Universidade Federal de Minas Gerais
Programa de Pós-Graduação em Engenharia Elétrica

Tese de Doutorado

AN INFEASIBILITY CERTIFICATE FOR NON-LINEAR PROGRAMMING BASED ON PARETO-CRITICALITY CONDITIONS

Author: Shakoor Muhammad

Advisor: Prof. Ricardo H. C. Takahashi
Co-Advisor: Prof. Frederico G. Guimarães

July 16th, 2015

"An Infeasibility Certificate for Non-Linear Programming Based on Pareto-Criticality Conditions"

Shakoor Muhammad

Tese de Doutorado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do grau de Doutor em Engenharia Elétrica.

Aprovada em 16 de julho de 2015.

Por:

Prof. Dr. Ricardo Hiroshi Caldeira Takahashi
DMAT (UFMG)

Prof. Dr. Frederico Gadelha Guimarães
DEE (UFMG) - Coorientador

Prof. Dr. João Antônio de Vasconcelos
DEE (UFMG)

Prof. Dr. Lucas de Souza Batista
DEE (UFMG)

Prof. Dr. Alexandre Cláudio Botazzo Delbem
SSC (USP)

Profa. Dra. Elizabeth Fialho Wanner
Departamento de Computação (CEFET-MG)

Acknowledgements

First and foremost, I thank to my supervisor, professor Ricardo Hiroshi Caldeira Takahashi without whose guidance, this work would not have been possible. I would always be deeply indebted for his encouragement, caring, patience, support and a source of inspiration throughout my research work in the Electrical Engineering Department at Universidade Federal de Minas Gerais.

I am also very grateful to my co-advisor, professor Frederico Gadelha Guimaraes for their invaluable advice, numerous stimulating discussions and suggestions.

I also thank to my professors in the Graduate Program in Electrical Engineering of the Universidade Federal de Minas Gerais, in particular, the professors Dr. Oriane Magela Neto, Dr. Felipe Campelo, Dr. Lucas de Souza Batista, Dr. Eduardo Gontijo Carrano, Dr. Rodney Rezende Saldanha and Dr. Reinaldo Martinez Palhares, who contributed to this work to some extent and to my education and stay in Brazil to a great extent. To the secretaries of the Graduate Program in Electrical Engineering, Anete Vieira, Arlete Vieira and Jeronimo Coelho, who were always kind and helpful with me and others.

On a more personal note, I thank all my colleagues and fellow graduate students at Universidade Federal de Minas Gerais for their friendship, advice, and help; many have graduated and gone their separate ways. I would especially like to thank to all, including Vitor Nazario, Regina Carla Lima Correa de Sousa, Denny Collina, Dimas Abreu Dutra, Victor Costa , Gilson Fernandes and Sajad Azizi.

The financial support of National Council for Scientific and Technological Development (CNPq) Brazil is gratefully acknowledged.

Finally, I would like to express my sincere gratitude to my mother, Sabira Muhammad, and my brothers and sisters for their loving support and encouragement. I thank my wife for her love, support and patience.

I dedicate this work to my mother and to the loving memory of my late father Mr. Ghani Muhammad.

Abstract

This thesis proposes a new necessary condition for the infeasibility of non-linear optimization problems (that becomes necessary under convexity assumption) which is stated as a Pareto-criticality condition of an auxiliary multiobjective optimization problem. This condition can be evaluated, in a given problem, using multiobjective optimization algorithms, in a search that either leads to a feasible point or to a point in which the infeasibility conditions holds. The resulting infeasibility certificate, which is built with primal variables only, has global validity in convex problems and has at least a local meaning in generic nonlinear optimization problems. In the case of noisy problems, in which gradient information is not available, the proposed condition can still be employed in a heuristic flavor, as a by-product of the expected features of the Pareto-front of the auxiliary multiobjective problem.

Key-words: nonlinear programming, multiobjective programming, infeasibility certificate, noisy problems.

Resumo

Esta tese propõe uma nova condição necessária para a infactibilidade de problemas de otimização não lineares (que se torna necessária sob suposição de convexidade) que é estabelecida como uma condição crítica de Pareto de um problema de otimização multi-objetivo auxiliar. Esta condição pode ser avaliada, em um dado problema, utilizando algoritmos de otimização multi-objetivo, em uma busca que leva ou para um ponto viável ou para um ponto em que as condições de inviabilidade são asseguradas. O certificado de inviabilidade resultante, que é construído somente com variáveis primais, possui validade global em problemas convexos e possui no mínimo um significado local em problemas genéricos de otimização não linear. No caso de problemas ruidosos, em que a informação de gradiente não é disponível, a condição proposta ainda pode ser aplicada sob uma noção heurística, como um produto das características da fronteira-Pareto do problema auxiliar multi-objetivo.

Palavras-chave: programação não linear, programação multi-objetivo, certificação de inviabilidade, problemas ruidosos.

List of Symbols

The following notations are employed here.

1. (. $\leq$.) Each coordinate of the first argument is less than or equal to the corresponding coordinate of the second argument.
2. (. $<$.) Each coordinate of the first argument is smaller than the corresponding coordinate of the second argument.
3. (. $\prec$.) Each coordinate of the first argument is less than or equal to the corresponding coordinate of the second argument, and at least one coordinate of the first argument is strictly smaller than the corresponding coordinate of the second argument.
4. the operators (. $\geq$.), (. $>$.) and (. $\succ$.) are defined in the analogous way.
5. $\mathbb{R}$ is the set of real numbers.
6. A^T is the transpose of matrix A.
7. $\mathcal{N}(\cdot)$, stands for the null space of a matrix.
8. $\mathcal{K}^+$, denotes the cone of the positive octant of suitable dimension.

Contents

Chapter 1

Introduction

What will happen when an optimization algorithm is unable to find a feasible solution? How could we know what went wrong? The question about feasibility and infeasibility of an optimization problem in order to know it's status is the main theme of this thesis.

There has been a lot of work related to feasibility and infeasibility in optimization in the last two decades. This effort is still going on even today. Why are we interested in feasibility and infeasibility of an optimization problem? Certainly it is most important according to the situation to find the best (optimum or efficient) solution, in place of any feasible solution. The detection of an optimization problem to be feasible or infeasible are indeed the two sides of the same coin. The existence of a feasible solution of a constrained optimization problem precedes the question of determining the best solution.

In recent years, the question of establishing if a problem is feasible or infeasible has grown in importance as the optimization models have grown larger and more complex in step with the phenomenal increase in expensive computing power. One of the approaches for such problems is to isolate an irreducible subset (IIS) of the constraints. In other words a subset of constraints that is itself infeasible, but that becomes feasible by removing one or more constraints. This type of approach is helpful in large optimization problems.

Certificates of infeasibility can be useful, within optimization algorithms, in order to allow the fast determination of the inconsistency of the problem constraints, avoiding spending large computational times in infeasible problems, and also providing a guarantee that a problem is indeed not solvable. A series of results in interior-point based linear programming has been related to the construction of infeasibility certificates (Ben-Tal and Nemirovskiĭ, 2001). The issue of detecting infeasibility in optimization problems has been particularly important in the context of mixed integer linear programming (Andersen et al., 2008). In recent convex analysis literature, some infeasibility certificates have been derived for conic programming (Nesterov et al., 1999; Ben-Tal and Nemirovskiĭ, 2001) and for the monotone complementarity problem (Andersen and Ye, 1999). This last result has been extended to general convex optimization problems (Andersen, 2000). More general studies involving general nonlinear programming were presented in (Nocedal et al., 2014).

The problem of quick infeasibility detection has been considered by Byrd et al. (2010) in the context of sequential quadratic programming (SQP) method. The detection of minimizer of infeasibility has been presented in (Benson et al., 2002), (Byrd et al., 2006), (Fletcher et al., 2002) and (Wächter and Biegler, 2006), which apply SQP filters or interior point procedures. In (Martínez and da Fonseca Prudente, 2012), an augmented Lagrangian algorithm is presented to enhance asymptotic infeasibility. Their algorithm preserve the property of convergence to stationary points of the sum of squares of infeasibility without harming the convergence to Karush-Kuhn-Tucker (KKT) points in the feasible cases. In (Byrd et al., 2010), a nonlinear programming algorithm is presented which provide fast local convergence guarantees regardless if a problem is feasible or infeasible.

The main purpose of this thesis is to characterize infeasibility of non-linear optimization problems as a Pareto-criticality of an auxiliary problem. It is shown here a structural similarity between the Kuhn-Tucker condition for Efficiency (KTE) and a new necessary condition for infeasibility (INF) which also becomes sufficient under the assumption of problem strict convexity. The infeasibility condition proposed in this thesis is a new infeasibility certificate in finite-dimensional spaces, and uses the original (primal) variable only. The application of the proposed certificate is straightforward even in the case of generic non-linear functions, without the assumption of convexity. In such cases, the certificate has local meaning only. The procedure used in this thesis will be carried in this way: (i) An auxiliary unconstrained multiobjective optimization problem is defined. (ii) A Pareto-critical point of this auxiliary problem is determined. (iii) This point is either a feasible point of the original problem or a point in which the (INF) condition holds. (iv) In the case of (INF) being satisfied at any point, a necessary condition for the problem infeasibility becomes established (such condition is also sufficient in convex problems). Such verification is straightforward, leading to a potentially useful primal variable infeasibility certificate.

The procedure used in this thesis will be carried in this way: (i) An auxiliary unconstrained multiobjective optimization problem is defined. (ii) A Pareto-critical point of this auxiliary problem is determined. (iii) This point is either a feasible point of the original problem or a point in which the (INF) condition holds. (iv) In the case of (INF) being satisfied at any point, a necessary condition for the problem infeasibility becomes established (such condition is also sufficient in convex problems). Such verification is straightforward, leading to a potentially useful primal variable infeasibility certificate.

Finally, this thesis also considers the situation in which no gradient information is available, what occurs, for instance, in noisy problems. In this case, no usual infeasibility certificate can be applied. However, the auxiliary multiobjective optimization problem still holds, and the verification of its *regularity* can still be performed (in a heuristic sense). This allows to define another version of the proposed infeasibility certificate.

1.1 Research Contributions and Objectives

This study consider a constrained non-linear optimization problem and characterize its infeasibility as a Pareto-criticality condition of an auxiliary problem. This condition is evaluated by the use of non-linear unconstrained algorithms, in a search that either leads to a feasible point or to a point in which the infeasibility condition holds. The proposed new infeasibility certificate showed global validity in the case of convex problems and has at least a local meaning in generic nonlinear optimization problems. The following, to the best of author's knowledge, are the basic contributions that this dissertations incorporates.

1. The proposed algorithm generates either a solution converging towards feasibility and complementarity simultaneously or a certificate providing infeasibility.
2. If the original optimization problem is feasible, then this algorithm provides a single feasible optimal solution. On the other hand, if the original problem is infeasible then the proposed algorithm provide a certificate of infeasibility.
3. This algorithm provides an infeasibility certificate on primal variables, different from other existing certificates of infeasibility, which may be important both for theoretical and practical reasons.

1.2 Thesis Outline

This thesis is divided into six chapters. In this chapter, the importance of infeasibility certificates in optimization problems is introduced. The present chapter also highlights the benefits of certification in order to avoid spending a lot of computational time on infeasible problems. In the next chapter, a brief introduction has been included about multiobjective optimization and about a numerical method for solving optimization problems. This material is necessary in order to provide the tools that are used in the problem formulation to be presented later. Chapter three describes some infeasibility certificates that were presented in literature. Chapter four presents the infeasibility certificates proposed here, which constitute the main results of this thesis. Chapter five presents the actual algorithms that are constructed on the basis of the proposed infeasibility certificate. Numerical tests are also conducted in that chapter. The sixth and final chapter summarizes and concludes this work. A few suggestions regarding further possible exploration of this research area have been included too.

Chapter 2

Prelliminary Discussion

This chapter presents some prelliminary material that is necessary for the development of the infeasibility certificates that are proposed in this thesis. The issue of multiobjective optimization is discussed first. A specific method of numerical optimization that will be employed in the numerical experiments to be presented later is also discussed.

2.1 Multiobjective Optimization

Problems which involve simultaneous optimization of more than one objective function that are competing are called multiobjective optimization problems. Mathematically, the general form of a multiobjective optimization problem (MOOP) is given by,

$$
\begin{aligned}
\text{(MOOP)} \quad \min/\max \quad & f_k(x) & k = 1, 2, \ldots, t, \\
s.t. \quad & g_j(x) \leq 0 & j = 1, 2, \ldots, \bar{m}, \\
& h_j(x) = 0 & j = \bar{m}+1, \ldots, m, \\
& x_i^L \leq x_i \leq x_i^U & i = 1, 2, \ldots, n.
\end{aligned} \tag{2.1}
$$

The vector x is a vector of n decision variables: $x = (x_1, x_2, ...x_n)^T$. The decision variable search region is bounded by a set of box constraints i.e. x_i^L and x_i^U are the lower and upper bounds for the decision variable x_i respectively. Those points which satisfy all the constraints and variables are said to be feasible solutions and in the case of violations of the constraints they are said to be infeasible solutions. The set of points which satisfy all constraints is said to be the feasible region.

The above MOOP has t objective functions $f(x) = (f_1(x), f_2(x), \ldots, f_t(x))^T$, each of them can be either minimized or maximized at the same time. By convention, and w.l.g., minimization problems will be considered here. A difference between single-objective and multi-objective problems is that, in the multi-objective case the objective functions constitute a multi-dimensional space (space Z). Each solution under any mapping have an image z in the objective space, where $f(x) = z = (z_1, z_2, ..., z_t)^T$. Under any mapping the n-dimensional solution vector from the decision space has a t-dimensional objective vector in the objective space as its image. A typical diagram explains the case as follows.

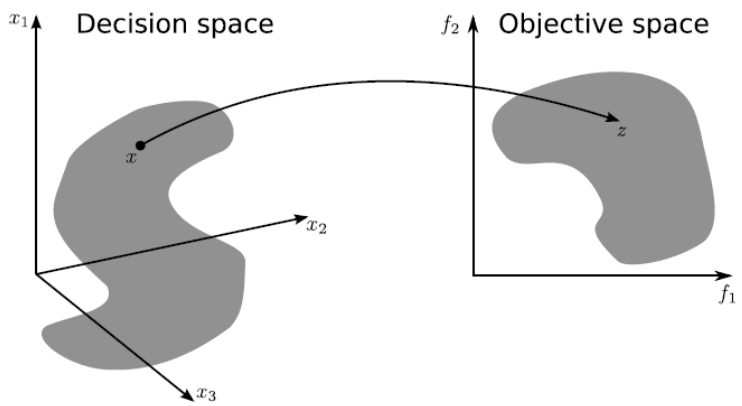

Figure 2.1: Mapping between decision and objective space

2.1.1 Dominance concept

The concept of dominance is central in multiobjective optimization, in order to define the solutions of the problems (Deb, 2001; Miettinen, 1999).

Definition 2.1 *A solution $x^{(1)}$ is said to dominate the other solution $x^{(2)}$ if both of the following conditions are true.*

1. *The solution $x^{(1)}$ is not worse than $x^{(2)}$ in all objectives. i.e.*

   $f(x^{(1)}) \leq f(x^{(2)})$

2. *The solution $x^{(1)}$ is strictly better than $x^{(2)}$ in at least one objective, or*

   $f_k(x^{(1)}) < f_k(x^{(2)})$ *for at least one k=1,2,. . . ,t.*

If either of the above conditions is violated, the solution $x^{(1)}$ does not dominate the solution $x^{(2)}$. The situation in which $x^{(1)}$ dominates the solution $x^{(2)}$ is denoted by $x^{(1)} \prec x^{(2)}$.

2.1.2 Pareto Optimality

The solutions of a multiobjective optimization problem are defined using the concept of dominance.

Definition 2.2 *Non-Dominated set: Considering a set of solutions $\mathcal{P}$, the non-dominated set $\mathcal{P}'$ contains those solutions that are not dominated by any member of the set $\mathcal{P}$.*

When $\mathcal{P}$ is the entire set of feasible solutions, the resulting non-dominated set $\mathcal{P}'$ is called the Pareto-optimal set. The solutions in $\mathcal{P}'$ are called Pareto-optimal solutions, or efficient solutions.

2.1.3 Methods for the Solution of Multiobjective Optimization

Several formulations can be used for dealing with multiobjective optimization problems. In this subsection, we present the ones which are relevant for the developments that are presented in this thesis.

2.1.3.1 The Scalarization Method

A multiobjective optimization problem can be approached by combining its multiple objectives into one single scalar objective function. This approach is known as scalarization or weighted sum approach. More specifically, the weighted sum method minimizes a positively weighted convex sum of the objectives, that is, that represents a new optimization problem with a unique objective function. The minimizer of this single objective function is an efficient solution for the original multiobjective problem, i.e. its image belongs to the Pareto curve.

$$
\begin{aligned}
min \qquad & \sum_{k=1}^{t} \gamma_k \cdot f_k(x) \\
& \sum_{k=1}^{t} \gamma_k = 1 \\
& \gamma_k \geq 0, \quad k = 1, \ldots, t \\
& x \in S
\end{aligned}
$$

Particularly we can say that if the γ weight vector is strictly greater than zero, then the minimizer of the problem is a strict Pareto optimum. While in the case of at least one $\gamma_k = 0$, then the minimizer of the problem may become a weak Pareto optimum.

The result by Geoffrion (1968) states necessary and sufficient conditions in the case of convexity as: If the solution set S is convex and the t-objectives f_k are convex on S, then x^* is a strictly Pareto optimum if and only if it exists γ such that x^* is an optimal solution of problem $P(\gamma)$. Similarly: If the solution set S is convex and the t objectives f_k are convex on S, x^* is a weakly Pareto optimum if and only if there exists γ, such that x^* is an optimal solution of problem $P(\gamma)$.

If the convexity hypothesis does not hold, then only the necessary condition remains valid, i.e., the optimal solutions of $P(\gamma)$ is strict Pareto optimum if $\gamma > 0$ and on the other hand it's weak Pareto optimum if at least one $\gamma \leq 0$.

2.1.3.2 ε-constraints Method

Another solution technique to multiobjective optimization is the ϵ-constraints method (Chankong and Haimes, 1983). Here, the decision maker chooses one objective out of t to be minimized; the remaining objectives are constrained to be less than or equal to given target values. In mathematical terms, if we let $f_1(x)$ to be the objective function chosen to be minimized, we have the following problem:

$$\begin{aligned} min \quad & f_1(x) \\ & f_k(x) \leq \epsilon_k, for\, all\, k \in \{1, \ldots, t\} \\ & x \in S \end{aligned}$$

The solution for this problem is called an weak solution, which may be, under additional conditions, an efficient solution.

2.2 The Nonlinear Simplex Search

In this section, we present the specific optimization method that will be employed in the numerical experiments that are conducted in this thesis.

The simplex search method was firstly proposed by Spendley, Hext, and Himsworth in 1962, (Spendley et al. (1962)) and later refined by Nelder and Mead (1965), it is also known as Nelder Simplex Search (NSS) or Downhill Simplex Search. The method was introduced for the minimization of multi-dimensional and non-linear unconstrained optimization problems. It is an algorithm based on the simplex algorithm of Spendley et al. (1962)). The geometrical structure of a simplex is composed of $(n + 1)$ points in n dimensions. If any point x of a simplex is taken as the origin, the n other points define vector directions which span the n-dimension vector space

(Durand and Alliot (1999)). By taking successive elementary geometric transformations, the initial simplex converges towards a minimum value at each iteration. This method is carried out by four movements, namely reflection, expansion, contraction and shrinkage in a geometric shape called simplex.

Definition 2.3 *A simplex or n-simplex Δ is a convex hull of a set of $n+1$ affine independent points Δ_i (i=1, ...,n+1), in some Euclidean space of dimension n.*

Definition 2.4 *A simplex is called non-degenerated, if and only if, the vectors in the simplex denote a linearly independent set. Otherwise, the simplex is called degenerated, and then, the simplex will be defined in a lower dimension than n.*

If the vertices of the simplex are all mutually equidistant, then the simplex is said to be regular. Thus, in two dimensions, a regular simplex is an equilateral triangle, while in three dimensions a regular simplex is a regular tetrahedron. The convergence towards a minimum value at each iteration of Nelder and Mead's method is conducted by four scalar parameters to control the movements performed in the simplex: Reflection (α), Expansion (γ), Contraction (β) and shrinkage σ. At each iteration, the $n+1$ vertices Δ_i of the simplex represent solutions which are evaluated and sorted according to monotonicity value $f(\Delta_1) \leq f(\Delta_2) \leq \ldots \leq f(\Delta_{n+1})$. In which $\Delta = \{\Delta_1, \Delta_2, \ldots \Delta_{n+1}\}$ is a set of vertices that define a nondegenerate simplex. According to Nelder and Mead, these parameters should satisfy:

$$\alpha > 0, \;\; \gamma > 1, \;\; \gamma > \alpha, \;\; 0 < \beta < 1 \;\; and \;\; 0 < \sigma < 1 \tag{2.2}$$

Actually, there is no method that can be used to establish these parameters. However, the nearly universal choices used in Nelder and Mead's method (Nelder and Mead (1965)) are:

$$\alpha = 1, \;\; \gamma = 2, \;\; \beta = 0.5 \;\; and \;\; \sigma = 0.5 \tag{2.3}$$

The transformations performed into the simplex by the Nelder and Mead method are defined as:

1. Reflection: $x_r = (1+\alpha)\Delta_c - \alpha\Delta_{n+1}$
2. Expansion: $x_e = (1+\alpha\gamma)\Delta_c - \alpha\gamma\Delta_{n+1}$
3. Contraction:
    a) Outside Contraction: $x_{oc} = (1+\alpha\beta)x_c - \alpha\beta\Delta_{n+1}$

b) Inside Contraction: $x_{ic} = (1-\beta)x_c + \beta\Delta_{n+1}$

4. Shrinkage: Each vertex of the simplex is transformed by the geometric shrinkage defined by: $\Delta_i = \Delta_1 + \sigma(\Delta_i - \Delta_1)$, i= 1, . . . , n+1, and the new vertices are evaluated, see figure (2.2).

Where $x_c = \frac{1}{n}\sum_{i=1}^{n}\Delta_i$ is the centroid of the n best points except Δ_{n+1}, which is the worst function value and Δ_1 is the best solution identified within the simplex. The figure (2.2) shows all the possible movements performed by the method.

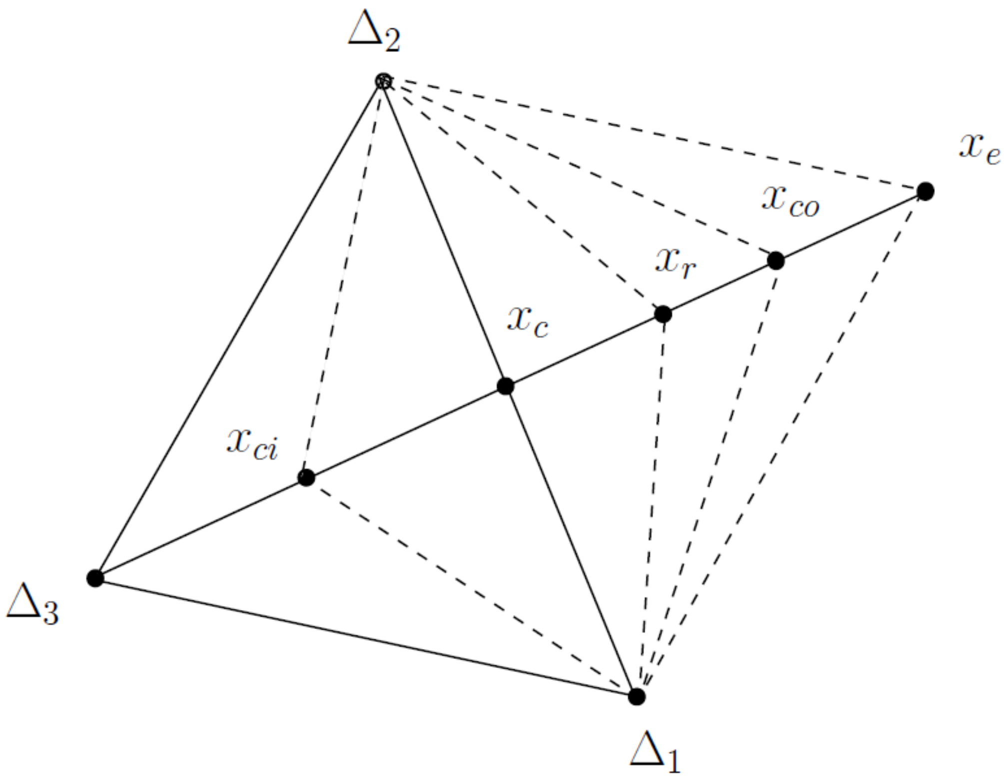

Figure 2.2: Geometrical representation illustrate all possible movements in the simplex performed by the NSS method. This simplex corresponds to an optimization problem with two decision variables.

The simplex corresponds to an optimization problem with two decision variables, where Δ_1 and Δ_3 are the best and worst points respectively. At each iteration, the simplex is modified by one of the above movements, according to the following rules:

1. If $f(\Delta_1) \leq f(x_r) \leq f(\Delta_n)$, then $\Delta_{n+1} = x_r$

2. If $f(x_e) < f(x_r) < f(\Delta_1)$, then $\Delta_{n+1} = x_e$, otherwise $\Delta_{n+1} = x_r$

3. If $f(\Delta_n) \leq f(x_r) < f(\Delta_{n+1})$ and $f(x_{oc}) \leq f(x_r)$ then $\Delta_{n+1} = x_{oc}$

4. If $f(x_r) \geq f(\Delta_{n+1})$ and $f(x_{ic}) < f(\Delta_{n+1})$, then $\Delta_{n+1} = x_{ic}$; otherwise, perform shrinkage.

The stopping criteria employed by Nelder and Mead, and commonly adopted in many optimization problems is defined by:

$$\sqrt{\frac{1}{n+1}\sum_{i=1}^{n+1}(f(\Delta_i) - \bar{f})^2} \leq \epsilon \tag{2.4}$$

In which $\bar{f} = \frac{1}{n+1}\sum_{i=1}^{n+1} f(\Delta_i)$ and ϵ is a predefined constant. When equation (2.4) satisfies, then Δ_c of the smallest simplex can be taken as the optimum point.

Chapter 3

Infeasibility Certificates

In this chapter, some existing approaches for the formulation of infeasibility certificates are presented.

As an optimization model becomes larger and more complex, infeasibility happens more often during the process of model formulation, and it becomes more difficult to diagnose the problem. A linear program may have thousands of constraints or even more: which of these are causing the infeasibility and how should the problem be repaired? In the case of nonlinear programs the issue becomes more complex. The problem may be entirely infeasible or the solver may just have been given a poor starting point from which it is unable to reach feasibility.

In modern optimization models, it is necessary to diagnose and repair infeasibility in face of the complexity of the models. In the last two decades, algorithmic approaches have been introduced for the solution of such problems. The following three main approaches are used to handle such issues (Greenberg, 1983):

(i) Identification of irreducible subset (IIS) within the larger set of constraints defining the model. This approach has the property that the IIS is irreducible, but it becomes feasible if one or more of its constraints are removed. Identifying an IIS permits the modeler to focus attention on a small set of conflicting functions within the larger model. Further improvement of the base algorithms try to return IISs that are of small cardinality.

(ii) The second approach of analyzing infeasibility is to identify the maximum feasible subset of constraints within the larger set of constraints defining the problem, or the minimum cardinality set of constraints that must be removed so that the remainder constitutes a feasible set.

(iii) The third approach seeks to suggest the best repair for the problem, where 'best' can be defined in various ways that can be handled algorithmically, e.g.

the fewest changes to constraint right hand side values. The suggested repair can of course be accepted, modified or rejected by the modeler.

The above methods for analyzing infeasibility as described above mostly depend on the ability of the solver to determine the feasibility or infeasibility of a problem subject to an arbitrary set of constraints with very high accuracy. This ability and skills are easily available for a linear problem, but on the other hand it is much more problematic for mixed integer and nonlinear problems.

This chapter starts from the result known as Farkas lemma, that certifies that an optimization problem is indeed infeasible. An easy example deals with the case in detail. The rest of the chapter presents different methods and infeasibility certificates for linear and non-linear optimization problems. This part is a gateway to our work which is described in the next chapters.

Example 3.1 *The linear optimization problem given by:*

$$\begin{aligned} minimize \quad & x_1 \\ subject\,to \quad & x_1 \leq 1, \\ & x_1 \geq 2, \end{aligned} \tag{3.1}$$

This problem is clearly infeasible, i.e. the problem has no solution. In other words the problem does not have a solution. To find the possible solution for the above problem, there are several possible ways to repair the problem. For example, the right hand side of the constraints may be changed appropriately, or one of the constraints may be removed. In the above simple example it is easy to discover the infeasibility and figure out a repair. Generally infeasibility problems are much more larger and complex to figure out it by hand.

3.1 The Farkas' certificate of primal infeasibility

If (3.1) is feasible, then it will be enough to have a feasible solution x to certify this claim. On the other hand, if (3.1) is not feasible, how can somebody claim that it is infeasible? Luckily there exist a well known certificate of infeasibility to answer this question, known as Farkas' lemma (Erling D. Andersen, 2011). Lets explain this certificate of infeasibility by taking a linear optimization problem.

$$
\begin{aligned}
(P)\ \ minimize \quad & c^T x \\
subject\,to \quad & Ax = b, \\
& x \geq 0,
\end{aligned} \tag{3.2}
$$

In which $b \in R^m$, $A \in R^{m \times n}$, and $c, x \in R^n$. The problem (3.2) is infeasible if and only if there exist a y such that

$$
\begin{aligned}
b^T y > 0 \\
A^T y \leq 0
\end{aligned} \tag{3.3}
$$

In other words any y satisfying (3.3) is a certificate of primal infeasibility. Notice that it is easy to verify that a Farkas' certificate y^* is valid because it corresponds to checking the conditions

$$
b^T y^* > 0 \tag{3.4}
$$

and

$$
A^T y^* \leq 0 \tag{3.5}
$$

Which shows that y^* is a certificate of infeasibility. It is easy to prove that indeed it is the case. Therefore, if an infeasibility certificate exists, then (3.2) can't be a feasible problem since we would have a contradiction of proving them. The generalization of (3.2) is given by

$$
\begin{aligned}
minimize \quad & c^T x \\
subject\,to \quad & A_1 x = b_1, \\
& A_2 x \leq b_2, \\
& A_3 x \geq b_3, \\
& x \geq 0,
\end{aligned} \tag{3.6}
$$

The equation (3.6) is infeasible if and only if there exists a (y_1, y_2, y_3) such that

$$
\begin{aligned}
b_1^T y_1 + b_2^T y_2 + b_3^T y_3 &> 0, \\
A_1^T y_1 + A_2^T y_2 + A_3^T y_3 &\leq 0, \\
y_2 &\leq 0, \\
y_3 &\geq 0.
\end{aligned} \tag{3.7}
$$

The above generalized Farkas' certificate of infeasibility for (3.1) is given by

$$
\begin{aligned}
y_1 + 2y_2 &> 0, \\
y_1 &\leq 0, \\
y_2 &\geq 0.
\end{aligned} \tag{3.8}
$$

And hence the valid certificate for it is $y_1 = -1$ and $y_2 = 1$. It is notable here that an infeasibility certificate is not unique because if it is multiplied by any strictly positive number, then it is still a certificate of infeasibility. Actually the infeasibility certificate of an optimization problem is a property of it rather than of the algorithm. Therefore it is good enough to request an infeasibility certificate from an algorithm whenever it claims a problem is infeasible. Since the properties of an infeasibility certificate are algorithm independent, decision based on infeasibility certificates will be similarly algorithm independent.

3.2 Validity of Farkas' certificate for relaxed problems

Farkas' certificate is not only used to certify that an optimization problem is infeasible, it can also be used to find out the case of infeasibility. In common practice if a problem is infeasible we would like to repair it, or on the other hand to know at least which part of the problem causing infeasibility. For example a simple approach is the case of problem (3.1). If we change the right-hand of the second constraint to

$x_1 \geq 1.2$ then the revised Farkas' conditions are

$$\begin{aligned} y_1 + 2y_2 &> 0, \\ y_1 &\leq 0, \\ y_2 &\geq 0. \end{aligned} \tag{3.9}$$

It is still very clear that the previous certificate of infeasibility $y_1 = -1$ and $y_2 = 1$ is still a valid certificate for the changed problem as well. If we further change the right-hand side of the second constraint to 1, then the Farkas' certificate of infeasibility is no longer valid. Generally, when repairing an infeasibility problem it should be change as much as the infeasibility certificate remain invalid because otherwise the problem stays infeasible. All y_i are non zero in the infeasibility certificate. If any $y_i = 0$, then the i-th constraint is not involved in the infeasibility since if the i-th constraint is removed from the problem and y_i is removed from the vector y, then the reduced y is still an infeasibility certificate.

3.3 Primal or Dual infeasibility in Linear Case

In this section, the certificate of primal and dual infeasibility (Andersen, 2001) is discussed. Furthermore, a definition of a basis certificate and strongly polynomial algorithm of Farkas' type for the computation of the basis certificate of infeasibility have been included.

Generally, if a linear program has an optimal solution, then the certificate of feasibility status are the primal and dual optimal solutions. It is well known, in the solvable cases that the linear program must have a basic optimal solution. We know that if a linear program is primal or dual infeasible, then the Farkas lemma (as discussed in the previous section) provides the basic infeasibility certificate.

Interior-point methods have emerged as an efficient alternative to simplex based solution methods for linear programming. Unluckily some of these methods such as the primal-dual algorithms discussed in (Wright, 1997) do not handle primal or dual infeasible linear programs very well, but interior-point methods based on the homogeneous model find a possible infeasible status both in theory and practice (Andersen and Andersen, 2000; Roos et al., 1997). In their approach they used Farkas' lemma to generate these infeasibility certificates from the interior-point methods.

Consider problem (3.2). For convenience and without loss of generality the rank of A is assumed as $rank(A) = m$. The dual problem corresponding to (3.2) is

$$\begin{aligned} (D) \qquad & maximize\ b^T y \\ & subject\ to\ A^T y + s = c \\ & s \geq 0, \end{aligned}$$

In which $y \in R^m$ and $s \in R^n$. Problem (3.2) is said to be feasible if a solution that satisfies the constraints of (3.2) exists. Similarly (D) is said to be feasible if (D) has at least one solution satisfying the constraints of (D). The following Lemma is a well-known fact of linear programming.

Lemma 3.3.1

a. (P) *has an optimal solution if and only if there exist* (x^*, y^*, s^*) *such that*

$$Ax^* = b,\ A^T y^* + s^* = c,\ c^T x^* = b^T y^*,\ x^*, s^* \geq 0.$$

b. (P) *is infeasible if and only if there exists* y^* *such that*

$$A^T y^* \leq 0,\ \ b^T y^* > 0. \tag{3.10}$$

c. (D) *is infeasible if and only if there exists* x^* *such that*

$$Ax^* = 0,\ \ c^T x^* < 0,\ \ x^* \geq 0. \tag{3.11}$$

Proof See (Roos et al., 1997).

Therefore, (P) has an optimal solution if and only if (P) and (D) are both feasible. Apart from this, a primal and dual optimal solution is a certificate that the problem has an optimal solution. If the problem is primal or dual infeasible, then any y^* satisfying (3.10) and any x^* satisfying (3.11) is a certificate of primal and dual infeasibility. A linear programming problem may be both primal and dual infeasible and in that case both a certificate for the primal and dual infeasibility exists. It is interesting to note that, if a linear optimization problem is solved by the method of column generation, and for example the first sub-problem is infeasible, then any column of this sub-problem having a positive inner product with the infeasibility certificate y^* of primal infeasibility is suitable to be included in the next sub problem. At the end of this search, if no such a column exists, then the whole problem can be concluded to be infeasible with the non-unique infeasibility certificate y^*.

In the following, the definition of an optimal basic partition of the indices of the variables for a primal infeasible program is taken from (Andersen, 2001) in order to know that whether an LP has a feasible or an infeasible solution. The phase 1 problem corresponding to (3.2) is

$$\begin{aligned} &maximize\ z_{1p} = e^T t^+ + e^T t^- \\ &subject\ to\ Ax + It^+ - It^- = b, \\ &x, t^+, t^- \geq 0. \end{aligned} \tag{3.12}$$

in which e is a vector of appropriate dimension containing all ones. Problem (3.13) has clearly a feasible solution and the purpose of the objective function in this form is to minimize the sum of infeasibility. The primal problem (3.2) has a feasible solution if and only if $z^*_{1p} = 0$. The basic partition (β, N) of the indices of the variables taken from (Andersen, 2001) is a certificate of primal infeasibility if it satisfies the following definition.

Definition 3.1 *A basic partition (β, N) of the indices of the variables to (3.2) is a certificate of primal infeasibility if*

$$\exists i : e_i^T B^{-1} A \geq 0,\ e_i^T B^{-1} b < 0 \tag{3.13}$$

It is notable that any infeasible linear program has a basic partition of the indices of the variables which satisfies the above definition. The following result is stated in (Andersen, 2001).

Theorem 3.3.2 *Given any certificate (y^*, s^*) of primal infeasibility (Andersen and Andersen, 2000), then a basis certificate satisfying definition (3.1) can be computed in strongly polynomial time.*

3.4 Infeasibility certificate for monotone complementarity problem

Andersen and Ye (1999) presented the generalization of a homogeneous self-dual linear programming (LP) algorithm for the solution of monotone complementarity problem (MCP). Their algorithm generates either a solution converging towards feasibility and complementarity or a certificate proving infeasibility of the problem. The monotone complementarity problem in the standard form is given by

$$(MCP) \qquad minimize\ x^T s \qquad (3.14)$$
$$subject\ to\ s = f(x),\ (x,s) \geq 0$$

In the above equation $f(x)$ is a continuous monotone mapping i.e. $f : R^n_+ \to R^n$, where $R^n_+ := \{x \in R^n : x \geq 0\}$ and $x, s \in R^n$. Equation (3.14) can be written as: for every $x^1, x^2 \in R^n_+$, we have

$$(x^1 - x^2)^T (f(x^1) - f(x^2)) \geq 0$$

The problem (3.14) is said to be (asymptotically) feasible if and only if there exist a bounded sequence $\{(x^t, s^t)\} \subset R^{2n}_{++}, \quad t = 1, 2 \ldots,$ such that

$$\lim_{t \to \infty} s^t - f(x^t) \longrightarrow 0,$$

Any limit point $(\hat{x}, \hat{s})$ of the above sequence is called an (asymptotically) feasible point for the monotone complementarity problem (3.14). Moreover the problem (3.14) has an interior feasible point if it has an (asymptotically) feasible point $(\hat{x} > 0, \hat{s} > 0)$. Equation (3.14) is called to be (asymptotically) solvable if there exist an (asymptotically) feasible point $(\hat{x} > 0, \hat{s} > 0)$ such that $\hat{x}^T \hat{s} = 0$, where $(\hat{x}, \hat{s})$ is called optimal or the monotone complementarity solution for (3.14). The monotone complementarity problem (3.14) is said to be strongly infeasible if and only if there is no sequence $\{(x^t, s^t)\} \subset R^{2n}_{++}, \quad t = 1, 2 \ldots,$ such that

$$\lim_{t \to \infty} s^t - f(x^t) \longrightarrow 0,$$

The monotone complementarity problem (MCP) algorithm (Andersen and Ye, 1999) has the following features:

- It achieves $O(\sqrt[n]{\log}(1/\epsilon))$ iteration complexity if f satisfies the scaled Lipschitz condition.
- It solves the problem without any regularity assumption concerning the existence of optimal, feasible, or interior feasible points.
- It can start at a positive point, feasible or infeasible, near the central ray of the positive orthant (cone), and it does not need to use any big-M penalty parameter or lower bound.

- If (MCP) has a solution, the algorithm generates a sequence that approaches feasibility and optimality simultaneously; if the problem is (strongly) infeasible, the algorithm generates a sequence that converges to a certificate proving infeasibility.

3.5 Infeasibility certificate in homogenous model for convex optimization

The previous section was about the certificate of monotone complementarity problem (MCP). The good thing about (MCP) is that it is either solvable or (strongly) infeasible, which provides a certificate of optimality or infeasibility. In (Andersen, 2000), the suggested formulation of (Andersen and Ye, 1999) is applied to the Karush-Kuhn-Tucker optimality condition corresponding to a homogenous model for convex optimization problem which provides an infeasibility certificate. This (MCP) corresponding certificate provides information about whether the primal or dual problem is infeasible given certain assumptions.

The convex optimization problems have an optimal solution that can be found by most of the interior point methods. If the problem is primal or dual infeasible, then the optimal solution is not possible. Andersen and Ye (1999) handled this issue and generalized it for the linear problems as a monotone complementarity problem (MCP). This larger class contains all convex optimization problems, because the Karush-kuhn-Tucker conditions corresponding to a convex optimization problems form an MCP. In the previous section, in the certificate of (Andersen and Ye, 1999), it is not stated whether an infeasibility certificate indicates primal or dual infeasibility when the homogenous model is applied to the optimality conditions of a convex optimization problem. This issue is handled in (Andersen, 2000), which shows that an infeasibility certificate in some cases indicates whether the primal or dual problem is infeasible. The optimization problem in (Andersen, 2000) is given by

$$\begin{aligned} &minimize\ c(x) \\ &subject\ to\ a_i(x) \geq 0, \ \ i = 1, \ldots, m, \end{aligned} \tag{3.15}$$

in which $x \in R^n$. The function $c : R^n \to R$ is assumed to be convex, and the component function $a_i : R^n \to R$, $i = 1, \ldots, m$, are assumed to be concave. All functions in (3.15) are assumed to be once differentiable. Hence, the problem (3.15) minimizes a convex function over a convex set. The Lagrange function is defined as:

$$L(x,y) := c(x) - y^T a(x)$$

The Wolf dual corresponding to (3.15) is defined as

$$\begin{aligned} & maximize\ L(x,y) \\ & subject\ to\ \nabla_x L(x,y)^T = 0 \\ & y \geq 0 \end{aligned} \tag{3.16}$$

The combined equations (3.15) and (3.16) give the MCP

$$\begin{aligned} & minimize\ y^T z \\ & subject\ to\ \nabla_x L(x,y)^T = 0 \\ & a(x) = z \\ & y, z \geq 0 \end{aligned} \tag{3.17}$$

in which $z \in R^m$ is a vector of slack variables. A solution to (3.17) is said to be complementarity if the corresponding objective value is zero.

In the following, the homogenous model suggested in (Andersen and Ye, 1997; Andersen, 2000) is applied to this problem, the obtained homogenized MCP is

$$\begin{aligned} & minimize\ z^T y + \tau\kappa \\ & subject\ to\ \tau \nabla_x L(x/\tau, y/\tau)^T = 0, \\ & \tau a(x/\tau) = z \\ & -x^T \nabla_x L(x/\tau, y/\tau)^T - y^T a(x/\tau) = \kappa, \\ & z, \tau, y, \kappa \geq 0 \end{aligned} \tag{3.18}$$

in which τ and κ are additional variables. From (Andersen and Ye, 1999), equation (3.18) is said to be asymptotically feasible if and only if a convergent sequence $(x^k, z^k, \tau^k, y^k, \kappa^k)$ exists for $k = 1, 2 \ldots$ such that

$$\lim_{k \to \infty} \begin{pmatrix} \tau^k \nabla_x L(x^k/\tau^k, y^k/\tau^k)^T, \\ \tau^k a(x^k/\tau^k) - z^k, \\ -(x^k)^T \nabla_x L(x^k/\tau^k, y^k/\tau^k)^T - (y^k)^T a(x^k/\tau^k) - \tau^k \end{pmatrix} = 0 \tag{3.19}$$

and

$$(x^k, z^k, \tau^k, z^k, \kappa^k) \in R^n \times R_+^m \times R_{++} \times R_+^m \times R_{++} \ \forall k, \tag{3.20}$$

in which the limit point of the $(x^k, z^k, \tau^k, y^k, \kappa^k)$ is called an asymptotically feasible point. Also this limit point is said to be asymptotically complementary if

$$(y^*)^T z^* + \tau^* \kappa^* = 0$$

Theorem 3.5.1 *Equation (3.18) is asymptotically feasible, and every asymptotically feasible point is an asymptotically complementarity solution.*

Proof . See (Andersen and Ye, 1997)

This result implies that the objective function (3.18) is redundant, and hence the problem is a feasibility problem. The following lemma from (Andersen, 2000) possibly concludes that either the primal or the dual problem is infeasible.

Lemma 3.5.2 *Let $(x^k, z^k, \tau^k, y^k, \kappa^k)$ be any bounded sequence satisfying (3.20) such that*

$$\lim_{k\to\infty} (x^k, z^k, \tau^k, y^k, \kappa^k) = (x^*, z^*, \tau^*, y^*, \kappa^*)$$

is an asymptotically feasible and maximally complementarity solution to (3.18). Given

$$\lim_{k\to\infty} -(x^k)^T \nabla_x L(x^k/\tau^k, y^k/\tau^k)^T - (y^k)^T a(x^k/\tau^k) = \kappa^* > 0, \tag{3.21}$$

then

$$\lim_{k\to\infty} \ sup \ (\nabla a(x^k/\tau^k)(y^k/\tau^k) - a(x^k/\tau^k))^T (y^k) > 0 \tag{3.22}$$

or

$$\lim_{k\to\infty} \ sup \ - \nabla c(x^k/\tau^k) x^k > 0 \tag{3.23}$$

holds true. Moreover, if

$$\lim_{k\to\infty} \tau^k \nabla c(x^k/\tau^k) = 0, \tag{3.24}$$

then the primal problem (3.15) is infeasible if (3.22) holds and the dual (3.16) is infeasible if (3.23) holds.

Proof . (Andersen, 2000)

3.6 Infeasibility detection in Nonlinear Optimization Problems

Byrd et al. (2010) address the need for optimization algorithms that can solve feasible problems and detect when a given optimization problem is infeasible. For this purpose an active-set sequential quadratic programming method was proposed, which is derived from an exact penalty approach that adjusts the penalty parameter appropriately to emphasize optimality over feasibility. In this approach, the updating process of penalty parameter is used in every iteration, particularly in the case of infeasible problems. The optimization problem in (Byrd et al., 2010) is given by

$$\begin{aligned} &\min_{x \in R^n} f(x) \\ &s.t.\;\; g_i(x) \geq 0,\;\; i \in I = \{1, \ldots, t\} \end{aligned} \tag{3.25}$$

in which $f : R^n \to R$ and $g_i : R^n \to R$ are smooth functions. When there is no feasible point of (3.25), then the algorithm returns a solution of the problem

$$\min_x \nu(x) \triangleq \sum_{x \in I} max\{-g_i(x), 0\} \tag{3.26}$$

When problem (3.25) is infeasible, the iterations converge quickly to an infeasible stationary point $\hat{x}$, which is defined as a stationary point of problem (3.26) such that $\nu(\hat{x}) \geq 0$. The problem (3.25) is locally infeasible if there is an infeasible stationary point $\hat{x}$ for it. The general form of the Penalty-SQP framework in (Byrd et al., 2010) is given below:

$$\phi(x; \rho) = \rho f(x) + \nu(x)$$

in which ν is the infeasibility measure as defined in (3.26) and $\rho > 0$ is a penalty parameter updated dynamically within the approach. If the penalty parameter is very small, then the stationary points of the non-linear program (3.25) are also stationary points of the penalty function ϕ as in (Han and Mangasarian, 1979). Given a value for ρ_k and an iterate x_k, the appropriate step d_k is defined as a solution to the subproblem

$$\min_{x \in R^n} q_k(d; \rho_k) \tag{3.27}$$

in which

$$\begin{aligned} q_k(d; \rho_k) &= \rho_k \nabla f(x_k)^T d + 1/2 d^T W(x_k, \lambda_K; \rho_k) d + \sum_{x \in I} max\{-g_i(x_k) \\ &\quad -\nabla g_i(x_k)^T d, 0\} \end{aligned} \tag{3.28}$$

is a local model of the penalty function $\phi(.; \rho)$ about x_k. $W(x_k, \lambda_k; \rho_k)$ is the Hessian matrix given as follows

$$W(x_k, \lambda_k; \rho_k) = \rho_k \nabla^2 f(x_k) - \sum_{x \in I} \lambda_k^i \nabla^2 g_i(x_k) \tag{3.29}$$

The Hessian used here is different from that used in the standard penalty methods (Nocedal and Wright, 2006) in that the penalty parameter only multiplies the Hessian of the objective but not the term involving the Hessian of the constraints. The smooth reformulation of (3.27) is given below

$$\min_{x \in R^n, s \in R^t} \rho_k \nabla f(x_k)^T + 1/2 d^T W(x_k, \lambda; \rho_k) d + \sum_{i \in I} s_i \tag{3.30a}$$

$$s.t.\ g_i(x_k) + \nabla g_i(x_k)^T d + s_i \geq 0, i \in I, \tag{3.30b}$$

$$s_i \geq 0, i \in I, \tag{3.30c}$$

in which s_i are slack variables. The sub-problem (3.30) is the focal point of this approach which seeks optimality and feasibility with evolution of the value for ρ. With a solution d_k to problem (3.30), the iterate is updated as

$$x_{k+1} = x_k + \alpha_k d_k$$

in which α_k is a steplength parameter that ensures sufficient reduction in $(\phi(.; \rho_k))$.

The constraints (3.30) are always feasible, which was one of the main motivations for the Sl_1QP approach proposed by Fletcher in 1980s.

The rules described by Byrd et al. (2008b) were designed to ensure global convergence (even in the infeasible case) and (Byrd et al., 2003, 2008a) show the effectiveness in practice. However the problem was that it did not produce a fast rate of convergence in the infeasible case.

Properties for penalty SQP algorithms applied to feasible problems have been studied in (Fletcher, 1987), but in (Byrd et al., 2010), the authors focused their analysis on the infeasible case. In the infeasible case, many others like Gill et al. (2002) make no attempt to make a fast rate of convergence to stationary points. For this, Byrd et al. (2010) focused their analysis on the infeasible case for the following penalty problem.

$$\begin{aligned} &\min_{x,r} \rho f(x) + \sum_{i\in I} r_i \\ &s.t.\ \ g_i(x) + r_i \geq 0,\ \ r_i \geq 0, i \in I \end{aligned} \tag{3.31}$$

in which r_i are slack variables. If x^ρ is defined as a first-order optimal solution of problem (3.31) for a given value of ρ, then there exist slack variables r^ρ and Lagrange multipliers λ^ρ, σ^ρ such that $(x^\rho, \lambda^\rho, r^\rho, \sigma^\rho)$ satisfy the KKT system

$$\rho\nabla f(x) - \sum_{i\in I} \lambda_i \nabla g_i(x) = 0, \tag{3.32a}$$

$$1 - \lambda_i - \sigma_i = 0,\ \ i \in I, \tag{3.32b}$$

$$\lambda_i(g_i(x) + r_i) = 0,\ \ i \in I, \tag{3.32c}$$

$$\sigma_i r_i = 0,\ \ i \in I, \tag{3.32d}$$

$$g_i(x) + r_i \geq 0, \tag{3.32e}$$

$$r, \lambda, \sigma \geq 0. \tag{3.32f}$$

Particularly, if $\rho > 0$ such a solution has $r^\rho = 0$, then x^ρ is a first order optimal solution for the non-linear problem (3.25). The following lemma (Byrd et al., 2010) is an alternative way of characterizing solutions of the penalty problem (3.31).

Lemma 3.6.1 *Suppose that* $(x^\rho, \lambda^\rho, r^\rho, \sigma^\rho)$ *is a primal-dual KKT point for problem* (3.31) *and that the strict complementarity conditions*

$$r^\rho + \sigma^\rho > 0,\ \ \lambda_i^\rho + (g_i(x^\rho) + r_i^\rho) > 0 \tag{3.33}$$

hold for all $i \in I$. Then (x^ρ, λ^ρ) satisfy the system

$$\rho \nabla f(x) - \sum_{i \in I} \lambda_i \nabla g_i(x) = 0, \tag{3.34a}$$

$$and\ either\ 1 - \lambda_i - \sigma_i = 0,\ \ i \in I, \tag{3.34b}$$

$$g_i(x) < 0,\ and\ \lambda_i = 1,\ or \tag{3.34c}$$

$$g_i(x) > 0,\ and\ \lambda_i = 0,\ or \tag{3.34d}$$

$$g_i(x) = 0,\ and\ \lambda_i \in (0, 1). \tag{3.34e}$$

Conversely, if (x, λ) satisfy (3.34), it also satisfies (3.32) together with $r_i = max(0, -g_i(x))$ and $\sigma_i = 1 - \lambda_i$

Proof See (Byrd et al., 2010)

3.7 Modified Lagrangian Approach

In (Martínez and da Fonseca Prudente, 2012), a modified augmented Lagrangian algorithm is presented for handling asymptotic infeasibility. Their modified algorithm preserves the property of convergence to stationary points of the sum of squares of infeasibility, while it does not affect those points in the feasible cases which converges to the KKT points.

Many of the global optimization algorithms converge to KKT in the best case, while in the worst case these algorithms converge to infeasible points (stationary points for some infeasibility measure). In this case, one expects that the problem is infeasible. However, every affordable optimization algorithm can converge to an infeasible point, even when the feasible points exist. Therefore optimizers that wish to find feasible and optimal solutions of practical problems usually change the initial approximation and/or the algorithmic parameters of the algorithm when an almost infeasible point is found. On the other hand, practical optimization algorithm should be effective not only for finding solutions of the problems but also for finding infeasibility certificates when there is no alternative.

Augmented Lagrangian type algorithms are studied in (Andreani et al., 2007; Rockafellar, 1974). In particular, in the algorithm introduced in (Andreani et al., 2007), the iterates x^k are computed as approximate minimizers of Augmented Lagrangian in which multipliers and penalty parameters are updated. The increasing precision requirements makes it very difficult to solve sub-problems when the penalty parameters go to infinity, which is necessarily the case when a feasible point is not found. In this paper it was observed that, in that case, the same convergence results are obtained using bounded away from zero tolerances for solving the sub-problems. This fact motivates the employment of dynamic adaptive tolerances that depend on the degree of infeasibility and complementarity at each iterate x^k. Adaptive precision control for optimality depending on infeasibility measures has been considered, with different purposes. The problem in (Martínez and da Fonseca Prudente, 2012) is given by:

$$\begin{aligned} &Minimize\ \ f(x) \\ &subject\ to\ \ h(x) = 0 \\ &\ g(x) \leq 0 \\ &x \in \Omega \end{aligned} \tag{3.35}$$

in which $h : R^n \to R^m$, $g : R^n \to R^p$, $f : R^n \to R$ are smooth and $\Omega \subset R^n$ is a bounded n-dimensional box given by:

$$\Omega = \{x \in R^n \mid a_i \leq x_i \leq b_i \ \forall \ i = 1, \ldots, n\}$$

The Augmented Lagrangian function given in (Rockafellar, 1974) is defined as:

$$L_p(x, \lambda, \mu) = f(x) + \frac{\rho}{2}\{\sum_{i=1}^{m}[h_i(x) + \frac{\lambda_i}{\rho}]^2 + \sum_{i=1}^{p}[max(0, g_i(x) + \frac{\mu_i}{\rho})]^2\}$$

for all $x \in \Omega$, $\rho > 0$, $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p_+$

This algorithm (Martínez and da Fonseca Prudente, 2012) is similar to the one in (Andreani et al., 2007), but the difference is in the stopping criterion for the sub-problem. The original algorithm imposes that the convergence tolerance $\{\epsilon_k\}$ for the sub-problems should tend to zero while this condition on stopping criteria is relaxed in (Martínez and da Fonseca Prudente, 2012).

Chapter 4

A new Certificate of Infeasibility for Non-linear Optimization Problems

In this chapter we propose a new necessary condition for the infeasibility of non-linear optimization problems which is stated as Pareto-criticality condition of an auxiliary multiobjective optimization problem.

4.1 Preliminary Statements

Consider the optimization problem defined by:

$$\begin{aligned} &\min_x f(x) \\ &\text{subject to: } g(x) \leq 0 \end{aligned} \tag{4.1}$$

in which $f(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^p$ and $g(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^m$ are vector functions. The set of feasible points is denoted by

$$\Omega \triangleq \{x \in \mathbb{R}^n \mid g(x) \leq 0\} \tag{4.2}$$

In particular case if $p = 1$, the problem (4.1) is a conventional mono-objective optimization problem. When $p > 1$ the problem becomes multi-objective. In this last case, a feasible point $\tilde{x} \in \mathbb{R}^n$ of the decision variable space is said to be dominated by another feasible point $\bar{x} \in \mathbb{R}^n$ if $f(\bar{x}) \prec f(\tilde{x})$. The solution set of the multiobjective optimization problem is defined as the set $\mathcal{P} \subset \Omega$ of feasible points that are not dominated by any other feasible point. This set is called the *efficient solution set*, or the *Pareto-optimal set*. In order to state general results, the solution set of a mono-objective problem is also denoted by $\mathcal{P}$.

The following compactness assumption will be necessary for the derivation of our results:

Assumption 4.1.1 *Assume that there is a subset of the constraint functions,* $\{g_1(\cdot), g_2(\cdot), \ldots, g_k(\cdot)\}$ *with* $k \leq m$*, such that the set* $\Omega_c \subset \mathbb{R}^n$ *defined by*

$$\Omega_c = \{x \mid g_1(x) \leq 0,\ g_2(x) \leq 0, \ldots,\ g_k(x) \leq 0\}$$

is a non-empty compact set. ♢

This assumption holds in a large class of problems, for instance when there is a "box" in the decision variable space in which the search is to be conducted. The issue of feasibility/infeasibility can become a difficult question only w.r.t. the other constraint functions, $g_{k+1}(\cdot), \ldots, g_m(\cdot)$.

4.2 Infeasibility Condition

Let $\lambda \in \mathbb{R}^p$ and $\mu \in \mathbb{R}^m$. The Kuhn-Tucker conditions for efficiency at a solution $\bar{x}$ of problem (4.1) is stated as (Luc, 1988; Marusciac, 1989):

$$\text{(KTE)} \begin{cases} F(\bar{x})\,\lambda + G(\bar{x})\,\mu \ = \ 0 \\ \lambda \succ 0\ ,\ \mu \geq 0 \\ g(\bar{x}) \ \leq \ 0 \\ \mu_i\, g_i(\bar{x}) \ = \ 0\ ;\ \forall\, i = 1, \ldots, m \end{cases} \tag{4.3}$$

Notice that the Karush-Kuhn-Tucker conditions for optimality of the single-objective case is a particular case of KTE.

For problem (4.1), given a point $\bar{x} \in \mathbb{R}^n$, one of the four possibilities below must happen (by exhaustion):

(a) $\bar{x} \in \mathcal{P}$, it means that the Kuhn-Tucker necessary conditions for Efficiency (KTE) hold

(b) $\bar{x} \in \Lambda$, with Λ defined as the set of points for which hold:

$$\text{(INF)} \begin{cases} \exists\ i \mid g_i(\bar{x}) > 0 \\ G(\bar{x})\mu \ = \ 0 \\ \mu \succ 0 \\ g_j(\bar{x}) < 0 \ \Rightarrow\ \mu_j \ = \ 0 \end{cases} \tag{4.4}$$

for some vector of multipliers $\mu \in \mathbb{R}^m$.

(c) $\bar{x} \in \Omega$ and $\bar{x} \notin \mathcal{P}$.

(d) $\bar{x} \notin \Omega$ and $\bar{x} \notin \Lambda$.

Points that satisfy the condition (KTE) are *Pareto-critical* for problem (4.1). The condition (INF) is very similar to (KTE). It will be shown that points that satisfy (INF) are also Pareto-critical w.r.t. another auxiliary problem. For this we define the following vector function $\hat{g}(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^m$ as follows:

$$\hat{g}_i(x) = \begin{cases} 0 \ , \ \forall \ x \mid g_i(x) \leq 0 \\ g_i(x) \ , \ \forall \ x \mid g_i(x) > 0 \end{cases} \quad i = 1, \ldots, m \tag{4.5}$$

The following unconstrained auxiliary problem is defined:

$$\min_x \hat{g}(x) \tag{4.6}$$

and the corresponding efficient solution set of this problem is denoted by $\mathcal{A}$:

$$\mathcal{A} = \{x \in \mathbb{R}^n \mid \ \nexists \ \bar{x} \in \mathbb{R}^n \text{ such that } \hat{g}(\bar{x}) \prec \hat{g}(x)\} \tag{4.7}$$

In the case of feasibility of problem (4.1), the efficient solution set $\mathcal{A}$ have feasible solution points only. On the other hand, if (4.1) is strictly infeasible then $\mathcal{A}$ is composed on those points for which the INF condition holds.

It should be noticed that, under assumption 4.1.1, it can be stated that: $\mathcal{A} \neq \emptyset$ and $\mathcal{A} \subset \Omega_c$. Denote by $\hat{g}(\mathcal{A})$ the image set of function $\hat{g}(\cdot)$ over $\mathcal{A}$. The following lemma comes directly from the definition of the function $\hat{g}(\cdot)$:

Lemma 4.2.1 *The following statements hold:*

(i) $\Omega \neq \emptyset \ \Rightarrow \ \hat{g}(\mathcal{A}) \equiv 0 \ , \ \Omega \equiv \mathcal{A}$

(ii) $\Omega = \emptyset \ \Rightarrow \ \hat{g}(x) \succ 0 \ \forall \ x \in \mathcal{A}$

◊

The next lemma states a relation between the set Λ and the set $\mathcal{A}$.

Lemma 4.2.2 *The following statement holds:*

$$\Omega = \emptyset \ \Rightarrow \ \Lambda \supset \mathcal{A}$$

◊

Proof The condition (INF), which holds for the points of Λ, corresponds to the KTE necessary condition for the Pareto-optimality w.r.t. problem (4.6) when problem (4.1) is infeasible, which holds for the points of $\mathcal{A}$. ■

Under convexity assumption, a stronger result can be obtained as:

Lemma 4.2.3 *Suppose that the functions $f(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^p$ and $g(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^m$ are convex. In this case, the following statements hold:*

(i) $\Omega = \emptyset \;\;\Rightarrow\;\; \Lambda \equiv \mathcal{A}$

(ii) $\Omega \neq \emptyset \;\;\Rightarrow\;\; \Lambda = \emptyset$

♢

Proof Both statements come from the fact that $\mathcal{A}$ is the Pareto-optimal set of the auxiliary problem (4.6), which means that the points in this set must satisfy a Pareto-criticality condition for this problem. If the problem (4.1) is infeasible, as in case (i), the Pareto-criticality condition becomes (INF). In this case, the problem convexity leads to the sufficiency of the Pareto-criticality for a point to belong to $\mathcal{A}$. Otherwise, in case (ii), the Pareto-criticality condition only holds for feasible points (see Lemma 4.2.1-(i)), and (INF) does not hold for any point. ■

The following Corollary of Lemmas 4.2.1 and 4.2.2 can be stated, without assuming convexity:

Corollary 4.2.4 *Consider any point $\bar{x} \in \mathbb{R}^n$. The following relations hold:*

(i) $\bar{x} \in (\Lambda \cap \mathcal{A}) \Rightarrow \Omega = \emptyset$

(ii) $\Omega = \emptyset \Rightarrow (\Lambda \cap \mathcal{A}) \neq \emptyset$

♢

It should be noticed that the verification of criticality condition $\bar{x} \in \Lambda$ depends only on local gradient evaluation on point $\bar{x}$. On the other hand, the efficiency condition $\bar{x} \in \mathcal{A}$ has a global meaning, and cannot be evaluated on the basis of local information only. However, an assumption of convexity leads to an equivalence of criticality and efficiency. In this way, the following Corollary of Lemmas 4.2.1 and 4.2.3 holds under convexity:

Corollary 4.2.5 *Consider any point $\bar{x} \in \mathbb{R}^n$. Suppose that the functions $f(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^p$ and $g(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^m$ are convex. The following relations hold:*

(i) $\bar{x} \in \Lambda \Rightarrow \Omega = \emptyset$

(ii) $\Omega = \emptyset \Rightarrow \mathcal{A} = \Lambda \neq \emptyset$

♢

The main result of this work is stated, in the version without convexity, as the conjunction of the Lemmas 4.2.1, 4.2.2 and Corollary 4.2.4:

Theorem 4.2.6 *Consider the optimization problem defined by (4.1). Then:*

$$\begin{aligned} &\mathcal{A} = \Omega \neq \emptyset \Leftrightarrow (\Lambda \cap \mathcal{A}) = \emptyset \\ &(\mathcal{A} \cap \Lambda) \neq \emptyset \Leftrightarrow \Omega = \emptyset \end{aligned} \tag{4.8}$$

$\diamondsuit$

The stronger version of this theorem, assuming convexity, is the conjunction of the Lemmas 4.2.1, 4.2.3 and Corollary 4.2.5:

Theorem 4.2.7 *Consider the optimization problem defined by (4.1), and assume that the functions $f(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^p$ and $g(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^m$ are convex. Then:*

$$\begin{aligned} &\mathcal{A} = \Omega \neq \emptyset \Leftrightarrow \Lambda = \emptyset \\ &\mathcal{A} = \Lambda \neq \emptyset \Leftrightarrow \Omega = \emptyset \end{aligned} \tag{4.9}$$

$\diamondsuit$

As a consequence of theorem 4.2.7, valid for the convex case, the search for a point inside the feasible set Ω of problem (4.1) can be stated as a multiobjective optimization on the auxiliary problem (4.6), which performs a search for Pareto-critical points $x_a \in \mathcal{A}$. Once *any* point $x_a \in \mathcal{A}$ has been found, there are two possibilities: (i) $x_a \in \Omega$, or (ii) $x_a \in \Lambda$. At this point, either x_a is feasible, or a certificate of infeasibility has been found.

Chapter 5

Results and Discussions

This chapter deals with the numerical implementation of the proposed infeasibility certificate.

5.1 Verification of INF Condition

In this work, we applied a scalarization strategy for solving the auxiliary problem which is related to the infeasibility certificate. A mathematical programming method is employed for finding a Pareto optimal solution.

5.1.1 Noise-free problems

First, consider the usual situation of noise-free problems (problems for which it is possible to obtain gradient information), which is the traditional setting of optimization problems. This means that it will be possible to calculate the function derivatives, which will allow the definition of gradient-based tests. In order to implement the search for a point x_a inside the set $\mathcal{A}$ leading either to a feasible point or to a certificate of infeasibility, it is enough to find a single Pareto-optimal solution to the auxiliary problem. A scalarized version of the multi-objective auxiliary problem (4.6) is stated as:

$$\min_{x} \max_{i} w_i \ \hat{g}_i(x) \tag{5.1}$$

Each optimal solution of (5.1) is a Pareto-optimal solution of (4.6). For each Pareto optimal point $\bar{x}$ there exists a weight vector w_i such that $\bar{x}$ is the optimum solution of (5.1). The meaning of the solutions of problem (5.1), in its non-convex flavor, is stated as:

Theorem 5.1.1 *Consider a solution x_a of problem (5.1). Then:*

$$x_a \notin \Omega \Rightarrow \Omega = \emptyset \tag{5.2}$$

♢

Under the assumption of convexity:

Theorem 5.1.2 *Consider a solution x_a of problem (5.1). Then:*

$$x_a \in \Omega \Leftrightarrow \Omega \neq \emptyset \tag{5.3}$$

♢

To check the feasibility or infeasibility of the original problem on the basis of this solution, it is necessary to check the conditions described in section 4.2. If the obtained solution $\bar{x}$ is feasible, then the feasibility problem is solved. Otherwise, if $\bar{x}$ solution is infeasible, it is necessary to check the INF conditions numerically on $\bar{x}$. Define:

$$\hat{G}(\bar{x}) = \left[\begin{array}{cccc} \nabla\hat{g}_1 & \nabla\hat{g}_2 & \dots & \nabla\hat{g}_k \end{array} \right] \tag{5.4}$$

in which k denotes the number of violated constraints in point $\bar{x}$, which were assumed to occupy the first k indices of the constraint set, w.l.g.. Suppose $\bar{x} \in \Lambda$, then it comes from equation (4.4):

$$\hat{G}(\bar{x})\ \mu = 0 \tag{5.5}$$

Suppose w.l.g. that $\nabla\hat{g}_1 \neq 0$, and assume that $\mu_1 = 1$. Then:

$$\left[\begin{array}{ccc} \nabla\hat{g}_2 & \dots & \nabla\hat{g}_k \end{array} \right] \left[\begin{array}{c} \mu_2 \\ \vdots \\ \mu_k \end{array} \right] = -\nabla\hat{g}_1 \tag{5.6}$$

Now, assume provisionally that matrix $H = \left[\begin{array}{ccc} \nabla\hat{g}_2 & \dots & \nabla\hat{g}_k \end{array} \right]$ has full column rank. The case in which $n = k - 1$ leads to a straightforward solution:

$$\left[\begin{array}{c} \mu_2 \\ \vdots \\ \mu_k \end{array} \right] = -\left[\begin{array}{ccc} \nabla\hat{g}_2 & \dots & \nabla\hat{g}_k \end{array} \right]^{-1} \nabla\hat{g}_1 \tag{5.7}$$

In the case of $n > k-1$ a least-squares solution can be expressed as:

$$\begin{bmatrix} \mu_2 \\ \vdots \\ \mu_k \end{bmatrix} = -\left(\begin{bmatrix} \nabla\hat{g}_2 & \dots & \nabla\hat{g}_k \end{bmatrix}^T \begin{bmatrix} \nabla\hat{g}_2 & \dots & \nabla\hat{g}_k \end{bmatrix} \right)^{-1} \begin{bmatrix} \nabla\hat{g}_2 & \dots & \nabla\hat{g}_k \end{bmatrix}^T \nabla\hat{g}_1 \tag{5.8}$$

In both cases, if all the components of μ given by equations (5.7) and (5.8) are positive, then the condition (INF) will hold and the problem will be strictly infeasible. On the other hand, if there is at least one negative multiplier, it is not possible to declare that the original problem is infeasible. In order to remove the assumption of full column rank of H, consider the submatrices composed by subsets of the columns of H, such that they attain the largest possible column rank. All such matrices can be evaluated, using the suitable criterion, either (5.7) or (5.8). If the INF condition is verified for any such a matrix, then the problem can be declared to be infeasible.

Now, in the case of $n < k-1$,

$$\begin{bmatrix} \nabla\hat{g}_1 & \nabla\hat{g}_2 & \dots & \nabla\hat{g}_k \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix} = 0 \tag{5.9}$$

which means that $\mu \in \mathcal{N}(\hat{G})$, in which $\mathcal{N}(\cdot)$ stands for the null space of the argument matrix. Let $\mathcal{K}^+$ denote the cone of the positive octant of suitable dimension. In this case, if

$$\mathcal{N}(\hat{G}) \cap \mathcal{K}^+ \neq 0 \tag{5.10}$$

then the condition (INF) will hold, and the problem will be strictly infeasible.

5.1.2 Noisy problems

Now, consider the situation of noisy problems, in which the function values are supposed to be corrupted by some noise. In this case, the computation of derivatives should be avoided, since the noise would be amplified in that computation. In this situation, the proposed indicator is still suitable for providing an infeasibility certificate. The kind of evidence to be employed, in this case, relies on the description of the Pareto-front of the auxiliary problem (4.6) (the image of the set $\mathcal{A}$, in the space of constraint function values), denoted by $g(\mathcal{A})$. The following theorem states the facts that support the proposed procedure.

Theorem 5.1.3 *Let $y = g(x)$ for some $x \in \mathcal{A}$. The following statements hold:*

(i) $y \prec 0 \;\Rightarrow\; x \in \Omega$

(ii) $0 \prec y \;\Rightarrow\; \Omega = \emptyset$

$\diamondsuit$

The reasoning that was implicit, in the former subsection, was that: (i) An optimization algorithm should be executed in order to solve problem (5.1). (ii) After the solution is found, if it is not feasible, a criticality test is performed, in order to ensure that it is indeed a solution. Once the criticality test returns a positive answer, the (INF) condition is considered to hold, and the infeasibility of the problem (4.1) is declared.

Now, in order to replace the criticality test as the additional evidence that supports the declaration of infeasibility of problem (4.1), it is adopted a *regularity test*.

Definition 5.1 *Let $\mathcal{F}$ denote the intersection of the Pareto-front of 4.6 with the positive orthant of the space of constraint function values:*

$$\mathcal{F} = g(\mathcal{A}) \cap \mathcal{K}^{+} \tag{5.11}$$

Let $C(\mathcal{F})$ denote the convex hull of $\mathcal{F}$, and let $\partial C(\mathcal{F})$ denote the boundary of the set $C(\mathcal{F})$. The surface $\mathcal{F}$ is said to be regular *if:*

(i) *It has no holes.*

(ii) *Every point $y \in \mathcal{F}$ belongs to $\partial C(\mathcal{F})$.*

$\diamondsuit$

Of course, the regularity is not a necessary attribute of $\mathcal{F}$. However, the features (i) and (ii) are rather usual in the context of constraint functions. Therefore, the *regularity* provides support to the declaration of infeasibility of problem (4.1).

The *regularity test* procedure is stated as:

(a) Find a description of the Pareto-front of the auxiliary problem (4.6) in the positive orthant.

(b) If such a description is consistent with a *regular* Pareto-front surface, then declare that the (4.1) is infeasible.

Step (a) should be performed using a derivative-free multiobjective optimization algorithm. For this purpose, the evolutionary multiobjective algorithms are well-suited. For instance, the NSGA-II algorithm (Deb et al., 2002) and the SPEA-2 algorithm (Zitzler et al., 2002) both provide suitable mechanisms for producing a uniform sampling of $\mathcal{F}$. However, for problems with a large number of constraint functions, a better choice would be a decomposition-based evolutionary algorithm, such as MOEA-D (Zhang and Li, 2007), due to the many-objective degradation effect on algorithms that use Pareto-based selection (such as NSGA-II or SPEA-2).

Step (b) can be performed according to the following steps: (i) The existence of holes is tested with the use of a goal attainment scalarization, defining a test direction pointing to the region in which there could be a hole. If a new point is found by this procedure, there is no hole, and the point is added to the Pareto-set sampling. Otherwise, there is a hole, and the surface is not regular. (ii) The condition that every point belonging to the Pareto-front sample also belongs to the boundary of its convex hull is tested by verifying if every point has a hyperplane that separates it from the other ones.

If step (b) finishes with a positive answer about the regularity of the $\mathcal{F}$ surface, the problem (4.1) is declared infeasible.

5.2 Illustrative Examples

In order to perform computational tests involving the proposed infeasibility condition, a simple computational framework has been defined. Since problem (5.1) is a single objective problem involving non-differentiable functions, the Non-linear Nelder-Mead Simplex Method Nelder and Mead (1965) can be used to solve it, as discussed in chapter two. For this purpose, a random starting point x_0 is considered. The non-linear programming problem results into a solution $\bar{x}$. Finally, the obtained solution $\bar{x}$ is checked for the infeasibility certificate.

Here, the proposed algorithm is explained graphically with the help of some simple examples. Each example show that the problem is feasible or strictly infeasible at the solution point $\bar{x}$ and consequently the obtained solutions belong to the feasible set Ω or to the set of points Λ respectively.

Example 5.1 *Consider the following optimization problem with three constraints:*

$$\min_x f(x)$$

subject to:

$$g_1(x) = -x_1 + {x_2}^2 \leq 0$$

$$g_2(x) = {x_1}^2 + 3{x_2}^2 - 4 \leq 0$$

$$g_3(x) = {x_1}^2 + (x_2 - 5)^2 - 4 \leq 0$$

In this example, f(x) could be a single or multiobjective function.
For this, we can define the vector function like (4.5) as follows;

$$\hat{g}_i(x) = \begin{cases} 0 \ , & \forall \ \ x \mid g_i(x) \leq 0 \\ g_i(x) \ , & \forall \ \ x \mid g_i(x) > 0 \end{cases}$$
$$i = 1, \ldots, 3$$

From the above vector function, the following auxiliary problem is defined as:

$$\min_x \hat{g}(x)$$

The auxiliary problem for the above example is given by

$$\min_x \hat{g}(x) = \min \begin{pmatrix} \hat{g}_1(x) \\ \hat{g}_2(x) \\ \hat{g}_3(x) \end{pmatrix}$$

Using minmax formulation we have

$$\min_x \max_i w_i \hat{g}_i(x)$$

Applying Nelder Mead Simplex method to solve this minmax problem, we get different solutions depending on the values of weights w_i. In the following figure, the dense area shows the INF solutions of this problem.

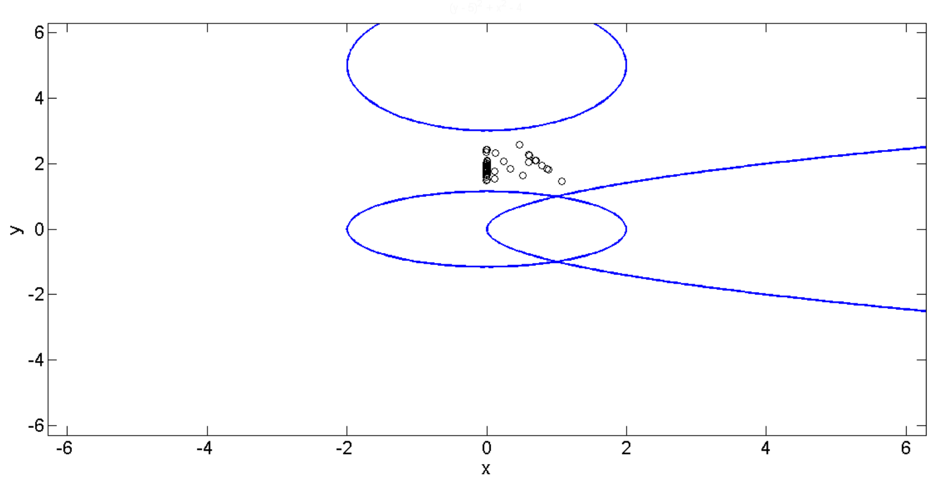

Figure 5.1: Pareto critical solutions of the problem

This problem is strictly infeasible with trade off solutions near to the boundaries of the constraints.

Example 5.2 *Consider the same example with the first two constraints:*

$$\min_x f(x))$$

Subject to:

$$g_1(x) = -x_1 + {x_2}^2 \leq 0$$

$$g_2(x) = {x_1}^2 + 3{x_2}^2 - 4 \leq 0$$

Applying the same procedure as in the previous example, we have the following figure.

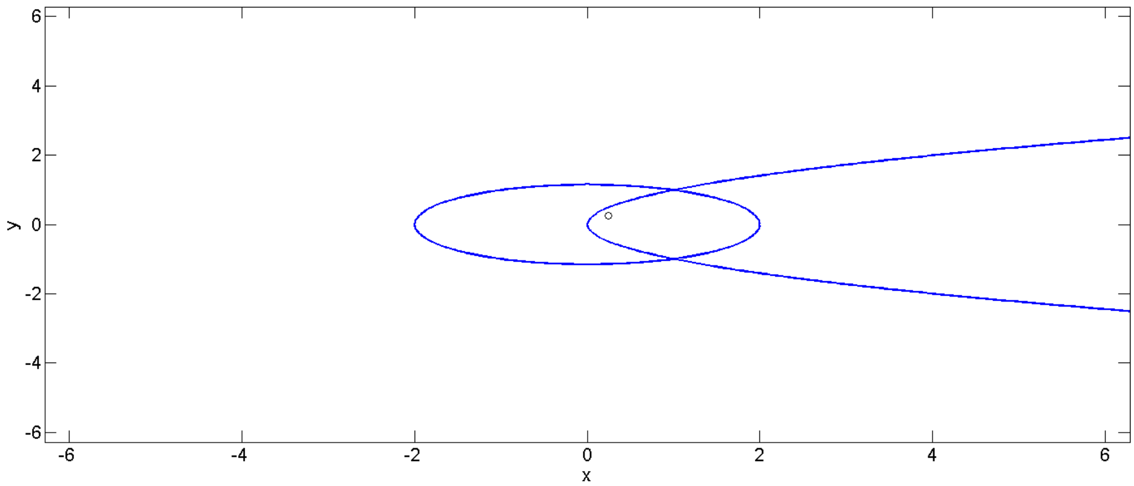

Figure 5.2: Shows a single feasible solution of the problem

Figure (5.2), shows that the problem is feasible and we end up with one feasible solution. Here we can write $\mathcal{A} = \Omega \neq \emptyset$ and $\Lambda = \emptyset$

Example 5.3 *Consider the same example with the last two constraints.*

$$\min_x f(x)$$

subject to:

$$g_2(x) = {x_1}^2 + 3{x_2}^2 - 4 \leq 0$$

$$g_3(x) = {x_1}^2 + (x_2 - 5)^2 - 4 \leq 0$$

Figure 5.3 shows a beam of INF solutions, and consequently the problem is strictly infeasible again.

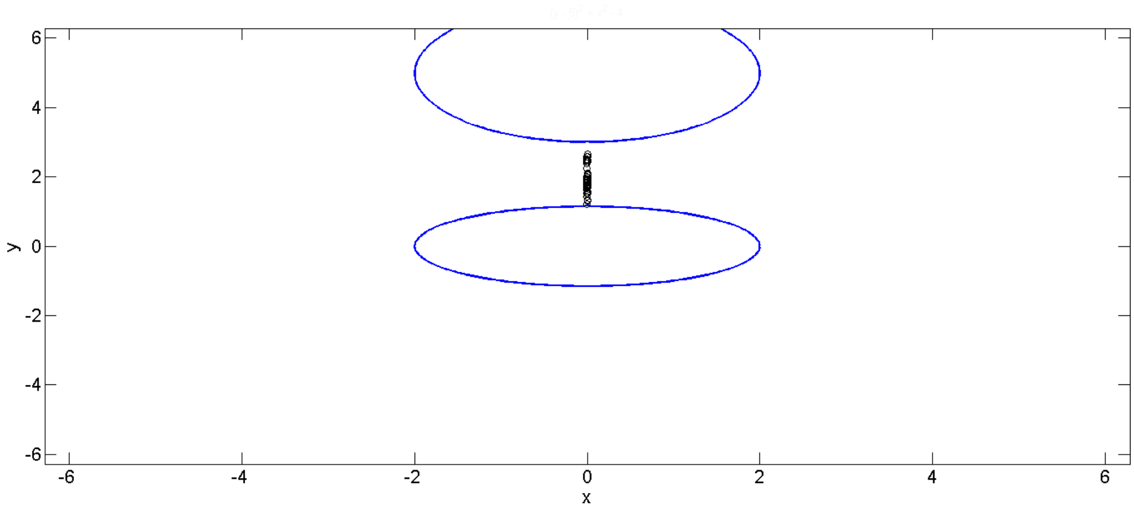

Figure 5.3: The dense area in the figure shows a beam of INF solutions.

Example 5.4 *Consider the following optimization problem taken from (Byrd et al., 2010) with four constraints:*

$$\min_x f(x)$$

subject to:

$$g_1(x) = {x_1}^2 + x_2 + 1 \leq 0$$
$$g_2(x) = {x_1}^2 + x_2 + 1 \leq 0$$
$$g_3(x) = -{x_1}^2 + {x_2}^2 + 1 \leq 0$$
$$g_4(x) = x_1 + {x_2}^2 + 1 \leq 0$$

Figure 5.4 shows that the problem is strictly infeasible with four objective functions.

5.3 Algorithms

The whole computational procedure is summarized in the following algorithms.

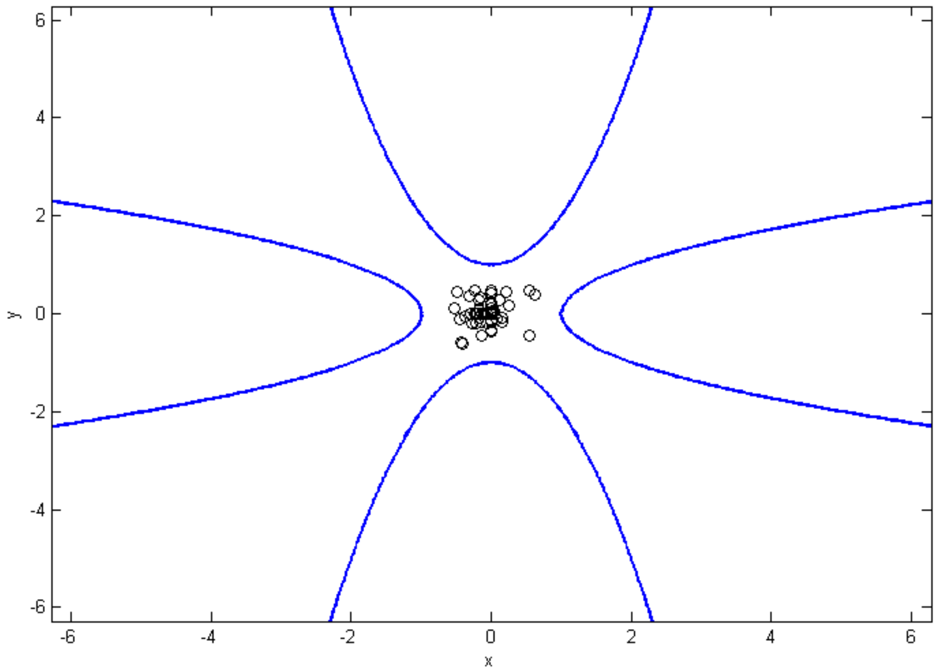

Figure 5.4: Infeasible (INF) solutions with four objective functions.

Input: Function $g(.)$, $iterMax$
Output: Certificate c

1 $iter \leftarrow 0$
2 **while** *Certificate* c *does not give a conclusion and* $iter <= iterMax$ **do**
3   Initialize weights w_i for the auxiliary problem
4   $x0$ is initialized at random
5   $x \leftarrow solver(auxObjFun(x, w), x0)$
6   $c \leftarrow checkingCertificate(x)$
7   $iter \leftarrow iter + 1$;
8 **end**
9 **return** *c* ;

Algorithm 1: Main Loop

In Line 5 of Algorithm 1, a solution x is obtained by the use of solver Nelder Mead Simplex (Nelder and Mead, 1965) from a random starting point x_0, with random weights w composing the auxiliary objective function. The obtained solution x is checked for certificate (line 6 of algorithm 1).

By arbitrarily choosing $\mu_1 = 1$, a linear system is solved for the Lagrange multipliers. If all the Lagrange multipliers are greater than 0 then the solution x will result into an infeasible certificate.

5.4 Performance on Test Problems

The proposed algorithm was implemented in MATLAB (2013). Computational experiments were carried out on a Pentium Core 2 Quad (Q6600) computer with 8GB of RAM and operating system of Windows 7. The algorithm is tested on five examples as introduced in Section 5.2 on a batch of different number of executions for each instance. The initial points were selected at random.

The computational results of batches of 30 executions are given in Tables 5.1, 5.2 and 5.3 for four different scenarios, with maximum number of iterations ($iterMax$) equal to 30, 50, 200 and 500. In the tables of computational results, lines "FC" and "IC" indicate the number of feasibility and infeasibility certificates obtained for each problem. Line "NC" indicates the number of iterations in which any certificate is obtained. Line "$\#\bar{x}$" indicates the average number of solutions obtained for achieving the feasibility or infeasibility certificates for each example.

Example 5.5 *Consider the following example with 6 constraints.*

$$\min_x f(x)$$

Subject to:

$$g_1(x) = x_1^2 - x_2 + 1 \leq 0$$
$$g_2(x) = x_1^2 - x_2^4 \leq 0$$
$$g_3(x) = 2x_1^2 x_2^{\exp x_1} - x_2^4 \leq 0$$
$$g_4(x) = -x_1 + x_2^2 \leq 0$$
$$g_5(x) = x_1^2 + (x_2 - 5)^2 - 4 \leq 0$$
$$g_6(x) = x_1^2 + 3x_2^2 - 4 \leq 0$$

Example 5.6 *Consider the following example with 7 constraints.*

$$\min_x f(x)$$
Subject to:
$$g_1(x) = x_4^{x_3} e^{x_4} \leq 0$$
$$g_2(x) = +x_1^2 + x_2 + x_3^5 + x_4^2 + 1 \leq 0$$
$$g_3(x) = x_1^2 - x_2^4 \leq 0$$
$$g_4(x) = 2x_1^2 x_2^{\exp x_4} - x_2^4 \leq 0$$
$$g_5(x) = -x_1 + x_2^2 \leq 0$$
$$g_6(x) = x_1^2 + (x_3 - 5)^2 - 4 \leq 0$$
$$g_7(x) = x_1^2 + 3x_2^2 - 4 \leq 0$$

Finally, three more examples are included in order to further validate the proposed algorithm. These quadratic examples are taken from Byrd et al. (2010) having 2, 4 and 5 constraints respectively, and so-known as example 1, 3 and 5, and called in this thesis as Byrd 1, 2 and 3 respectively.

Table 5.1: Computational results for maximum 30 iterations

	Ex. 5	Ex. 6	Byrd 1	Byrd 2	Byrd 3
FC	0	0	0	0	0
IC	0	6	30	28	30
NC	30	24	0	2	0
$\#\bar{x}$	3,00	2,90	1,00	1,86	1,00

Table 5.2: Computational results for maximum 50 iterations

	Ex. 5	Ex. 6	Byrd 1	Byrd 2	Byrd 3
FC	0	0	0	0	0
IC	15	29	30	30	30
NC	15	1	0	0	0
$\#\bar{x}$	36,93	13,83	1,00	2,03	1,00

Analyzing tables 5.1, 5.2 and 5.3 it can be seen that the proposed algorithm is capable to provide feasibility or infeasibility certificates for the given maximum numbers of iterations. When the number of iteration is 30 i.e. $iterMax = 30$, the algorithm produces 100% of accuracy for instances Byrd 1 and Byrd 3. Considering 50 iterations, the algorithm results into 50% of its executions in infeasibility for Ex.5 and this percentage is 99% in case of example Ex.6. Table 5.3 is considered for 500

Table 5.3: Computational results for maximum 200 and 500 iterations

	Ex. 5	Ex. 6	Ex. 5
FC	0	0	0
IC	26	30	30
NC	4	0	0
$\#\bar{x}$	88.46	15.36	74.73
$maxIter$	200	200	500

iterations, where EX.5 shows different results in case of 200 and 500 iterations. In the case of 500 iterations, EX.5 produces 100% infeasibility.

Chapter 6

Conclusion

In this thesis, a new infeasibility certificate for non-linear optimization problems on the basis of Pareto-criticality condition of an auxiliary multiobjective optimization problem was developed. The main result presented here is a new necessary condition (that becomes necessary and sufficient under convexity assumptions) for the infeasibility of finite-dimensional optimization problems, which is related to the Kuhn-Tucker conditions for efficiency in multi-objective problems. By defining a suitable auxiliary vector function, the search for the feasible set can be stated as the search for a point that is Pareto-critical for such an auxiliary problem. Once a Pareto-critical point w.r.t. the auxiliary problem is found, it is either a feasible solution of the original problem or brings a certificate of infeasibility (which is globally valid for convex problems). Differently from other existing certificates of infeasibility, the proposed one relies on primal variables only.

Another important difference of the proposed methodology is that it admits a modified version that does not rely on gradient information. In this case, the criticality test is replaced by a heuristic regularity test.

The performance of the proposed methodology was tested on some functions, and it delivered promising results.

Bibliography

Andersen, E. D. (2000). On primal and dual infeasibility certificates in a homogeneous model for convex optimization. *SIAM Journal on Optimization*, 11(2):380–388.

Andersen, E. D. (2001). Certificates of primal or dual infeasibility in linear programming. *Computational Optimization and Applications*, 20(2):171–183.

Andersen, E. D. and Andersen, K. D. (2000). The MOSEK interior point optimizer for linear programming: an implementation of the homogeneous algorithm. In *High Performance Optimization*, pages 197–232. Springer.

Andersen, E. D. and Ye, Y. (1997). On a homogeneous algorithm for a monotone complementarity problem with nonlinear equality constraints. In *Complementarity and Variational Problems: State of the Art*, pages 1–11. SIAM Philadelphia, PA.

Andersen, E. D. and Ye, Y. (1999). On a homogeneous algorithm for the monotone complementarity problem. *Mathematical Programming*, 84(2):375–399.

Andersen, K., Louveaux, Q., and Weismantel, R. (2008). Certificates of linear mixed integer infeasibility. *Operations Research Letters*, 36(6):734–738.

Andreani, R., Birgin, E. G., Martínez, J. M., and Schuverdt, M. L. (2007). On augmented Lagrangian methods with general lower-level constraints. *SIAM Journal on Optimization*, 18(4):1286–1309.

Ben-Tal, A. and Nemirovskiĭ, A. S. (2001). *Lectures on Modern Convex Optimization: analysis, algorithms, and engineering applications*. SIAM.

Benson, H. Y., Vanderbei, R. J., and Shanno, D. F. (2002). Interior-point methods for nonconvex nonlinear programming: Filter methods and merit functions. *Computational Optimization and Applications*, 23(2):257–272.

Byrd, R. H., Curtis, F. E., and Nocedal, J. (2010). Infeasibility detection and SQP methods for nonlinear optimization. *SIAM Journal on Optimization*, 20(5):2281–2299.

Byrd, R. H., Gould, N. I., Nocedal, J., and Waltz, R. A. (2003). An algorithm for nonlinear optimization using linear programming and equality constrained subproblems. *Mathematical Programming*, 100(1):27–48.

Byrd, R. H., Lopez-Calva, G., and Nocedal, J. (2008a). A line search penalty method for nonlinear optimization. Technical report, Technical Report 08/05, Optimization Technology Center, Northwestern University.

Byrd, R. H., Nocedal, J., and Waltz, R. A. (2006). Knitro: An integrated package for nonlinear optimization. In *Large-scale nonlinear optimization*, pages 35–59. Springer.

Byrd, R. H., Nocedal, J., and Waltz, R. A. (2008b). Steering exact penalty methods for nonlinear programming. *Optimization Methods and Software*, 23(2):197–213.

Chankong, V. and Haimes, Y. (1983). *Multiobjective Decision Making: theory and methodology*. Noth-Holland.

Deb, K. (2001). *Multi-objective Optimization Using Evolutionary Algorithms*, volume 16. John Wiley & Sons.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *Evolutionary Computation, IEEE Transactions on*, 6(2):182–197.

Durand, N. and Alliot, J.-M. (1999). A combined Nelder-Mead simplex and genetic algorithm. In *GECCO'99: Proceedings of the Genetic and Evolutionary Computation Conference*, pages 921–928.

Erling D. Andersen, M. A. (2011). How to use Farkas' lemma to say something important about infeasible linear problems.

Fletcher, R. (1987). *Practical Methods of Optimization*. Wiley.

Fletcher, R., Gould, N. I., Leyffer, S., Toint, P. L., and Wächter, A. (2002). Global convergence of a trust-region SQP-filter algorithm for general nonlinear programming. *SIAM Journal on Optimization*, 13(3):635–659.

Geoffrion, A. M. (1968). Proper efficiency and the theory of vector maximization. *Journal of Mathematical Analysis and Applications*, 22(3):618–630.

Gill, P. E., Murray, W., and Saunders, M. A. (2002). SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Journal on Optimization*, 12(4):979–1006.

Greenberg, H. (1983). A functional description of ANALYZE: A computer-assisted analysis system for linear programming models. *ACM Transactions on Mathematical Software (TOMS)*, 9(1):18–56.

Han, S.-P. and Mangasarian, O. L. (1979). Exact penalty functions in nonlinear programming. *Mathematical programming*, 17(1):251–269.

Luc, D. T. (1988). *Theory of Vector Optimization*. Springer-Verlag, Berlin.

Martínez, J. M. and da Fonseca Prudente, L. (2012). Handling infeasibility in a large-scale nonlinear optimization algorithm. *Numerical Algorithms*, 60(2):263–277.

Marusciac, I. (1989). On Fritz-John-type optimality criterion in multi-objective optimization. *Revue d 'Analyse Numerique et de Theorie de l'Approximation*, 11:109–114.

MATLAB (2013). *version 7.10.0 (R2013a)*. The MathWorks Inc., Natick, Massachusetts.

Miettinen, K. (1999). *Nonlinear Multiobjective Optimization*. Springer.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4):308–313.

Nesterov, Y., Todd, M., and Ye, Y. (1999). Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems. *Mathematical Programming*, 84(2):227–267.

Nocedal, J., Öztoprak, F., and Waltz, R. A. (2014). An interior point method for nonlinear programming with infeasibility detection capabilities. *Optimization Methods and Software*, 29(4):837–854.

Nocedal, J. and Wright, S. J. (2006). *Penalty and Augmented Lagrangian Methods*. Springer.

Rockafellar, R. T. (1974). Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM Journal on Control*, 12(2):268–285.

Roos, C., Terlaky, T., and Vial, J.-P. (1997). *Theory and algorithms for linear optimization: an interior point approach*. Wiley Chichester.

Spendley, W., Hext, G. R., and Himsworth, F. (1962). Sequential application of simplex designs in optimisation and evolutionary operation. *Technometrics*, 4(4):441–461.

Wächter, A. and Biegler, L. T. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57.

Wright, S. J. (1997). *Primal-Dual Interior-Point Methods*. SIAM.

Zhang, Q. and Li, H. (2007). MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *Evolutionary Computation, IEEE Transactions on*, 11(6):712–731.

Zitzler, E., Laumanns, M., and Thiele, L. (2002). SPEA 2: Improving the Strenght Pareto Evolutionary Algorithms. In *EUROGEN 2001. Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems*, pages 95–100.