $\kappa_0 = \max_{1,\ldots,m}\{\|a_i\|\}$. The strict positive constants $\eta_* \ll 1$, $\tau < 1$, $\alpha_\eta < 1$, and $\beta_\eta < 1$ are specified.
Set $\alpha_0 = \max_{1,\ldots,q_{\mathcal{E}}+q_{\mathcal{I}}} \mu_{0,j}$, $\omega_0 = \alpha_0$, $\eta_0 = \alpha_0^{\alpha_\eta}$,
and $k = 0$. In addition, specify strictly positive constants $\delta_* \ll 1$ and $\theta_{\text{tol}} \gg 1$.
Set $\delta_0 = \omega_0/(\beta_{\max}\theta(\lambda_0, \nu_0, \mu_0))$.

Step 1. **[Inner iteration]**. Find $x_k \in \Omega$ that approximately solves (4.17), that means such that (4.22) holds.

Step 2. **[Test of convergence]**. If $\delta_k \leq \delta_*$ and $\|c_{\mathcal{E}}(x_k)\| \leq \eta_*$ and $c_{\mathcal{I}}(x_k) \geq 0$, stop.

Step 3. **[Disaggregated updates]**. For $j = 1, \ldots, q_{\mathcal{E}}$ and for $l = q_{\mathcal{E}} + 1 \ldots, q_{\mathcal{I}}$ execute Step 3a if $\|c_{\mathcal{E}}(x_k)_{[\mathcal{Q}_j]}\| \leq \eta_*$ and $c_{\mathcal{I}}(x_k)_{[\mathcal{Q}_l]} \geq 0$
or Step 3b otherwise.
- Step 3a. **[Update the Lagrange multiplier estimates]**. Set (from (4.13) and (4.15))

$$\begin{aligned}
\lambda_{k+1,[\mathcal{Q}_j]} &= \lambda_{k,[\mathcal{Q}_j]} + \tfrac{1}{\mu_{k,j}} c_{\mathcal{E}}(x_k(\lambda_k, \nu_k; \mu_k))_{[\mathcal{Q}_j]}, \\
\nu_{k+1,[\mathcal{Q}_l]} &= \nu_{k,[\mathcal{Q}_l]} + \tfrac{1}{\mu_{k,l}} c_{\mathcal{I}}^{-}\left[x_k(\lambda_k, \nu_k; \mu_k), \nu_k; \mu_k\right]_{[\mathcal{Q}_j]}, \\
\mu_{k+1,j} &= \mu_{k,j}, \\
\mu_{k+1,l} &= \mu_{k,l}.
\end{aligned}$$

- Step 3b. **[Reduce the penalty parameter]**. Set

$$\begin{aligned}
\lambda_{k+1,[\mathcal{Q}_j]} &= \lambda_{k,[\mathcal{Q}_j]}, \\
\nu_{k+1,[\mathcal{Q}_l]} &= \nu_{k,[\mathcal{Q}_l]}, \\
\mu_{k+1,j} &= \tau_{k,j}\mu_{k,j}, \\
\mu_{k+1,l} &= \tau_{k,l}\mu_{k,l},
\end{aligned}$$

where $\tau_{k,s} = \begin{cases} \tau & \text{if } \mu_{k,s} = \alpha_k \\ \min(\tau, \alpha_k) & \text{otherwise} \end{cases}$ for $s = j, l$.

Step 4. **[Aggregated updates]**. Define

$$\alpha_{k+1} = \max_{j=1,\ldots,q_{\mathcal{E}}+q\mathcal{I}} \mu_{k+1,j}.$$

If $\alpha_{k+1} < \alpha_k$, then set

$$\begin{aligned}
\omega_{k+1} &= \alpha_{k+1}, \\
\eta_{k+1} &= \alpha_{k+1}^{\alpha_\eta}, \\
\delta_{k+1} &= \omega_{k+1}/(\beta_{\max}\theta(\lambda_{k+1}, \nu_{k+1}, \mu_{k+1})).
\end{aligned}$$

Otherwise set

$$\begin{aligned}
\omega_{k+1} &= \omega_k\alpha_{k+1}, \\
\eta_{k+1} &= \eta_k\alpha_{k+1}^{\beta_\eta}, \\
\delta_{k+1} &= \omega_{k+1}/(\beta_{\max}\theta(\lambda_{k+1}, \nu_{k+1}, \mu_{k+1})).
\end{aligned}$$

Increment k by one and go to Step 1.

To apply the convergence analysis from [5], we need to tie the sequence of stopping criterion

tolerance δ_k to the original sequence of stopping tolerance of [5], ω_k and to ensure a stable relationship between the stopping criterion (4.22) in the GSS solution of the subproblem and the stationary condition defined in [5]. To do so, let $\theta_{\text{tol}} \gg 1$ be given and we define

$$\theta(\lambda, \nu, \mu) = \max\left\{1, \frac{1}{\theta_{\text{tol}}}\left(1 + \|\lambda\| + \|\nu\| + \sum_{j=1}^{q_{\mathcal{E}}+q_{\mathcal{I}}} \frac{1}{\mu_j}\right)\right\}. \tag{4.24}$$

Observe that the difference between this and the θ define in [4] is just the terms $\|\nu\|$ (which is added) and the sum which now runs from 1 to $q_{\mathcal{E}} + q_{\mathcal{I}}$. We note also that any function $\theta(\lambda, \nu, \mu)$ such that

$$\left(\|\lambda\| + \|\nu\| + \sum_{j=1}^{q_{\mathcal{E}}+q_{\mathcal{I}}} \frac{1}{\mu_j}\right) = O(\theta(\lambda, \nu, \mu)) \text{ as } \left(\|\lambda\| + \|\nu\| + \sum_{j=1}^{q_{\mathcal{E}}+q_{\mathcal{I}}} \frac{1}{\mu_j}\right) \to \infty$$

suffices for the purposes of establishing global convergence properties.

Finally, we need to take into account the fact that in algorithm 3.3 we took

$$\varepsilon = \min\{\varepsilon_{\max}, \beta_{\max}\Delta\}. \tag{4.25}$$

This leads to $\Delta \geq \frac{\varepsilon}{\beta_{\max}}$ and suggests the following update for δ_k :

$$\delta_{k+1} = \frac{\omega_{k+1}}{\beta_{\max}\theta(\lambda_{k+1}, \nu_{k+1}, \mu_{k+1})}. \tag{4.26}$$

It is relevant, as we will see in the proof of convergence, to note that the initialization of δ_0 and the definition of $\theta(\lambda, \nu, \mu)$, together with the update rule (4.26), ensure that for all $k \geq 0$,

$$\delta_k \leq \min\left\{1, \theta_{\text{tol}}\left(1 + \|\lambda\| + \|\nu\| + \sum_{j=1}^{q_{\mathcal{E}}+q_{\mathcal{I}}} \frac{1}{\mu_j}\right)^{-1}\right\} \frac{\omega_k}{\beta_{\max}} \leq \frac{\omega_k}{\beta_{\max}}, \tag{4.27}$$

where we use the simple result of analysis $\frac{1}{\max(a,b)} = \min(\frac{1}{a}, \frac{1}{b})$ for all a, b positive.

Relation Between $\theta(\lambda, \nu, \mu)$ and Ill Conditioning

The quantity $\theta(\lambda, \nu, \mu)$ defined in (4.24) provides a mechanism for dealing with ill conditioning. There is ill conditioning when the Lipschitz constant of the augmented Lagrangian takes very large values. This occurred when the penalty parameters $\mu_{k,j}$ are very small or when the Lagrange multipliers λ_k and ν_k are very large in magnitude. As we can see in Proposition 4.3, the Lipschitz constant M_k of the augmented Lagrangian depends on an essential way on λ_k, ν_k and $\mu_{k,j}$. To counter the effect of ill conditioning, we must tighten the stopping criterion δ_k accordingly. When we look at the update rule for δ_k given in (4.26), we see that δ_k quickly becomes very small if the problem is ill conditioned (this can force the algorithm to converge to a point which is not a stationary point of the original problem). It is for this reason that we defined $\theta(\lambda, \nu, \mu)$ in (4.24) in such a way that it becomes active only when $\left(1 + \|\lambda_k\| + \|\nu_k\| + \sum_{j=1}^{q_{\mathcal{E}}+q_{\mathcal{I}}} \frac{1}{\mu_{k,j}}\right)$ exceeds a prescribed threshold θ_{tol}. In this way, if θ_{tol} is sufficiently large, and the penalty parameter remains uniformly bounded away from zero and the Lagrange multipliers converge to their real values, then we have $\delta_k = \frac{\omega_k}{\beta_{\max}}$ for all k, and avoid a rapid decrease of δ_k.

4.5 Proof of Convergence of the Inner Iteration

We now need to adapt the results on the convergence of linearly constrained GSS to our augmented Lagrangian sub-problem and make sure that the stopping criterion are related in a stable way to the original stopping criterion of [5]. Before we proceed, we need to recall some important aspects of the original algorithm of Conn, Gould, Sartenaer and Toint [5] which leads to the proof of convergence of the algorithm given in [4].

They defined the set of linear constraints that are nearly binding[1] at a point $x \in \Omega = \{x \,|\, Ax \geq b\}$ with respect to ω (we will called it the ω-binding constraints) by

$$D(x,\omega) = \left\{ i \in \{1,\ldots,m\} \,|\, a_i^T x - b_i \leq \kappa_0 \omega \right\}.$$

From this, we define the ω-normal cone $N(x,\omega)$ as the cone generated by the rows the opposite $-a_i^T$ of the rows a_i^T of the matrix of constraints A for which $i \in D(x,\omega)$ and the ω-tangent $T(x,\omega)$ cone as its dual, this means that $T(x,\omega) = N(x,\omega)^\circ$. Note that there is a relationship between the ω-binding constraints and what we call in the context of GSS the ε-binding constraints $I(x,\varepsilon)$. The ε-binding constraints were defined in terms of the distance to the binding constraints at a near boundary point. Recall that $I(x,\varepsilon) = \{i \,|\, d(x,A_i) = \frac{a_i^T x - b_i}{\|a_i\|} \leq \varepsilon\}$, and we defined, in a similar way, the notions of ε-normal and ε-tangent cones $N(x,\varepsilon)$ and $T(x,\varepsilon)$. This motivated the choice of $\kappa_0 = \max_{1,\ldots,m} \|a_i\|$ at Step 0 of Algorithm 4.4.1. Note that with this choice of κ_0 if $i \in I(x,\varepsilon)$, and $\varepsilon \leq \omega$, then we have $a_i^T x - b_i \leq \|a_i\|\varepsilon \leq \kappa_0 \omega$. Thus $D(x,\omega) \supset I(x,\varepsilon)$, so $N(x,\omega) \supset N(x,\varepsilon)$ and taking the polar, we have $T(x,\omega) \subset T(x,\varepsilon)$ and then

$$P_{T(x,\omega)} \leq P_{T(x,\varepsilon)}. \tag{4.28}$$

The stopping criterion for the subproblem (4.17) were defined in term of the projection of the direction of the steepest descent of the augmented Lagrangian onto the ω_k-tangent cone at x_k:

$$\|P_{T(x_k,\omega_k)}(-\nabla_x L(x_k,\lambda_k,\nu_k;\mu_k))\| \leq \omega_k, \tag{4.29}$$

where the scalar $\omega_k > 0$ is a suitable tolerance that is updated at each outer iteration in a way to ensure $\omega_k \to 0$ as $k \to \infty$.

Now we have all the ingredients to prove convergence results of the inner iteration of algorithm 4.4.1.

Proposition 4.4. *Suppose that condition 6, 7 and 8 hold, and let M be the Lipschitz constant for $\nabla f(x)$, $J_{\mathcal{E}}(x)$, $J_I(x)$, $J_{\mathcal{E}}(x)^T c_{\mathcal{E}}(c)$ and $J_{\mathcal{I}}(x)^T c_{\mathcal{I}}(x)$ on Ω, where $J_{\mathcal{E}}(x)$ is the Jacobian matrix of $c_{\mathcal{E}}(x)$ and so on. Then the following bound holds at the outer iteration k of Algorithm 4.4.1:*

$$\|P_{T(x_k,\omega_k)}(-\nabla_x L(x_k,\lambda_k,\nu_k;\mu_k))\| \leq \frac{1}{\hat{\kappa_{\min}}} \left(M_k \beta_{\max} + \frac{\hat{\alpha}}{\beta_{\min}} \right) \delta_k, \tag{4.30}$$

where

$$M_k = M \left(1 + \|\lambda\| + \|\nu\| + \sum_{j=1}^{q_{\mathcal{E}}+q_{\mathcal{I}}} \frac{1}{\mu_j} \right).$$

[1] It is also called working set.

Proof. We have seen from Proposition 4.3 that M_k is the Lipschitz constant of $\nabla_x L(x_k, \lambda_k, \nu_k; \mu_k)$. Since the stopping criterion (4.22) is invoked at unsuccessful iterations of the GSS solution of (4.17), we apply Theorem 3.1 with ρ given by (4.21) to obtain

$$\|P_{T(x_k,\varepsilon_k)}(-\nabla_x L(x_k, \lambda_k, \nu_k; \mu_k))\| \leq \frac{1}{\hat{\kappa}_{\min}} \left(M_k \beta_{\max} + \frac{\hat{\alpha}}{\beta_{\min}} \right) \delta_k, \tag{4.31}$$

where we have used (4.22). From (4.25), (4.22) and (4.27), we know that $\varepsilon_k \leq \beta_{\max} \Delta_{k,u} \leq \beta_{\max} \delta_k \leq \omega_k$, so that (4.28) holds. Therefore,

$$\|P_{T(x_k,\omega_k)}(-\nabla_x L(x_k, \lambda_k, \nu_k; \mu_k))\| \leq P_{T(x_k,\varepsilon_k)}(-\nabla_x L(x_k, \lambda_k, \nu_k; \mu_k))\|.$$

Combining this with (4.31) yields the result. ■

The next theorem is central to this approach. It says that the asymptotic behavior of $\|P_{T(x_k,\omega_k)}(-\nabla_x L(x_k, \lambda_k, \nu_k; \mu_k))\|$ in Algorithm 4.4.1 is like the behavior of the original algorithm in [5].

Proposition 4.5. *Suppose that condition 6, 7 and 8 hold. Then there exists a constant $C > 0$, independent of k, such that the following holds at the outer iteration k of Algorithm 4.4.1:*

$$\|P_{T(x_k,\omega_k)}(-\nabla_x L(x_k, \lambda_k, \nu_k; \mu_k))\| \leq C\omega_k. \tag{4.32}$$

Proof. From Proposition 4.4, we have that

$$\|P_{T(x_k,\omega_k)}(-\nabla_x L(x_k, \lambda_k, \nu_k; \mu_k))\| \leq \frac{1}{\hat{\kappa}_{\min}} \left(M_k \delta_k \beta_{\max} + \frac{\hat{\alpha}}{\beta_{\min}} \delta_k \right).$$

The upper bounds on δ_k from (4.27) say that $\delta_k \leq \frac{\omega_k}{\beta_{\max}}$ and $M_k \delta_k \beta_{\max} \leq M_k \theta_{\text{tol}} \omega_k$. Therefore,

$$\|P_{T(x_k,\omega_k)}(-\nabla_x L(x_k, \lambda_k, \nu_k; \mu_k))\| \leq \frac{1}{\hat{\kappa}_{\min}} \left(M_k \theta_{\text{tol}} + \frac{\hat{\alpha}}{\beta_{\max}\beta_{\min}} \right) \omega_k;$$

the result follows with $C = \frac{1}{\hat{\kappa}_{\min}} \left(M_k \theta_{\text{tol}} + \frac{\hat{\alpha}}{\beta_{\max}\beta_{\min}} \right)$. ■

4.6 Global Convergence Analysis

The importance of Proposition 4.5 is that the convergence analysis for the original algorithm of [5] can be applied and the original proofs still hold with minor changes on some constants that appear and some hypotheses involving the inequalities constraints. We now discuss some convergence properties of the augmented Lagrangian algorithm from [5] that hold for the GSS adaptation. Before doing so, we set a notation and one additional condition.

Suppose $\{x_k\}_{k \in \mathcal{K}}$ is a sequence that converges to x_*. We denote by A_* the matrix whose rows are the linear constraints that are binding at x_*, and denote by Z_* a matrix whose columns form an orthonormal basis for the nullspace of A_*. If $J_{\mathcal{E}}(x)$ is the Jacobian matrix of $c_{\mathcal{E}}(x)$ that is

$$J_{\mathcal{E}}(x)^T = [\nabla c_1(x), \ldots \nabla c_{n_{\mathcal{E}}}(x)],$$

and $J_{\mathcal{I}}(x)$ is the Jacobian matrix of $c_{\mathcal{I}}(x)$ that is

$$J_{\mathcal{I}}(x)^T = [\nabla c_1(x), \ldots \nabla c_{n_{\mathcal{I}}}(x)] .$$

We define

$$\bar{J}_{\mathcal{I}}(x)^T = \left[J_{\mathcal{I}}(x)^T, I_{n_{\mathcal{I}}\times n_{\mathcal{I}}}\right] \text{ and } \bar{Z}_* = \begin{bmatrix} Z_* \\ 0_{n_{\mathcal{I}}\times p} \end{bmatrix},$$

where $I_{n_{\mathcal{I}}\times n_{\mathcal{I}}}$ is the identity matrix of order $n_{\mathcal{I}}$, $0_{n_{\mathcal{I}}\times p}$ is the zero matrix of order $n_{\mathcal{I}} \times p$ and p is the number of columns of Z_*.
The least-square Lagrange multipliers estimates at x_* corresponding to A_* are defined to be

$$\lambda(x_*) \equiv -((J_{\mathcal{E}}(x_*)Z_*)^+)^T Z_*^T \nabla f(x_*), \tag{4.33}$$

and

$$\nu(x_*) \equiv -((\bar{J}_{\mathcal{I}}(x_*)\bar{Z}_*)^+)^T \bar{Z}_*^T \nabla f(x_*), \tag{4.34}$$

where for a matrix C with full rank, its right inverse C^+ is defined by

$$C^+ = C^T(CC^T)^{-1}.$$

We stress that, as stated, the Lagrange multipliers estimates (4.33) can not be calculated directly as they required a priori knowledge of x_*. They are merely introduced as analytical devices.

We need the following condition to ensure that the least-square Lagrange multiplier estimates for the equalities constraints are well defined at each iteration of our algorithm.

Condition 9. The matrix $J_{\mathcal{E}}(x_*)Z_*$ has column rank no smaller than $n_{\mathcal{E}}$ at any limit point x_* of the sequence $\{x_k\}$.

The fundamental convergence result for Algorithm 4.4.1 is an adaptation of Theorem 4.6 in [5] and discussed in [4] as Theorem 5.1. Note that this is a global convergence result.

Theorem 4.1. *Let x_* be any limit point of the sequence $\{x_k\}$ generated by Algorithm 4.4.1 for which Condition 8 and 9 hold, and let $\mathcal{K}$ be the set of indices of an infinite subsequence of $\{x_k\}$ whose limit is x_*. Finally, let $\lambda_* = \lambda(x_*)$ and $\nu_* = \nu(x_*)$. Then*

(i) *there are positive constants κ_2, κ_3, κ_4, and κ_4 such that*

$$\begin{aligned}
&\|\bar{\lambda}(x_k, \lambda_k, \mu_k) - \lambda_*\| \leq \kappa_2\omega_k + \kappa_3\|x_k - x_*\|,\\
&\|\bar{\nu}(x_k, \lambda_k, \mu_k) - \nu_*\| \leq \kappa_4\omega_k + \kappa_5\|x_k - x_*\|,\\
&\|\lambda(x_k) - \lambda_*\| \leq \kappa_3\|x_k - x_*\|,\\
&\|\nu(x_k) - \nu_*\| \leq \kappa_5\|x_k - x_*\|.
\end{aligned}$$

where $\bar{\lambda}(x_k, \lambda_k, \mu_k) = \lambda_{k+1}$ and $\bar{\nu}(x_k, \lambda_k, \mu_k) = \nu_{k+1}$ given in (4.13) and (4.15) and

$$\|c_{\mathcal{E}}(x_k)_{\mathcal{Q}_j}\| \leq \kappa_2\omega_k\mu_{k,j} + \mu_{k,j}\|(\lambda_k - \lambda_*)_{\mathcal{Q}_j}\| + \kappa_3\mu_{k,j}\|x_k - x_*\|$$

for all $j = 1, \ldots, q_{\mathcal{E}}$ and all $k \in \mathcal{K}$ sufficiently large.

(ii) *x_* is Karush-Kuhn-Tucker point (first-order stationary point) for the problem (4.1)-(4.4), λ_* and ν_* the corresponding vectors of Lagrange multipliers, the sequence $\{\bar{\lambda}(x_k, \lambda_k, \mu_k)\}$ converges to λ_* for $k \in \mathcal{K}$ and the sequence $\{\bar{\nu}(x_k, \lambda_k, \mu_k)\}$ converges to ν_* for $k \in \mathcal{K}$.*

5. Conclusion

In this essay, we investigated a class of direct search methods, the Generating Set Search method, for linearly and nonlinearly constrained global optimization problems. For the linearly constrained case, the constraints were kept explicit and treated geometrically. The notions of ε-tangent and ε-normal cones helped to define at each iteration the set of search directions conforming to the geometry of the feasible region and lead to convergence results involving a continuous measure of stationarity. For the general nonlinear case with explicit linear constraints, we used an augmented Lagrangian approach. This helped us to reduce the problem to a sequence of linearly-constrained minimizations of an augmented Lagrangian possessing the same dimension as the original problem. We were then able to propose a general GSS algorithm and to extend the existing convergence results in the case of nonlinear equality constraints to that of nonlinear inequality and equality constraints.

Future research directions might emcompass investigating contrained global optimization. Precisely, it may be interresting to look at multistart algorithms like population set based direct search methods and see if we can analyse their global convergence properties theoritically and numerically. The importance of these methods in Large scale optimization and in engineering is a motivation for this analysis.

Appendix A. Finding the Lipschitz constant of the gradient of the augmented Lagrangian

We start by recalling some results on the projection onto a convex set. Let $X \subset \mathbb{R}^n$ be convex, closed, and non-empty. For $z \in \mathbb{R}^n$, we defined $P_X(z)$ called the projection of z onto X to be the optimum for

$$\min_{x \in X} \|x - z\|^2. \tag{A.1}$$

That is the minimizer is the closest point in X to z. The projection function has several properties, which we present in the following theorem. The last properties is more important for the purpose of this essay.

Theorem A.1. *Let $X \subset \mathbb{R}^n$ be convex, closed, and non-empty. Then*

1. *For all $z \in \mathbb{R}^n$, there is a unique optimum for (A.1), which we call $\bar{x} = P_X(z)$.*
2. *$\bar{x}$ is the unique point in X such for all $x \in X,\ (z - \bar{x})^T(x - \bar{x}) \leq 0$.*
3. *If X is a subspace, then $z - \bar{x}$ is orthogonal to X (i.e., $(z - \bar{x})^T y = 0\ \forall y \in X$).*
4. *The projection function is non-expensive:*

$$\forall x,\, y \in \mathbb{R}^n,\ \|P_X(x) - P_X(y)\| \leq \|x - y\|.$$

In other words, P_X is Lipschitz with constant 1.

We recall that in Section 4.3, we found the gradient of the augmented Lagrangian as

$$\begin{aligned} \nabla_x L(x, \lambda, \nu, \mu) &= \nabla_x f(x) + \sum_{j=1}^{q_{\mathcal{E}}} \sum_{i \in \mathcal{Q}_j} \left[\lambda_i \nabla_x c_i(x) + \tfrac{1}{\mu_j} c_i(x) \nabla_x c_i(x) \right] \\ &\quad + \sum_{l=q_{\mathcal{E}}+1}^{q_{\mathcal{I}}} \sum_{i \in \mathcal{Q}_l} \widetilde{P}_+ \left(\nu_i \nabla_x c_i(x) + \tfrac{1}{\mu_l} c_i(x) \nabla_x c_i(x) \right), \end{aligned} \tag{A.2}$$

where

$$\left[\widetilde{P}_+ \left(\nu_i \nabla_x c_i(x) + \frac{1}{\mu_l} c_i(x) \nabla_x c_i(x) \right) \right]_\ell = \begin{cases} \nu_i \frac{\partial c_i(x)}{\partial x_\ell} + \frac{1}{\mu_l} c_i(x) \frac{\partial c_i(x)}{\partial x_\ell} \text{ if } \nu_i + \frac{1}{\mu_l} c_i(x) \leq 0 \text{ and } \frac{\partial c_i(x)}{\partial x_\ell} \leq 0 \\ 0 \text{ otherwise.} \end{cases}$$

Let E_+ denote the non-negative half space of $\mathbb{R}^n$:

$$E_+ = \{y \,|\, y \geq 0\}.$$

E_+ is a closed, convex and non empty set. So by Theorem A.1 the projection onto E_+ that we denote for simplicity P_+ exist and it is non-expansive, that is it satisfies item 4 of Theorem A.1. Note that

$$P_+(x) = \max(x, 0),\ \forall x \in \mathbb{R}^n$$

where the i^{th} component of $\max(x, 0)$ is

$$[\max(x, 0)]_i = \begin{cases} x_i \text{ if } x_i \geq 0 \\ 0 \text{ otherwise.} \end{cases}$$

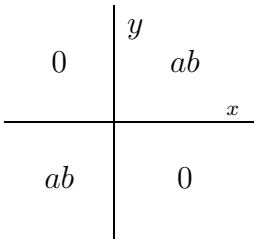
Now we prove that $\widetilde{P}_+$ is the restriction onto a convex set that we will define later. To shorten the presentation, we consider the following general setting. We assume that $\widetilde{P}_+$ is defined as

$$\widetilde{P}_+(ab) = \begin{cases} ab \text{ if } a \leq 0 \text{ and } b \leq 0 \\ 0 \text{ otherwise.} \end{cases}$$

Let us restrict ourself for a moment in one dimension. The projection of the product ab is given by

$$P_+(ab) = \begin{cases} ab \text{ if } (a \leq 0 \text{ and } b \leq 0) \text{ or } (a \geq 0 \text{ and } b \geq 0) \\ 0 \text{ otherwise.} \end{cases}$$

In the xy plane, the values of $P_+(ab)$ are then as shown in the graph below



So the projection of ab view as an application defined from $\mathbb{R}^2$ takes the values ab for $(a, b) \in (\mathbb{R}_- \times \mathbb{R}_-) \bigcup (\mathbb{R}_+ \times \mathbb{R}_+)$. This set is not convex but we can see that $\widetilde{P}_+(ab)$ is the restriction of $P_+(ab)$ on $\mathbb{R}_- \times \mathbb{R}_-$ which is convex. So we have proven that $\widetilde{P}_+(ab)$ is the restriction of $P_+(ab)$ on a convex set. This is the key element of our analysis. Now see that with this analysis for a, b, a', $a' \in \mathbb{R}$,

$$|\widetilde{P}_+(ab) - \widetilde{P}_+(a'b')| = |P_+(ab) - P_+(a'b')| \leq |ab - a'b'| \tag{A.3}$$

The extension of this result (A.3) in higher dimension derives directly from the fact that the norm in R^n can be considered as a sum of norm in $\mathbb{R}$. We used all these materials now to prove Proposition 4.3.

A.1 Proof of Proposition 4.3

Let $x,\ y \in \Omega$. we have

$$
\begin{aligned}
&\|\nabla_x L(x,\lambda,\nu,\mu) - \nabla_x L(y,\lambda,\nu,\mu)\| \\
&\leq \|\nabla_x f(x) - \nabla_x f(y)\| \\
&+ \sum_{j=1}^{q_{\mathcal{E}}} \sum_{i\in\mathcal{Q}_j} \left[|\lambda_i| \|\nabla_x c_i(x) - \nabla_x c_i(y)\| + \tfrac{1}{\mu_j} \|c_i(x)\nabla_x c_i(x) - c_i(y)\nabla_x c_i(y)\| \right] \\
&+ \sum_{l=q_{\mathcal{E}}+1}^{q_I} \sum_{i\in\mathcal{Q}_l} \|\widetilde{P}_+\left(\nu_i\nabla_x c_i(x) + \tfrac{1}{\mu_l} c_i(x)\nabla_x c_i(x)\right) - \widetilde{P}_+\left(\nu_i\nabla_x c_i(y) + \tfrac{1}{\mu_l} c_i(y)\nabla_x c_i(y)\right)\| \\
&\leq M\|x-y\| \\
&+ M(\|\lambda\| + \sum_{j=1}^{q_{\mathcal{E}}} \tfrac{1}{\mu_j})\|x-y\| \\
&+ M(\|\nu\| + \sum_{l=q_{\mathcal{E}}+1}^{q_I} \tfrac{1}{\mu_l})\|x-y\| \text{ we have used here the result (A.3) componentwise.} \\
&= M(1 + \|\lambda\| + \|\nu\| + \sum_{j=1}^{q_{\mathcal{E}}+q_{\mathcal{I}}} \tfrac{1}{\mu_j})\|x-y\|.
\end{aligned}
$$

And the complete the proof.

Acknowledgments

I am very thankful towards my supervisor Prof Montaz Ali who drew my attention to this essay topic and guide my work. I am indebted to my friend and college Fabrice Talla with who I discussed some orientations of this work. Laure Gouba, Ambrose Chongo and henry Amuasi proofread the manuscript and made comments that help to improve the presentation, I greatly appreciate their efforts.

My sincere gratitude goes to the AIMS funder and director for the wonderful idea of AIMS. It was a rewarding experience for me to study at AIMS. AIMS staff, lecturers, tutors and students of the 2006-2007 set contributed a lot to this unique experience. I gratefully acknowledge them for that.

I am extremely grateful to my former lecturers at the University of Yaounde I in Cameroon for the mathematical background they gave me. To my family, my friends and my fiance, I say thank you for their unconditional encouragement and love. Thank you to Muizenberg community church for the spiritual support.

Bibliography

[1] Dimitri P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic Press, 1982.

[2] Garth P. McCormick. *Nonlinear Programming, Theory, Algorithms, and Applications*. Wiley & Sons, 1983.

[3] Nicholas Gould, Dominique Orban, and Philippe L. Toint. An ℓ_1 interior-point penalty method for nonlinear optimization. 2003.

[4] Tamara G. Kolda, Robert Michael Lewis, and Virginia Torczon. A generating set direct search augmented lagrangian algorithm for optimization with a combination of general and linear constraints. Technical report, Sandia National laboratories Albuquerque, New Mexico 87185 and Livermore, California 94550, 2006.

[5] A. Conn, N. Gould, A. Sartenaer, and P. L. Toint. Convergence properties of an augmented lagrangian algorithm with a combination of general equality and linear constraints. *SIAM J. Optim.*, 6:674–703, 1996.

[6] Tamara G. Kolda, Robert Michael Lewis, and Virginia Torczon. Optimization by direct search: new perspectives on some classical and modern methods. *SIAM Review*, 45:385–482, 2003.

[7] Robert Michael Lewis, Anne Shepherd, and Virginia Torczon. Implementing generating set search methods for linearly-constrained minimization. Technical report, Departement of Computer Science, College of William and Mary, Williamsburg, Virginia 23187, 2007.

[8] Tamara G. Kolda and Virginia Torczon. Understanding asynchronuous parallel pattern search. *High Performance Algorithms and sofware for Nonlinear optimization*, pages 316–335, 2003. available at http://www.cs.wm.edu/∼va/research/.

[9] Patricia D. Hough, Tamara G. Kolda, and Virginia J. Torczon. Asynchronous parallel pattern search for nonlinear optimization. *SIAM Journal on Scientific Computing*, 23(1):134–156, June 2001.

[10] Genetic algorithm and direct search toolbox 2.1, 2007. http://www.mathworks.com/products/gads/.

[11] Tamara G. Kolda, Robert Michael Lewis, and Virginia Torczon. Stationary results for generating set search for linearly constrained optimization. *SIAM Review*, 17:943–968, 2006.

[12] Stefano Lucidi and Marco Sciandrone. On the global convergence of derivative-free methods for unconstrained optimization. *SIAM J. Optim.*, 13:97–116, 2002.

[13] G. Liuzzi, S. Lucidi, and M. Sciandrone. A derivative free algorithm for linearly constrained finite minimax problems. *SIAM J. Optim.*, 16:1054–1075, 2006.

[14] Robert Michael Lewis and Virginia Torczon. Pattern search methods for linearly constrained minimization. *SIAM J. Optim.*, 10:917–941, 2000.

[15] Virginia Torczon. On the convergence of pattern search algorithms. *SIAM J. Optim*, 7:1–25, 1997.

[16] S. M. H. Janesch. Exact penalty fuctions: a lower bound to the penalty parameter. *International Mathematical Forum*, 2:75–86, 2007.

[17] Robert Michael Lewis and Virginia Torczon. A globally convergent augmented lagragian pattern search algorithm for optimization with general constraints and simple bounds. *SIAM J. Optim*, 12(4):1075–1089, 2002.

[18] Robert Michael Lewis and Virginia Torczon. Pattern search methods for bound constrained minimization. *SIAM J. Optim.*, 9:1082–1099, 1999.

[19] Jeffrey G. Lagarias, James A. Reeds, Margaret H. Wright, and Paul E. Wright. Convergence properties for the nedler-mead simplex method in low dimensions. *SIAM J. Optim.*, 9(1):112–147, 1998.

[20] C. Bogani, M.G. Gasparo, and A. papini. Generalized pattern search methods for a class of nonsmooth problems with structure, 2006.

[21] Virginia Torczon. *Multi-Directional search: A direct search algorithm for parallel machines*. Phd, Departement of Mathematical Sciences, Rice University, Houston, 1989.

[22] Liz Dolan, Adam Gurson, Anne Shepherd, Chris Siefert, and Virginia Torczon. http://www.cs.wm.edu/∼va/software/directsearch/direct_code/.

[23] G.R. Walsh. *Methods of Optimization*, chapter Search methods for unconstrained optimization. Wiley & Sons, 1975. Chapter 3.