

PARAMETRIC NONLINEAR OPTIMIZATION PROBLEMS



ADDIS ABABA UNIVERSITY
COLLEGE OF COMPUTATIONAL AND NATURAL SCIENCES
DEPARTMENT OF MATHEMATICS

A project Submitted in partial fulfillment of the requirements of the degree of master of science in mathematics

By: Dejene Gizaw
Advisor: Semu Mitiku(PhD)

June, 2014
Addis Ababa, Ethiopia

ADDIS ABABA UNIVERSITY
DEPARTMENT OF MATHEMATICS

The undersigned hereby certify that they have read and recommend to the department of mathematics for acceptance of this project entitled “**PARAMETRIC NONLINEAR OPTIMIZATION PROBLEMS**” by **Dejene Gizaw** in partial fulfillment of the requirements for the degree of Master of Science in mathematics.

Advisor: Dr. Semu Mitiku Kassa

Signature:-----

Date-----

Examiner 1: Dr. _____

Signature:-----

Date-----

Examiner 2: Dr. _____

Signature:-----

Date-----

Contents

Abstract	i
Acknowledgement	ii
1 Introduction	1
2 Optimality Conditions for Nonlinear Programs	3
2.1 Existence of a Solution	4
2.2 Conical Approximation of Sets	5
2.3 First Order Necessary Optimality Conditions of Firtz-John Type	6
2.4 Constraint Qualifications	7
2.5 Second Order Sufficient Optimality Conditions	10
3 Parametric Nonlinear Optimization Problems	12
3.1 Stability of the Optimal Value and Optimal Solutions	13
3.1.1 Directional differentiability of optimal value functions	15
3.1.2 Implicit Function Theorem	16
3.2 Sensitivity Analysis for Parametric Nonlinear Optimization Problems	20
3.2.1 Optimality Conditions for Nonlinear Optimization problems	20
3.2.2 Post-optimal Calculation of Lagrange Multipliers	21
3.2.3 Second Order Sufficient Optimality Conditions	21
3.2.4 First Order Sensitivity Analysis of Optimal Solution	22
3.2.5 First and Second-Order Parameter Derivatives of the Optimal Value Function	26
3.2.6 Linear Perturbations/ RHS perturbations	27
4 Perturbed Optimization Problem in Real-Time	29
4.1 Approximation of Perturbed Solutions	29
4.2 Prediction of Sensitivity Domain	31
4.3 Handling Larger Perturbations	31
4.4 Solution Method for Parametric Quadratic Optimization	33
4.4.1 Background on (mpQP)	33
4.4.2 Characterization of the Partition	36
Bibliography	41

Abstract

In this project work a smooth parametric nonlinear optimization problems subject to equality and inequality constraints are considered. Emphasis is given on those conditions under which the optimal solutions are differentiable functions of parameters. In theory these conditions are related to regularity conditions and to second order sufficient conditions. The interest in conditions for solution differentiability originates in the real-time computation (approximation) of perturbed optimal solutions under parameter changes through first order Taylor expansions. We study the explicit formulae and methods for computing for the sensitivity derivatives of the solution vector and the associated multipliers with respect to parameters. We discuss post-optimal evaluations of sensitivity derivatives and their numerical implementation. The purpose of this work is to describe the application of sensitivity analysis to approximate perturbed solutions in view of optimality and admissibility by using Taylor expansion and extend the sensitivity theory to provide a complete map of the optimal solution in the space of varying parameters for the case of multi-parametric quadratic programming(mpQP).

Keywords. Nonlinear Optimization, Parametric nonlinear programming, Parametric quadratic programming

Acknowledgement

I would like to express my sincere gratitude to **Dr. Semu Mitiku**, my project advisor for his steady efforts in providing relevant advice, guidance and material supply through the course of my work.

Next I would like to appreciate and record my great full thanks to my families for giving me encouragement and continued support.

Chapter 1

Introduction

In this project work we are concerned mainly with a parametric nonlinear programming (NLP) problem of the form

$$\begin{cases} \min_x & f(x, p), \quad x \in \mathbb{R}^{n_x}, p \in \mathbb{R}^{n_p} \\ \text{subject to} & g_i(x, p) \leq 0, \quad i = 1, \dots, g_{n_g} \\ & h_j(x, p) = 0, \quad j = 1, \dots, h_{n_h} \end{cases}$$

where the components of the vector p can be seen as parameters or perturbations. Herein $f, g_1, \dots, g_{n_g}, h_1, \dots, h_{n_h} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \rightarrow \mathbb{R}$ are sufficiently smooth functions. For each value of p , the set of feasible solutions is

$$\Phi(p) := \{x \in \mathbb{R}^{n_x} | g_i(x, p) \leq 0, i = 1, \dots, g_{n_g}, h_j(x, p) = 0, j = 1, \dots, h_{n_h}\}$$

The optimal value of the program for each value of p is

$$v(p) := \inf_{x \in \Phi(p)} f(x, p)$$

The set of optimal solutions is

$$S(p) := \{\hat{x} \in \Phi(p) | f(\hat{x}, p) = v(p)\}.$$

Let p_0 denote a fixed *nominal parameter*. Our main interest will be the behavior of the optimal solutions and optimal value function as functions of p in a neighborhood of the nominal parameter p_0 . We shall also be interested in the application of sensitivity analysis to approximate perturbed optimal solutions due to small changes in the fixed parameters of the formulation.

The investigation of the local behavior of an optimal solution for a parametric optimization problem is one of the central topics in sensitivity and stability analysis [1]. Especially quantitative results on the dependency of a solution on the problem parameters (input data) can be applied to construct algorithms computing local(global) optimal solutions or to give an approximation of the set of efficient points in multi-objective optimization problems. Literature on optimization uses two different notions of sensitivity analysis. One conception of sensitivity analysis appears in the calculation of the objective function and constraint partial derivatives for determining search directions and optimality conditions. The other conception is related *parameter sensitivity analysis* studies estimating the changes in the objective

function and the optimal solution vector due to small changes in the fixed parameters of the formulation. The *stability analysis* of optimization problems is concerned with the Lipschitz continuity of optimal solutions with respect to the parameters [4],[7]. Basic results for sensitivity analysis of parametric nonlinear programming problems (NLP) give emphasis on those conditions that ensure the differentiability of the optimal solution vector with respect to the parameter involved in the problem. In this project work, more emphasis is given on those conditions which ensure that the optimal solution is differentiable with respect to the design parameters.

The fundamental results in this area is derived through application of the implicit function theorem (IFT) to the Karush-Kuhn-Tucker (KKT) conditions of the parametric (NLP). As shown in Fiacco [1] and Robinson [3], sensitivities can be obtained from a solution with suitable regularity conditions, only by solving a linearization of the KKT conditions. This approach also provide explicit formulae for the *parameter sensitivity derivatives* of optimal solution vector, the associated Lagrange multiplier and optimal value function. Conceptually, these formulae are tailored to solution algorithm calculations in [1]. However, there are numerical obstacles that prevent these expressions from being a direct byproduct of the current solution algorithms. For example, in [10] it was studied that the explicit formulae for sensitivity derivatives computation through SQP-methods(Sequential Quadratic Programming) can't be tied directly to iterative calculations in SQP-methods since the employed low rank updates of the Hessian usually do not converge to the exact Hessian. This means that the information gained in the solution process can't be used directly for an accurate computation of sensitivity derivatives. As a result post-optimal analysis of sensitivity derivatives is proposed.

In the present work, we take the idea of post-optimal evaluation sensitivity differential, the numerical aspects of (i) checking the second order sufficient conditions which constitute the theoretical basis of sensitivity analysis and (ii) computing the Hessian with high accuracy and using them to approximate the perturbed optimal solutions by their first order Taylor expansion with respect to the parameters.

The organization of the paper is the following. First in chapter 2 we revise some basic optimality conditions of nonlinear programming problems. We discuss the necessary and sufficient optimality conditions and the regularity conditions. In chapter 3 we study the stability and differential properties of optimal value function, optimal solutions and multipliers of nonlinear programming problems when the data of the problems are subjected to small perturbations. In the last chapter using mathematical background of the methodology, we discuss approximations of perturbed optimal solutions and finalizes with a solution methods for parametric quadratic optimization problems. The convexity or continuity properties of the solution are also proven.

Chapter 2

Optimality Conditions for Nonlinear Programs

Throughout this chapter we consider a nonlinear programming problem of determining a local solution x of

$$\left\{ \begin{array}{ll} \min_x & f(x) \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, \dots, n_g \\ & h_j(x) = 0, \quad j = 1, \dots, n_h \\ & x \in S \end{array} \right. \quad (\text{NLP})$$

where $f : \mathbb{R}^{n_x} \rightarrow \mathbb{R}$, $g = (g_1, \dots, g_{n_g})^\top : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_g}$, $h = (h_1, \dots, h_{n_h})^\top : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_h}$ be continuously differentiable functions and $S \subseteq \mathbb{R}^{n_x}$ a closed, convex set. Henceforth the symbol, $(\cdot)^\top$ denotes the transpose. Without loss of generality we will exclusively consider minimization problems, since maximization problems always can be transformed in to equivalent minimization problems.

The admissible set is given by

$$\Phi := \{x \in S : g_i(x) \leq 0, \quad i = 1, \dots, n_g, \quad h_j(x) = 0, \quad j = 1, \dots, n_h\}$$

The set

$$A(x) := \{i : g_i(x) = 0, \quad i = 1, \dots, n_g\}$$

is called the *index set of active inequality constraints* at $x \in \Phi$

An admissible $\hat{x} \in \Phi(p)$ is called a *local minimum* of the problem (NLP) if a neighbourhood U of \hat{x} exists, such that

$$f(\hat{x}) \leq f(x) \quad \forall x \in \Phi \cap U$$

A point $\hat{x} \in \Phi$ is called a *strong (strict) local minimum* of the problem (NLP) if a neighbourhood U of \hat{x} exists, such that

$$f(\hat{x}) < f(x) \quad \forall x \in \Phi \cap U, \quad (x \neq \hat{x})$$

A point $\hat{x} \in \Phi$ is called a *global minimum* of the problem (NLP) if

$$f(\hat{x}) \leq f(x) \quad \forall x \in \Phi$$

The problem (NLP) is called *convex* if f is convex and the constraint set, Φ is convex.

Claim: In convex optimization problems every local minimum is also a global one.

Proof. Assume that $\hat{x} \in \Phi$ is a local minimum point of a convex function $f : \Phi \rightarrow \mathbb{R}$. Then there exists $\epsilon > 0$ and a ball $U_\epsilon(\hat{x})$ so that \hat{x} is a global minimum of f on $\Phi \cap U_\epsilon(\hat{x})$. Now we consider an arbitrary $x \in \Phi$ with $x \notin U_\epsilon(\hat{x})$. Then it is $\|x - \hat{x}\| < \epsilon$. For $\alpha := \frac{\epsilon}{\|x - \hat{x}\|} \in (0, 1)$ we obtain $x_\alpha := \alpha x + (1 - \alpha)\hat{x} \in \Phi$ and $\|x_\alpha - \hat{x}\| = \|\alpha x + (1 - \alpha)\hat{x} - \hat{x}\| = \alpha\|x - \hat{x}\| = \epsilon$. i.e., it is $x_\alpha \in \Phi \cap U_\epsilon(\hat{x})$. Therefore, we get

$$f(\hat{x}) \leq f(x_\alpha) = f(\alpha x + (1 - \alpha)\hat{x}) \leq \alpha f(x) + (1 - \alpha)f(\hat{x}).$$

Hence

$$f(\hat{x}) \leq f(x)$$

Consequently, \hat{x} is the global minimum of f on Φ .

We will use the *Lagrangian function* □

$$L(x, \beta_0, \lambda, \mu) := \beta_0 f(x) + \sum_{i=1}^{n_g} \lambda_i g_i(x) + \sum_{j=1}^{n_h} \mu_j h_j(x)$$

2.1 Existence of a Solution

The existence of a solution for the optimization problem (NLP) where $\Phi \subseteq \mathbb{R}^n$ is a compact set and f is lower semi-continuous, is ensured by Weierstrass Theorem.

Definition 2.1.1. A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is upper semi-continuous at x_0 , if for every sequence $x_k \rightarrow x_0$ it holds

$$\limsup_{k \rightarrow \infty} f(x_k) \leq f(x_0)$$

A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is lower semi-continuous at x_0 , if for every sequence $x_k \rightarrow x_0$ it holds

$$f(x_0) \leq \liminf_{k \rightarrow \infty} f(x_k)$$

Theorem 2.1.1. (Weierstrass Extreme Value Theorem)

Let Φ be a compact subset of \mathbb{R}^{n_x} and let $f : \mathbb{R}^{n_x} \rightarrow \mathbb{R}$ be lower semi-continuous. Then, f achieves its minimum on Φ

Proof. Assume, that f is not bounded from below on Φ . Then, there is a sequence $x_i \in \Phi$ with $f(x_i) \leq -i$. Since Φ is compact, there exists a convergent subsequence $\lim_{k \rightarrow \infty} x_{i_k} = \hat{x}$ with $f(x_{i_k}) \leq -i_k$ for all $k \in \mathbb{N}$. Since f is lower semi-continuous it follows $f(\hat{x}) \leq \liminf_{k \rightarrow \infty} f(x_{i_k})$. Hence, $f(x_{i_k})$ is bounded from below by $f(\hat{x}) \in \mathbb{R}$. This contradicts $f(x_{i_k}) \leq -i_k \rightarrow -\infty$. This shows that f is bounded from below on Φ .

This in turn implies that $\hat{f} = \inf_{x \in \Phi} f(x)$ is a real number and for any $i \in \mathbb{N}$ there exists $x_i \in \Phi$ with $f(x_i) \leq \hat{f} + \frac{1}{i}$. Since Φ is compact, there exists a convergent subsequence $x_{i_k} \rightarrow \hat{x}$ with $f(x_{i_k}) \leq \hat{f} + \frac{1}{i_k}$ for all $k \in \mathbb{N}$. Since f is lower semi-continuous it follows $\hat{f} \leq f(\hat{x}) \leq \liminf_{k \rightarrow \infty} f(x_{i_k}) \leq \hat{f}$. Hence, f assumes its minimum on Φ . □

Theorem 2.1.2. Let $\Phi \subseteq \mathbb{R}^{n_x}$ and let $f : \mathbb{R}^{n_x} \rightarrow \mathbb{R}$ be a lower semi-continuous function on Φ . Let the level set

$$\text{lev}(f, f(y)) \cap \Phi = \{x \in \Phi : f(x) \leq f(y)\}$$

be non empty and compact for some $y \in \Phi$. Then, f attains its minimum on Φ

2.2 Conical Approximation of Sets

Conical approximations to sets play an important role in the formulation of necessary conditions for constrained optimization problems. We will summarize some important cones.

Definition 2.2.1. Let Φ be nonempty subset in \mathbb{R}^{n_x} . The (sequential) tangent cone to Φ at $x \in \Phi$ is given by

$$T(\Phi, x) := \left\{ d \in \mathbb{R}^{n_x} \mid \begin{array}{l} \text{there exist sequences } \{\alpha_k\}, \alpha_k \downarrow 0 \text{ and } \{x_k\}, x_k \in \Phi \\ k \in \mathbb{N} \text{ with } \lim_{k \rightarrow \infty} x_k = x, \lim_{k \rightarrow \infty} \frac{(x_k - x)}{\alpha_k} = d \text{ holds} \end{array} \right\} \quad (2.1)$$

Lemma 2.2.1. The tangent cone is a closed cone with vertex at zero.

Remark 2.2.1. By definition of tangent cone, $T(\Phi, x)$

- If x happens to be an interior point of Φ , then $T(\Phi, x) = \mathbb{R}^{n_x}$.
- If the line segment $[y, x] = \{\alpha y + (1 - \alpha)x : 0 \leq \alpha \leq 1\}$ is contained in Φ , then we have $y - x \in T(\Phi, x)$.
- If Φ is a convex set, then the tangent cone is convex and can be written as

$$\begin{aligned} T(\Phi, x) &:= \{d \in \mathbb{R}^{n_x} : \exists \alpha_k \downarrow 0, ; d_k \rightarrow d : x + \alpha_k d_k \in \Phi\} \\ &= \overline{\{d \in \mathbb{R}^{n_x} : \exists \alpha > 0 : x + \alpha d \in \Phi\}} \end{aligned}$$

As shown in [10], the tangent cone can be written as

$$T(\Phi, x) := \left\{ d \in \mathbb{R}^{n_x} \mid \begin{array}{l} \text{there exist } \delta > 0 \text{ and a mapping } r : (0, \delta] \rightarrow \mathbb{R}^{n_x} \\ \text{with } \lim_{\varepsilon \downarrow 0} \frac{r(\varepsilon)}{\varepsilon} = 0 \text{ and a sequence } \alpha_k, \alpha_k \downarrow 0 \text{ with} \\ x + \alpha_k d + r(\alpha_k) \in \Phi, \quad \forall k \in \mathbb{N} \end{array} \right\}$$

Definition 2.2.2. Let g be Fréchet differentiable. Then the linearizing cone of K and S at x is given by

$$T_{lin}(K, S, x) := \{d \in \text{cone}(S, x) \mid \nabla_x g_i(x)(d) \leq 0, \quad i \in A(x), \quad \nabla_x h_j(x)(d) = 0, \quad j \in \{1, \dots, n_h\}\}$$

where $K = \{y \in \mathbb{R}^{n_g} : y \leq 0_{n_g}\}$

A relation between tangent cone and linearized tangent cone is given by

Corollary 2.2.1. Let g and h be Fréchet differentiable. Then it holds

$$T(\Phi, x) \subseteq T_{lin}(K, S, x)$$

Here, the tangent cone $T(\Phi, x)$ is independent of the representation of the set Φ by inequality and equality constraints, whereas the linearizing cone $T_{lin}(K, S, x)$ depends on the functions g_i and h_j describing Φ .

Let x_0 be a local minimum of the problem (NLP). We consider the linearized problem associated with the problem (NLP)

$$NLP_{lin} \begin{cases} \min_x & f(x_0) + \nabla_x f(x_0)(x - x_0) \\ w.r.t. & x \in S \\ s.t. & g_i(x_0) + \nabla_x g_i(x_0)(x - x_0) \leq 0, \quad i = 1, \dots, n_g \\ & h_j(x_0) + \nabla_x h_j(x_0)(x - x_0) = 0, \quad j = 1, \dots, n_h \end{cases}$$

The feasible set of solutions is denoted by

$$\Phi_{lin} := \left\{ x \in S : \begin{array}{l} g_i(x_0) + \nabla_x g_i(x_0)(x - x_0) \leq 0, \quad i = 1, \dots, n_g, \\ h_j(x_0) + \nabla_x h_j(x_0)(x - x_0) = 0, \quad j = 1, \dots, n_h \end{array} \right\}$$

The local minimum x_0 for the problem (NLP) is a global minimum for NLP_{lin} , if and only if $\nabla_x f(x_0)(d) \geq 0$ for all $d \in T_{lin}(\Phi, x_0)$.

2.3 First Order Necessary Optimality Conditions of Firtz-John Type

Theorem 2.3.1. *Let $f : \mathbb{R}^{n_x} \rightarrow \mathbb{R}$ be Fréchet differentiable at x_0 and let x_0 be a local minimum of the problem (NLP). Then*

$$\nabla_x f(x_0)(d) \geq 0 \quad \forall d \in T(\Phi, x_0)$$

Proof. Let $d \in T(\Phi, x_0)$. Then there are sequences $\alpha_k \downarrow 0$, $x_k \rightarrow x_0, x_k \in \Phi$ with $d = \lim_{k \rightarrow \infty} \frac{x_k - x_0}{\alpha_k}$. Since x_0 is a local minimum and f is Fréchet differentiable at x_0 it follows

$$0 \leq f(x_k) - f(x_0) = \nabla_x f(x_0)(x_k - x_0) + o(\|x_k - x_0\|)$$

Division by α_k yields

$$0 \leq \nabla_x f(x_0) \underbrace{\left(\frac{x_k - x_0}{\alpha_k} \right)}_{\rightarrow d} + \underbrace{\| \frac{x_k - x_0}{\alpha_k} \|}_{\|d\|} \cdot \underbrace{\frac{o(\|x_k - x_0\|)}{\|x_k - x_0\|}}_{\rightarrow 0}$$

□

Theorem 2.3.2. *Suppose that $f : \mathbb{R}^{n_x} \rightarrow \mathbb{R}$ be twice continuously Fréchet differentiable at x_0 and let x_0 be a local minimum of (NLP). Then $\forall d \in T(\Phi, x_0)$*

(i) *if $\nabla_x f(x_0) = 0$, then $d^\top \nabla_x^2 f(x_0)d \geq 0$*

(iii) *if $\nabla_x f(x_0) = 0$, and $d^\top \nabla_x^2 f(x_0)d > 0$, then x_0 is a strict local minimum of f on Φ .*

Theorem 2.3.3. *Suppose that Φ is convex and that $f : \Phi \rightarrow \mathbb{R}$ is convex. Then a point x_0 minimizes f iff*

$$\nabla_x f(x_0)(d) \geq 0 \quad \forall d \in T(\Phi, x_0)$$

Theorem 2.3.4. (First Order Necessary Optimality Conditions)

Let x_0 be a local minimum of the problem (NLP) and S closed and convex set with $\text{int}(S) \neq \emptyset$. Then there exist multipliers $\beta_0 \geq 0$, $\lambda = (\lambda_1, \dots, \lambda_{n_g})^\top \in \mathbb{R}^{n_g}$ and $\mu = (\mu_1, \dots, \mu_{n_h})^\top \in \mathbb{R}^{n_h}$ not all zero such that

$$\nabla_x L(x_0, \beta_0, \lambda, \mu)(x - x_0) \geq 0 \text{ for } x \in S \quad (2.2)$$

$$g_i(x_0) \leq 0, \quad i = 1, \dots, n_g \quad (2.3)$$

$$h_j(x_0) = 0, \quad j = 1, \dots, n_h \quad (2.4)$$

$$\lambda_i g_i(x_0) = 0, \quad i = 1, \dots, n_g \quad (2.5)$$

$$\lambda_i \geq 0, \quad i = 1, \dots, n_g \quad (2.6)$$

Every point $(x, \beta_0, \lambda, \mu) \neq \Theta$ satisfying the Fritz-John Conditions is called the *Fritz-John point* of (NLP). The multipliers β_0 , λ and μ are called *Lagrange multipliers*. The main statement of the theorem is that there exists a nontrivial vector $(\beta_0, \lambda, \mu) \neq \Theta$. Where Θ is the generic zero element of some space. Unfortunately, the case $\beta_0 = 0$ may occur. In this case, the objective function f does not enter in to the Fritz-John conditions. In case of $\beta_0 \neq 0$ we call the Fritz-John point $(x, \beta_0, \lambda, \mu)$ the *Karush-Kuhn-Tucker (KKT) point*.

2.4 Constraint Qualifications

Conditions which ensure that the multiplier β_0 in Theorem 2.3.4 is not zero are called *regularity conditions* or *constraint qualifications*. In this case, without loss of generality β_0 can be normalized to one due to the linearity in the multipliers.

Definition 2.4.1. (Mangasarian-Fromowitz Constraint Qualification (MFCQ))

A feasible point $x_0 \in \Phi$ is said to be M-F regular if:

i) There exist a direction $\hat{d} \in \mathbb{R}^n$ such that $\nabla_x g_i(x_0)\hat{d} < 0$, $i \in A(x_0)$, $\nabla_x h_j(x_0)\hat{d} = 0$.

ii) The Jacobian matrix $[\nabla_x h_j(x_0)]$ has a full row rank n_h .

Theorem 2.4.1. (Karush-Kuhn-Tucker (KKT) Conditions I)

Let the assumptions of Theorem 2.3.4 be satisfied and let the MFCQ hold at x_0 . Then the assertions of Theorem 2.3.4 hold with $\beta_0 = 1$

Proof. Assume, that the MFCQ holds and that the Fritz-John conditions hold at x_0 with $\beta_0 = 0$, that is, there exist multipliers $\lambda_i \geq 0$, $i = 1, \dots, n_g$, μ_j , $j = 1, \dots, n_h$ not all zero with $\lambda_i g_i(x_0) = 0$ for all $i = 1, \dots, n_g$ and

$$\left(\sum_{i=1}^{n_g} \lambda_i \nabla_x g_i(x_0) + \sum_{j=1}^{n_h} \mu_j \nabla_x h_j(x_0) \right) (d) \geq 0 \text{ for all } d \in S - \{x_0\} \quad (2.7)$$

Let \hat{d} denote the vector in the constraint MFCQ, then we find

$$\sum_{i=1}^{n_g} \lambda_i \nabla_x g_i(x_0)(\hat{d}) = \sum_{i \in A(x_0)} \lambda_i \nabla_x g_i(x_0)(\hat{d}) \geq 0$$

Since $\lambda_i \geq 0$, $\nabla_x g_i(x_0)(\hat{d}) < 0$ for $i \in A(x_0)$ this inequality can only be valid if $\lambda_i = 0$ holds for every $i \in A(x_0)$. Since $\lambda_i = 0$ for $i \notin A(x_0)$ for all $i = 1, \dots, n_g$ Thus, inequality (2.7)

reduces to

$$0 \leq \sum_{j=1}^{n_h} \mu_j \nabla_x h_j(x_0)(d) = \sum_{j=1}^{n_h} (\mu_j \nabla_x h_j(x_0)(d) - \mu_j \nabla_x h_j(x_0)(\hat{d})) = \sum_{j=1}^{n_h} \mu_j \nabla_x h_j(x_0)(d - \hat{d})$$

for all $d \in S - \{x_0\}$. Since $\hat{d} \in \text{int}(S - \{x_0\})$ and $\nabla_x h_j(x_0), j = 1, \dots, n_h$ are linearly independent this inequality can only hold for $\mu_j = 0, j = 1, \dots, n_h$.

Hence, we derived $(\beta_0, \lambda, \mu) = 0_{1+n_{n_g}+n_h}$ which is a contradiction to the statement that not all multipliers are zero. \square

Definition 2.4.2. For a feasible point $x_0 \in \Phi$, $\Lambda(x_0)$ is the set of Lagrange multipliers associated with x_0 is defined as the set of nonzero vectors $(\lambda, \mu) \in \mathbb{R}^{n_g} \times \mathbb{R}^{n_h}$ satisfying the first order optimality conditions:

$$\begin{aligned} \nabla_x L(x_0, \lambda, \mu) &= \nabla_x f(x_0) + \sum_{i=1}^{n_g} \lambda_i \nabla_x g_i(x_0) + \sum_{j=1}^{n_h} \mu_j \nabla_x h_j(x_0) = 0_{n_x}, \\ \lambda_i &\geq 0, \quad i = 1, \dots, n_g \\ \lambda_i g_i(x_0) &= 0, \quad \forall i \in \{1, \dots, n_g\} \end{aligned}$$

That is

$$\Lambda(x_0) = \{(\lambda, \mu) : \nabla_x L(x_0, \lambda, \mu) = 0_{n_x}, \lambda \geq 0, \lambda^\top g(x_0) = 0\}$$

Theorem 2.4.2. Let x_0 be a local optimal solution for the problem (NLP). Then the set $\Lambda(x_0)$ of Lagrange multipliers corresponding to x_0 is nonempty, compact and convex if and only if the MFCQ holds at x_0

Definition 2.4.3. (Linear Independence Constraint Qualification (LICQ))

A feasible point $x_0 \in \Phi$ is said satisfy the LICQ if

- $x_0 \in \text{int}(S)$
- the gradients $\{\nabla_x g_i(x_0), i \in A(x_0), \nabla_x h_j(x_0), j = 1, \dots, n_h\}$ are linearly independent.

Satisfaction of LICQ guarantees that the Lagrange multipliers λ and μ are unique if x_0 is a local optimal solution of the program NLP, i.e, the set $\Lambda(x_0)$ reduces to singleton.

Theorem 2.4.3. (Karush-Kuhn-Tucker (KKT) Conditions II)

Let the assumptions of Theorem 2.3.4 be satisfied and let the LICQ hold at x_0 . Then the assertions of Theorem 2.3.4 hold with $\beta_0 = 1$ and in particular

$$\nabla_x L(x_0, \lambda, \mu) = 0_{n_x}$$

Furthermore, the multipliers λ and μ are unique.

Proof. Again, we assume that the Fritz-John conditions hold with $\beta = 0$. Since $x_0 \in \text{int}(S)$, inequality in theorem 2.3.4 has to hold for all $x_0 \in \mathbb{R}^{n_x}$ and thus

$$\sum_{i=1}^{n_g} \lambda_i \nabla_x g_i(x_0) + \sum_{j=1}^{n_h} \mu_j \nabla_x h_j(x_0) = \sum_{i \in A(x_0)} \lambda_i \nabla_x g_i(x_0) + \sum_{j=1}^{n_h} \mu_j \nabla_x h_j(x_0) = 0_{n_x}$$

The LICQ of the derivatives implies $\lambda_i = \mu_j = 0$ for all $i = 1, \dots, n_g, j = 1, \dots, n_h$. Again, this is a contradiction to $(\beta_0, \lambda, \mu) = 0_{1+n_{n_g}+n_h}$ which is a contradiction to the statement that

not all multipliers are zero.

The uniqueness of the Lagrange multipliers follows from the following considerations. Assume, that there are Lagrange multipliers λ_i , $i = 1, \dots, n_g$, μ_j , $j = 1, \dots, n_h$ and $\tilde{\lambda}_i$, $i = 1, \dots, n_g$, $\tilde{\mu}_j$, $j = 1, \dots, n_h$ satisfying the KKT conditions. Again, $x_0 \in \text{int}(S)$ particularly implies

$$\begin{aligned}\nabla_x f(x_0) + \sum_{i=1}^{n_g} \lambda_i \nabla_x g_i(x_0) + \sum_{j=1}^{n_h} \mu_j \nabla_x h_j(x_0) &= 0_{n_x} \\ \nabla_x f(x_0) + \sum_{i=1}^{n_g} \tilde{\lambda}_i \nabla_x g_i(x_0) + \sum_{j=1}^{n_h} \tilde{\mu}_j \nabla_x h_j(x_0) &= 0_{n_x}\end{aligned}$$

Subtracting these equations leads to

$$\sum_{i=1}^{n_g} (\lambda_i - \tilde{\lambda}_i) \nabla_x g_i(x_0) + \sum_{j=1}^{n_h} (\mu_j - \tilde{\mu}_j) \nabla_x h_j(x_0) = 0_{n_x}$$

For inactive inequality constraints we have $\lambda_i = \tilde{\lambda}_i$, $i \notin A(x_0)$. Since the gradients of the active constraints are assumed to be linearly independent, it follows $0 = \lambda_i - \tilde{\lambda}_i$, $i \in A(x_0)$ and $0 = \mu_j - \tilde{\mu}_j$, $j = 1, \dots, n_h$. Hence, the Lagrange multipliers are unique. \square

To decide, whether a given point that fulfills the necessary conditions is optimal we need *sufficient conditions*. We need the so-called critical cone

$$T_C(x_0) := \left\{ d \in \mathbb{R}^{n_x} \left| \begin{array}{l} \nabla_x g_i(x_0)(d) \leq 0, \quad i \in A(x_0), \lambda_i = 0, \\ \nabla_x g_i(x_0)(d) = 0, \quad i \in A(x_0), \lambda_i > 0, \\ \nabla_x h_j(x_0)(d) = 0, \quad j = 1, \dots, n_h \end{array} \right. \right\}$$

The term 'critical cone' is due to the following reasoning. Directions d with $\nabla_x g_i(x_0)(d) > 0$ for some $i \in A(x_0)$ or $\nabla_x h_j(x_0)(d) \neq 0$ for some $j \in \{1, \dots, n_h\}$ are infeasible directions. So, we consider only the feasible directions d with $\nabla_x g_i(x_0)(d) \leq 0$. For such directions the KKT conditions yield

$$\nabla_x f(x_0)^\top d + \sum_{i \in A(x_0)} \underbrace{\lambda_i \nabla_x g_i(x_0)^\top d}_{\leq 0} + \sum_{j=1}^{n_h} \underbrace{\mu_j \nabla_x h_j(x_0)^\top d}_{=0} = 0_{n_x}$$

and thus $\nabla_x f(x_0)d \geq 0$. If even $\nabla_x f(x_0)d > 0$ holds, then d is a direction of ascent and the direction d is not interesting for the investigation of sufficient conditions. So, let $\nabla_x f(x_0)d = 0$. This is the critical case. In this case it holds

$$\sum_{i \in A(x_0)} \lambda_i \nabla_x g_i(x_0)d = \sum_{i \in A(x_0), \lambda_i > 0} \lambda_i \nabla_x g_i(x_0)d = 0,$$

and thus $\nabla_x g_i(x_0)d = 0 \quad \forall i \in A(x_0)$ with $\lambda_i > 0$. Hence, $d \in T_C(x_0)$ and for such directions we need additional assumptions about the curvature (second derivative).

2.5 Second Order Sufficient Optimality Conditions

Theorem 2.5.1. (Second Order Sufficient Conditions)

Let f, g_i and h_j be twice continuously differentiable. Let $S = \mathbb{R}^{n_x}$ and let (x_0, λ_0, μ_0) be a KKT point of (NLP) with

$$\nabla_{xx}^2 L(x_0, \lambda_0, \mu_0)(d, d) > 0, \quad \forall d \in T_C(x_0), \quad d \neq 0_{n_x} \quad (2.8)$$

Then there exists a neighbourhood U of x_0 and some $\alpha > 0$ such that

$$f(x) \geq f(x_0) + \alpha \|x - x_0\|^2 \quad \forall x \in \Phi \cap U$$

Proof. (a) Let $d \in T(\Phi, x_0)$, $d \neq 0_{\mathbb{R}^{n_x}}$. Then there exist sequences $x_k \in \Phi$, $x_k \rightarrow x_0$ and $\alpha_k \downarrow 0$ with

$$\lim_{k \rightarrow \infty} \frac{x_k - x_0}{\alpha_k} = d$$

For $i \in A(x_0)$ we have

$$0 \geq \frac{g_i(x_k) - g_i(x_0)}{\alpha_k} = g'_i(\eta_k) \frac{x_k - x_0}{\alpha_k} \rightarrow g'_i(x_0)d$$

by the mean-value theorem. Similarly, we show $h_j(x_0)d = 0$ for $j = 1, \dots, n_h$. Since (x_0, λ_0, μ_0) is a KKT point with $\lambda_{0i} = 0$, if $g_i(x_0) < 0$, we obtain

$$\nabla_x f(x_0)d = - \sum_{i=1}^{n_g} \lambda_{0i} \nabla_x g_i(x_0)d - \sum_{j=1}^{n_h} \mu_{0j} \nabla_x h_j(x_0)d \geq 0.$$

Hence, x_0 fulfills the first order necessary condition $\nabla_x f(x_0)d \geq 0$ for all $d \in T(\Phi, x_0)$

(b) Assume, that the statement of the theorem is wrong. Then for any ball around x_0 with radius $\frac{1}{i}$ there exists a point $x_i \in \Phi$ with $x_i \neq x_0$ and

$$f(x_i) - f(x_0) < \frac{1}{i} \|x_i - x_0\|^2, \quad \|x_i - x_0\| \leq \frac{1}{i} \quad \forall i \in \mathbb{N} \quad (2.9)$$

Since the unit ball w.r.t. $\|\cdot\|$ is compact in \mathbb{R}^{n_x} , there exists a convergent subsequence $\{x_{i_k}\}$ with

$$\lim_{k \rightarrow \infty} \frac{x_{i_k} - x_0}{\|x_{i_k} - x_0\|} = d, \quad \lim_{k \rightarrow \infty} \|x_{i_k} - x_0\| = 0.$$

Hence, $d \in T(\Phi, x_0) \setminus \{0_{n_x}\}$. Taking the limit in (2.9) yields

$$\nabla_x f(x_0)d = \frac{f(x_{i_k}) - f(x_0)}{\|x_{i_k} - x_0\|} \leq 0$$

Together with (a) we have

$$\nabla_x f(x_0)d = 0$$

(c) Since x_0 is a KKT point, it follows

$$\nabla_x f(x_0)d = - \sum_{i \in A(x_0)} \underbrace{\lambda_{0i}}_{\geq 0} \underbrace{\nabla_x g_i(x_0)d}_{\leq 0} - \sum_{j=1}^{n_h} \mu_{0j} \underbrace{\nabla_x h_j(x_0)d}_{=0} \geq 0.$$

Thus, it is $\nabla_x g_i(x_0)d = 0$, if $\lambda_{0i} > 0$. Hence, $d \in T_C(x_0)$.
According to (2.9) it holds

$$\frac{f(x_{i_k}) - f(x_0)}{\|x_{i_k} - x_0\|^2} \leq \frac{1}{i_k} = 0 \quad (2.10)$$

for the direction d .

Furthermore, it is ($\beta_0 = 1$)

$$L(x_{i_k}, \beta_0, \lambda_0, \mu_0) = f(x_{i_k}) + \sum_{i=1}^{n_g} \lambda_{0i} g_i(x_{i_k}) + \sum_{j=1}^{n_h} \mu_{0j} h_j(x_{i_k}) \leq f(x_{i_k}),$$

$$L(x_0, \beta_0, \lambda_0, \mu_0) = \nabla_x f(x_0) + \sum_{i=1}^{n_g} \lambda_{0i} g_i(x_0) + \sum_{j=1}^{n_h} \mu_{0j} h_j(x_0) = f(x_0),$$

$$\nabla_x L(x_0, \beta_0, \lambda_0, \mu_0) = \nabla_x f(x_0) + \sum_{i=1}^{n_g} \lambda_{0i} \nabla_x g_i(x_0) + \sum_{j=1}^{n_h} \mu_{0j} \nabla_x h_j(x_0) = 0_{n_x}^\top.$$

Taylor expansion of L with $\beta_0 = 1$ w.r.t. to x at x_0 yields

$$\begin{aligned} f(x_{i_k}) &\geq L(x_{i_k}, \beta_0, \lambda_0, \mu_0) = L(x_0, \beta_0, \lambda_0, \mu_0) + \nabla_x L(x_0, \beta_0, \lambda_0, \mu_0)(x_{i_k} - x_0) \\ &\quad + \frac{1}{2} \nabla_{xx}^2 L(\eta_k, \beta_0, \lambda_0, \mu_0)(x_{i_k} - x_0, x_{i_k} - x_0) \\ &= f(x_0) + \frac{1}{2} \nabla_{xx}^2 L(\eta_k, \beta_0, \lambda_0, \mu_0)(x_{i_k} - x_0, x_{i_k} - x_0) \end{aligned}$$

where η_k is some point between x_0 and x_{i_k} . Division by $\|x_{i_k} - x_0\|^2$ and taking the limit, yields together with (2.10)

$$0 \geq \frac{1}{2} \nabla_{xx}^2 L(x_0, \beta_0, \lambda_0, \mu_0)(d, d).$$

This contradicts the assumption $\nabla_{xx}^2 L(x_0, \beta_0, \lambda_0, \mu_0)(d, d) > 0$, $\forall d \in T_C(x_0)$, $d \neq 0_{n_x}$

□

Chapter 3

Parametric Nonlinear Optimization Problems

We consider parametric nonlinear programming problems (NLP-problems) involving a parameter $p \in \mathbb{R}^{n_p}$. The *parametric NLP-problem* with equality and inequality constraints is given by

$$\text{NLP}(p) \quad \begin{cases} \min_x & f(x, p), & x \in \mathbb{R}^{n_x} \\ \text{subject to} & g_i(x, p) \leq 0, \quad i = 1, \dots, g_{n_g} \\ & h_j(x, p) = 0, \quad j = 1, \dots, h_{n_h} \end{cases} \quad (\text{NLP}(p))$$

where $f, g_1, \dots, g_{n_g}, h_1, \dots, h_{n_h} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \rightarrow \mathbb{R}$ are assumed throughout to be a class of C^2 on $\mathbb{R}^{n_x} \times \mathbb{R}^{n_p}$.

We assume that the problem parameters p , are held fixed during the course of the optimization, and the optimal solution point, x_0 , satisfies the first order Karush-Kuhn-Tucker optimality conditions.

For a fixed *nominal* parameter/reference point $p_0 \in \mathbb{R}^{n_p}$, the program

$$\begin{cases} \min_x & f(x, p_0), \\ \text{subject to} & g_i(x, p_0) \leq 0, \quad i = 1, \dots, g_{n_g} \\ & h_j(x, p_0) = 0, \quad j = 1, \dots, h_{n_h} \end{cases} \quad (\text{NLP}(p_0))$$

is called the *unperturbed* or *nominal* problem and assumed to coincide with the (NLP). We are interested in the behavior (continuity and differentiability properties) and the effects of variations in p on the optimal solutions

$$S(p) := \{x \in \Phi(p) \mid f(x, p) = v(p)\}$$

and the related optimal values of the objective function

$$v(p) := \inf_{x \in \Phi(p)} f(x, p)$$

of the perturbed problem (NLP(p)) in a neighborhood of the nominal parameter p_0 . For each value of p , the set of feasible solutions of (NLP(p)) is defined by

$$\Phi(p) := \{x \in \mathbb{R}^{n_x} \mid g_i(x, p) \leq 0, i = 1, \dots, g_{n_g}, h_j(x, p) = 0, j = 1, \dots, h_{n_h}\}$$

We associate *active sets* with \hat{x} of indices given by

$$A(\hat{x}, p) := \{i | g_i(\hat{x}, p) = 0, \quad i = 1, \dots, n_g\}$$

The Lagrangian function $L : \mathbb{R}^{n_x} \times \mathbb{R}^{n_g} \times \mathbb{R}^{n_h} \times \mathbb{R}^{n_p} \rightarrow \mathbb{R}$ for constrained nonlinear optimization problem NLP(p) is defined as

$$L(x, \lambda, \mu, p) := f(x, p) + \lambda^\top g(x, p) + \mu^\top h(x, p)$$

where we have used the functions

$g(x, p) := (g_1(x, p), \dots, g_{n_g}(x, p))^\top$, $h(x, p) := (h_1(x, p), \dots, h_{n_h}(x, p))^\top$ and the *Lagrange multipliers* $\lambda = (\lambda_1, \dots, \lambda_{n_g})^\top \in \mathbb{R}^{n_g}$, and $\mu = (\mu_1, \dots, \mu_{n_h})^\top \in \mathbb{R}^{n_h}$.

We introduce the active constraints and the corresponding multipliers to active constraints by

$$g^a := (g_i)_{i \in A(x, p)}, \quad \lambda^a \in \mathbb{R}^{n_a}, \quad n_a := n(A(x, p))$$

3.1 Stability of the Optimal Value and Optimal Solutions

Problem of Parametric Optimization: How do the optimal value function $v(p)$ and the mappings $\Phi : P \subseteq \mathbb{R}^{n_p} \rightarrow \mathbb{R}^{n_x}$ and, $S : P \subseteq \mathbb{R}^{n_p} \rightarrow \mathbb{R}^{n_x}$ change with the parameter p . (Continuously, smoothly?)

In this section we discuss continuity properties of the optimal value function $v(p)$ and the optimal solution set $S(p)$ of the parameterized problem (NLP(p)). We shall see that the lower- and upper semi-continuity of the marginal function $v(p)$ depend on different assumptions. Obviously to assure the lower semi-continuity of v at p_0 the feasible set $\Phi(p)$ should not become essentially larger by a small perturbation p of p_0 and to assure the upper semi-continuity of v at p_0 the feasible set $\Phi(p)$ should not become essentially smaller after a perturbation. Thus we will need some compactness assumptions and Constraint Qualifications. The results in this section are collected from [1, 5, 9]

Definition 3.1.1. *The mapping $\Phi : P \subseteq \mathbb{R}^{n_p} \rightarrow 2^{\mathbb{R}^{n_x}}$ is upper semi-continuous at $p_0 \in P$, if for any open set Ω containing $\Phi(p_0)$ there is an open neighborhood $V_\epsilon(p_0)$ of p_0 such that $\Phi(p_0) \subseteq \Omega$ for each $p_0 \in V_\epsilon(p_0) \cap P$.*

Definition 3.1.2. *Φ is said to be closed at $p_0 \in P$ if $\{p_k\} \subseteq P, p_k \rightarrow p_0, x_k \in \Phi(p_k)$, and $x_k \rightarrow x_0$ imply that $x_0 \in \Phi(p_0)$.*

Definition 3.1.3. *Φ is said to be uniformly compact near $p_0 \in P$ if there is a neighborhood $V_\epsilon(p_0)$ of p_0 such that the closure of $\cup_{p_0 \in P} \Phi(p_0)$ is compact.*

Proposition 3.1.1. *Let Φ be uniformly compact near $p_0 \in P$. Then Φ is closed at p_0 if and only if $\Phi(p_0)$ is compact and Φ is upper semi-continuous at p_0 .*

Lemma 3.1.1. *If $\Phi(p_0)$ is nonempty and $\Phi(p)$ is uniformly compact near p_0 , then $\Phi(p)$ and the marginal function $v(p)$ are upper semi-continuous at p .*

Proof. Since the functions $g_i(h, p)$ and $h_j(h, p)$ are continuous, $\Phi(p)$ is closed at p_0 . It follows from proposition(3.1.1) that $\Phi(p_0)$ is compact and $\Phi(p)$ is upper semi-continuous at p_0 . \square

Theorem 3.1.1. *Suppose that $\Phi(p_0)$ is nonempty, $\Phi(p)$ is uniformly compact near p_0 and the (MFCQ) holds at some, $x_0 \in \Phi(p)$. Then the marginal function $v(p)$ is continuous at p_0 .*

In the next theorems it is demonstrated that the MFCQ is preserved under small variations of the parameters and that the set of KKT multipliers is locally well behaved.

Theorem 3.1.2. *Assume that the MFCQ holds at some $x \in S(p_0)$. Let $\{p_k\}$ and $\{x_k\}$ be sequences such that $p_k \rightarrow p_0$, $x_k \in S(p_k)$ and $x_k \rightarrow x_0$. Then for k sufficiently large MFCQ holds at x_k and there exist subsequences $\{(\lambda_{k_m}, \mu_{k_m})\}$, $\{x_{k_m}\}$ with $(\lambda_{k_m}, \mu_{k_m}) \in \Lambda(x_{k_m}, p_{k_m})$ such that $(\lambda_{k_m}, \mu_{k_m}) \rightarrow (\lambda_0, \mu_0)$ for some $(\lambda_0, \mu_0) \in \Lambda(x_0, p_0)$*

Proof. Since $x_k \rightarrow x_0$, for sufficiently large k it follows that $A(x_k, p_k) \subseteq A(x_0, p_0)$ and $\{\nabla_x h(x_k, p_k)\}$ still have full row rank. Let \hat{d} given by MFCQ at (x_0, p_0) and let

$$\hat{d}_k = [I - \nabla_x h(x_k, p_k)^* \nabla_x h(x_k, p_k)] \hat{d} \quad (3.1)$$

where $\nabla_x h(x_k, p_k)^* = \nabla_x h(x_k, p_k)^\top [\nabla_x h(x_k, p_k) \nabla_x h(x_k, p_k)^\top]^{-1}$ is the pseudo-inverse of the full row rank Jacobian matrix $\nabla_x h(x_k, p_k)$.

For sufficiently large k , we have as $\hat{d}_k \rightarrow \hat{d}$,

$$\begin{aligned} \nabla_x g_i(x_k, p_k) \hat{d}_k &< 0 \quad i \in A(x_0, p_0), \\ \nabla_x h(x_k, p_k) \hat{d}_k &= 0 \end{aligned} \quad (3.2)$$

Since $A(x_k, p_k) \subseteq A(x_0, p_0)$, we have MFCQ satisfied at (x_k, p_k) . For each (x_k, p_k) take $(\lambda_k, \mu_k) \in \Lambda(x_k, p_k)$. Then

$$\nabla_x f(x_k, p_k) = \lambda_k^\top \nabla_x g(x_k, p_k) + \mu_k^\top \nabla_x h(x_k, p_k) \quad (3.3)$$

$$\mu_k^\top = [\nabla_x f(x_k, p_k) - \lambda_k^\top \nabla_x g(x_k, p_k)] \nabla_x h(x_k, p_k)^* \quad (3.4)$$

From (3.2) and (3.3) we have

$$\nabla_x f(x_k, p_k) \hat{d}_k = \lambda_k^\top \nabla_x g(x_k, p_k) \hat{d}_k \leq \lambda_i^{k\top} \nabla_x g_i(x_k, p_k) \hat{d}_k \quad \text{for any } i \in A(x_0, p_0).$$

Then

$$\begin{aligned} \lambda_i^{k\top} &\leq \frac{\nabla_x f(x_k, p_k) \hat{d}_k}{\nabla_x g_i(x_k, p_k) \hat{d}_k} && \text{if } i \in A(x_0, p_0) \\ \lambda_i^{k\top} &= 0, && \text{if } i \notin A(x_0, p_0) \end{aligned} \quad (3.5)$$

From (3.4) and (3.5) we have the sequence $\{(\lambda_k, \mu_k)\}$ bounded. Therefore, there exists a subsequence converging to some $(\lambda_0, \mu_0) \in \Lambda(\lambda_0, \mu_0)$ \square

Corollary 3.1.1. *If $\Phi(p_0)$ is nonempty and compact and if MFCQ holds at each $x_0 \in S(p_0)$, then $\Lambda(p_0)$ is compact.*

Proof. ([9]) Take a sequence $(\lambda_k, \mu_k) \in \Lambda(p_0)$. There exist a sequence $\{x_k\} \subseteq S(p_0)$ such that $(\lambda_k, \mu_k) \in \Lambda(x_k, p_0)$. Since $S(p_0)$ is compact, there exist subsequence x_{k_m} and $x_0 \in S(p_0)$ such that $x_{k_m} \rightarrow x_0$. From 3.1.2 choosing $p_{k_m} = p_0$, there exist subsequence $(\lambda_{k_j}, \mu_{k_j}) \in \Lambda(x_{k_j}, p_0)$ such that $(\lambda_{k_j}, \mu_{k_j}) \rightarrow (\lambda_0, \mu_0) \in \Lambda(\lambda_0, \mu_0) \subseteq \Lambda(p_0)$. \square

Corollary 3.1.2. *Suppose $\Phi(p_0)$ is nonempty and $\Phi(p)$ uniformly compact near (p_0) and that the MFCQ holds at each $x_0 \in S(p_0)$. Then for any sequences $\{p_k\}$, $\{x_k\}$ and $\{(\lambda_k, \mu_k)\}$, with $x_k \in S(p_k)$, $(\lambda_k, \mu_k) \in \Lambda(x_k, p_k)$, and $p_k \rightarrow p_0$, there exist subsequences $\{p_{k_m}\}$, $\{x_{k_m}\}$ $(\lambda_{k_m}, \mu_{k_m}) \in \Lambda(x_{k_m}, p_{k_m})$ such that $x_{k_m} \rightarrow x_0 \in S(p_0)$ and $(\lambda_{k_m}, \mu_{k_m}) \rightarrow (\lambda_0, \mu_0) \in \Lambda(x_0, p_0)$*

Theorem 3.1.3. *If $\Phi(p_0)$ is nonempty and $\Phi(p)$ is uniformly compact near (p_0) and if MFCQ holds at every $x_0 \in S(p)$, then for some $\sigma > 0$, MFCQ holds at each $x \in S(p)$ with $\|p - p_0\| \leq \sigma$. Furthermore the point-to-set mapping $\Lambda(p)$ is closed at p_0 and uniformly compact near p_0 and therefore upper semi-continuous at p_0 .*

Theorem 3.1.4. ([4], proposition 4.4, p. 263) *Let $p_0 \in \mathbb{R}^{n_p}$ be a given point in the parameter space \mathbb{R}^{n_p} . Suppose that (i) the function $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \rightarrow \mathbb{R}$ is continuous, (ii) the map $\Phi(\cdot) : \mathbb{R}^{n_p} \rightarrow \mathbb{R}^{n_x}$ is closed, (iii) (inf-compactness assumption) there exist $\alpha \in \mathbb{R}$ and a compact set $C \subseteq \mathbb{R}^{n_x}$ such that for every p in a neighborhood of p_0 , the level set*

$$\text{lev}_\alpha f(\cdot, p) := \{x \in \Phi(p) : f(x, p) \leq \alpha\}$$

is nonempty and contained in C , (iv) for any neighborhood U of the set $S(p_0)$ there is a neighborhood $V_\epsilon(p_0)$ such that $U \cap \Phi(p) \neq \emptyset$ for all $p \in V_\epsilon(p_0)$. Then:

- (a) *the optimal value function $v(p)$ is continuous at $p = p_0$, and*
- (b) *the optimal solution mapping $p \mapsto S(p)$ is upper semi-continuous at p_0 .*

3.1.1 Directional differentiability of optimal value functions

Definition 3.1.4. *The (one-sided) lower and upper Dini directional derivative of the function $v(p)$ at p_0 in the direction $z \in \mathbb{R}^{n_p}$ are respectively defined to be*

$$v'(p_0; z)_+ = \liminf_{t \rightarrow 0^+} \frac{v(p_0 + tz) - v(p_0)}{t}$$

$$v'(p_0; z)^+ = \limsup_{t \rightarrow 0^+} \frac{v(p_0 + tz) - v(p_0)}{t}$$

Theorem 3.1.5. *If MFCQ holds at some optimal point $x_0 \in S(p_0)$, then for any direction $z \in \mathbb{R}^{n_p}$*

$$v'(p_0; z)_+ \geq \min_{(\lambda, \mu) \in \Lambda(x_0, p_0)} \{\nabla_p L(x_0, p_0, \lambda, \mu)z\}$$

Corollary 3.1.3. *Suppose $\Phi(p_0)$ is nonempty, $\Phi(p)$ is uniformly compact near p_0 and the MFCQ holds at every optimal point $x_0 \in S(p_0)$. Then for any direction $z \in \mathbb{R}^{n_p}$, there is an $x_0 \in S(p_0)$. such that*

$$v'(p_0; z)_+ \geq \inf_{x \in S(p_0)} \min_{(\lambda, \mu) \in \Lambda(x_0, p_0)} \{\nabla_p L(x_0, p_0, \lambda, \mu)z\}$$

Corollary 3.1.4. *If $\Phi(p_0)$ is nonempty, $\Phi(p)$ is uniformly compact near p_0 and the MFCQ holds at every optimal point $x_0 \in S(p_0)$, then for any direction $z \in \mathbb{R}^{n_p}$, we have*

$$v'(p_0; z)^+ \leq \inf_{x \in S(p_0)} \max_{(\lambda, \mu) \in \Lambda(x_0, p_0)} \{\nabla_p L(x_0, p_0, \lambda, \mu)z\}$$

Corollary 3.1.5. *Let $\Phi(p_0)$ is nonempty, $\Phi(p)$ is uniformly compact near p_0 and the MFCQ holds for each $x_0 \in S(p_0)$. If $f, g_i, i = 1, \dots, n_g$ convex functions in x , and the functions $h_j, j = 1, \dots, n_h$ are affine with all functions continuously differentiable with respect to (x, p) , then the directional derivative exists for each $z \in \mathbb{R}^{n_p}$, and*

$$v'(x; z) = \inf_{x \in S(p_0)} \max_{(\lambda, \mu) \in \Lambda(x, p_0)} \{\nabla_p L(x, p_0, \lambda, \mu)z\}$$

Corollary 3.1.6. *If $\Phi(p_0)$ is nonempty, $\Phi(p)$ is uniformly compact near p_0 and the LICQ holds at every optimal point $x_0 \in S(p_0)$, then for any direction $z \in \mathbb{R}^{n_p}$, the directional derivatives exist and given by*

$$v'(p; z) = \inf_{x \in S(p_0)} \nabla_p L(x, \lambda(x), \mu(x), p_0)$$

where $(\lambda(x), \mu(x))$ is the unique Lagrange multiplier vector associated with $x \in S(p_0)$

Theorem 3.1.6. *Suppose the inf-compactness assumption holds and $\Phi(p_0)$ is nonempty, $\Phi(p)$ is uniformly compact near p_0 . suppose further the LICQ holds at every optimal point $x \in S(p_0)$, then the optimal value function $v(p)$ is Lipschitz and directionally differentiable near p_0 and its directional derivative at p_0 in the direction $z \in \mathbb{R}^{n_p}$ is given by*

$$v'(x; z) = \min_{x_0 \in S(p_0)} \{\nabla_p L(x_0, p_0, \lambda_0, \mu_0)z\} \quad (3.6)$$

where (λ_0, μ_0) denotes the Lagrangian multiplier vector of $(NLP(p))$ at the optimal solution $x_0 \in S(p_0)$

Proof. First, it is easy to see that there exists a neighborhood $V_{\epsilon_1}(p_0) \subset V_{\epsilon}(p_0)$ of p_0 such that $v(p) < \alpha$ for $p \in V_{\epsilon_1}(p_0)$ under the assumptions of the theorem since $(NLP(p))$ is feasible for all p near p_0 due to the LICQ at all solutions of $(NLP(p_0))$, whose solution set is not empty under the assumptions. Then we consider the following NLP:

$$NLP^\alpha(p) \quad \begin{cases} \text{minimize}_x & f(x, p), \\ \text{subject to} & g_i(x, p) \leq 0, i = 1, \dots, g_{n_g} \\ & h_j(x, p) = 0, j = 1, \dots, h_{n_h} \\ & f(x, p) \leq \alpha \end{cases}$$

where the constraint $f(x, p) \leq \alpha$ is inactive at any optimal solution of $NLP^\alpha(p)$ when $p \in V_{\epsilon_1}(p_0)$. In particular, the LICQ holds for $NLP^\alpha(p_0)$ and the uniform compactness holds for $NLP^\alpha(p)$ with p near p_0 . Let $v^\alpha(p)$ denote the optimal value function of $NLP^\alpha(p)$. Then $v^\alpha(p)$ is locally Lipschitz by [9, Theorem 5.1] near p_0 and $v^\alpha(p)$ is directionally differentiable [9, Corollary 4.4]. Moreover its directional derivative [9, Corollary 4.4] is given by the formula 3.6 at p_0 . Note that the conditions of the theorem remain valid under small perturbations of p_0 , hence $v^\alpha(p)$ is directionally differentiable near p_0 . Now, for $p \in V_{\epsilon_1}(p_0)$, we have $v(p) = v^\alpha(p)$. So, the theorem follows. \square

3.1.2 Implicit Function Theorem

It is important to discuss under what conditions a parameter dependent nonlinear equation solution will actually persist when problem parameters are changed.

For $x \in \mathbb{R}^n$, we denote by $U_r(x_0) = \{x \mid \|x - x_0\| \leq r\}$ the closed ball of radius r centered at x_0 .

Theorem 3.1.7. (Contraction Theorem) Consider a continuous function $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and suppose that for some $z_0 \in \mathbb{R}^n, r > 0$, and some κ with $0 \leq \kappa < 1$, we have

$$\|F(x) - F(y)\| \leq \kappa \|x - y\| \quad \forall x, y \in U_r(z_0)$$

$$\|F(z_0) - z_0\| \leq (1 - \kappa)r$$

Then the equation

$$z = F(z), \quad z \in \mathbb{R}^n$$

has one and only one solution, $\hat{z} \in U_r(z_0)$ and \hat{z} is the limit of the sequence

$$z_{k+1} = F(z_k), \quad k = 0, 1, 2, \dots$$

Theorem 3.1.8. (Implicit Function Theorem) Let $G : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ be a function that satisfy

- $G(x_0, p_0) = 0, \quad \forall x_0 \in \mathbb{R}^n$ and $p_0 \in \mathbb{R}^m$
- $G_x(x_0, p_0)$ is nonsingular with bounded inverse,

$$\|G_x(x_0, p_0)^{-1}\| \leq M \text{ for some } M > 0$$

- G and G_x are Lipschitz continuous, i.e., for all $x, y \in U_r(x_0)$ and for all $p, \hat{p} \in U_r(p_0)$, and some $L > 0$;

$$\|G(x, p) - G(y, \hat{p})\| \leq L(\|x - y\| + \|p - \hat{p}\|)$$

$$\|G_x(x, p) - G_x(y, \hat{p})\| \leq L(\|x - y\| + \|p - \hat{p}\|)$$

Then

- (I) there exists ϵ , with $0 < \epsilon \leq r$, and a unique continuous function $x(p) : V_\epsilon(p_0) \rightarrow \mathbb{R}^n$, with $x(p_0) = x_0$, such that

$$G(x(p), p) = 0, \quad \text{for all } p \in V_\epsilon(p_0) \tag{3.7}$$

- (II) Moreover, if G is continuously differentiable w.r.t. p in $U_r(x_0) \times V_r(p_0)$, then $x(\cdot)$ has a continuous derivative given by

$$\frac{d}{dp}x(p) = -G_x(x(p), p)^{-1}G_p(x(p), p) \quad \text{on } U_\epsilon(x_0) \times V_\epsilon(p_0) \tag{3.8}$$

,

Proof. (I) For unknown (x, p) one can rewrite the problem as

$$\begin{aligned} G(x, p) = 0 &\Leftrightarrow G_x(x_0, p_0)x = G_x(x_0, p_0)x - G(x, p) \\ &\Leftrightarrow x = \underbrace{G_x(x_0, p_0)^{-1}[G_x(x_0, p_0)x - G(x, p)]}_{\equiv F(x, p)} \end{aligned}$$

Hence $G(x, p) = 0$ iff x is the fixed point of $F(\cdot, p) = 0$. We must verify the conditions of the contraction theorem. Take $x, y \in U_{r_1}(x_0)$ and any fixed $p \in U_{r_1}(x_0)$ where r_1 chosen latter. Then

$$F(x, p) - F(y, p) = G_x(x_0, p_0)^{-1} \{G_x(x_0, p_0)[x - y] - [G(x, p) - G(y, p)]\} \quad (3.9)$$

By the Fundamental Theorem of Calculus, we have

$$\begin{aligned} F(x, p) - F(y, p) &= \int_0^1 \frac{d}{dt} G(tx + (1-t)y, p) dt \\ &= \int_0^1 G_x(tx + (1-t)y, p) dt [x - y] \\ &= \bar{G}_x(x, y, p) \end{aligned}$$

where in the last step we used the Mean Value Theorem to get \bar{G} . Then (3.9) becomes

$$\begin{aligned} \|F(x, p) - F(y, p)\| &\leq M \|G_x(x_0, p_0) - \bar{G}_x(x, y, p)\| \|x - y\| \\ &= M \left\| \int_0^1 G_x(x_0, p_0) - G_x(tx + (1-t)y, p) dt \right\| \|x - y\| \\ &\leq M \int_0^1 \|G_x(x_0, p_0) - G_x(tx + (1-t)y, p)\| dt \|x - y\| \\ &\leq M \int_0^1 L(\|x_0 - \underbrace{(tx + (1-t)y)}_{\in U_{r_1}(x_0)}\| + \|p_0 - p\|) dt \|x - y\| \\ &\leq \underbrace{ML2r_1}_{\equiv \kappa} \|x - y\| \\ &= \kappa \|x - y\| \end{aligned}$$

Therefore, if we take

$$r_1 < \frac{1}{2ML}, \text{ then } \kappa < 1.$$

The second condition of the Contraction Theorem is also satisfied. Let $y_0 = (x_0, p_0)$

$$\begin{aligned} \|F(x_0, p) - x_0\| &= \|F(x_0, p) - F(x_0, p_0)\| \\ &= \|G_x(y_0)^{-1}[G_x(y_0)x_0 - G(x_0, p)] - [G_x(y_0)^{-1}[G_x(y_0)x_0 - G(y_0)]]\| \\ &= \|G_x(y_0)^{-1}[G(y_0) - G(x_0, p)]\| \\ &\leq ML\|p - p_0\| \\ &\leq MLr \end{aligned}$$

where r (with $0 < r \leq r_1$) is to be chosen. We want the above to be less than or equal to $(1 - \kappa)r_1$, so we choose $r \leq \frac{(1-\kappa)r_1}{ML}$.

Hence for each $p \in V_\epsilon(p_0)$ we have a unique solution $x(p)$.

We now show that $x(p)$ is continuous in p . Let $p, \bar{p} \in V_\epsilon(p_0)$, with corresponding solutions $x(p)$ and $x(\bar{p})$. Then

$$\begin{aligned} \|x(p) - x(\bar{p})\| &= \|F(x(p), \bar{p}) - F(x(\bar{p}), \bar{p})\| \\ &\leq \|F(x(p), p) - F(x(\bar{p}), p)\| + \|F(x(\bar{p}), p) - F(x(\bar{p}), \bar{p})\| \\ &\leq \kappa \|x(p) - x(\bar{p})\| + \\ &\quad \|G_x(y_0)^{-1}[G_x(y_0)x(\bar{p}) - G(x(\bar{p}), p)] - G_x(y_0)^{-1}[G_x(y_0)x(\bar{p}) - G(x(\bar{p}), \bar{p})]\| \\ &\leq \underbrace{\kappa}_{<1} \|x(p) - x(\bar{p})\| + ML\|p - \bar{p}\| \end{aligned}$$

Hence,

$$\|x(p) - x(\bar{p})\| \leq \frac{ML}{1 - \kappa} \|p - \bar{p}\|$$

which concludes the (IFT). □

To prove the second result of (IFT) the following lemmas are useful.

Lemma 3.1.2. (Banach Lemma)

Let $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a linear operator with $\|A\| < 1$. Then $\|(I + A)^{-1}\|$ exist and

$$\|(I + A)^{-1}\| \leq \frac{1}{1 - \|A\|}$$

Lemma 3.1.3. Under the conditions of (IFT), there exists $M_1 > 0$ and $r > 0$ such that $G_p(x, p)^{-1}$ exist and $\|G_p(x, p)^{-1}\| \leq M_1$ in $U_r(x_0) \times V_r(p_0)$.

We are now ready to prove differentiability of the solution branch.

Proof. (II) Using the definition of (Fréchet) derivative, we are given that there exists $G_x(x, p)$ such that $G(x, p) - G(y, p) = G_x(x, p)(x - y) + O_1(x, y, p)$, where $O_1(x, y, p)$ is such that

$$\frac{\|O_1(x, y, p)\|}{\|x - y\|} \rightarrow 0 \text{ as } \|x - y\| \rightarrow 0 \quad (3.10)$$

Similarly, there exists $G_p(x, p)$ such that $G(x, p) - G(x, \hat{p}) = G_p(x, p)(p - \hat{p}) + O_2(x, p, \hat{p})$, where $O_2(x, p, \hat{p})$ such that

$$\frac{\|O_2(x, p, \hat{p})\|}{\|p - \hat{p}\|} \rightarrow 0 \text{ as } \|p - \hat{p}\| \rightarrow 0 \quad (3.11)$$

We must show that there exists $\frac{d}{dp}(x(p))$ such that

$$x(p) - x(\hat{p}) = \frac{d}{dp}(x(p))(p - \hat{p}) + o(p, \hat{p}),$$

with

$$\frac{\|o(p, \hat{p})\|}{\|p - \hat{p}\|} \rightarrow 0 \text{ as } \|p - \hat{p}\| \rightarrow 0.$$

Now

$$\begin{aligned} 0 &= G(x(p), p) - G(x(\hat{p}), \hat{p}) \\ &= G(x(p), p) - G(x(\hat{p}), p) + G(x(\hat{p}), p) - G(x(\hat{p}), \hat{p}) \\ &= G_x(x(p), p)(x(p) - x(\hat{p})) + O_1(x(p), x(\hat{p}), p) + G_p(x(\hat{p}), p)(p - \hat{p}) + O_2(x(\hat{p}), p, \hat{p}) \end{aligned}$$

Lemma(3.1.3) guarantees the existence of $G_x(x(p), p)^{-1}$, and we find

$$\begin{aligned} x(p) - x(\hat{p}) &= -G(x(p), p)^{-1}[G_p(x(\hat{p}), p)(p - \hat{p}) - (O_1 + O_2)] \\ &= -G(x(p), p)^{-1}[G_p(x(p), p)(p - \hat{p}) - o] \end{aligned}$$

where

$$o = [G_p(x(p), p) - G_p(x(\hat{p}), p)](p - \hat{p}) + O_1 + O_2$$

Let us, for the moment, ignore the harmless factor $G_x(x(p), p)^{-1}$ and consider each term of o . Since x and g_p are continuous, we have

$$\frac{\|[G_p(x(p), p) - G_p(x(\hat{p}), p)](p - \hat{p})\|}{\|p - \hat{p}\|} \rightarrow 0 \text{ as } \|p - \hat{p}\| \rightarrow 0$$

Also the existence of G_p implies (3.10)

$$\frac{\|O_2(x(p), p, \hat{p})\|}{\|p - \hat{p}\|} \rightarrow 0 \text{ as } \|p - \hat{p}\| \rightarrow 0$$

Using (3.11), we have

$$\frac{\|O_1(x(p), x(\hat{p}), p)\|}{\|p - \hat{p}\|} = \frac{\|O_1(x(p), x(\hat{p}), p)\|}{\|x(p) - x(\hat{p})\|} \frac{\|x(p) - x(\hat{p})\|}{\|p - \hat{p}\|} \rightarrow 0$$

as $\|p - \hat{p}\| \rightarrow 0$ because the second factor is bounded due to continuity of $x(p)$. Thus,

$$\frac{d}{dp}x(p) = G_x(x(p), p)^{-1}G_p(x(p), p)$$

To prove that $\frac{d}{dp}x(p)$ is continuous it suffices to show that $G_x(x(p), p)^{-1}$ is continuous. Indeed,

$$\begin{aligned} \|G_x(x(p), p)^{-1} - G_x(x(\hat{p}), \hat{p})^{-1}\| &= -G_x(x(p), p)^{-1}[G_x(x(\hat{p}), \hat{p}) - G_x(x(p), p)]G_x(x(\hat{p}), \hat{p})^{-1} \\ &\leq M_1^2 L(\|x(p) - x(\hat{p})\| + \|p - \hat{p}\|). \end{aligned}$$

□

3.2 Sensitivity Analysis for Parametric Nonlinear Optimization Problems

3.2.1 Optimality Conditions for Nonlinear Optimization problems

We shall recall the first order optimality conditions for an optimal solution x_0 of $(NLP(p))$ which is called the KKT-conditions under the *strong regularity condition* that the LICQ holds.

i.e., $rank(\nabla_x g^a(x_0, p)) = n_a$ and $rank(\nabla_x h(x_0, p)) = n_h$.

Theorem 3.2.1. (Strong Necessary Optimality Conditions for NLP(p))

Let x_0 be an optimal solution of $NLP(p)$ for which the Jacobian in $\nabla_x g^a(x_0, p)$, $\nabla_x h(x_0, p)$ have full rank n_a and n_h respectively. Then there exist a uniquely determined multipliers $(\lambda, \mu) \in \mathbb{R}^{n_g} \times \mathbb{R}^{n_h}$ satisfying

$$\begin{aligned} \nabla_x L(x_0, \lambda, \mu, p) &= 0_{n_x}, \\ \lambda_i g_i(x_0, p) &= 0, \quad \forall i \in \{1, \dots, n_g\}, \\ h_j(x_0, p) &= 0, \quad \forall j \in \{1, \dots, n_h\}, \\ \lambda_i &\geq 0, \quad \forall i \in \{1, \dots, n_g\} \end{aligned} \tag{3.12}$$

3.2.2 Post-optimal Calculation of Lagrange Multipliers

An accurate value of λ and μ can be calculated post-optimally once the optimal solution x_0 has been determined. The procedure uses the appropriate QR factorization of a matrix as described in Fiacco [1]. Under the assumption of linear independence of the binding constraint gradients, $P = \begin{pmatrix} \nabla_x g^a(x_0, p) \\ \nabla_x h(x_0, p) \end{pmatrix}$ has rank $= n_a + n_h := n_b$, and without loss of generality, it can be assumed that the sub-matrix involving the first $n_a + n_h := n_b$ columns of P is nonsingular. Hence there exist $n_b \times n_x$ matrix R and an orthogonal $n_x \times n_x$ matrix Q with

$$P = \begin{pmatrix} \nabla_x g^a(x_0, p) \\ \nabla_x h(x_0, p) \end{pmatrix} = RQ \quad (3.13)$$

The matrices R and Q can be partitioned into

$$R = [R_1 : \Theta], \quad Q = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} \quad (3.14)$$

with an upper triangular $n_b \times n_b$ invertible matrix R_1 and $n_b \times n_x$ matrix Q_1 respectively, $(n_x - n_b) \times n_x$ matrix Q_2 . From the fact $\lambda_i = 0, \forall i \notin A(x_0, p)$, the KKT conditions (3.12) gives

$$-\nabla_x f(x_0, p) = \begin{bmatrix} \lambda^a \\ \mu \end{bmatrix}^\top (R_1 Q_1) \quad (3.15)$$

Since the matrix R_1 is regular, we get the explicit expression

$$\begin{bmatrix} \lambda^a \\ \mu \end{bmatrix} = -(\nabla_x f(x_0, p) Q_1^\top R_1^{-1})^\top \quad (3.16)$$

3.2.3 Second Order Sufficient Optimality Conditions

Second order sufficient conditions (SOSC) are needed to ensure that any point x_0 that satisfies the KKT conditions (3.12) is indeed an optimal solution of problem NLP(p) [1]. Another important aspect of SOSC appears in the analysis of the problem NLP(p) where SOSC are indispensable for showing that the optimal solutions are differentiable functions of the parameter p .

Theorem 3.2.2. (Strong Second Order Sufficient Conditions for NLP(p))

For a given parameter p , let x_0 be an admissible point for the problem NLP(p) which satisfies the KKT conditions (3.12). Assume that

(i) *the gradients in $\nabla_x g^a(x_0, p), \nabla_x h(x_0, p)$ are linearly independent.
i.e., $\text{rank}(\nabla_x g^a(x_0, p)) = n_a, \text{rank}(\nabla_x h(x_0, p)) = n_h$*

(ii) *strict complementarity $\lambda^a > 0$ of the Lagrange multipliers holds*

(iii) *the Hessian of the Lagrangian is positive definite on*

$$T_C(x_0, p) = \{d : \nabla_x g^a(x_0, p)d = 0, \nabla_x h(x_0, p)d = 0\},$$

$$d^\top \nabla_x^2 L(x_0, \lambda, \mu, p) d > 0, \quad \forall d \neq 0_{n_x} \text{ with } d \in T_C(x_0, p) \quad (3.17)$$

Then there exist $\epsilon > 0$ and a constant $\alpha > 0$ such that

$$f(x, p) \geq f(x_0, p) + \alpha \|x - x_0\|^2, \quad \forall x \in \Phi(p), \quad \|x - x_0\|^2 \leq \epsilon.$$

In particular, x_0 is a strict local minimizer of $NLP(p)$.

Since the positive definiteness of the Hessian restricted to the null space of the Jacobian of the binding constraints a numerical check of the SOSC may be performed as follows. Consider the $n_x \times (n_x - (n_a + n_h))$ matrix H with full column rank whose column span the kernel of binding constraints. Any vector $v \in T_C(x_0, p)$ can be written as $v = Hw$ for some vector $w \in \mathbb{R}^{n_x - n_b}$, $n_b = n_a + n_h$. The condition may be restated as

$$w^\top H^\top \nabla_x^2 L(x_0, \lambda, \mu, p) H w > 0, \quad \forall w \in \mathbb{R}^{n_x - n_b}, \quad w \neq 0 \quad (3.18)$$

The matrix $H^\top \nabla_x^2 L(x_0, \lambda, \mu, p) H$ is called the *projected Hessian*. It follows from (3.18) that the positive definiteness of the projected Hessian on the whole space $\mathbb{R}^{n_x - n_b}$ is equivalent to the positive definiteness of the Hessian $\nabla_x^2 L(x_0, \lambda, \mu, p)$ on $T_C(x_0, p)$. Thus the test for SOSC proceeds by showing that the projected Hessian has only *positive* eigenvalues.

One efficient method to compute the matrix H is the *RQ* factorization [1]. Suppose that a RQ-factorization (3.13) and (3.14) is given. Then $H := Q_2^\top$ form an orthogonal basis for $T_C(x_0, p)$ which follows from

$$PH = RQQ_2 = R \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} Q_2^\top = [R_1 : \Theta] \begin{bmatrix} \Theta \\ I_{n_x - n_b} \end{bmatrix} = 0 \quad (3.19)$$

where $I_{n_x - n_b}$ denotes the identity matrix of dimension $n_x - n_b$. The matrices R_1 and Q_1 in (3.14) are unique while Q_2 is not. If \hat{Q}_2 is another orthogonal basis for the null space of binding constraints, then the corresponding projected Hessian is a matrix which is similar to the matrix $H^\top \nabla_x^2 L(x_0, \lambda, \mu, p) H$. Thus, the eigenvalues of the projected Hessian are independent of the orthogonal basis for $T_C(x_0, p)$.

3.2.4 First Order Sensitivity Analysis of Optimal Solution

Theorem 3.2.3. (Sensitivity Theorem, [1])

Let the triple (x_0, λ_0, μ_0) be an admissible point and corresponding Lagrangian multipliers which satisfy the strong SOSC of Theorem (3.2.2) for the nominal problem $(NLP(p_0))$. Then there exists neighborhood $V_\epsilon(p_0) \subset \mathbb{R}^{n_p}$ and continuously differentiable functions $x : V_\epsilon(p_0) \rightarrow \mathbb{R}^{n_x}$, $\lambda : V_\epsilon(p_0) \rightarrow \mathbb{R}^{n_g}$ and $\mu : V_\epsilon(p_0) \rightarrow \mathbb{R}^{n_h}$ with the following properties:

- (i) $x(p_0) = p_0$, $\lambda(p_0) = \lambda_0$, $\mu(p_0) = \mu_0$
- (ii) $A(x(p), p) = A(x_0, p_0)$ i.e., the active sets remain constant in $V_\epsilon(p_0)$
- (iii) the gradients in $\nabla_x g^a(x(p), p)$ and $\nabla_x h(x(p), p)$ are linearly independent,

- (iv) for all $p \in V_\epsilon(p_0)$, $(x(p), \lambda(p), \mu(p))$ satisfies the strong SOSC for the perturbed problem $NLP(p)$. In particular, $(x(p), \lambda(p), \mu(p))$ is a strict local minimum of $NLP(p)$ with gradient

$$\frac{d}{dp} \begin{pmatrix} x(p_0) \\ \lambda(p_0) \\ \mu(p_0) \end{pmatrix} = - \begin{pmatrix} \nabla_x^2 L & (\nabla_x g)^\top & (\nabla_x h)^\top \\ \Lambda_0 \cdot \nabla_x g & \Gamma_0 & \Theta \\ \nabla_x h & \Theta & \Theta \end{pmatrix}^{-1} \cdot \begin{pmatrix} \nabla_{xp}^2 L \\ \Lambda_0 \cdot \nabla_{pg} \\ \nabla_p h \end{pmatrix} \quad (3.20)$$

where $\Lambda_0 = \text{diag}(\lambda_{01}, \dots, \lambda_{0n_g})$, $\Gamma_0 = \text{diag}(g_1, \dots, g_{n_g})$. All functions and their derivatives are evaluated at $(x_0, \lambda_0, \mu_0, p_0)$.

Proof. The KKT conditions (3.12) for the unknown triple $(x, \lambda, \mu) = (x(p), \lambda(p), \mu(p))$ can be written in the form

$$F(x, \lambda, \mu, p) := \begin{pmatrix} \nabla_x L(x, \lambda, \mu, p)^\top \\ \Lambda \cdot g(x, p) \\ h(x, p) \end{pmatrix} = 0_{n_x+n_g+n_h}. \quad (3.21)$$

where $\Lambda := \text{diag}(\lambda_1, \dots, \lambda_{n_g})$. F is continuously differentiable and it holds

$$F(x_0, \lambda_0, \mu_0, p_0) = 0_{n_x+n_g+n_h}$$

At the nominal solution, the Jacobian of the mapping $F(x, \lambda, \mu, p)$ with respect to the variable (x, λ, μ) is given by

$$\frac{\partial}{\partial(x, \lambda, \mu)} F(x_0, \lambda_0, \mu_0, p_0) = \begin{pmatrix} \nabla_x^2 L(x_0, \lambda_0, \mu_0, p_0) & (\nabla_x g(x_0, p_0))^\top & (\nabla_x h(x_0, p_0))^\top \\ \Lambda_0 \cdot \nabla_x g(x_0, p_0) & \Gamma_0 & \Theta \\ h(x_0, p_0) & \Theta & \Theta \end{pmatrix} \quad (3.22)$$

We call this matrix the *Kuhn-Tucker matrix*. We intend to apply the implicit function theorem to show the non-singularity of the Kuhn-Tucker matrix. In order to show this, we assume without loss of generality, that the index set of active inequality constraint is given by $A(x_0, p_0) = r + 1, \dots, n_g$, where r is the number of inactive inequality constraints. Then, the strict complementarity condition implies

$$\Lambda_0 = \begin{pmatrix} \Theta & \Theta \\ \Theta & \Lambda^a \end{pmatrix} \quad \Gamma_0 = \begin{pmatrix} \Gamma_0^1 & \Theta \\ \Theta & \Theta \end{pmatrix}$$

with non-singular matrices

$$\Lambda^a := \text{diag}(\lambda_{r+1}^a, \dots, \lambda_{n_g}^a) \quad \text{and} \quad \Gamma_0^1 := \text{diag}(g_1(x_0, p_0), \dots, g_r(x_0, p_0))$$

Now consider the linear equation

$$\begin{pmatrix} \nabla_x^2 L(x_0, \lambda_0, \mu_0, p_0) & (\nabla_x g(x_0, p_0))^\top & (\nabla_x h(x_0, p_0))^\top \\ \Lambda_0 \cdot \nabla_x g(x_0, p_0) & \Gamma_0 & \Theta \\ h(x_0, p_0) & \Theta & \Theta \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 0_{n_x} \\ 0_{n_g} \\ 0_{n_h} \end{pmatrix}$$

for $u_1 \in \mathbb{R}^{n_x}$, $u_2 = (u_{21}, u_{22})^\top \in \mathbb{R}^{r+(n_g-r)}$ and $u_3 \in \mathbb{R}^{n_h}$. Exploitation of the special structure of Λ_0 and Γ_0 yields $\Gamma_0^1 u_{21} = 0_{\mathbb{R}^r}$. Since Γ_0^1 is nonsingular it follows that $u_{21} = 0_r$. With this,

by multiplying the second block equation with $(\Lambda^a)^{-1}$ it remains to investigate the reduced system

$$\begin{pmatrix} A & B^\top & C^\top \\ B & \Theta & \Theta \\ C & \Theta & \Theta \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 0_{n_x} \\ 0_{n_g-r} \\ 0_{n_h} \end{pmatrix}$$

with $A := \nabla_x^2 L(x_0, \lambda_0, \mu_0, p_0)$, $B := (\nabla_x g_i(x_0, p_0))$, $i = r+1, \dots, n_g$, and $C := \nabla_x h(x_0, p_0)$. The last two block equations yield $Bu_1 = 0_{n_g-r}$ and $Cu_1 = 0_{n_h}$. Multiplication of the first block equation from the left by u_1^\top yields

$$0 = u_1^\top Au_1 + (Bu_1)^\top u_2 + (Cu_1)^\top u_3 = u_1^\top Au_1$$

Since A is positive definite on $T_C(x_0) \setminus 0_{n_x}$, i.e. it holds $d^\top Ad > 0$ for all $d \neq 0_{n_x}$, with $Bd = 0_{n_g-r}$ and $Cd = 0_{n_h}$, it follows that $u_1 = 0_{n_x}$. Taking this property into account, the first block equation reduces to $(Bu_1)^\top u_2 + (Cu_1)^\top u_3 = 0_{n_x}$. By the linear independence of the gradient $\nabla_x g_i(x_0, p_0)$ $i \in A(x_0, p_0)$ and $\nabla_x h_j(x_0, p_0)$, $j \in \{1, \dots, n_h\}$ we obtain $u_2 = 0_{n_g-r}$, $u_3 = 0_{n_h}$. Putting all together, the above linear equation has the unique solution $u_1 = 0_{n_x}$, $u_2 = 0_{n_g-r}$, $u_3 = 0_{n_h}$, which implies that the Kuhn-Tucker matrix is non-singular and the implicit function theorem is applicable.

By the implicit function theorem there exist neighborhoods $V_\epsilon(p_0)$ and $U_\delta(x_0, \lambda_0, \mu_0)$, and a uniquely defined functions

$$(x(\cdot), \lambda(\cdot), \mu(\cdot)) : V_\epsilon(p_0) \longrightarrow U_\delta(x_0, \lambda_0, \mu_0)$$

satisfying

$$F(x(p), \lambda(p), \mu(p), p) = 0_{n_x+n_g+n_h}, \quad \forall p \in V_\epsilon(p_0) \quad (3.23)$$

Furthermore, these functions are continuously differentiable and the differentiation of the identity $F(x(p), \lambda(p), \mu(p), p) = 0_{n_x+n_g+n_h}$ at the nominal parameter p_0 then yields the following system of linear equations for the *sensitivity differentials* of optimal solutions and multipliers:

$$\begin{pmatrix} \nabla_x^2 L(\cdot) & \nabla_x g(\cdot)^\top & \nabla_x h(\cdot)^\top \\ \Lambda_0 \cdot \nabla_x g(\cdot) & \Gamma_0 & \Theta \\ h(\cdot) & \Theta & \Theta \end{pmatrix} \frac{d}{dp} \begin{pmatrix} x(p_0) \\ \lambda(p_0) \\ \mu(p_0) \end{pmatrix} + \begin{pmatrix} \nabla_{xp}^2 L(\cdot) \\ \Lambda_0 \cdot \nabla_p g(\cdot) \\ \nabla_p h(\cdot) \end{pmatrix} = 0$$

All functions and their derivatives are evaluated at $(x_0, \lambda_0, \mu_0, p_0)$. Hence we obtain the explicit formulae (3.20) for the sensitivity differentials.

It remains to verify, that $x(p)$ actually is a strong(strict) local minimum of $(\text{NLP}(p))$. The continuity of the functions $x(p), \lambda(p)$ and g together with $\lambda_i(p_0) > 0, i = r+1, \dots, n_g$ and $g_i(x(p_0), p_0) = g_i(x_0, p_0) < 0, i = 1, \dots, r$ guarantees $\lambda_i(p) > 0, i = r+1, \dots, n_g$ and $g_i(x_0, p_0) < 0, i = 1, \dots, r$ for p sufficiently close to p_0 . From (3.23) it follows that $g_i(x(p), p) = 0, i = r+1, \dots, n_g$ and $h_j(x(p), p) = 0, j = 1, \dots, n_h$. Thus, $x(p) \in \Phi(p)$ and the KKT conditions are satisfied. Furthermore, the index set $A(x(p), p) = A(x_0, p_0)$ remains unchanged in a neighborhood of p_0 . Finally, we have to show that $\nabla_x^2 L(x(p), \lambda(p), \mu(p), p)$ remains positive definite on $T_C(x(p))$ for p sufficiently close to p_0 . Notice, that the critical cone $T_C(x(p))$ varies with p . By now, we only know that $\nabla_x^2 L(p_0) = \nabla_x^2 L(x_0, \lambda_0, \mu_0, p_0)$ is positive

definite on $T_C(x(p_0))$ which is equivalent to the existence of $c > 0$ with $d^\top \nabla_x^2 L(p_0) d \geq c \|d\|^2$ for all $d \in T_C(x(p_0))$. Owing to the strict complementarity in a neighborhood of p_0 it holds

$$T_C(x(p)) = \{d \in \mathbb{R}^{n_x} \mid \nabla_x g_i(x(p), p) d = 0, i \in A(x_0, p_0), \nabla_x h_j(x(p), p) d = 0, j = 1 : n_h\}$$

around p_0 . Assume that for every $i \in \mathbb{N}$ there exist some $p^i \in \mathbb{R}^{n_p}$ with $\|p^i - p_0\| \leq \frac{1}{i}$ such that for all $j \in \mathbb{N}$ there exist some $d^{ij} \in T_C(x(p^i))$, $d^{ij} \neq 0_{n_x}$, with

$$d^{ij\top} \nabla_x^2 L(x(p^i)) d^{ij} < \frac{1}{j} \|d^{ij}\|^2$$

Since the unit ball w.r.t $\|\cdot\|$ is compact in \mathbb{R}^{n_x} , there exist a convergent subsequence $\{p^{i_k}\}$ with $\lim_{k \rightarrow \infty} p^{i_k} = p_0$ and

$$\lim_{k \rightarrow \infty} \frac{d^{i_k j}}{\|d^{i_k j}\|} = \hat{d}, \quad \|\hat{d}\| = 1, \quad \hat{d} \in T_C(x(p_0))$$

and

$$\hat{d}^\top \nabla_x^2 L(p_0) \hat{d} \leq 0.$$

This contradicts the positive definiteness of $\nabla_x^2 L(p_0)$. □

Theorem 3.2.4. (Solution Differentiability without Strict Complementarity, [6])

At a local solution x_0 of the nominal problem ($NLP(p_0)$), assume that

a) the LICQ holds, i.e., the gradients $\{\nabla_x g_i(x_0, p_0), i \in A(x_0, p_0), \nabla_x h_j(x_0, p_0), j = 1, \dots, n_h\}$ are linearly independent.

b) the strong SOSC hold. Then

(i) x_0 is local isolated minimizing point of ($NLP(p_0)$) and the associated multipliers λ_0 and μ_0 are unique

(ii) for $p \in V_\epsilon(p_0)$, there exists a unique 'continuous' vector function $(x(p), \lambda(p), \mu(p))$ satisfying the strong SOSC for a local solution of the problem ($NLP(p)$) such that $(x(p_0), \lambda(p_0), \mu(p_0)) = (x_0, \lambda_0, \mu_0)$ and hence $x(p)$ is a locally unique minimizer of $NLP(p)$ with associated unique multipliers $\lambda(p)$ and $\mu(p)$

(iii) the gradients $\{\nabla_x g_i^a(x(p), p), \nabla_x h_j(x(p), p), j = 1, \dots, n_h\}$ are linearly independent.

(iv) there exist $0 < \alpha, \beta, \gamma < \infty$ and $\epsilon > 0$ such that, for all p with $\|p - p_0\| < \epsilon$,

- a) $\|x(p) - x_0\| \leq \alpha \|p - p_0\|$
- b) $\|\lambda(p) - \lambda_0\| \leq \beta \|p - p_0\|$
- c) $\|\mu(p) - \mu_0\| \leq \gamma \|p - p_0\|$

(v) in any direction $z \neq 0$, the (uniquely determined) directional derivative of the components of $(x(p), \lambda(p), \mu(p))$ exists at $p = p_0$

(vi) $v(p) = f(x(p), p)$ is differentiable w.r.t. p at p_0 with $\nabla_p v(p_0) = \nabla_p L(x_0, \lambda_0, \mu_0, p_0)$

Corollary 3.2.1. (First-Order Estimation of $(x(p), \lambda(p), \mu(p))$ in $V_\epsilon(p_0)$)

Under the assumptions of Theorem (3.2.3) a first order estimation of $(x(p), \lambda(p), \mu(p))$ for $p \in V_\epsilon(p_0)$ is given by

$$\begin{pmatrix} x(p) \\ \lambda(p) \\ \mu(p) \end{pmatrix} = \begin{pmatrix} x_0 \\ \lambda_0 \\ \mu_0 \end{pmatrix} + \nabla_p \begin{pmatrix} x(p_0) \\ \lambda(p_0) \\ \mu(p_0) \end{pmatrix} (p - p_0) + o(\|p - p_0\|) \quad (3.24)$$

and $\phi(p) = o(\|p\|)$ means that $\phi(p)/\|p\| \rightarrow 0$ as $p \rightarrow p_0$

And an approximation of optimal solution $x(p)$ is obtained by the linear approximation

$$x(p) \approx x_0 + \nabla_p x(p_0)(p - p_0) \quad (3.25)$$

3.2.5 First and Second-Order Parameter Derivatives of the Optimal Value Function

Let $(x(p), \lambda(p), \mu(p))^\top$ be a KKT triple, where $x(p)$ solves problem NLP(p) for $p \in V_\epsilon(p_0)$. Then the "optimal value Lagrangian" is defined as

$$\bar{L}(p) := L(x(p), \lambda(p), \mu(p), p) = f(x(p), p) + \lambda(p)^\top g(x(p), p) + \mu(p)^\top h(x(p), p)$$

Theorem 3.2.5. (First and Second-Order Changes in the Optimal Value Function of NLP(p))

If the conditions of Theorem (3.2.3) hold for the problem NLP(p) and if the problem functions are twice continuously differentiable in (x, p) near (x_0, p_0) , then in the neighbourhood of p_0

(a) $v(p) = \bar{L}(p)$

(b) $\nabla_p v(p) = \nabla_p L = \nabla_p f(x(p), p) + \lambda(p)^\top \nabla_p g(x(p), p) + \mu(p)^\top \nabla_p h(x(p), p)$

(c) $\nabla_p^2 v(p) = \nabla_p [\nabla_p L(x(p), \lambda(p), \mu(p), p)]$

Proof. a) Recall that for $p \in V_\epsilon(p_0)$, $\lambda_i(p)g_i(x(p), p) = 0$, $i = 1, \dots, n_g$, strict complementary slackness holds, $h_j(x(p), p) = 0$, $j = 1, \dots, n_h$, and $(x(p), \lambda(p), \mu(p))^\top$ is continuously differentiable. It follows that

$$v(p) = f(x(p), p) = L(x(p), \lambda(p), \mu(p), p) = \bar{L}(p)$$

b) $v(p) = f(x(p), p)$

$$\begin{aligned} \nabla_p v(p) &= \nabla_p f(x(p), p) = \nabla_p \bar{L}(p) \\ &= \nabla_p L(x(p), \lambda(p), \mu(p), p) \\ &= \nabla_x L(x(p), \lambda(p), \mu(p), p) + \nabla_\lambda L(x(p), \lambda(p), \mu(p), p) \nabla_p \lambda(p) \\ &\quad + \nabla_\mu L(x(p), \lambda(p), \mu(p), p) \nabla_p \mu(p) + \nabla_p L(x(p), \lambda(p), \mu(p), p). \end{aligned}$$

Since the KKT conditions hold at $(x(p), \lambda(p), \mu(p))$ for $p \in V_\epsilon(p_0)$, it follows that $\nabla_x L(x(p), \lambda(p), \mu(p), p) = 0_{n_x}$. Strict complimentary slackness, continuity, and differentiability the imply (i) $g_i(x_0, p_0) < 0$, implying $g_i(x(p), p) < 0$, implying $\lambda_i(p) = 0$, implying $\nabla_p \lambda_i(p) = 0$ for $p \in V_\epsilon(p_0)$; or (ii) $\lambda_i(p_0) > 0$, implying $\lambda_i(p) = 0$, implying $g_i(x(p), p) = 0$ for $p \in V_\epsilon(p_0)$. From (i) and (ii) it follows that

$\nabla_\lambda L(x(p), \lambda(p), \mu(p), p) \nabla_p \lambda(p) = (g_1(x(p), p), \dots, g_{n_g}(x(p), p)) \nabla_p \lambda(p) = 0$. Also since $h_j(x(p), p) = 0$ for $p \in V_\epsilon(p_0)$
 $\nabla_\mu L(x(p), \lambda(p), \mu(p), p) \nabla_p \mu(p) = (h_1(x(p), p), \dots, h_{n_h}(x(p), p)) \nabla_p \mu(p) = 0$. Thus

$$\nabla_p v(p) = \nabla_p f(x(p), p) = \nabla_p L(x(p), \lambda(p), \mu(p), p) \quad (3.26)$$

c) Differentiating the result obtained in (b) with respect to p gives

$$\begin{aligned} \nabla_p^2 v(p) &= \nabla_{xp}^2 L(\cdot) \nabla_p x(p) + \nabla_{\lambda p}^2 L(\cdot) \nabla_p \lambda(p) + \nabla_{\mu p}^2 L(\cdot) \nabla_p \mu(p) + \nabla_p^2 L(\cdot) \\ &= \nabla_x \left\{ \nabla_p f(x(p), p)^\top + \sum_{i=1}^{n_g} \lambda_i(p) \nabla_p g_i(x(p), p) + \sum_{j=1}^{n_h} \mu_j(p) \nabla_p h_j(x(p), p) \right\} \nabla_p x(p) \\ &\quad + \left\{ \nabla_p^2 f(x(p), p) + \sum_{i=1}^{n_g} \lambda_i(p) \nabla_p^2 g_i(x(p), p) + \sum_{j=1}^{n_h} \mu_j(p) \nabla_p^2 h_j(x(p), p) \nabla_p x(p) \right\} \\ &\quad + \left\{ \sum_{i=1}^{n_g} \nabla_p g_i(x(p), p)^\top \nabla_p \lambda_i(p) + \sum_{j=1}^{n_h} \nabla_p h_j(x(p), p)^\top \nabla_p \mu_j(p) \right\}. \end{aligned}$$

After simplification we find

$$\nabla_p^2 v(p) = \nabla_{xp}^2 L(\cdot) \nabla_p x(p) + \sum_{i=1}^{n_g} \nabla_p g_i^\top \nabla_p \lambda_i(p) + \sum_{j=1}^{n_h} \nabla_p h_j^\top \nabla_p \mu_j(p) + \nabla_p^2 L(\cdot) \quad (3.27)$$

The numerical advantage of this second order representation is its independence from the Jacobian of the constraints and the derivative of the multipliers and can be used to estimate the error of first order approximation give in (3.27).

□

3.2.6 Linear Perturbations/ RHS perturbations

The sensitivity differentials of optimal solutions, simplifies considerably if the constraints in (NLP(p)) involve linear perturbations in p in the form

$$\begin{cases} \min_x & f(x), \\ \text{subject to} & g_i(x) - p \leq 0, \quad i = 1, \dots, n_g \\ & h_j(x) - p = 0, \quad j = 1, \dots, n_h \end{cases} \quad (3.28)$$

Then we have $\nabla_{xp} L(x_0, \lambda_0^a, \mu_0, p_0) = 0$, $\nabla_p g^a(x_0, p_0) = -I_{n_a}$, $\nabla_p h(x_0, p_0) = -I_{n_h}$. Let

$$\begin{pmatrix} m_{i,j} \end{pmatrix}_{i,j=1,\dots,n_x, n_a+n_h} = \begin{pmatrix} \nabla_x^2 L(\cdot) & \nabla_x g^a(\cdot)^\top & \nabla_x h(\cdot)^\top \\ \nabla_x g^a(\cdot) & \Theta & \Theta \\ \nabla_x h(\cdot) & \Theta & \Theta \end{pmatrix}^{-1}$$

be the inverse of the Kuhn-Tucker matrix. Then with $p = p_1 \dots, p_{n_a+n_h}$ it follows that

$$\frac{dx_i}{dp_j}(p_0) = m_{i,j+n_x}, \quad i = 1, \dots, n_x, \quad j = 1, \dots, n_a + n_h$$

$$\frac{d\lambda_i^a}{dp_j}(p_0) = m_{i+n_x,j+n_x}, \quad i = 1, \dots, n_a, \quad j = 1, \dots, n_a$$

$$\frac{d\mu_i^a}{dp_j}(p_0) = m_{i+n_x,j+n_x}, \quad i = 1, \dots, n_h, \quad j = 1, \dots, n_h$$

Moreover, the sensitivity differential of the optimal value function reduces to

$$\frac{dv}{dp}(p_0) = \begin{pmatrix} \lambda^a(p_0) \\ \mu(p_0) \end{pmatrix} \quad (3.29)$$

we may differentiate again and obtain the second order derivative

$$\frac{d^2v}{dp^2}(p_0) = \frac{d}{dp} \begin{pmatrix} \lambda^a(p_0) \\ \mu(p_0) \end{pmatrix} \quad (3.30)$$

Chapter 4

Perturbed Optimization Problem in Real-Time

4.1 Approximation of Perturbed Solutions

The differentiability properties of optimal solutions to the parametric nonlinear programming problems are important in approximation of perturbed solutions. The knowledge of the nominal solution $x_0 = x(p_0)$ and the sensitivity differentials $\frac{dx}{dp}(p_0)$ allow to approximate the perturbed solutions $x(p)$ by its first order Taylor expansion

$$x(p) \approx x(p_0) + \frac{dx}{dp}(p_0)(p - p_0) \quad (4.1)$$

Here, parametric programming detaches from the sensitivity analysis theory. While sensitivity analysis stops here, it is where we know what happens if the process conditions deviate from the nominal values to some value in its neighborhood, parametric programming is concerned with the whole range of the parametric variability.

Should the first order not accurate enough, a further improvement real-time approximations in views of optimality and feasibility can be achieved by the idea in [2]

One way to estimate the quality of approximate perturbed solutions of (NLP(p)) in (4.1) by their first Taylor's series expansions with respect to the parameter is to evaluate the Taylor's expansion of the objective function. If $x(p)$ is a solution of (NLP(p)) satisfying the strong SOSC, then a first-order estimate of optimal value function $f(x(p), p)$ for $p \in V_\epsilon(p_0)$ is given by

$$v(p) = f(x(p), p) \approx f(x_0, p_0) + \nabla_p f(p_0)(p - p_0) \quad (4.2)$$

and the error in the first order approximation can be further estimated from the second-order Taylor's expansion given by

$$v(p) = f(x(p), p) \approx f(x_0, p_0) + \nabla_p f(p_0)(p - p_0) + \frac{1}{2}(p - p_0)^\top \nabla_p^2 f(p_0)(p - p_0) \quad (4.3)$$

Illustrative Example

Consider an optimization problem in two variables $x = (x_1, x_2)$

$$\begin{aligned} \min_x f(x, p) &= -(0.5 + p)\sqrt{x_1} - (0.5 - p)x_2 \\ \text{s.t. } g_1(x, p) &= x_1 + x_2 \leq 1, \\ g_2(x, p) &= 0.10 - x_1 \leq 0 \end{aligned}$$

The nominal parameter $p_0 = 0$.

The first order optimality conditions that $x(p)$ solves this problem require that for a given p , there exist $\lambda_1(p), \lambda_2(p) \geq 0$ such that $x(p)$ feasible and satisfy for

$$\begin{aligned} L(x, \lambda, \mu, p) &= -(0.5 + p)\sqrt{x_1} - (0.5 - p)x_2 - \lambda_1(x_1 + x_2 - 1) + \lambda_2(0.1 - x_1) \\ -0.5(0.5 + p)x^{-\frac{1}{2}} + \lambda_1 - \lambda_2 &= 0 \\ -0.5 + p + \lambda_1 &= 0 \\ \lambda_1(x_1 + x_2 - 1) &= 0 \\ \lambda_2(0.1 - x_1) &= 0 \end{aligned}$$

Hence the optimal solution for parameters p in small neighborhood of $p_0 = 0$ is given by

$$x_1(p) = \left(\frac{0.5 + p}{1 - 2p}\right)^2, \quad x_2(p) = 1 - x_1(p), \quad \lambda_1(p) = \lambda^a(p) = 0.5 - p$$

The active set $A(x(p), p) = \{1\}$ and hence the Lagrangian is

$$L(x, \lambda^a, p) = -\left(\frac{1}{2} + p\right)\sqrt{x_1} - \left(\frac{1}{2} - p\right)x_2 - \lambda^a(x_1 + x_2 - 1).$$

The nominal solution is $x_0 = (0.25, 0.75)$ and $\lambda_0^a = 0.5$. Though the Hessian is $\nabla_x^2 L(x_0, \lambda_0^a, p_0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ is no positive definite on the whole space \mathbb{R}^2 , the strong SOSC are satisfied in view of $\nabla_x g(x_0, p_0) = (1, 1)$, and $\ker(\nabla_x g^a(x_0, p_0)) = \mathbb{R} \cdot (1, -1)$. The matrices in (3.20) are

$$\begin{pmatrix} \nabla_x^2 L & \nabla_x (g^a)^\top \\ \nabla_x g^a & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} \nabla_{xp} L \\ \nabla_p g^a \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}.$$

Then the sensitivity differentials is $\frac{d}{dp}(x_1, x_2, \lambda^a) = (2, -2, -1)^\top$ from which we can get the approximation

$$\begin{pmatrix} x_1(p) \\ x_2(p) \\ \lambda^a(p) \end{pmatrix} \approx \begin{pmatrix} 0.25 \\ 0.75 \\ 0.5 \end{pmatrix} + \begin{pmatrix} 2 \\ -2 \\ -1 \end{pmatrix} p$$

to the quality of this approximation, let us consider the perturbation $p = 0.05$. The exact Hessian computed is $(x(0.05), \lambda^a(0.05)) = (0.3735, 0.6265, 0.45)$.

The first order estimation yields the value $(0.35, 0.65, 0.45)$ which is acceptable with the exact value. The second order analysis of the optimal value function is as follows. The exact value of the function is $f(0.05) = -0.6181$. The first order sensitivity derivative is $\frac{df}{dp} = \nabla_p L = -\sqrt{x_1} + x_2 = 0.25$ and the second order sensitivity derivative is $\frac{d^2 f}{dp^2} = \frac{d\lambda^a}{dp} = -1$. Then the first order expansion(4.2) gives us a value -0.6125 which can be improved by second order expansion(4.3) where the value is -0.615.

4.2 Prediction of Sensitivity Domain

The linearization in (3.24) is only justified in some neighborhood of the nominal parameter p_0 . Unfortunately, the sensitivity theorem does not indicate how large this neighborhood is. In particular, if the index set of active inequality constraints changes, then sensitivity theorem is not applicable. When dealing with construction of approximate optimal solution of perturbed problem in real time of the form (4.1) one has to ensure that a change of parameter p does not change the active of active constraints $A(x_0, p_0)$. In general we need to deal with two cases: a constraint that enters the active set and a constraint that leave the active set. Thus we need to determine the sensitivity domain when the perturbation is too large in order to apply Theorem (3.2.3). The following explanation is based on [5].

The first order Taylor expansion of the active Lagrange multipliers is give by

$$\lambda^a(p) \approx \lambda^a(p_0) + \nabla_p \lambda^a(p_0)(p - p_0) \quad (4.4)$$

A constraint will leave the active set when the corresponding multiplier will go from nonzero value to zero. Hence if one of the multipliers in (4.4) approximates zero.

$$0 = \lambda_i^a(p) \approx \lambda_i^a(p_0) + \nabla_p \lambda_i^a(p_0)(p^i - p_0), \quad i \in A(x_0, p_0) \quad (4.5)$$

From this it follows that the approximate $p^i = (p_1^i, \dots, p_{n_p}^i)^\top$ that causing a constraint g_i to leave the active set is given by

$$p_j^i \approx (p_0)_j - \frac{\lambda_i^a(p_0)}{\nabla_{p_j} \lambda_i^a(p_0)}, \quad i \in A(x_0, p_0), \quad j \in \{1, \dots, n_p\} \quad (4.6)$$

provided that $\nabla_{p_j} \lambda_i^a(p_0) \neq 0$

In similar manner a constraint g_i $i \notin A(x_0, p_0)$ is zero.

$$g_i(x(p), p) \approx g_i(x_0, p_0) + \nabla_p g_i(x_0, p_0)(p^i - p_0), \quad i \notin A(x_0, p_0) \quad (4.7)$$

where $\nabla_p g_i(x_0, p_0) = \frac{\partial}{\partial x} g_i(x_0, p_0) \nabla x(p_0) + \frac{\partial}{\partial p} g_i(x_0, p_0)$, $i \notin A(x_0, p_0)$. Hence an approximation of perturbation $p^i = (p_1^i, \dots, p_{n_p}^i)^\top$ causing the constraint g_i to enter the active set is give by

$$p_j^i \approx (p_0)_j - \frac{g_i(x_0, p_0)}{\nabla_{p_j} g_i(x_0, p_0)}, \quad i \notin A(x_0, p_0), \quad j \in \{1, \dots, n_p\} \quad (4.8)$$

provided that $\nabla_{p_j} g_i(x_0, p_0) \neq 0$

Therefore, the sensitivity domain P_0 is determined by those values p_j^i which are closest to the the nominal perturbation $(p_0)_j$ is

$$\begin{aligned} P_0 &\approx P_0^1 \times P_0^2 \cdots \times P_0^{n_g}, \\ P_0^j &= \left[\max_{\bar{p}_j < (p_0)_j} \{\bar{p}_j \in \bar{P}_j\}, \min_{\bar{p}_j > (p_0)_j} \{\bar{p}_j \in \bar{P}_j\} \right], \quad j = 1, \dots, n_p \\ \bar{P}_j &= \{p_j^i : i = 1, \dots, n_g\} \cup (-\infty, \infty) \end{aligned} \quad (4.9)$$

4.3 Handling Larger Perturbations

In general, sensitivity derivatives do not exist at points where the active set changes. However, we can show that at least the directional derivatives exist [6], [7]. Hence, one strategy

for dealing with changes in the active set is based on calculating the directional derivatives. The following method for dealing with constraints entering or leaving the active set is proposed by C. Büsken and H. Maurer in [5].

1. Calculate the optimal solution $(x(p_0^1), \lambda^a(p_0^1))$ and the sensitivity differentials $(\frac{d}{dp}x(p_0^1), \frac{d}{dp}\lambda^a(p_0^1))$ at the nominal value $p_0^1 = p_0$
2. Calculate the sensitivity domain as in (4.9). Let p_0^2 denote the perturbation that cause a constraint to enter or to leave the active set.
3. Calculate the sensitivity differentials $(\frac{d}{dp}x(p_0^2), \frac{d}{dp}\lambda^a(p_0^2))$ at the value p_0^2 with the active set updated to reflect the change in step 2. Let λ^c denote the updated Lagrange multiplier.
4. Calculate the first order changes as

$$\Delta x := \frac{d}{dp}x(p_0^1)(p_0^2 - p_0^1) + \frac{d}{dp}x(p_0^2)(p - p_0^2) \quad (4.10)$$

$$\Delta \lambda := \frac{d}{dp}\lambda^c(p_0^1)(p_0^2 - p_0^1) + \frac{d}{dp}\lambda^c(p_0^2)(p - p_0^2) \quad (4.11)$$

5. Calculate the new first order approximations by

$$x(p) \approx x(p_0) + \Delta x, \quad \lambda^c(p) \approx \lambda^c(p_0) + \Delta \lambda^c, \quad \text{if } p - p_0^2 \geq 0 \quad (4.12)$$

otherwise by (3.20).

The approximation is also leads to the first and second order Taylor expansions of optimal value function.

Numerical Example of Predicting Sensitivity Domain

Consider an optimization problem in two variables $x = (x_1, x_2)^\top \in \mathbb{R}^2$ and $p \in \mathbb{R}$

$$\begin{aligned} \min_x f(x, p) &= (x_1 + 1)^2 + (x_2 - 2)^2 \\ \text{s.t. } g_1(x, p) &= -x_1 + p \leq 0, \\ g_2(x, p) &= 2x_1 + x_2 - 6 \leq 0 \end{aligned}$$

The nominal parameter $p_0 = 1$.

The necessary optimality conditions of theorem (3.2.1) gives the optimal solution for parameters p in small neighborhood of $p_0 = 1$

$$x(p) = (p, 2), \quad \lambda_1(p) = \lambda^a(p) = 2p + 2$$

The active set $A(x(p), p) = \{1\}$ and $f(x(p), p) = (p + 1)^2$. Hence the optimal candidate is $x_0 = (1, 2)$, $\lambda_0^a = (4, 0)$. The nominal value is $f(x_0, p_0) = 4$. The SOSC holds in view of $\nabla_x^2 L = I_2$.

To compare this exact value with the approximate solutions (4.1)-(4.3), the sensitivities of the optimization variables using (3.20) and Lagrange multipliers are computed as $\frac{dx}{dp}(1) = (1, 0)$

and $\frac{d\lambda^a}{dp}(1) = 2$. Since the perturbation appears linearly applying (4.2) and (4.3) we get the first and second order sensitivity derivatives for the objective function $\frac{df}{dp}(1) = 4$ and $\frac{d^2f}{dp^2}(1) = 2$. The sensitivity domain calculated from (4.6) and (4.8) as $P_0 = [-1, 2]$.

Now applying the idea of approximating perturbed solution for $p = 2$. Since the perturbation is inside the sensitivity domain P_0 , we can use (4.1) for real time optimization. we get $(x(2), 2) \approx (2, 2)$ which agree with the exact solution. The first order Taylor expansion (4.2) of the objective function gives $f(x(2), 2) \approx 8$ and the second order Taylor expansion (4.3) yields $f(x(2), 2) \approx 9$ which agree with the exact optimal value.

Next, we investigate a perturbation $p = 2.5$ which is outside the sensitivity domain. Using (4.1) we obtain an infeasible point $x(2.5) = (2.5, 2) \notin S(2.5)$. Using step 2 of the strategy handling larger perturbation, the constraint g_2 enters the active set when $p = p_0^2 = 2$. For value of p greater than p_0^2 new search direction along constraints g_1 and g_2 is determined in step 3 where

$$x(p) = (p, 6 - 2p), \quad \lambda_1(p) = 10p - 14, \quad \lambda_2(p) = 4p - 8$$

The updated active set $A(x(p), p) = \{1, 2\}$ and $f(x(p), p) = (p + 1)^2 + (4 - 2p)^2$. The sensitivity differentials is $\frac{dx}{dp}(2) = (1, -2)$ and $\frac{d\lambda}{dp}(2) = (10, 4)$. The formula (4.12) gives an estimate of the new optimum $x(2.5) \approx (2.5, 1)$, which agrees with the optimal solution.

4.4 Solution Method for Parametric Quadratic Optimization

In this section we will consider the following class of problems

$$\begin{aligned} v(p) = \min & \frac{1}{2}x^\top Qx + p^\top F^\top x + c^\top x \\ \text{s.t.} & G_i x \leq b_i + S_i p, \quad i \in \mathcal{I} \\ & x \in X \subseteq \mathbb{R}^{n_x} \\ & p \in P \subseteq \mathbb{R}^{n_p} \end{aligned} \tag{mpQP}$$

where $Q \in \mathbb{R}^{n_x \times n_x}$ is a symmetric positive definite constant matrix, $F \in \mathbb{R}^{n_x \times n_p}$, $c \in \mathbb{R}^{n_x \times 1}$, $G \in \mathbb{R}^{m \times n_x}$, $b \in \mathbb{R}^{n_p \times 1}$, $S \in \mathbb{R}^{m \times n_p}$ are matrices, X and P are compact polyhedral convex sets. In the sequel, let the subscript index denote a subset of the rows of a matrix or vector

The set $\mathcal{A}(x, p) = \{i \in \{1, \dots, m\} : G_i x = b_i + S_i p\}$ is the set of indices of the active constraints.

4.4.1 Background on (mpQP)

As in [2] we solve the (mpQP) by formulating the KKT conditions

$$\begin{aligned} Qx + Fp + c + G^\top \lambda &= 0, \quad \lambda \in \mathbb{R}^m \\ \lambda_i(G_i x - b_i - S_i p) &= 0, \quad \forall i \in \mathcal{I} \\ G_i x - b_i - S_i p &\leq 0, \quad \forall i \in \mathcal{I} \\ \lambda_i &\geq 0, \quad \forall i \in \mathcal{I} \end{aligned} \tag{4.13}$$

Since Q has full rank, the first system in (4.13) gives

$$x = -Q^{-1}[Fp + c + G^T \lambda] \quad (4.14)$$

Definition 4.4.1. Let $\hat{x}(p)$ be the optimal solution to (mpQP) for a given p . We define active constraints the constraints with $G_i \hat{x}(p) = b_i + S_i p$, and inactive constraints the constraints with $G_i \hat{x}(p) - b_i - S_i p < 0$. The optimal active set is the set of indices of active constraints at the optimum, $\hat{\mathcal{A}}(p) = \mathcal{A}(\hat{x}(p), p) = \{i : G_i \hat{x}(p) = b_i + S_i p\}$. We say that the i^{th} inequality constraint is weakly active constraint if $i \in \mathcal{A}(\hat{x}(p), p)$ and $\hat{\lambda}_i = 0$ for all $\hat{\lambda}$ satisfying the KKT conditions, (4.13). We say that an inequality constraint is strongly active constraint if $i \in \mathcal{A}(\hat{x}(p), p)$ and there exists some $\hat{\lambda}_i > 0$ satisfying the KKT conditions, (4.13).

Definition 4.4.2. A set $C \subseteq \mathbb{R}^n$ is a polyhedron if and only if there exist an $m \times n$ matrix Z and a vector z of real numbers such that $C = \{x \in \mathbb{R}^n : Zx \leq z\}$ where $0 < m < \infty$

Assume for the moment that we know the set \mathcal{A} of constraints that are active at the optimum for a given p . We can now form matrices G^a , b^a and S^a , and the Lagrange multipliers $\lambda^a \geq 0$, corresponding to the optimal active set. Assume that LICQ holds, then the KKT conditions lead to

$$\begin{pmatrix} Q & (G^a)^T \\ G^a & 0 \end{pmatrix} \begin{pmatrix} \hat{x} \\ \hat{\lambda} \end{pmatrix} = \begin{pmatrix} -c \\ b^a \end{pmatrix} + \begin{pmatrix} -F \\ S^a \end{pmatrix} p. \quad (4.15)$$

An application of Theorem (3.2.3) to (mpQP) at $[x(\hat{p}), \hat{p}]$ gives the following result.

$$\frac{d}{dp} \begin{pmatrix} x(\hat{p}) \\ \lambda(\hat{p}) \end{pmatrix} = -(\hat{M})^{-1} \hat{N} \quad (4.16)$$

where, \hat{M} and \hat{N} are computed from (4.15)

Thus, in the quadratic optimization problem, the Jacobian reduce to an algebraic manipulation of the matrices declared in (mpQP). In the neighborhood of the KKT point, $[x(\hat{p}), \hat{p}]$, Corollary (3.2.1) writes as:

$$\begin{pmatrix} \hat{x}(p) \\ \hat{\lambda}(p) \end{pmatrix} = -(\hat{M})^{-1} \hat{N}(p - \hat{p}) + \begin{pmatrix} x(\hat{p}) \\ \lambda(\hat{p}) \end{pmatrix}. \quad (4.17)$$

Here the time consuming computation of the nominal solution $x(\hat{p})$ and the sensitivity differentials $\frac{dx}{dp}(\hat{p})$ is done off-line. Hence, if the parameter p deviates from the nominal parameter \hat{p} to some value to in its neighborhood the equation provide an online approximation whose computation is very cheap since only a matrix-vector product and two vector additions are necessary.

We have now characterized the solution to (mpQP) for a given optimal active set, $\mathcal{A} \subseteq \{1, \dots, m\}$ and a fixed p . However, as long as \mathcal{A} remains the optimal active set in a neighborhood of p , the solution (4.17) remains optimal, when \hat{x} is viewed as a function of p . Such a neighborhood where \mathcal{A} is optimal is determined by imposing that $\hat{x}(p)$ must remain feasible

$$G\hat{x}(p) \leq b + Sp \quad (4.18)$$

and that the Lagrange multipliers λ must remain nonnegative

$$\hat{\lambda}(p) \geq 0 \quad (4.19)$$

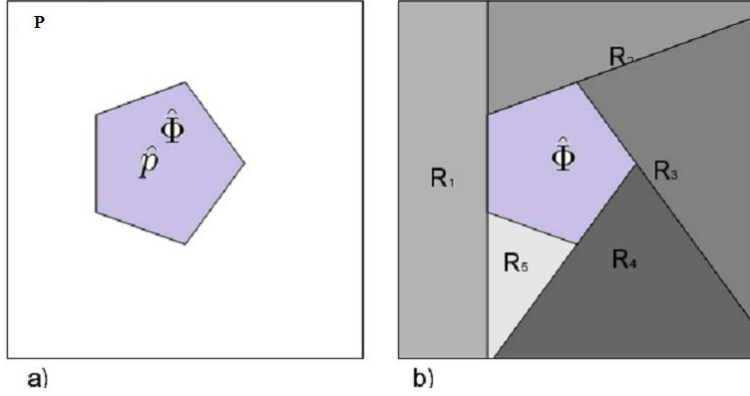


Figure 4.1: Division of critical regions, *A. Bemporad et al. (2002)*, a snapshot from [11]

Equations (4.18) and (4.19) describe a polyhedron in the parameter space. This region is denoted as the **critical region**, $\hat{\Phi}$, corresponding to the given set \mathcal{A} of active constraints, is a convex polyhedral set, and represents the largest set of parameters p such that the combination \mathcal{A} of active constraints at the minimizer is optimal [2], [11].

After having characterized a critical region one needs a method for partitioning the rest of the parameter space. The recursive algorithm proposed in [11] is summarized as follows.

- 1 Solve an LP to obtain a feasible $\hat{p} \in P$, where P is the range of parameters for which the (mpQP) is to be solved.
- 2 Fix $p = \hat{p}$ and solve (mpQP) to find the optimal active set \mathcal{A} for \hat{p}
- 3 Use (4.17)-(4.19) to characterize the solution and critical region $\hat{\Phi}$ corresponding to \mathcal{A}
- 4 Divide the parameter space as in Figure (4.1b) by reversing one by one the hyperplanes defining the critical region
- 5 Iteratively subdivide each new region R_i in a similar way as was done with $\hat{\Phi}$

Theorem 4.4.1. (Continuity and convexity properties)

Consider the (mpQP) and let Q be positive definite, P a polyhedron (convex). Then the set of feasible parameters $P_f \subseteq P$ is convex, the optimizer $x(p) : P_f \rightarrow \mathbb{R}^{n_x}$ is continuous and piecewise affine, and the optimal solution $v(p) : P_f \rightarrow \mathbb{R}$ is continuous, convex, and piecewise quadratic.

Proof. We first prove convexity of P_f and $v(p)$. Take generic $p_1, p_2 \in P_f$ and let $v(p_1), v(p_2)$ and x_1, x_2 be the corresponding optimal values and minimizers. Let $\alpha \in [0, 1]$ and define $x_\alpha \triangleq \alpha x_1 + (1 - \alpha)x_2$, $p_\alpha \triangleq \alpha p_1 + (1 - \alpha)p_2$. By feasibility, x_1, x_2 satisfy the constraints $Gx_1 \leq b + Sp_1$, $Gx_2 \leq b + Sp_2$. These inequalities can be linearly combined to obtain $Gx_\alpha \leq b + Sp_\alpha$ and therefore x_α is feasible for the optimization problem (mpQP). Since a feasible solution, $x(p_\alpha)$, exists at p_α , an optimal solution exists at p_α and hence P_f is convex. The optimal solution at p_α will be less than or equal to the feasible solution:

$$v(p_\alpha) \leq c^\top x_\alpha + \frac{1}{2}x_\alpha^\top Qx_\alpha + p_\alpha^\top Fx_\alpha$$

and hence,

$$\begin{aligned}
v(p_\alpha) &- [\alpha(c^\top x_1 + \frac{1}{2}x_1^\top Qx_1 + p_\alpha^\top Fx_1) + (1-\alpha)(c^\top x_2 + \frac{1}{2}x_2^\top Qx_2 + p_\alpha^\top Fx_2)] \\
&\leq c^\top x_\alpha + \frac{1}{2}x_\alpha^\top Qx_\alpha + p_\alpha^\top Fx_\alpha - [\alpha(c^\top x_1 + \frac{1}{2}x_1^\top Qx_1 + p_\alpha^\top Fx_1) \\
&\quad + (1-\alpha)(c^\top x_2 + \frac{1}{2}x_2^\top Qx_2 + p_\alpha^\top Fx_2)] \\
&= \frac{1}{2}[\alpha^2 x_1^\top Qx_1 + (1-\alpha)^2 x_2^\top Qx_2 + 2\alpha(1-\alpha)x_2^\top Qx_1 - \alpha x_1^\top Qx_1 - (1-\alpha)x_2^\top Qx_2] \\
&= -\frac{1}{2}\alpha(1-\alpha)(x_1 - x_2)^\top Q(x_1 - x_2) \leq 0 \\
v(\alpha p_1 + (1-\alpha)p_2) &\leq \alpha v(p_1) + (1-\alpha)v(p_2), \quad \forall p_1, p_2 \in P_f, \quad \alpha \in [0, 1]
\end{aligned}$$

Within the closed polyhedral regions, $\hat{\Phi}$, in P_f the solution $x(p)$ is affine. The boundary between two regions belongs to both closed regions. Because the optimum is unique the solution must be continuous across the boundary. The fact that $v(p)$ is continuous and piecewise quadratic follows trivially. \square

Remark 4.4.1. Multi-parametric linear program(mpLP): *Note that when Q is a null matrix, (mpQP) reduces to a multi-parametric linear program (mpLP). This does not affect the solution procedure remains the same. This is because the results presented in Theorem (3.2.3) continue to hold true and SOS is valid in spite of the fact that Q is a null matrix as discussed on page 110 in [1]. For mpLPs x is an affine function of p and λ remains constant in a critical region and therefore Corollary (3.2.1) can be used. Whilst the results of Theorem (3.2.3) regarding P_f and $x(p)$ are still valid, $v(p)$ simplifies to a continuous, convex, and piecewise linear function of p .*

Hence, at the end of the the solution procedure the solution obtained is a conditional piecewise function of the parameters and Theorem (3.2.3) implies that the optimal function computed, $v(p)$, is continuous and convex.

4.4.2 Characterization of the Partition

Below, we denote by $\hat{x}_k(p)$ the linear expression of the piecewise affine(PWA) function $\hat{x}(p)$ over the critical region Φ^{R_k} , where k is an index enumerating the optimal active sets.

Definition 4.4.3. *Let a polyhedron $P \subseteq \mathbb{R}^{n_p}$ be represented by the linear inequalities $Zx \leq z$. Let the i^{th} hyperplane $Z_i x \leq z_i$ be denoted by \mathcal{H} . If $P \cap \mathcal{H}$ is $(p-1)$ -dimensional, then $P \cap \mathcal{H}$ is called a facet of the polyhedron.*

Definition 4.4.4. *Two polyhedra are called neighboring polyhedra if they have a common facet.*

Definition 4.4.5. *Let a polyhedron P be represented by $Zp \leq z$. We say that $Z_i p \leq z_i$ is redundant if $Z_j p \leq z_j, \forall j \neq i \Rightarrow Z_i p \leq z_i$ (i.e., it can be removed from the description of the polyhedron). The inequality i is redundant with degree s if it is redundant and there exists an s -dimensional subset Y of P such that $Z_i p \leq z_i$ for all $p \in Y$*

Definition 4.4.6. *A representation of a polyhedron (4.18)-(4.19) is r-minimal if all redundant constraints have degree $s \geq r$. It is minimal if there are no redundant constraints.*

It is clear that, when we fix the active set, it is fairly easy to characterize the optimal solution and Lagrange multipliers corresponding to this active set, and the region in the parameter space in which this active set is optimal. The main task for an (mpQP) solver is therefore to find every active set which is optimal in some full-dimensional region in the parameter space. We do this by, for each critical region Φ^{R_i} we identify finding the optimal active set in every full-dimensional neighboring critical region.

Theorem 4.4.2. *Consider an optimal active set $\{i_1, i_2, \dots, i_k\}$ and its corresponding minimal representation of the critical region $\hat{\Phi}$ obtained by (4.18)-(4.19) after removing all redundant inequalities. Let Φ^{R_i} be a full-dimensional neighboring critical region to $\hat{\Phi}$ and assume LICQ holds on their common facet $\mathcal{F} = \hat{\Phi} \cap \mathcal{H}$ where \mathcal{H} is the separating hyperplane between $\hat{\Phi}$ and Φ^{R_i} . Moreover, assume that there are no constraints which are weakly active at the optimizer $\hat{x}(p)$ for all $p \in \hat{\Phi}$. Then:*

- a) **Type I:** *If \mathcal{H} is given by $G_{i_{k+1}}\hat{x}_0(p) = b_{i_{k+1}} + S_{i_{k+1}}p$, then the optimal active set in Φ^{R_i} is $\{i_1, i_2, \dots, i_k, i_{k+1}\}$*
- b) **Type II:** *If \mathcal{H} is given by $\hat{\lambda}_{i_k}(p) = 0$, then the optimal active set in Φ^{R_i} is $\{i_1, i_2, \dots, i_{k-1}\}$*

Proof. a) **Type I:** In order for some constraint $i_j \in \{i_1, i_2, \dots, i_k\}$ not to be in the optimal active set in Φ^{R_i} , by continuity of $\hat{\lambda}(p)$ (because of theorem (4.4.1) and LICQ), it follows that $\hat{\lambda}_{i_j}(p) = \hat{\lambda}_{i_j}^0(p) = 0$, $p \in \mathcal{F}$. Since there are no constraints which are weakly active for all $p \in \hat{\Phi}$, this would mean that constraint i_j becomes non-active at \mathcal{F} . But this contradicts the assumption of minimality since $\hat{\lambda}_{i_j} \geq 0$ and $G_{i_{k+1}}\hat{x}(p) = b_{i_{k+1}} + S_{i_{k+1}}p$ would be coincident. On the other hand $\{i_1, \dots, i_k\}$ can not be the optimal active set on Φ^{R_i} because $\hat{\Phi}$ is the largest set of p 's such that $\{i_1, \dots, i_k\}$ is the optimal active set. Then, the optimal active set in Φ^{R_i} is the superset of $\{i_1, \dots, i_k\}$. Now assume that another constraint i_{k+2} is active in Φ^{R_i} . That means $G_{i_{k+2}}\hat{x}_i(p) = b_{i_{k+2}} + S_{i_{k+2}}p$ in Φ^{R_i} and by continuity of $\hat{x}(p)$, the equality also holds for $p \in \mathcal{F}$. However, $G_{i_{k+2}}\hat{x}_0(p) = b_{i_{k+2}} + S_{i_{k+2}}p$, which contradicts the assumption of minimality. Therefore, only $\{i_1, \dots, i_k, i_{k+1}\}$ can be an optimal active set in Φ^{R_i} . The proof of Type II is similar. \square

Corollary 4.4.1. *Consider the same assumptions as in Theorem (4.4.2), except that the assumption of minimality is relaxed into $(p-1)$ -minimality, i.e., two or more hyperplanes can coincide. Let $\mathcal{I} \subseteq \{i_1, i_2, \dots, i_k\}$ be the set of indices corresponding to coincident hyperplanes in the $(p-1)$ -minimal representation of (4.18)-(4.19) of $\hat{\Phi}$*

- every constraint i_j where $i_j \in \{i_1, i_2, \dots, i_k\} \setminus \mathcal{I}$ is active in Φ^{R_i}
- every constraint i_j where $i_j \notin \{i_1, i_2, \dots, i_k\} \cup \mathcal{I}$ is inactive in Φ^{R_i}

Based on the above discussion and results, the main steps of the (mpQP) solver are outlined in the following algorithm[2].

(mpQP) Algorithm

1. In a given region solve (mpQP) by treating p as a free variable to obtain \hat{p}
2. Fix $p = \hat{p}$ and solve (mpQP) to obtain $[x(\hat{p}), \hat{p}]$
3. Compute $[-(\hat{M})^{-1}\hat{N}]$ from (4.17)
4. Obtain $[\hat{x}(p), \hat{\lambda}(p)]$
5. Form a set of inequalities, Φ^R as

$$\Phi^R := \{G\hat{x}(p) \leq b + Sp, \hat{\lambda}(p) \geq 0, P\}$$

where G, b, S corresponds to inactive inequalities

6. Remove redundant inequalities from the set Φ^R of inequalities and define the corresponding $\hat{\Phi} = \Delta(\Phi^R)$, where Δ is an operator which removes redundant constraints
7. Define the rest of the region $\Phi^{rest} := P - \hat{\Phi}$
8. If no more regions to explore, go to next step, otherwise go to Step 1
9. Collect all the solutions and unify the regions having the same solution to obtain a compact representation.

Numerical Example

Here we consider an illustrative example by using the algorithm outlined. Consider an (mpLP) which can be considered a special case of positive semi-definite (mpQP), namely the one with $Q = 0$.

$$\text{LP}(p) \left\{ \begin{array}{l} \max_x \quad 0.8x_1 + 10.8x_2 \\ \text{s.t.} \quad 0.8x_1 + 0.44x_2 \leq 24000 + p_1 \\ \quad \quad 0.05x_1 + 0.1x_2 \leq 2000 + p_2 \\ \quad \quad 0.1x_1 + 0.36x_2 \leq 6000 \\ \quad \quad x_1 \geq 0, x_2 \geq 0 \\ \quad \quad 0 \leq p_1 \leq 6000 \\ \quad \quad 0 \leq p_2 \leq 500 \end{array} \right. \quad (\text{LP})$$

Step 1. Solve the program by treating p_1 and p_2 as a free free variables. A feasible point obtained is $[p_1, p_2] = [0, 0]^\top$.

Step 2. Fix $\hat{p} = [0, 0]^\top$ and solve the problem (LP). The nominal solution is

$$x(\hat{p}) = \begin{pmatrix} 26207 \\ 6896.6 \end{pmatrix} \text{ and } \lambda(\hat{p}) = \begin{pmatrix} 4.555 \\ 87.52 \\ 0 \end{pmatrix}$$

Step 3. Compute $-\hat{M}^{-1}\hat{N}$.

$$\text{The solution is } -\hat{M}^{-1}\hat{N} = \begin{pmatrix} 1.724 & -7.586 \\ -0.8621 & 13.79 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Step 4. Compute $[\hat{x}(p), \hat{\lambda}(p)]$ from (4.17):

$$\begin{pmatrix} \hat{x}_1(p) \\ \hat{x}_2(p) \\ \hat{\lambda}_1(p) \\ \hat{\lambda}_2(p) \\ \hat{\lambda}_3(p) \end{pmatrix} = \begin{pmatrix} 1.724 & -7.586 \\ -0.8621 & 13.79 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} (p - \hat{p}) + \begin{pmatrix} 26207 \\ 6896.6 \\ 4.55 \\ 87.52 \\ 0 \end{pmatrix} \quad \text{or}$$

$$\begin{cases} \hat{x}_1(p) = 1.724p_1 - 7.586p_2 + 26207, \\ \hat{x}_2(p) = -0.8621p_1 + 13.79p_2 + 6896.6, \\ \hat{\lambda}_1(p) = 4.555, \\ \hat{\lambda}_2(p) = 87.52, \\ \hat{\lambda}_3(p) = 0 \end{cases}$$

Step 5. Form a set of inequalities corresponding to Φ^R

$$\Phi^R = \begin{cases} G\hat{x}(p) \leq b + Sp : -0.138p_1 + 4.206p_2 \leq 896.5 \\ \hat{\lambda}(p) \geq 0 : \begin{cases} \hat{\lambda}_1(p) = 4.555 \geq 0, \\ \hat{\lambda}_2(p) = 87.52 \geq 0, \\ \hat{\lambda}_3(p) = 0 \geq 0 \end{cases} \\ P : \begin{cases} 0 \leq p_1 \leq 6000, \\ 0 \leq p_2 \leq 500 \end{cases} \end{cases}$$

Step 6. Remove redundant constraints

$$\hat{\Phi} = \begin{cases} -0.138p_1 + 4.206p_2 \leq 896.5 \\ 0 \leq p_1 \leq 6000 \\ 0 \leq p_2 \end{cases}$$

Step 7. Define the rest of the region

$$\Phi^{rest} = \begin{cases} -0.138p_1 + 4.206p_2 \geq 896.5 \\ 0 \leq p_1 \leq 6000 \\ p_2 \leq 500 \end{cases} \quad (4.20)$$

Step 8. Region (4.20) is a region to be explored. We return to **Step 1** and include constraints (4.20) in the optimization problem (LP). This problem terminates in the next iteration ending with two critical regions.

Step 9. Collect the two regions. Since they have different solutions, we can not merge them.

$$\Phi^{R_1} = \begin{cases} -0.14p_1 + 4.21p_2 \leq 896.5 \\ 0 \leq p_1 \leq 6000 \\ 0 \leq p_2 \end{cases} \quad \text{Optimal Soln : } \begin{cases} \hat{x}_1(p) = 1.72p_1 - 7.59p_2 + 26207, \\ \hat{x}_2(p) = -0.86p_1 + 13.79p_2 + 6896.6 \\ v(p) = 4.66p_1 + 87.52p_2 + 2686758.6 \end{cases}$$

$$\Phi^{R_2} = \begin{cases} -0.14p_1 + 4.21p_2 \geq 896.5 \\ 0 \leq p_1 \leq 6000 \\ p_2 \leq 500 \end{cases} \quad \text{Optimal Solution : } \begin{cases} \hat{x}_1(p) = 1.48p_1 + 24590.2, \\ \hat{x}_2(p) = -0.41p_1 + 9836.1 \\ v(p) = 7.53p_1 + 305409.8 \end{cases}$$

Therefore, we can conclude the following:

- (i) A complete map of all the optimal solutions as a function of p_1 and p_2 is available.
- (ii) The space of p_1 and p_2 has been divided into two regions, Φ^{R_1} and Φ^{R_2} , where the solution remain optimal and hence
 - (a) one does not have to exhaustively enumerate the complete space of p_1 and p_2
 - (b) the optimal solution can be obtained by simply substituting the value of p_1 and p_2 into the parametric profiles without any further optimization calculations.
- (iii) The sensitivity of the optimal value to the parameters can be identified.

Bibliography

- [1] Fiacco, A.V., *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, volume 165 of Mathematics in Science and Engineering. Academic Press, 1983.
- [2] E. N., Pistikopoulos, M.C., Georgiadis, and Vivek Dua, Diago Narciso, Pinky Dua, K.I., Kouramas *Multi-Parametric Programming: Theory, Algorithms, and Applications*, Edited by E. N. Pistikopoulos, Michael C.Georgiadis, and Vivek Dua, Wiley-VCH Verlag GmbH and Co.KGaA, Weinheim, 2007.
- [3] S. M. Robinson, *Strongly regular generalized equations.*, Math. of Oper. Research 5, pp. 43-62, 1980.
- [4] Bonnans, J. F. and Shapiro, A. *Perturbation Analysis of Optimization Problems.*, Springer Series in Operations Research., Springer, New York, 2000.
- [5] Büskens, C. and Maurer, H. *Sensitivity Analysis and Real-Time Optimization of Parametric Nonlinear Programming Problems.* In *Online Optimization of Large Scale Systems*, (M. Grötschel, S. O. Krumke and J. Rambau, editors), pp. 3-16., Springer, 2001.
- [6] K. Jittorntrum, *Solution point differentiability without strict complementarity in nonlinear programming.* In a Publication of the Mathematical Programming Society, *Sensitivity, Stability and Parametric Analysis.*, Edited by A.V. Fiacco, pp. 127-138, Amsterdam, 1984.
- [7] S. Dempe, *On Generalized Differentiability of Optimal Solutions in Nonlinear Parametric Optimization*, In mathematical programming with data perturbations, edited by Anthony V. Fiacco, Marcel Dekker, Inc., pp. 47-64, New York, 1998.
- [8] Jacques Gauvin, François Dubeau: *Differential Properties of the Marginal Function in the Mathematical Programming.* Mathematical Programming Study 19, pp. 101-119, 1982
- [9] M.R. Hestenes, *Optimization Theory: The Finite Dimensional Case*, J.Wiley Publication, New York, 1975
- [10] T. J. Beltracchi, G. A. Gabriele, *An Investigation of Using an RQP Based Method to Calculate Parameter Sensitivity Derivatives.*, pp.3-24, New York, 1989
- [11] Alberto Bemporad, Manfred Morari, Vivek Dua, E.N. Pistikopoulos *The explicit linear quadratic regulator for constrained systems*, Automatica 38(1) 3-20, 2002
- [12] P. Tøndel, T.A. Johansen, A. Bemporad, *An algorithm for multi-parametric quadratic programming and explicit MPC solutions* Tech. report 2001-7-W, Available at http://www.itk.ntnu.no/ansatte/Johansen_Tor.Arne/mpQPA1g.pdf, 2001