

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF MATHEMATICS



Graduate Seminar Report

On

Sequential Quadratic Programming

(SUBMITTED TO THE PARTIAL FULFILMENT OF THE M.Sc. DEGREE IN MATHEMATICS)

Compiled by: Arega Kumlign

Advisor: Dr. Berhanu Guta

June, 2010

Addis Ababa

Table of content

Content	page
Acknowledgment	
1. Introduction.....	1
1.1 Preliminaries.....	2
1.1.1 Basic definitions.....	2
1.1.2 Necessary and sufficient optimality conditions.....	3
2. Sequential quadratic problem.....	7
2.1 Concept of SQP.....	7
2.2 Equality constrained problem.....	8
2.3 Quadratic programming sub problem.....	11
2.3.1 Alternative formulation of SQP.....	12
2.3.2 SQP algorithm.....	13
2.4. Equality and inequality constrained problem.....	14
2.5. Quasi Newton positive definite Approximation.....	17
2.6. Penalty merit function.....	18
Reference	

Acknowledgment

First and foremost I would like to express my heart-felt gratitude to my respected advisor

Dr. Berhanu Guta for his genuine advice and constructive comments in the preparation of this seminar report.

Next, I would like to thank many persons for the support and assistance that I have received while writing this paper as well as in my study. Especially I thank Ato Dagnachew Eyachew, Ato Asmamaw Addis, and Ato Abebe Ambachew for their attention and generous support in my life. I will never forget how much I learned from them.

Finally, my thanks go to my family, I could have never arrived at this level without the love, care, and support of my families. I would like to thank them for all the joy, happiness and others they brought in to my life.

Chapter One

Introduction

Sequential quadratic programming (SQP) method have proved highly effective for solving constrained optimization problems with non linear functions both in the objective and constraint functions. It was first introduced by Wilson [1963] and Fletcher [1970] proposed to solve constrained minimization problems.

We consider general problems with general equality and inequality constrained functions which are twice differentiable.

$$(P) \quad \text{Minimize } f(x)$$

$$\text{Subject to } g_i(x) \leq 0, \quad \text{for } i = 1, \dots, m.$$

$$h_i(x) = 0, \quad \text{for } i=1, 2, \dots, l$$

f and g_i are twice differentiable at and each h_i for $i=1, 2, \dots, l$ are continuously twice differentiable.

Sequential quadratic programming method were mainly developed having in mind to ensure rapid convergence which close to the solution. Its basic idea is to use Newton's method to find a stationary point of the Lagrange function.

Its implementation is based on the following three main stages:

1. Solving Quadratic programming sub problem.
2. Updating the Hessian matrix
3. Line search and merit functions. (We consider only the L1 merit penalty function).

The paper starts by mentioning some of the basic definitions, necessary and sufficient optimality conditions, and important theorems that are essential for the development of sequential quadratic programming algorithms, and continues by describing the basic features of SQP method both in equality and inequality constrained non linear problems, followed by BFGS positive definite approximation of the Hessian the Lagrange function and L1 merit penalty function.

1.1 Preliminaries

1.1.1 Basic Definitions

Definition 1.1.1: Let S be a nonempty subset in E_n . S is said to be convex if

$$\lambda x_1 + (1 - \lambda)x_2 \in S, \forall x_1, x_2 \in S, \lambda \in [0, 1].$$

Definition 1.1.2: Let $f: S \rightarrow E_1, S \subseteq E_n$. A function f is said to be convex on S if

$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \quad \forall x_1, x_2 \in S, \lambda \in [0, 1]$. f is said to be strictly convex on S if the above inequality is true as strict inequality for each distinct x_1, x_2 .

Definition 1.1.3: Consider the problem of minimize $f(x)$ over E_n . Let $\bar{x} \in S$ then, \bar{x} is called

1. global minimum if $f(\bar{x}) \leq f(x), \forall x \in E_n$.
2. Local minimum point if there exists $\varepsilon > 0, \bar{x}$ is a minimum point of f on $N_\varepsilon(\bar{x}) \cap E_n$. i. e $f(\bar{x}) \leq f(x), \forall x \in N_\varepsilon(\bar{x}) \cap E_n$.
3. Strict local minimum if $f(\bar{x}) < f(x), \forall x \in N_\varepsilon(\bar{x}) \cap E_n$.

Definition 1.1.4: Let S be a non empty subset of E_n , and let $f: S \rightarrow E_1$ then, f is said to be differentiable at \bar{x} if there exist a vector $\nabla f(\bar{x})$, called the gradient vector, and a function $\alpha: E_n \rightarrow E_1$ such that $f(x) = f(\bar{x}) + \nabla f(\bar{x})^t(x - \bar{x}) + \|x - \bar{x}\| \alpha(\bar{x}, x - \bar{x}) \quad \forall x$ where $\lim_{x \rightarrow \bar{x}} \alpha(\bar{x}, x - \bar{x}) \rightarrow 0$.

Definition 1.1.5: Consider the problem to minimize $f(x)$ subject to $x \in S$, where $f: E_n \rightarrow E_1$, and S is a non empty set in E_n , a nonzero vector d is called a feasible direction at $x \in S$ if there exists a $\delta > 0$ such that $x + \lambda d \in S$ and $\lambda \in (0, \delta)$. furthermore; d is called an improving feasible direction at $x \in S$ if there exists a $\delta > 0$ such that

$$f(x + \lambda d) < f(x) \text{ and } x + \lambda d \in S \text{ for all } \lambda \in (0, \delta).$$

Definition 1.1.6: A symmetric matrix H is said to be positive semi definite everywhere if $x^t H(\bar{x})x \geq 0, x \in E_n, \bar{x} \in S$. if strict inequality holds then it is said to be positive definite.

Theorem 1.1.7: Let S be a nonempty open convex set in E_n , and let $f: E_n \rightarrow E_1$ and S is a non empty set in E_n . then f is convex if and only if the Hessian matrix is positive semi definite at each point in S .

Definition: A point \bar{x} is called regular for problem (p) given in the following problem if the set of vectors $\{\nabla h_i(\bar{x}) \mid i = 1, \dots, l\} \cup \{\nabla g_i(\bar{x}) \mid i = 1, \dots, m\}$ is linearly independent.

1.1.2. Necessary and Sufficient Optimality Conditions

In this sub section we will describe some of the optimality conditions for general constrained optimization problems

Theorem 1.1.2 :(Karush Kuhn Tucker Necessary Condition)

Let X be a non empty open set in E_n , and let

$f: E_n \rightarrow E_1, g_i: E_n \rightarrow E_1, \text{ for } i = 1, \dots, m. h_i: E_n \rightarrow E_1 \quad i=1, 2, \dots, l$

Consider the problem

(P) Minimize $f(x)$

Subject to $g_i(x) \leq 0, \text{ for } i = 1, \dots, m.$

$h_i(x) = 0, \text{ for } i=1, 2, \dots, l$

Suppose f and g_i are differentiable at \bar{x} and each $h_i \text{ for } i=1, 2, \dots, l$ are continuously differentiable at \bar{x} . Further, suppose that $\nabla g_i \text{ for } i \in \{1, \dots, m\}$ and $\nabla h_i \text{ for } i \in \{1, \dots, l\}$

are linearly independent. If \bar{x} solves (P) locally, then there exists unique scalars u_i, v_i such that

$$\text{i) } \nabla f(\bar{x}) + \sum_{i=1}^m u_i \nabla g_i(\bar{x}) + \sum_{i=1}^l v_i \nabla h_i(\bar{x}) = 0$$

$$\text{ii) } u_i \geq 0, i = 1, \dots, m.$$

$$\text{iii) } u_i g_i(\bar{x}) = 0, i = 1, \dots, m.$$

$$\text{iv) } g_i(\bar{x}) \leq 0, i = 1, 2, \dots, m$$

$$\text{v) } h_i(\bar{x}) = 0, i = 1, \dots, l.$$

Remark 1.1.2.1: Consider the following problem

P: minimize $\{f(x): x \in S\}$, Where $S = \{x: g_i(x) \leq 0, \text{ for } i = 1, \dots, m. h_i(x) = 0 \text{ for } i=1,2,\dots,l\}$ each functions are twice differentiable, and X is nonempty subset of E_n .

The Lagrange function is given by:

$$L(x, u, v) = f(x) + \sum_{i=1}^m u_i g_i(x) + \sum_{i=1}^l v_i h_i(x)$$

Now, let \bar{x} be a KKT point for the given Problem P with the associated multipliers \bar{u} and \bar{v} corresponding to the equality and inequality constraints respectively.

Define the restricted Lagrange

$$\phi(x) = L(x, \bar{u}, \bar{v}) = f(x) + \sum_{i \in I} \bar{u}_i g_i(x) + \sum_{i=1}^l \bar{v}_i h_i(x),$$

where $I = \{i \in (1, \dots, m): g_i(\bar{x}) = 0\}$

Observe that the dual feasibility condition

$$\nabla f(\bar{x}) + \sum_{i \in I} \bar{u}_i \nabla g_i(\bar{x}) + \sum_{i=1}^l \bar{v}_i \nabla h_i(\bar{x}) = 0$$

in the KKT system is equivalent to the statement $\nabla \phi(\bar{x})$ vanishes. Moreover we have $\phi(x) \leq f(x), \forall x \in S$ while $\phi(\bar{x}) = f(\bar{x})$.

If \bar{x} turns out to be a local minimize for ϕ , then it will also be a local minimizer for problem P.

Lemma 1.1.2.2: consider the problem (P)

Suppose that \bar{x} is a KKT point for problem P with Lagrange multipliers \bar{u} and \bar{v} corresponding to the equality and inequality constraints respectively.

Define the restricted Lagrange as above and denote its hessian by $\nabla^2 \phi$

a) If $\nabla^2 \phi(\bar{x})$ is PSD, for all $x \in S$, then \bar{x} is a global minimum for problem P, on the other hand if $\nabla^2 L$ is PSD for all $x \in S \cap N_\epsilon(\bar{x})$ for some ϵ neighborhood $N_\epsilon(\bar{x})$ about \bar{x} , then \bar{x} is a local minimize point for P.

b) If $\nabla^2 \phi(\bar{x})$ is PD, for all $x \in S \cap N_\epsilon(\bar{x})$, then \bar{x} is a strict local minimum for problem P.

Theorem 1.1.2.3: (KKT Second Order Sufficient Condition)

Consider the following problem P: minimize $\{f(x) : x \in S \text{ Where } S = \{x : g_i(x) \leq 0, \text{ for } i = 1, \dots, m, h_i(x) = 0 \text{ for } i=1, 2, \dots, l\}$

Each function is twice differentiable, and X is nonempty subset of E_n .

Suppose \bar{x} is a KKT point for problem P, with Lagrange multipliers \bar{u}, \bar{v} associated with the inequality and equality constraints, respectively. Let $I = \{i : g_i(\bar{x}) = 0\}$, and denote $I^+ = \{i \in I : \bar{u}_i > 0\}$,

$$I^0 = \{i \in I : \bar{u}_i = 0\}$$

Define the restricted Lagrange function $\phi(x) = f(x) + \sum_{i \in I} \bar{u}_i g_i(x) + \sum_{i=1}^l \bar{v}_i h_i(x)$ and its Hessian at \bar{x} by $\nabla^2 \phi(\bar{x}) = \nabla^2 f(\bar{x}) + \sum_{i \in I} \bar{u}_i \nabla^2 g_i(\bar{x}) + \sum_{i=1}^l \bar{v}_i \nabla^2 h_i(\bar{x})$.

$$\text{Define } C = \left\{ \begin{array}{l} d \neq 0 : \nabla g_i(\bar{x})^t d = 0, i \in I^+, \nabla g_i(\bar{x})^t d \leq 0, i \in I^0 \\ \nabla h_i(\bar{x})^t d = 0, i = 1, \dots, l \end{array} \right\}$$

Then, if $d^t \nabla^2 \phi(\bar{x}) d > 0$. for all $d \in C$ we have that \bar{x} is a strict local minimum for P.

Proof:

Suppose that \bar{x} is not a strict local minimum. Then, there exists a sequence $\{x_k\}$ in S converging to \bar{x} such that $x_k \neq \bar{x}$ and $f(x_k) \leq f(\bar{x})$ for all k. define $d_k = (x_k - \bar{x}) / \|x_k - \bar{x}\|$ and $\lambda_k = \|x_k - \bar{x}\|$ for all k, we have $x_k = \bar{x} + \lambda_k d_k$, where $\|d_k\| = 1$ for all k, and we have

$\{\lambda_k\} \rightarrow 0$ as $k \rightarrow \infty$ since $\|d_k\| = 1$ for all k, a convergent subsequence exists. Assume

without loss of generality, that the given sequence itself represent this subsequence. Hence

$\{d_k\} \rightarrow d$, where $\|d\| = 1$. moreover, we have

$$0 \geq f(\bar{x} + \lambda_k d_k) - f(\bar{x}) = \lambda_k \nabla f(\bar{x})^t d_k + \frac{1}{2} \lambda_k d_k^t \nabla^2 f(\bar{x}) d_k + \lambda_k^2 \alpha_f(\bar{x}; \lambda_k d_k) \quad (a)$$

$$0 \geq g_i(\bar{x} + \lambda_k d_k) - g_i(\bar{x}) = \lambda_k \nabla g_i(\bar{x})^t d_k + \frac{1}{2} \lambda_k d_k^t \nabla^2 g_i(\bar{x}) d_k + \lambda_k^2 \alpha_{g_i}(\bar{x}; \lambda_k d_k) \quad (b)$$

$$0 \geq h_i(\bar{x} + \lambda_k d_k) - h_i(\bar{x}) = \lambda_k \nabla h_i(\bar{x})^t d_k + \frac{1}{2} \lambda_k d_k^t \nabla^2 h_i(\bar{x}) d_k + \lambda_k^2 \alpha_{h_i}(\bar{x}; \lambda_k d_k) \quad (c)$$

Where α_f, α_{g_i} for $i \in I$ and α_{h_i} $i=1, \dots, l$ approach zero as $k \rightarrow \infty$.

Dividing each expression in (a), (b) and (c) by λ_k and taking the limit as $k \rightarrow \infty$, we obtain

$$\nabla f(\bar{x})^t d \leq 0, \quad \nabla g_i(\bar{x})^t d \leq 0, i \in I, \quad \nabla h_i(\bar{x})^t d = 0 \quad i = 1, \dots, l \dots \dots \quad (e)$$

Now, since \bar{x} is a KKT point, we have

$$\nabla f(\bar{x}) + \sum_{i \in I} \bar{u}_i \nabla g_i(\bar{x}) + \sum_{i=1}^l \bar{v}_i \nabla h_i(\bar{x}) = 0$$

Taking the inner product with d and using equation (e), we conclude that

$$\nabla f(\bar{x})^t d = 0, \quad \nabla g_i(\bar{x})^t d = 0, i \in I^+, \quad \nabla g_i(\bar{x})^t d \leq 0, i \in I^0 \text{ and } \nabla h_i(\bar{x})^t d = 0 \quad i = 1, \dots, l$$

Hence, in particular $d \in C$. Furthermore, multiplying each of (b) by \bar{u}_i for $i \in I$, and each of (c)

by \bar{v}_i for $i=1, \dots, l$ and adding, we get, using

$$\nabla f(\bar{x})^t d_k + \sum_{i \in I} \bar{u}_i \nabla g_i(\bar{x})^t d_k + \sum_{i=1}^l \bar{v}_i \nabla h_i(\bar{x})^t d_k = 0$$

$$0 \geq \frac{\lambda_k^2}{2} d_k^t \nabla^2 L(\bar{x}) d_k + \lambda_k^2 \left[\alpha_f(\bar{x}; \lambda_k d_k) + \sum_{i \in I} \bar{u}_i \alpha_{g_i}(\bar{x}; \lambda_k d_k) + \sum_{i=1}^l \bar{v}_i \alpha_{h_i}(\bar{x}; \lambda_k d_k) \right]$$

Dividing by $\lambda_k^2 > 0$ and taking the limit $k \rightarrow \infty$ we obtain

$$d^t \nabla^2 L(\bar{x}) d \leq 0, \text{ where } \|d\| = 1 \text{ and } d \in C.$$

Which is a contradiction to the fact that $d^t \nabla^2 L(\bar{x}) d$ is positive definite for all $d \in C$.

Therefore, \bar{x} must be a strict local minimum point for problem P.

Chapter Two

Sequential Quadratic Programming Problem

In this seminar report we focus on a particular solution method of non linear constrained programming problems; Sequential quadratic programming method.

We consider the following non linear optimization problems:

$$\begin{aligned} & \text{minimize } f(x) \\ & \text{s.t } g_i(x) \leq 0, \quad i = 1, 2, \dots, m. \\ & h_i(x) = 0, \quad i = 1, 2, \dots, l \dots \dots \dots \dots \dots \dots (1) \end{aligned}$$

where $f: E_n \rightarrow E_1, g_i: E_n \rightarrow E_1, i = 1 \dots, m, h_i: E_n \rightarrow E_1, i = 1, \dots, l$ are assumed to be twice continuously differentiable functions.

2.1 Concept of SQP

Sequential quadratic programming is an efficient and powerful algorithm to solve nonlinear programming problems which makes a quadratic model of the objective function and linear model of the constraints and solves them at each iterates. SQP is also called sequential or recursive quadratic programming which employ Newton's method for solving the KKT condition for the original problem.

The problem formulated in this way is called quadratic programming.

SQP is developed based on:

1. To directly solve the first order necessary condition (FOC) for the second order approximation of the Lagrange function.
2. Finding the solution of FOC using Newton's method.

The process can be seen as finding search direction towards sequentially through minimizing the quadratic approximation of the Lagrange function with the linear approximation of the constraints.

2.2 Equality Constrained Problems

For the derivation of sequential quadratic programming we shall use the equality constrained problem. An understanding of this problem is essential in the design of SQP methods for general non linear programming problems.

Problem of statement

$$\begin{aligned} \text{ECP (p): } & f(x) \rightarrow \text{minimize} \\ & \text{s.t } h_i(x) = 0, \quad i = 1, \dots, l \end{aligned}$$

Where $x \in E_n$ and all functions are assumed to be continuously twice differentiable.

The Lagrange function associated with problem P is given by $L(x, v) = f(x) + h(x)^t v$.

The KKT optimality condition for problem P requires in addition to primal solution x , a dual Lagrange multiplier vector V such that the following holds true:

- i) $\nabla f(x) + \sum_{i=1}^l v_i \nabla h_i(x) = 0$
- ii) $h_i(x) = 0, \quad i = 1, \dots, l \dots \dots \dots (2).$

If we use the Lagrange function we can write the KKT condition (2) as follows:

$$\begin{pmatrix} \nabla_x L(x, v) \\ \nabla_v L(x, v) \end{pmatrix} = 0$$

Now we can write (2) compactly as

$$W(x, v) = 0$$

We now use the Newton's method to minimize a function for which (2) represents the first order condition that equates the gradient to zero.

Remark 2.1:

Newton's method is a method for solving non linear equations by linearizing the equations at each iterates.

Suppose that we wish to solve the equation system:

$$h_1(x) = 0.$$

$$h_2(x) = 0.$$

.

.

.

$$h_n(x) = 0.$$

$$\text{Or } h(x) = 0$$

Then from Taylor's theorem we have:

$$h(\bar{x} + \Delta x) \approx \tilde{h}(\bar{x} + \Delta x) = h(\bar{x}) + H(\bar{x})\Delta x.$$

Where $H(x)$ is the Jacobian matrix of $h(x)$

The function $\tilde{h}(\bar{x} + \Delta x)$ is the linear approximation of $h(x)$ around the point $x = \bar{x}$

Hence given an iterate (x_k, v_k)

We solve the first order approximation:

$$W(x_k, v_k) + \nabla W(x_k, v_k) \begin{bmatrix} x - x_k \\ v - v_k \end{bmatrix} = 0 \dots \dots \dots (3).$$

To determine the next iterate (x_{k+1}, v_{k+1}) or to conclude it is the solution of the original problem. Where ∇W is the jacobian of W and

$$\nabla^2 L(x_k) = \nabla^2 f(x_k) + \nabla^2 h(x_k)^t V$$

is the hessian of Lagrange at x_k with Lagrangian multiplier vector v_k .

∇h denote the jacobians of h comprises of rows transpose of $\nabla h_i(x_k)$ $i = 1, \dots, l$

we have:

$$\nabla W(x_k, v_k) = \begin{bmatrix} \nabla^2 L(x_k) & \nabla h(x_k)^t \\ \nabla h(x_k) & 0 \end{bmatrix} \dots \dots \dots (4)$$

Now using (2) and (4) we can write (3) as follows

$$\nabla^2 L(x_k)[x - x_k] + \nabla h(x_k)^t[v - v_k] = -\nabla f(x_k) - \nabla h(x_k)^t v_k$$

$$\nabla h(x_k)[x - x_k] = -h(x_k).$$

Letting $d = x - x_k$ we have the following

$$\begin{aligned} \nabla^2 L(x_k)d + \nabla h^t(x_k)v &= -\nabla f(x_k) . \\ h(x_k)d &= -h(x_k) \dots \dots \dots (5) \end{aligned}$$

Equivalently we can write as $\begin{bmatrix} \nabla^2 L(x_k) & \nabla h(x_k)^t \\ \nabla h(x_k) & 0 \end{bmatrix} \begin{pmatrix} d \\ v \end{pmatrix} = \begin{pmatrix} -\nabla f(x_k) \\ -h(x_k) \end{pmatrix}$

Which is system of linear equations with respect to d and v . We can solve for (d, v) using this system.

If a solution d_k exists, and $d_k \neq 0$ put $x_{k+1} = x_k + d_k$ and repeat the process until $d=0$ solves equation (5). When this occurs, at all we have found a KKT solution for the problem P.

In the absence of convexity, a KKT point can be a global minimum, a local minimum, "Saddle point", even a local or global maximum point.

Equation (5) is sometimes called the Newton –Lagrange method (LNS) which is well defined when the matrix $\begin{bmatrix} \nabla^2 L(x_k) & \nabla h(x_k)^t \\ \nabla h(x_k) & 0 \end{bmatrix}$ is non singular.

Assumption 2.2:

- a). The constraint Jacobian $\nabla h(x_k)$ has full row rank at a point x_k .
- b). The matrix $\nabla^2 L(x_k)$ is positive definite on the null space of the Jacobian of constraints . i.e. $Z^t \nabla^2 L(x_k) Z > 0$ where Z is a basis for a null space of $\nabla h(x_k)$. which holds true whenever (x, v) is close to the optimum (x^*, v^*) and the second order sufficient condition is satisfied at the solution.

Under these assumptions the KKT matrix $\begin{bmatrix} \nabla^2 L(x_k) & \nabla h(x_k)^t \\ \nabla h(x_k) & 0 \end{bmatrix}$ is nonsingular.

Proof:

Suppose there exists vectors d_1 and d_2 such that

$$\begin{bmatrix} \nabla^2 L(x_k) & \nabla h(x_k)^t \\ \nabla h(x_k) & 0 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = 0 \dots \dots \dots (1)$$

This gives the following

$$\begin{aligned} \nabla^2 L(x_k)d_1 + \nabla h(x_k)^t d_2 &= 0. \\ \nabla h(x_k)d_1 &= 0. \\ 0 &= \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}^t \begin{bmatrix} \nabla^2 L(x_k) & \nabla h(x_k)^t \\ \nabla h(x_k) & 0 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = d_1^t \nabla^2 L(x_k) d_1 \end{aligned}$$

Since d_1 lies in the null space of ∇h , we have $d_1 = Zu$ for some vector $u \in R^{n-m}$.

Thus, we have $d_1^t \nabla^2 L(x_k) d_1 = u^t Z^t \nabla^2 L(x_k) Z u = 0$ which gives $u=0$. therefore $d_1 = 0$

And $\nabla^2 L(x_k) d_1 + \nabla h(x_k)^t d_2 = 0$ gives us $d_2 = 0$ as $\nabla h(x_k)$ is linearly independent.

thus the equation (1) is satisfied only if $d_1 = 0$ and $d_2 = 0$.

Hence we have the matrix $\begin{bmatrix} \nabla^2 L(x_k) & \nabla h(x_k)^t \\ \nabla h(x_k) & 0 \end{bmatrix}$ is non singular.

2.3 Quadratic Programming Sub Problem

Problem ECP is modeled by a quadratic programming sub problem (QPs for short), whose optimality condition are the same as in the Newton Lagrange system (LNS). The algorithm is the same as Newton Lagrange method, but instead of solving the system LNS (equation 5) we have to solve the following QPs, whose optimality condition might tend to drive the process towards beneficial KKT solution.

$$\text{ECP (p): } f(x) \rightarrow \text{minimize}$$

$$\text{s. t } h_i(x) = 0, \quad i = 1, \dots, l$$

$$\text{QPs}(x_k, v_k) = \text{minimize } f(x_k) + \nabla f(x_k)^t d + \frac{1}{2} d^t \nabla^2 L(x_k) d$$

s. t

$$h_i(x_k) + \nabla h_i(x_k)^t d = 0, \quad i = 1, 2 \dots l. \quad \dots \dots \dots \dots \dots \dots \dots \dots \quad (6)$$

If LICQ are satisfied and $\nabla^2 L(x_k)$ is positive definite on the null space of Jacobian of the constraints the problem (6) would have a unique solution which satisfies:

$$\begin{bmatrix} \nabla^2 L(x_k) & \nabla h(x_k)^t \\ \nabla h(x_k) & 0 \end{bmatrix} \begin{pmatrix} d_k \\ \bar{v}_k \end{pmatrix} = \begin{pmatrix} -\nabla f(x_k) \\ -h(x_k) \end{pmatrix}$$

- The KKT condition for problem (6) requires in addition to primal d , a Lagrange multiplier \bar{v} such that the following holds true:

$$\nabla f(x_k) + \nabla^2 L(x_k, v_k) d + \sum_{i=1}^l \bar{v}_i \nabla h_i(x_k)$$

$$h_i(x_k) + \nabla h_i(x_k)^t d = 0, \quad i = 1, 2 \dots l.$$

This is equivalent to the following:

$$\nabla^2 L(x_k, v_k) d + \nabla h(x_k)^t \bar{v} = -\nabla f(x_k)$$

$$\nabla h(x_k) d = -h(x_k), \quad i = 1, 2 \dots l.$$

Hence by the non singularity of the coefficient matrix we have $d_k = d_{k+1}$, $\bar{v}_k = v_{k+1}$

Here, the Newton iterate can be defined either a solution of the quadratic problem or as the iterate generated by the Newton's method applied to the optimality condition of the problem.

2.3.1 An Alternative Formulation of SQP Method

Let the Lagrange function be given by

$$L(x, v) = f(x) + v^t h(x)^t$$

$$\begin{aligned} \nabla L(x_k, v_k)d &= \nabla f(x_k)d + v_k^t \nabla h(x_k)d \\ &= \nabla f(x_k)d - v_k^t h(x_k)d \end{aligned}$$

It follows the two quadrates

$$\begin{aligned} \frac{1}{2} d^t \nabla^2 L(x_k, v_k)d + \nabla L(x_k, v_k)d + L(x_k, v_k) \quad \text{and} \\ \frac{1}{2} d^t \nabla^2 L(x_k, v_k)d + \nabla f(x_k, v_k)d + f(x_k, v_k) \end{aligned}$$

Differs only by constants, and therefore the quadratic sub problems

$$\begin{aligned} \text{Minimize } & \frac{1}{2} d^t \nabla^2 L(x_k, v_k)d + \nabla L(x_k, v_k)d + L(x_k, v_k) \\ \text{s. t } & h(x_k) + \nabla h(x_k)d = 0 \quad (2.3.1) \\ & \text{and} \end{aligned}$$

$$\begin{aligned} \text{Minimize } & \frac{1}{2} d^t \nabla^2 L(x_k, v_k)d + \nabla f(x_k, v_k)d + f(x_k, v_k) \\ \text{s. t } & h(x_k) + \nabla h(x_k)d = 0 \quad (2.3.2) \end{aligned}$$

Have the same solutions d_k . however the Lagrange multipliers for the two QPs differ since the gradients of the objective function are different

Optimality condition for (2.3.2)

$$\begin{aligned} \nabla^2 L(x_k, v_k)d + \nabla f(x_k) + \nabla h(x_k)^t w_k &= 0 \\ h(x_k) + \nabla h(x_k)d &= 0 \end{aligned}$$

Which is equivalent to

$$\begin{aligned} \nabla^2 L(x_k, v_k)d + \nabla h(x_k)^t w_k &= -\nabla f(x_k) \\ h(x_k) + \nabla h(x_k)d &= 0 \\ \Rightarrow \nabla^2 L(x_k, v_k)d + \nabla h(x_k)^t w_k - \nabla h(x_k)\lambda_k &= -\nabla f(x_k) - \nabla h(x_k)\lambda_k \end{aligned}$$

$$h(x_k) + \nabla h(x_k)d = 0$$

$$\Rightarrow \nabla^2 L(x_k, v_k)d + \nabla h(x_k)(w_k - \lambda_k) = -\nabla L(x_k, \lambda_k)$$

$$h(x_k) + \nabla h(x_k)d = 0$$

$$\Rightarrow w_k - \lambda_k = w_{k+1}$$

$$\Rightarrow w_k = w_{k+1} + \lambda_k = \lambda_{k+1} \quad (2.3.3)$$

Since we assume that the two quadratic sub problems have the same solution d_k

Comparing the optimality conditions we get (2.3.3) for equality constrained NLPs; this is the only significant difference using the two quadratic sub problems.

There are several comments regarding the linearly constrained Quadratic sub problem (QP)

1. An optimum to QP, if it exists is a KKT point for QP and satisfies equation (5), however, the minimization process of QP derives towards a desirable KKT point satisfying equation (5).
2. The objective function QP represents not just a quadratic approximation for $f(x)$, but also an additional term $\frac{1}{2} \sum_{i=1}^l v_{ki} d^t \nabla^2 h(x_k) d$ to represent the curvature of the constraints.
3. QP may be unbounded or infeasible, whereas P is not and this difficulty can be handled by bounding the variation in d .

2.3.2 The SQP algorithm

1. Initialization: Select an initial (feasible) starting primal dual solution (x_k, v_k) .

If termination criteria are satisfied stop with (x_k, v_k) a KKT point. else go to 2

2. Main step: Solve the corresponding quadratic sub problem to obtain a solution d_k

with a set Lagrange multiplier v_{k+1} . If $d_k = 0$ then

(x_k, v_{k+1}) ; satisfies the KKT condition for the original problem. stop

otherwise put $x_{k+1} = x_k + d_k$ and repeat the main step.

2.4 Equality and Inequality Constrained Problems.

$$\begin{aligned} \text{Problem of statement: } P = \min f(x) \\ \text{s.t } g_i(x) \leq 0, \quad i = 1, \dots, m \\ h_i(x) = 0 \quad i = 1, \dots, l. \end{aligned} \quad (8)$$

The corresponding Lagrange function is given by

$$L(x, u, v) = f(x) + g(x)^t u + h(x)^t v.$$

The KKT condition for this problem is there exists $(\bar{x}, \bar{u}, \bar{v})$ for which the following conditions hold.

- i) $\nabla f(\bar{x}) + \sum_{i=1}^m \bar{u}_i \nabla g_i(\bar{x}) + \sum_{i=1}^l \bar{v}_i \nabla h_i(\bar{x}) = 0$
- ii) $\bar{u}_i \geq 0, i = 1, \dots, m. \quad \bar{v}_i \text{ unrestricted } i=1, \dots, l$
- iii) $\bar{u}_i g_i(\bar{x}) = 0, i = 1, \dots, m.$
- iv) $g_i(\bar{x}) \leq 0, i = 1, 2, \dots, m$
- v) $h_i(\bar{x}) = 0, i = 1, 2, \dots, l$

For any feasible solution x_k . Let $G(x_k)$ denote the Jacobian of the function corresponding to active constraints (i.e. rows of $G(x_k)$ are $\nabla h_{i(x)}^t$ and active constraints $g_i(x_k)$). we will denote by x^* any particular local solution. We assume the following holds for any such solution.

- The first order necessary condition is satisfied. i.e. there exists an optimal multipliers vector (u^*, v^*) together which satisfies first order KKT condition.
- $G(x^*)$ has full row rank.
- Strict complementary holds at x^* (i.e. if $g_i(x^*) = 0$, then $u_i^* > 0$).
- $\nabla^2 L(x^*)$ is positive definite on the null space of $G(x^*)$ [second order sufficient condition for strict local minimum.]

At the current iterate x_k we need to locally approximate the problem (P) by a quadratic sub problem. i.e. an optimization problem with a quadratic objective function and linear constraints. Thus, we will construct the sub problem by linearizing the constraints of (P) around x_k .

The general form of the quadratic sub problem has the following form:

$$\mathbf{QP}(x_k, \bar{u}_k, \bar{v}_k) = (\nabla f(x_k)^t d + \frac{1}{2} d^t \nabla^2 L(x_k, \bar{u}_k, \bar{v}_k) d) \rightarrow \text{minimization}$$

Subject to

$$g_i(x_k) + \nabla g_i(x_k)^t d \leq 0, i = 1, 2, \dots, m$$

$$h_i(x_k) + \nabla h_i(x_k)^t d = 0, i = 1, 2, \dots, l \quad \dots \dots \dots (9)$$

where $\nabla^2 L(x_k, \bar{u}_k, \bar{v}_k) = \nabla^2 f(x_k) + \sum_{i=1}^m \bar{u}_{ki} \nabla^2 g_i(x_k) + \sum_{i=1}^l \bar{v}_{ki} \nabla^2 h_i(x_k)$

The KKT condition for this problem requires the primal feasibility x_k and Lagrange multipliers u and v such that the following conditions are satisfied:

- i) $\nabla f(x_k) + \nabla^2 L(x_k) d + \sum_{i=1}^m u_{ki} \nabla g_i(x_k) + \sum_{i=1}^l v_{ki} \nabla h_i(x_k) = 0$
- ii) $u_i [g_i(x_k) + \nabla g_i(x_k)^t d] = 0, i = 1, \dots, m.$
- iii) $u \geq 0,$ and v is unrestricted
- iv) $g_i(x_k) + \nabla g_i(x_k)^t d \leq 0, i = 1, 2, \dots, m$
- v) $h_i(x_k) + \nabla h_i(x_k)^t d = 0, i = 1, 2, \dots, l \quad \dots \dots \dots (10)$

Solving problem (10). If d_k solves QP with the Lagrange multipliers (u_{k+1}, v_{k+1}) and $d_k = 0$, then x_k along with (u_{k+1}, v_{k+1}) yields a KKT solution to the original problem (p). Otherwise, set $x_{k+1} = x_k + d_k$. and repeat the process.

It is not in general true that SQP is a feasible point method. ie it's iterate need not be feasible.

In other words the sub problem QP need not to be feasible where as a result of $G(x^*)$ has full row rank will be the case at points x_k which are close to x^* , but it will not necessarily true at "non local points".

The algorithm has to be designed in such a way that its iterate converges to a desire point.

If this happens to occur, we need to recall methods that will solve this difficulty. In the following sections we will discuss method focus on this difficulties and problems that caused

the nature of the function. i.e. f may not be always convex (that the hessian matrix is not positive definite) or the hessian matrix may be PD at the first iterate but it may not be positive definite in the later iterations.

mark 2.4

Locally convergent: concerns the behavior of the iterates, if the algorithm is started (or has found a point) near the optimum x^* , and typically rates of convergence.

Global convergence: attempts to describe when the iteration of the algorithm will converge to an optimum solution starting from any point. Usually in the line search and trust region methods.

Definition 2.4.1. suppose $\{x^k\}$ is a sequence in \mathbb{R}^n that converges to x^*

1. $\{x^k\}$ is said to be converge linearly if there exists $c \in (0,1)$ such that $\|x^* - x^{(k+1)}\| \leq c \|x^* - x^{(k)}\|$, for all k sufficiently large.

2. $\{x^k\}$ is said to be converge super linearly if $\frac{\|x^* - x^{(k+1)}\|}{\|x^* - x^{(k)}\|} \rightarrow 0$, as $k \rightarrow \infty$.

3. $\{x^k\}$ is said to be converge quadratically if there exists $c \in (0, \infty)$ such that

$$\|x^* - x^{(k+1)}\| \leq c \|x^* - x^{(k)}\|^2 \text{.for all } k \text{ sufficiently large}$$

2.5 Quasi Newton Positive Definite Approximation

The SQP method described above has several negative sides. among these:

- It requires evaluation of second order derivative at each iteration which can be computationally expensive, and even the more difficult; the derivatives may not be easy to compute.
- The Hessian $\nabla^2 L(x_k)$ might not be positive definite and as result, the corresponding QP will be hard to solve. These complexities can be overcome by using a positive definite approximation B_k of the Hessian matrix $\nabla^2 L(x_k)$ which is easy to compute and update.

There is a lot of theoretical work that goes in to designing an update scheme for the matrices B_k we would like these matrices to be positive definite at least in the null space of $G(x^*)$.

Now we can solve the above problem by replacing B_k for $\nabla^2 L(x_k)$ and update the matrix B_k using the BFGS method and update the matrices as follows:

Let $p_k = x_{k+1} - x_k$ and

$$q_k = \nabla' L(x_{k+1}) - \nabla' L(x_k)$$

$$\nabla' L(x) = \nabla f(x_k) + \sum_{i=1}^m u_{ki} \nabla g_i(x_k) + \sum_{i=1}^l v_{ki} \nabla h_i(x_k)$$

$$B_{k+1} = B_k + \frac{q_k q_k^t}{q_k^t p_k} - \frac{B_k p_k p_k^t B_k}{p_k^t B_k p_k}$$

Remark 2.2: the matrix B_k remains positive definite so long as $p^t q > 0$. we can also add a constant multiple of identity matrix to B_k to make it remains positive definite in the case of the QPs is infeasible.

ALGORITHM

- i) Initialization: Select an initial starting primal dual solution (x_k, u_k, v_k) and a positive definite approximation B_k for $\nabla^2 L(x_k)$. if termination criteria satisfied stop with (x_k, u_k, v_k) a KKT point. else go to (ii)
- ii) Main step: Formulate the corresponding quadratic sub problem and Solve, with $\nabla^2 L(x_k)$ replaced by B_k to obtain a solution d_k with Lagrange multiplier (u_{k+1}, v_{k+1}) . If $d_k=0$ then, (x_k, u_{k+1}, v_{k+1}) satisfies the KKT condition for the original problem. Stop. Otherwise put $x_{k+1} = x_k + d_k$ and repeat (ii)

2.6 Penalty Merit Function

A principal disadvantage of SQP method described thus far is that convergence guaranteed only when the algorithm is initialized sufficiently close to a desirable solution.

To ensure that SQP method converges from remote starting point it is common to use a merit function ϕ to control the size of the steps (in line search method), or to determine whether a step is acceptable and the trust region radius needs to be modified (trust region methods).these are utilized only for the primal variables, not for the dual ones. Merit function is a combination of the objective and constraint functions which simultaneously minimized with the objective function at the solution of the problem. Another attribute of the merit function is that it guides the iterates of the algorithm and provide a measure of progress by means of exhibiting a descent. Questions may arise how should be the parameter μ selected so that a coherent merit function results. Choose an iteration $\mu > \max\{\|u_k\|, \|v_k\|\}$ yields d_k is a descent μ is not to change any more once the iterations apparently start to converge

Let us consider the I_1 merit function

$$F_E(x) = f(x) + \mu \left[\sum_{i=1}^m g_i^+(x) + \sum_{i=1}^l |h_i(x)| \right]$$

Lemma 2.6.1:

Given an iterate x_k . Consider the quadratic sub problem (9), where $\nabla^2 L(x_k)$ is replaced by a positive definite approximation B_k . let d_{x_k} solves this problem with Lagrange multipliers u and v . If $d_{x_k} \neq 0$ and $\mu \geq \max \{u_1, \dots, u_m, |v_1|, \dots, |v_l|\}$ then, d_{x_k} is a descent direction at $x = x_k$ for $F_E(x) = f(x) + \mu [\sum_{i=1}^m g_i^+(x) + \sum_{i=1}^l |h_i(x)|]$.

Proof: Using the primal feasibility, the dual feasibility and the complementary slackness conditions [9],[10] for the quadratic sub problem we have:

$$\begin{aligned} \nabla f(x_k)^t d &= -d^t B_k d - \sum_{i=1}^m u_i \nabla g_i(x_k)^t d - \sum_{i=1}^l v_i \nabla h_i(x_k)^t d \\ &= -d^t B_k d + \sum_{i=1}^m u_i g_i(x_k) + \sum_{i=1}^l v_i h_i(x_k) \\ &\leq -d^t B_k d + \sum_{i=1}^m u_i \max \{0, g_i(x_k)\} + \sum_{i=1}^l |v_i h_i(x_k)| \\ &\leq -d^t B_k d + \mu \left\{ \sum_{i=1}^m \max \{0, g_i(x_k)\} + \sum_{i=1}^l |h_i(x_k)| \right\} \end{aligned}$$

Now from equation 12, for a step length $\lambda \geq 0$

$$\begin{aligned} F_E(x_k) - F_E(x_k + \lambda d) &= [f(x_k) - f(x_k + \lambda d)] \\ &+ \mu \left\{ \sum_{i=1}^m [\max \{0, g_i(x_k)\} - \max \{0, g_i(x_k + \lambda d)\}] \right. \\ &\left. + \sum_{i=1}^l [|h_i(x_k)| - |h_i(x_k + \lambda d)|] \right\} \end{aligned}$$

Letting $O_i(\lambda)$ an appropriate function that approach zero as $\lambda \rightarrow 0$ for $i = 1, \dots, m + l$

We have for $\lambda > 0$ and sufficiently small

$$f(x_k + \lambda d) = f(x_k) + \lambda \nabla f(x_k)^t d + \lambda O_i(\lambda)$$

$$\begin{aligned} g_i(x_k + \lambda d) &= g_i(x_k) + \lambda \nabla g_i(x_k)^t d + \lambda O_i(\lambda) \\ &\leq g_i(x_k) - \lambda g_i(x_k) + \lambda O_i(\lambda) \end{aligned}$$

Hence $\max\{0, g_i(x_k + \lambda d)\} \leq (1 - \lambda) \max\{0, g_i(x_k)\} + \lambda O_i(\lambda)$

Similarly

$$\begin{aligned} h_i(x_k + \lambda d) &= h_i(x_k) + \lambda \nabla h_i(x_k)^t d + \lambda O_{m+i}(\lambda) \\ &= (1 - \lambda) h_i(x_k) + \lambda O_{m+i}(\lambda) \end{aligned}$$

And hence $|h_i(x_k + \lambda d)| \leq (1 - \lambda) |h_i(x_k)| + \lambda |O_{m+i}(\lambda)|$

Thus, we obtain for $\lambda \geq 0$ and sufficiently small

$$\begin{aligned} F_E(x_k) - F_E(x_k + \lambda d) \\ \geq \lambda \left[-\nabla f(x_k)^t + \mu \left\{ \sum_{i=1}^m \max\{0, g_i(x_k)\} + \sum_{i=1}^l |h_i(x_k)| \right\} + O(\lambda) \right] \end{aligned}$$

Where $O(\lambda) \rightarrow 0$, as $\lambda \rightarrow 0$

This gives $F_E(x_k) - F_E(x_k + \lambda d) \geq \lambda [d^t B_k d + O(\lambda)] > 0$, $\forall \lambda \in (0, \delta)$ for some $\delta > 0$. From the positive definiteness of B_k we have d is a descent direction at $x = x_k$ for the penalty merit function F_E .

Theorem 2.6.2: Assume that the sequence of points (x_k) generated by the algorithm SQP is contained in a compact subset of R^n , and that for any point x in X , and $\beta > 0$ the corresponding QP has a unique solution and unique KKT multipliers (u, v) satisfies $\mu \geq \max\{u_1, \dots, u_m, |v_1|, \dots, |v_l|\}$. Furthermore assume that all the matrices B_k are uniformly bounded and uniformly positive definite. i.e. $\exists \beta_1 > 0, \beta_2 > 0$ such that $\|B_k\| \leq \beta_1, d^t B_k d \geq \beta_2 \|d\|^2, \forall k$ then every limit point of (x_k) is a KKT point of (P).

To obtain local convergence we had to assume that the step size of 1 is used near the optimum. the merit function, therefore, must allow step length of 1 near the optimum when the matrix B_k are chosen appropriately. A significant disadvantage of the merit function $F_E(x)$ is that it may not allow the step length $\lambda = 1$ near the solution, this phenomena known as the Maratos effect, precludes us from obtaining quadratic or super linear convergence.

Ways to deal with this phenomenon, one may require the merit function to exhibit a decrease over a sequence of several iterations, but not necessarily at every iteration.

One of the ways of implementing this problem is to accept the step length $\lambda=1$, even if it leads to an increase in $F_E(x)$.

If taking such step size does not lead to decrease in $F_E(x)$ after a small number of iterations, then restore the original iterates and perform a line search.

Implementation of a line search algorithm is a critical issue when implementing a non linear programming algorithm, and has significant effect on the overall efficiency of the resulting code. We need a line search to stabilize the algorithm.

Algorithm for L1 merit function

- Initialization: Put the iteration counter $k=1$ and select a suitable starting solution with respect to some Lagrange multipliers $u_k \geq 0$ and v_k associated with the inequality and equality constraints respectively.
- Main step: Solve the quadratic programming Sub problem QP of the original problem with $\nabla^2 L(x_k)$ replaced by a positive definite approximation B_k and obtaining d_k along with the Lagrangian multipliers (u_{k+1}, v_{k+1}) . If $d_k = 0$ stop with x_k as a KKT solution for the problem P. Otherwise find $x_{k+1} = x_k + \lambda_k d_k$. Where λ_k minimizes $F_E(x_k + \lambda d_k)$ over $\lambda \in R, \lambda \geq 0$ update B_k to a positive definite matrix B_{k+1} (if necessary) and repeat the main step.

Example 1: solve the following equality constrained minimization problem:

$$\begin{aligned} & \text{minimize } (x_2 - 2)^2 - x_1^2 \\ & \text{s. t } 4x_1^2 + x_2^2 - 1 = 0 \end{aligned}$$

$$L(x, \lambda) = (x_2 - 2)^2 - x_1^2 + \lambda(4x_1^2 + x_2^2 - 1)$$

$$\nabla f(x) = \begin{pmatrix} -2x_1 \\ 2(x_2 - 2) \end{pmatrix} \quad \nabla h(x) = \begin{pmatrix} 8x_1 \\ 2x_2 \end{pmatrix} \quad \nabla^2 f(x) = \begin{bmatrix} -2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\nabla L(x, \lambda) = \begin{pmatrix} -2x_1 - 8x_1\lambda \\ 2(x_2 - 2) + 2\lambda x_2 \end{pmatrix}, \quad \nabla^2 L(x, \lambda) = \begin{bmatrix} -2 + 8\lambda & 0 \\ 0 & 2 + 2\lambda \end{bmatrix}$$

The corresponding quadratic sub problem is

$$\begin{aligned} & \text{minimize } -2x_1 d_1 + 2(x_2 - 2)d_2 + \frac{1}{2}((-2 + 8\lambda)d_1^2 + (2 + 2\lambda)d_2^2) \\ & \text{s. t } 8x_1 d_1 + 2x_2 d_2 + 4x_1^2 + x_2^2 - 1 = 0 \end{aligned}$$

Let the initial point be $(x_0, \lambda_0) = (2, 4, 0.5)$.

Step 0: The quadratic sub problem at $(x^0, \lambda^0) = (2, 4, 0.5)$ is given by

$$\begin{aligned} & \text{minimize } -4d_1 + 4d_2 + \frac{1}{2}(2d_1^2 + 3d_2^2) \\ & \text{s. t } 16d_1 + 8d_2 + 31 = 0 \end{aligned}$$

The corresponding KKT system is give by:

$$\begin{aligned} 2d_1 + 16\lambda^1 &= 4 \\ 3d_2 + 8\lambda^1 &= -4 \\ 16d_1 + 8d_2 &= -31 \end{aligned}$$

Solving the KKT system gives the solution

$$\begin{aligned} d_1^{(0)} &= -0.804 \\ d_2^{(0)} &= -2.267 \\ \lambda^{(1)} &= 0.35 \end{aligned}$$

The next iteration point is $x_1^{(1)} = x_1^{(0)} + d_1^{(0)} = 2.000 - 0.804 = 1.196$

$$x_2^{(1)} = x_2^{(0)} + d_2^{(0)} = 4.000 - 2.267 = 1.732.$$

Step 1: the quadratic sub problem corresponding to $(x^{(1)}, \lambda^{(1)}) = (1.196, 1.732, 0.350)$

$$\begin{aligned} & \text{minimize } -2.932d_1 - 0.536d_2 + \frac{1}{2}(0.80d_1^2 + 2.70d_2^2) \\ & \text{s. t } 9.568d_1 + 3.464d_2 + 31 = 0 \end{aligned}$$

The KKT system is given by:

$$\begin{aligned} 0.8d_1 + 9.568\lambda^{(2)} &= 2.392 \\ 2.7d_2 + 3.464\lambda^{(2)} &= 0.536 \\ 9.568d_1 + 3.464d_2 &= -7.721 \end{aligned}$$

Solving the KKT system gives the solution

$$\begin{aligned} d_1^{(1)} &= -0.734 \\ d_2^{(1)} &= -0.201 \\ \lambda^{(2)} &= 0.312 . \end{aligned}$$

The next iteration point is $x_1^{(2)} = x_1^{(1)} + d_1^{(1)} = 1.196 - 0.734 = 0.462$

$$x_2^{(2)} = x_2^{(1)} + d_2^{(1)} = 1.732 - 0.201 = 1.531 .$$

$$\lambda^{(2)} = 0.312 .$$

Step 2: the quadratic sub problem corresponding to $(x^{(2)}, \lambda^{(2)}) = (0.462, 1.531, 0.312)$

$$\begin{aligned} & \text{minimize } -0.924d_1 - 0.938d_2 + \frac{1}{2}(0.496d_1^2 + 2.624d_2^2) \\ & \text{s. t } 3.696d_1 + 3.062d_2 + 2.198 = 0 \end{aligned}$$

The KKT system is given by:

$$\begin{aligned} 0.496d_1 + 3.696\lambda^{(3)} &= 0.924 \\ 2.624d_2 + 3.062\lambda^{(3)} &= 0.938 \\ 3.696d_1 + 3.062d_2 &= -2.198 \end{aligned}$$

Solving the KKT system gives the solution:

$$\begin{aligned} d_1^{(2)} &= -0.575 \\ d_2^{(2)} &= -0.024 \\ \lambda^{(3)} &= 0.327 . \end{aligned}$$

The next iteration point is $x_1^{(3)} = x_1^{(2)} + d_1^{(2)} = 0.462 - 0.575 = -0.113$

$$x_2^{(3)} = x_2^{(2)} + d_2^{(2)} = 1.531 - 0.024 = 1.507.$$

Step 3: the quadratic sub problem corresponding to $(x^{(3)}, \lambda^{(3)}) = (-0.113, 1.507, 0.327)$

$$\begin{aligned} \text{minimize } & 0.226d_1 - 0.986d_2 + \frac{1}{2}(0.616d_1^2 + 2.654d_2^2) \\ \text{s. t } & -0.904d_1 + 3.014d_2 + 1.322 = 0 \end{aligned}$$

The KKT system is given by:

$$\begin{aligned} 0.661d_1 - 0.904\lambda^{(4)} &= -0.226 \\ 2.654d_2 + 3.014\lambda^{(4)} &= 0.986 \\ -0.904d_1 + 3.014d_2 &= -1.322 \end{aligned}$$

Solving the KKT system gives the solution:

$$\begin{aligned} d_1^{(3)} &= 0.490 \\ d_2^{(3)} &= -0.292 \\ \lambda^{(4)} &= 0.584 . \end{aligned}$$

The next iteration point is:

$$x_1^{(4)} = x_1^{(3)} + d_1^{(3)} = -0.113 + 0.490 = 0.377$$

$$x_2^{(4)} = x_2^{(3)} + d_2^{(3)} = 1.507 - 0.292 = 1.215.$$

Step 4: the quadratic sub problem corresponding to $(x^{(4)}, \lambda^{(4)}) = (0.377, 1.215, 0.584)$

$$\begin{aligned} \text{minimize } & -0.754d_1 - 1.57d_2 + \frac{1}{2}(2.672d_1^2 + 3.168d_2^2) \\ \text{s. t } & 3.016d_1 + 2.43d_2 + 1.045 = 0 \end{aligned}$$

The KKT system is given by:

$$\begin{aligned} 2.672d_1 + 3.016\lambda^{(5)} &= 0.754 \\ 3.168d_2 + 2.43\lambda^{(5)} &= 1.570 \\ 3.016d_1 + 2.43d_2 &= -1.045 \end{aligned}$$

Solving the KKT system gives the solution:

$$\begin{aligned} d_1^{(4)} &= -0.382 \\ d_2^{(4)} &= 0.044 \\ \lambda^{(5)} &= 0.588 . \end{aligned}$$

The next iteration point is:

$$x_1^{(5)} = x_1^{(4)} + d_1^{(4)} = 0.377 - 0.382 = -0.005$$
$$x_2^{(5)} = x_2^{(4)} + d_2^{(4)} = 1.215 + 0.044 = 1.259.$$

Step5: the quadratic sub problem at the point $(x^{(5)}, \lambda^{(5)}) = (-0.005, 1.259, 0.588)$ is :

$$\text{minimize } 0.01d_1 - 1.482d_2 + \frac{1}{2}(2.704d_1^2 + 3.176d_2^2)$$
$$\text{s. t } -0.04d_1 + 2.518d_2 + 0.585 = 0$$

The KKT system is given by:

$$2.704d_1 - 0.04\lambda^{(6)} = -0.01$$
$$3.176d_2 + 2.518\lambda^{(6)} = 1.482$$
$$-0.04d_1 + 2.518d_2 = -0.585$$

Solving the KKT system gives the solution:

$$d_1^{(5)} = 0.009$$
$$d_2^{(5)} = -0.232$$
$$\lambda^{(6)} = 0.881 .$$

The next iteration point is:

$$x_1^{(6)} = x_1^{(5)} + d_1^{(5)} = -0.005 + 0.009 = 0.004$$
$$x_2^{(6)} = x_2^{(5)} + d_2^{(5)} = 1.259 - 0.232 = 1.027.$$

Step 6: the quadratic sub problem at the point $(x^{(6)}, \lambda^{(6)}) = (0.004, 1.027, 0.881)$ is :

$$\text{minimize } -0.008d_1 - 1.946d_2 + \frac{1}{2}(5.048d_1^2 + 3.762 d_2^2)$$
$$\text{s. t } 0.032d_1 + 2.054d_2 + 0.055 = 0$$

The KKT system is given by:

$$5.048d_1 + 0.032\lambda^{(7)} = 0.008$$
$$3.762d_2 + 2.054\lambda^{(7)} = 1.946$$
$$0.032d_1 + 2.054d_2 = -0.055$$

Solving the KKT system gives the solution:

$$d_1^{(6)} = -0.005$$
$$d_2^{(6)} = -0.027$$
$$\lambda^{(7)} = 0.996 .$$

The next iteration point is:

$$\begin{aligned}x_1^{(7)} &= x_1^{(6)} + d_1^{(6)} = 0.004 - 0.005 = -0.001 \\x_2^{(6)} &= x_2^{(6)} + d_2^{(6)} = 1.027 - 0.027 = 1.000.\end{aligned}$$

Step 7: The quadratic sub problem at the point $(x^{(7)}, \lambda^{(7)}) = (-0.001, 1.00, 0.996)$ is :

$$\begin{aligned}\text{minimize } & 0.002d_1 - 2d_2 + \frac{1}{2}(5.968d_1^2 + 3.992d_2^2) \\ \text{s.t. } & -0.008d_1 + 2d_2 = 0\end{aligned}$$

The KKT system is given by:

$$\begin{aligned}5.968d_1 + -0.008\lambda^{(8)} &= -0.002 \\ 3.992d_2 + 2\lambda^{(8)} &= 2.00 \\ 0.008d_1 + 2d_2 &= 0\end{aligned}$$

Solving the KKT system gives the solution:

$$\begin{aligned}d_1^{(7)} &= 0.001 \\ d_2^{(7)} &= 0.000 \\ \lambda^{(8)} &= 1.000.\end{aligned}$$

The next iteration point is:

$$\begin{aligned}x_1^{(8)} &= x_1^{(7)} + d_1^{(7)} = -0.001 + 0.001 = 0.000 \\ x_2^{(8)} &= x_2^{(7)} + d_2^{(7)} = 1.000.\end{aligned}$$

Step 8: The quadratic sub problem at the point $(x^{(8)}, \lambda^{(8)}) = (0.000, 1.00, 1.000)$ is :

$$\begin{aligned}\text{minimize } & -2d_2 + \frac{1}{2}(6d_1^2 + 4d_2^2) \\ \text{s.t. } & 2d_2 = 0\end{aligned}$$

This immediately gives

$$\begin{aligned}d_1^{(8)} &= 0.000 \\ d_2^{(8)} &= 0.000\end{aligned}$$

Thus the algorithm terminates, with optimal solution $(x_1^*, x_2^*, \lambda^*) = (0.0, 1.0, 1.0)$.

References

1. Bazaraa, S.M. and et.al.(1993). Non Linear programming, theory and application. 2nd edition. John Wiley and Sons, Inc. Canada.
2. Boggs, T.P and Tolle, W.J. (2007). Sequential quadratic programming. Accessed on (15/01/2010 from (<http://www.personal.umich.edu/~mepelman/teaching/.../511notes07-12pdf>).
3. Mark, S.G. Introduction to Sequential Quadratic Programming. Accessed on 25/04/2010 from (<http://www.math.mtu.edu/~mspockeh/mas5630spring2003/.../511notes07-12>
4. Nocedal, J. and. Wright, J.S. (1999). Numerical Optimization.
5. Robert .M.F. (2004). Issues in Non convex Optimization. (Accessed on 28/10/ 2010 from (<http://www.springerlink.com/index/F85022116228k8450.pdf>
6. Simmons, M.D. Non linear programming for operation research.
7. _____. (2007). SQP and PDID algorithm for non linear programming. Accessed on 20/04/2010 from (<http://www.doc.ic.ac.uk/~br/berc/CODsqpipnotes.pdf>)