Topics in Constrained Optimization

Tejas M. Natu

Department Of Mathematics
Florida State University

Candidacy Exam - Fall 20

Candidacy Committee

-
-
-

Research Objectives

- To study and expand on the theory of constrained optimization and convex analysis on Riemannian manifolds.

- Some recent works as a motivation and starting point,
  - Bergman and Herzog (2019) - generalization of classical KKT conditions and constraint qualification using intrinsic concepts on smooth manifolds.
  - Bergman and Herzog (2019) - notion of the Fenchel conjugate on Riemannian manifold.
  - Boumal and Liu (2019) - Barrier and Augmented Lagrangian methods for constrained optimization on Riemannian manifolds.

Candidacy Objectives

- To study the classical theory of constrained optimization and constraint qualifications in the Euclidean geometric setting.
- To study the fundamentals of convex analysis and to conduct a study of some salient features of convex optimization problems and linear programming as special classes of constrained optimization problems.
- To conduct a basic study of the fundamentals of penalty methods and interior point methods as major classes of algorithms to solve the constrained optimization problem.

Candidacy Objectives

- To learn the basics of differential geometry as the first step towards studying optimization on Riemannian manifolds.
- To understand the recent generalizations of the KKT conditions and constraint qualifications on the smooth manifolds (Bergmann and Herzog,2019).
- Organize all the material studied in the form of simplified and detailed set of write ups made available on https://github.com/TejasNatuOpt/Candidacy-Material

A Word on Citations

- An overview of the topics covered in these slides follows on the next page. The references cited there are the primary resources used for conducting this whole study.
- These citations are included there to be consistent and to avoid clutter in the main slides (starting page number 8).
- There are some citations in the main slides. These are the ones which are not included on the overview page and are typically the ones which were used for a specific example, observation or as a motivation for a figure.
- Also included in the main slides are references to the write ups mentioned about earlier to point out details that have not been included in these slides but have been studied nevertheless.

An Overview

- Classical Karush Kuhn Tucker theory with an emphasis on constraint qualifications [1,2,3,4].
- Convex optimization problems and the linear programming problem. Some insights into the dual problem and its importance in algorithms to solve these problems [1,8,10,12].
- Basics of the barrier and primal dual interior point methods (the short step algorithm) for linear programming problem [7,8,10,12].
- Computational proof of concept for the short step algorithm.
- Translation of the KKT theory to the setting of a smooth manifold [14,15,16].

Constrained Optimization

- Given a constrained optimization problem $\mathcal{P}$

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && g_j(x) \leq 0,\ \ j = 1, 2, ..., l \\ & && h_i(x) = 0,\ \ i = 1, 2, ..., m \qquad\qquad (\mathcal{P}) \\ & && x \in \mathbb{R}^n \end{aligned}$$

- The functions $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^l$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ are general nonlinear functions and assumed to be continuously differentiable.

- **Goal**- To characterize the optimal solutions to the problem $(\mathcal{P})$.

- The classical Karush Kuhn Tucker conditions provide a convenient characterization which allows for the development of algorithms.

Preliminaries

- **Feasible Set-** The feasible set for the problem $\mathcal{P}$ is denoted by C and given as

$$C = \left\{x \in \mathbb{R}^n : g(x) \leq \bar{0},\ h(x) = \bar{0}\right\}$$

- **Local optimal solution-** For the problem $\mathcal{P}$, a feasible point $\bar{x} \in C$ is called constrained local optimal/minimizer if there exists $\delta > 0$ such that

$$f(\bar{x}) \leq f(x),\ \text{for all } x \in \mathcal{B}(\bar{x}, \delta)$$

where $\mathcal{B}(\bar{x}, \delta)$ is an open ball of radius δ around $\bar{x}$.

- **Feasible directions.** The set of directions at $\bar{x}$, denoted as $F(\bar{x})$ such that movement along these with appropriate step length ensures feasibility.

$$F(\bar{x}) = \left\{d : d \neq 0,\ \text{and } \bar{x} + \lambda d \in \mathbb{C},\ \lambda \in (0, \delta) \text{ for some } \delta > 0\right\}$$

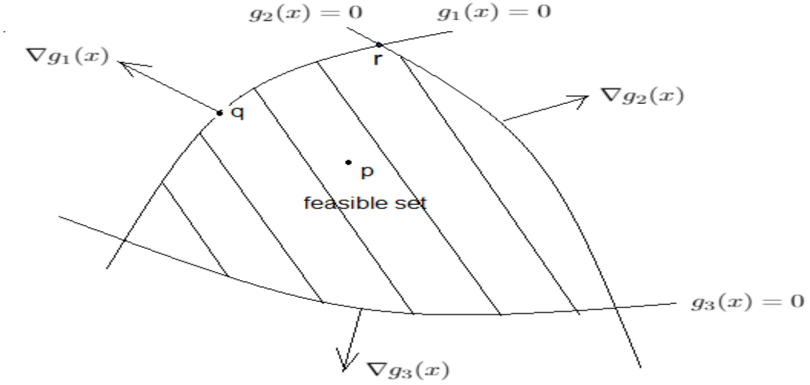
Active Constraints

- **Active constraints**- These is the set of inequality constraints that are satisfied with an equality at a given point $\bar{x} \in C$ and defined as
$$\mathcal{A}(\bar{x}) = \{j : g_j(\bar{x}) = 0\}$$
We denote indices corresponding to equality constraints by $\mathcal{E}$. denote the cardinality of $\mathcal{A}(\bar{x})$ by a.

- The figure on the next page shows a feasible set consisting of three inequality constraints.

- At p, none of the constraints are active, at q only $g_1(x)$ is active and at r, both $g_1(x)$ and $g_2(x)$ are active.

Active Constraints



Feasible Set/Active constraints

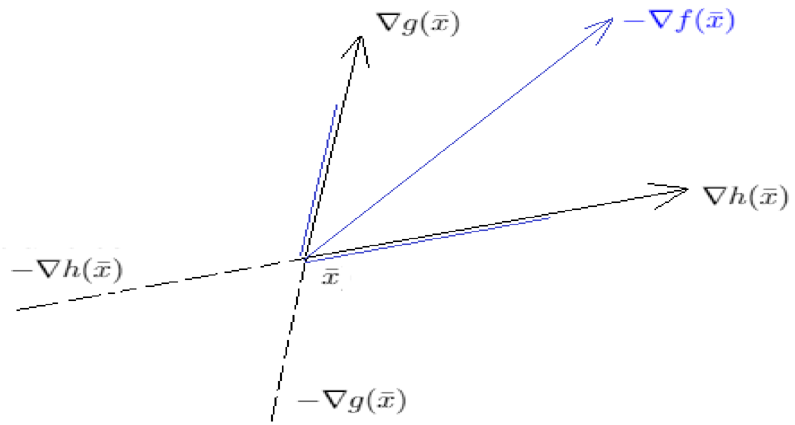
The KKT Conditions

Theorem

For the problem $\mathcal{P}$, let $\bar{x}$ be a local optimal solution. Then, differentiablity of (f,g,h) + Constraint Qualifications on (g,h) assure the existence of scalers $\bar{u}_j$ and $\bar{v}_i$ such that

$$\boxed{\begin{aligned} -\nabla f(\bar{x}) = \sum_{j=1}^{l} \bar{u}_j \nabla g_j(\bar{x}) + \sum_{i=1}^{m} \bar{v}_i \nabla h_i(\bar{x}) \ \textit{(stationarity)} \\ u_j g_j(\bar{x}) = 0,\ j = 1, \ldots, l \ \textit{(complementarity)} \\ \bar{u} \geq 0 \ \textit{(non negativity)} \\ h(\bar{x}) = 0,\ g(x) \leq 0 \ \textit{(feasibility)} \end{aligned}}$$

The scalers $\bar{u}$ and $\bar{v}$ are called the KKT multipliers.



Geometry of KKT conditions

The KKT Conditions- Some Comments

- **Necessary condition for unconstrained optimization**-When there are no constraints, the KKT necessary conditions reduce to $\nabla f(\bar{x}) = 0$. This is nothing but the necessary condition for an unconstrained optimization problem.

- **Generalization of the Lagrange Multipliers**-The KKT conditions are a generalization of the well known fact from the Lagrange multiplier technique which states that the gradient of objective function should be **parallel** to the gradient of constraint function.

- **Parallel becomes Conic Combination**- The negative of the gradient of objective function must be a linear combination of the gradients of active constraints and equality constraints split into two less than equal to type inequality constraints with positive weights. This is called as a conic combination.

The KKT Conditions- Some Comments

- **Complimentarity**-The complimentarity condition allows us to include inactive constraints into the stationarity to avoind keeping a track of active and inactive constraints separately.

- The fundamental geometric idea is that at a local optimal there cannot exist a direction that is simultaneously both feasible and reduces the function value. This is discussed in detail in the write up titled.

- The directions given by $\pm\nabla h_1(\bar{x}), \ldots, \pm\nabla h_1(\bar{x})$ are the directions along which we move outside of the feasible region and hence infeasible. The stationarity condition suggests that the vector $-\nabla f(\bar{x})$ is allowed to have a component along these directions as $-\nabla f(\bar{x})$ can have a component along infeasible directions but not along feasible directions.

- For a similar reason $-\nabla f(\bar{x})$ does not have a component along $-\nabla g(\bar{x})$ direction because $-\nabla g(\bar{x})$ is a feasible direction. However it can possess a component along the $\nabla g(\bar{x})$ as suggested by the stationarity condition.

- **The Lagrangian-** We define the Lagrangian for the problem $\mathcal{P}$ as

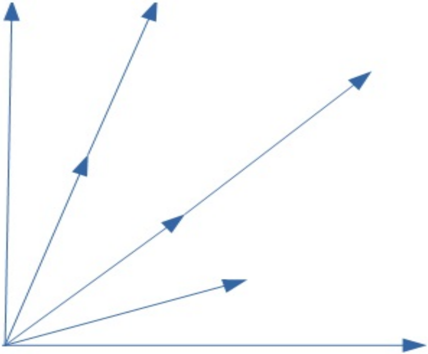
$$L(x, u, v) = f(x) + u^T g(x) + v^T h(x)$$

  The stationarity condition can also be expressed as $\nabla L_x(\bar{x}, \bar{u}, \bar{v}) = 0$

- **KKT point-** A local optimal $\bar{x}$ is called a KKT point if it satisfies the KKT conditions.

Cone

- **Cone-** A subset $S \in \mathbb{R}^n$ is called a cone if $td \in C$ for all $t \geq 0$ and $d \in S$.



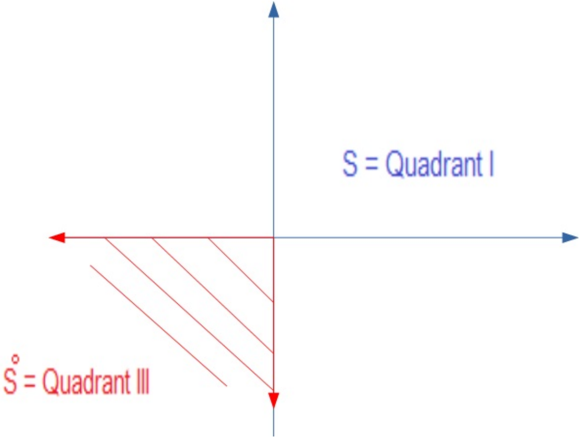
Cone - Quadrant I

Polar of a set/cone

- **Polar of a set.** The polar of a set $S \subseteq \mathbb{R}^n$ is defined as

$$S^o = \{p \in \mathbb{R}^n | p^T s \leq 0 \ \forall s \in S\}$$

Ex. The polar of quadrant 1 is quadrant 3



Two Important Cones - Tangent Cone ($T_C(\bar{x})$)

- Two cones related to the feasible set for the problem $\mathcal{P}$ shall be of interest to us, namely the **Bouligand tangent cone** $T_{\mathcal{C}}(\bar{x})$ and the **cone of linearized feasible directions** $(F_0(\bar{x}))$.

- **Tangent (d) to the Feasible Set at** $\bar{x} \in C$

$$d = \lim_{k \to \infty} \frac{x_k - \bar{x}}{\|x_k - \bar{x}\|}, \{x_k\} \to \bar{x}, x_k \in C,\ x_k \neq \bar{x},\ \forall k.$$

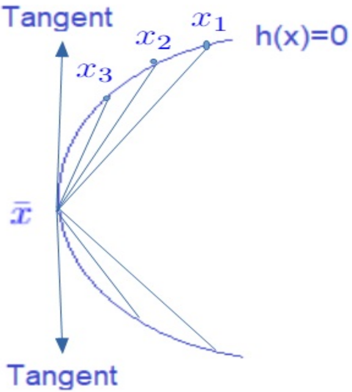
- **Bouligand Tangent Cone-**

$$T_C(\bar{x}) = \left\{ d \in \mathbb{R}^n : \text{d is a tangent to } C \; at \; \bar{x} \right\}.$$

- It can be shown that $T_C(\bar{x})$ is a cone. The details have been included in the write up [19].

Tangent Cone

Shown in the figure [23] are two tangential directions at a feasible point $\bar{x}$ for an equality constraint $h(x) = 0$



The Cone of linearized feasible directions $F_0(\bar{x})$

- **Cone of Linearized Feasible Directions** is defined as,

$$F_0(\bar{x}) = \left(d \neq 0 : \begin{matrix} \nabla g_j(\bar{x})^T d \leq 0 \\ \nabla h_i(\bar{x})^T d = 0 \end{matrix} \right)$$

  $j \in \mathcal{A}(\bar{x})$ and $i \in \mathcal{E}$

- **Example 1.** Consider the constraints

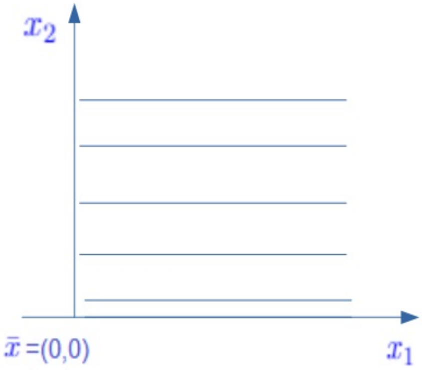
$$g_1(x) = -x_1 \leq 0,$$
$$g_2(x) = -x_2 \leq 0$$

  At the feasible point $\bar{x} = (0, 0)$ both the constraints are active and,

$$\nabla g_1(\bar{x}) = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \text{ and } \nabla g_2(\bar{x}) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

Example 1- Constraint Set

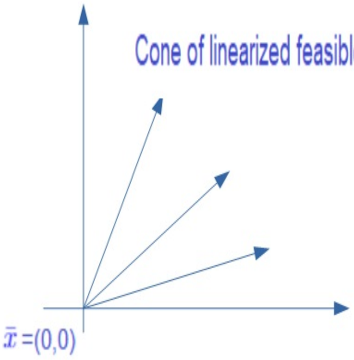
- The constraint set is essentially the first quadrant including the x_1 and x_2 axes.



Constraint Set - Quadrant I

Example 1- Cone of linearized feasible directions

- The cone of linearized feasible directions at $\bar{x} = (0, 0)$ is the set of all directions $d \in \mathcal{R}^2$ such that $\nabla g_1^T(\bar{x})d \leq 0$ and $\nabla g_2^T(\bar{x})d \leq 0$ and can be easily evaluated to be the first quadrant itself including the axes.



Relationship between the two cones.

We will look at an example that demonstrates the relationship between the tangent cone and cone of linearized feasible directions.

- **Example 2.** Consider the constraints

$$\begin{aligned} g_1(x) &= -x_1 - x_2 \leq 0, \\ h_1(x) &= x_1 x_2 = 0 \end{aligned}$$

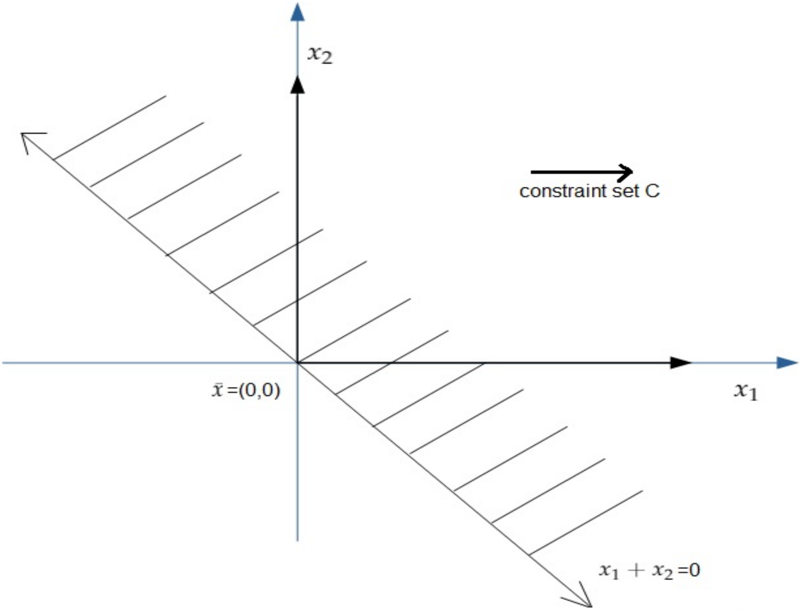
at the feasible point $\bar{x} = (0,0)$.

$$\nabla g_1(\bar{x}) = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \text{ and } \nabla h_1(\bar{x}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\text{Constraint Set } C = \{(x_1, x_2) \,|\, x_1 \geq 0,\ x_2 \geq 0,\ x_1 x_2 = 0\}.$$

The details of the calculations involved in this example can be found in the write up.

Example 2- Constraint Set



Example 2 Contd.

- The cone of linearized feasible directions $F_0(\bar{x})$ can be calculated as the set of directions $d \in \mathbb{R}^2$ such that $\nabla h_1(\bar{x})^T d = 0$ and $\nabla g_1(\bar{x})^T d \leq 0$.

- $F_0(\bar{x})$ is given as

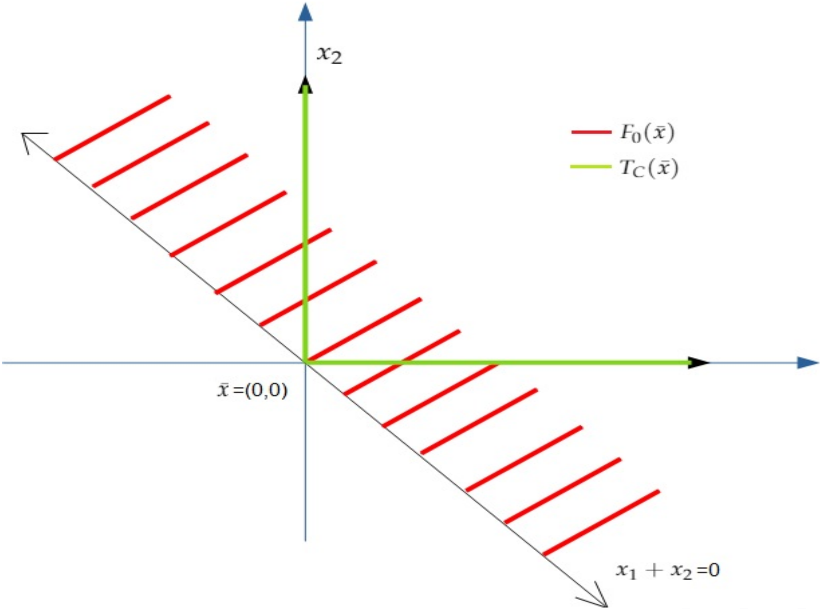
$$F_0(\bar{x}) = \{d \in \mathbb{R}^2 | -d_1 - d_2 \leq 0\}$$

- The tangent set to the constraints at $\bar{x}$

$$T_C(\bar{x}) = \{(d_1, d_2) \,|\, d_1 \geq 0,\ d_2 \geq 0,\ d_1 d_2 = 0\}.$$

- The picture on the next slide depicts the two cones.

Example 2- The Two Cones



Relationship between the Cones $T_C(\bar{x})$ and $F_0(\bar{x})$

- Notice in the previous example $T_C(\bar{x}) \subseteq F_0(\bar{x})$ and $T_C(\bar{x}) \neq F_0(\bar{x})$. And this holds true in general and is stated as a lemma.

Lemma

Let $\bar{x}$ be a feasible point of a feasible set C. Then we have

$$T_C(\bar{x}) \subseteq F_0(\bar{x})$$

The proof of this lemma is included in the write up [19].

Theorem

Given a non empty set constraint set C in R^n and $\bar{x} \in C$. Also assume f is differentiable at $\bar{x}$. If $\bar{x}$ locally solves the problem $\mathcal{P}$ then

$$\nabla f(\bar{x})^T d \geq 0, \text{ for all } d \in T_C(\bar{x}).$$

Proof Consider a direction $d \in T_C(\bar{x})$. Using the differentiablility of f at $\bar{x}$, Taylor's expansion gives,

$$f(x_k) - f(\bar{x}) = \nabla f(\bar{x})^T (x_k - \bar{x}) + o\big(\|x_k - \bar{x}\|\big)$$

divide by $\|x_k - \bar{x}\|$ to obtain

$$\frac{f(x_k) - f(\bar{x})}{\|x_k - \bar{x}\|} = \frac{\nabla f(\bar{x})^T (x_k - \bar{x})}{\|x_k - \bar{x}\|} + \frac{o(\|x_k - \bar{x}\|)}{\|x_k - \bar{x}\|}$$

Taking the limiting value as $\|x_k - \bar{x}\| \to 0$ the second term goes to 0, whereas $\frac{(x_k - \bar{x})}{\|x_k - \bar{x}\|} \to d \in T_{\mathbb{C}}(\bar{x})$. The term on the left always remains positive for sufficiently small $\|x_k - \bar{x}\|$ as $\bar{x}$ is a local minimum. So,

$$\boxed{\nabla f(\bar{x})^T d \geq 0, \text{ for all } d \in T_C(\bar{x})}$$

- **Note**- Although this theorem provides a geometric necessary condition, it does not provide an algebraic characterization for the local optimal $\bar{x}$ for the problem $\mathcal{P}$.

$F_0(\bar{x})$ in Matrix Form

- Form a $p \times n$ matrix with $p = a + 2m$

$$A = \begin{bmatrix} \nabla g_1^T(\bar{x}) \\ \vdots \\ \nabla g_a^T(\bar{x}) \\ \nabla h_1^T(\bar{x}) \\ \vdots \\ \nabla h_m^T(\bar{x}) \\ -\nabla h_1^T(\bar{x}) \\ \vdots \\ -\nabla h_m^T(\bar{x}) \end{bmatrix}_{p \times n},$$

- $Ad \leq 0$ for every $d \in F_0(\bar{x})$.

Important Observations

- At optimal $\bar{x}$, $Ad \leq 0$ for every $d \in F_0(\bar{x})$ and $\nabla f(\bar{x})^T d \geq 0$ for every $d \in T_C(\bar{x})$
- **If** $T_C(\bar{x}) = F_0(\bar{x})$, **for every** $d \in F_0(\bar{x})$, $Ad \leq 0$, $\nabla f(\bar{x})^T d \geq 0$.
- **Equivalently, no** d **such that** $Ad \leq 0$ **and** $-\nabla f(\bar{x})^T d > 0$.
- Let $c = -\nabla f(\bar{x})$, then the system $Ad \leq 0$ and $c^T d > 0$ has no solution. This renders system 1 in the following lemma unsolvable and causes system 2 to have a solution.

Farkas' Lemma (A theorem of alternative)

Lemma

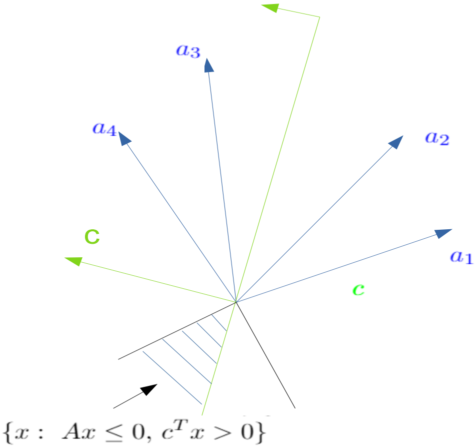
Given a matrix A of size $m \times n$ and $c \in R^n$. Then exactly one of the following systems has a solution,

1. $Ax \leq 0$ *and* $c^T x > 0$ *for some* $x \in R^n$
2. $A^T y = c$ *and* $y \geq 0$ *for some* $y \in R^m$

A geometric depiction of systems 1 and 2 is shown in the slides that follow. The figures are such that a_js are to be understood as the columns of matrix A^T.

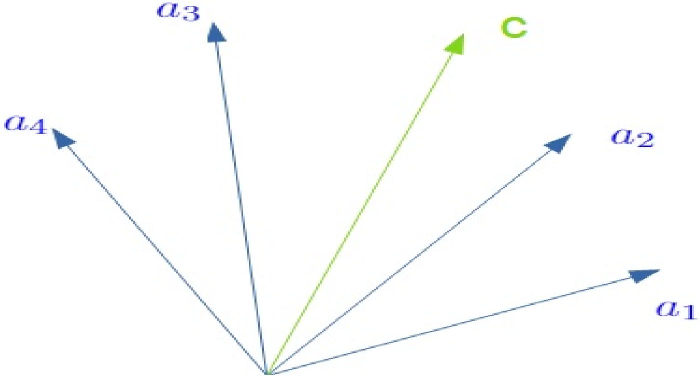
Farkas' Lemma - System 1

- System 1 has a solution



Farkas' Lemma - System 2

- **OR** System 2 has a solution



The KKT Conditions

- Following some details described in the write up [19], we have for $\bar{u} \geq 0$

$$-\nabla f(\bar{x}) = \sum_{j=1}^{a} \bar{u}_j \nabla g_j(\bar{x}) + \sum_{i=1}^{m} \bar{v}_i \nabla h_i(\bar{x})$$

- Remaining inequality constraints can be included by imposing the complimentarity condition, that is

$$\bar{u}_j g_j(\bar{x}) = 0,\ j = 1,\ \ldots,\ l \text{ (complementary slackness)}$$

$$\nabla f(\bar{x}) + \sum_{j=1}^{l} \bar{u}_j \nabla g_j(\bar{x}) + \sum_{i=1}^{m} \bar{v}_i \nabla h_i(\bar{x}) = 0 \text{ (Stationarity)}$$

- Thus $\bar{x}$ is a KKT point with $\bar{u}$ and $\bar{v}$ as the KKT multipliers or Lagrange multipliers.

But ...

- But $T_C(\bar{x}) \subseteq F_0(\bar{x})$ and not necessarily equal as seen in an earlier example.

- **Impose** $F_0(\bar{x}) \subseteq T_C(\bar{x})$ **as the constraint qualification** so that we have $T_C(\bar{x}) = F_0(\bar{x})$.

Abadie's constraint qualification

$$T_C(\bar{x}) = F_0(\bar{x})$$

Introduced by J.Abadie in 1967.

For the problem $\mathcal{P}$ with $(f, g, h \in C^1) +$

Abadie Constraint Qualification

$\bar{x}$-local optimal $\Rightarrow$ $\bar{x}$-KKT point

Other Constraint Qualifications

- **Linear Independence Constraint Qualification (LICQ)**

The set $\{\nabla g_1(\bar{x}), \ldots, \nabla g_a(\bar{x}), \nabla h_1(\bar{x}), \ldots, \nabla h_m(\bar{x})\}$ is linearly independent.

- **Mangasarian Fromovitz constraint qualifications (MFCQ)**

The set $\{\nabla h_1(\bar{x}), \ldots, \nabla h_m(\bar{x})\}$ is linearly independent and there exists a $d \in R^n$ such that $\nabla h_i(\bar{x})^T d = 0,\ i \in \mathcal{E}$ *and* $\nabla g_j(\bar{x})^T d < 0,\ j \in \mathcal{A}(\bar{x})$

- **Guinard's constraint qualification (GCQ)**

$$T_C(\bar{x})^o = F_0(\bar{x})^o \iff \left(T_C(\bar{x})^o \subseteq F_0(\bar{x})^o\right)$$

Relationship between Constraint Qualifications

- If LICQ holds at a point $\bar{x} \in C$ the the following chain of implications holds at $\bar{x}$ (proof in the write ups)

$$\boxed{\text{LICQ} \Rightarrow \text{MFCQ} \Rightarrow \text{ACQ} \Rightarrow \text{GCQ}}$$

- The KKT result can be restated with any one of these constraint qualifications in place of ACQ

$$\boxed{\begin{array}{c} \text{For the problem } \mathcal{P} \text{ with } (f, g, h \in C^1) + \\ \text{Any Constraint Qualification} \\ \bar{x}\text{-local optimal} \Rightarrow \bar{x}\text{-KKT point} \end{array}}$$

- The chain of implications does not hold in the opposite direction

$$\boxed{\text{GCQ} \not\Rightarrow \text{ACQ} \not\Rightarrow \text{MFCQ} \not\Rightarrow \text{LICQ}}$$

A counterexample for every case has been included in the write up [19].

Constraint Qualifications- Some Comments

Following the chain of implications on constraint qualifications,

- From LICQ to GCQ - it becomes more difficult to algebraically verify the validity of constraint qualifications at a feasible point. However,

- From LICQ to GCQ - the constraint qualifications become more general in the ability to characterize a local optimal as a KKT point.

- GCQ is the weakest constraint qualification.

GCQ weakest - in what sense ? This discussion will follow the discussion on the relationship between constraint qualifications and KKT multipliers.

Following the chain of implications on constraint qualifications, different CQs provide different information on the KKT multipliers, some are mentioned below (detailed discussion in the write ups)

- LICQ implies unique KKT multipliers - The matrix corresponding to the stationarity condition would possess full rank.

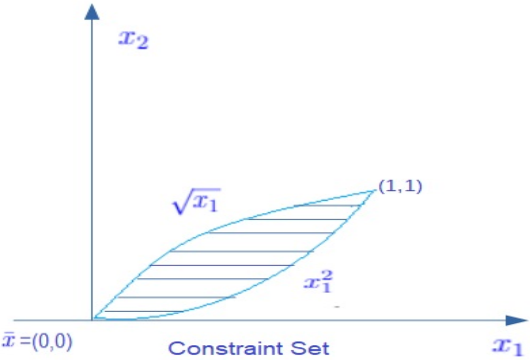
- ACQ and GCQ imply the existence of KKT multipliers - That ACQ implies existence of KKT multipliers has been explained earlier, the corresponding result for GCQ has been described in detail in the write ups. The multipliers in this case may not be unique.

What follows next are an example each for unique and non unique multipliers.

LICQ implies unique multipliers

- **Example 3.** Consider the problem

$$
\begin{aligned}
\text{minimize} \quad & -x_1x_2 \\
\text{subject to} \quad & g_1 = -x_2 + (x_1)^2 \leq 0, \\
& g_2 = x_2 - \sqrt{x_1} \leq 0
\end{aligned}
$$



LICQ implies unique multipliers

- The local optimal is $\bar{x} = (0, 0)^T$ where both the constraints are active and $\nabla g_1(\bar{x}) = (2, -1)^T$ and $\nabla g_1(\bar{x}) = (-1/2, 1)^T$ while $\nabla f(\bar{x}) = (-1, -1)^T$.

- Clearly for unique values of $u_1 = 1$ and $u_2 = 2$, the stationarity condition $-\nabla f(\bar{x}) = u_1 \nabla g_1(\bar{x}) + u_2 \nabla g_2(\bar{x})$ is satisfied. Moreover $u_1, u_2 > 0$ and complimentarity exists as both the constraints are active.

No LICQ, But ACQ

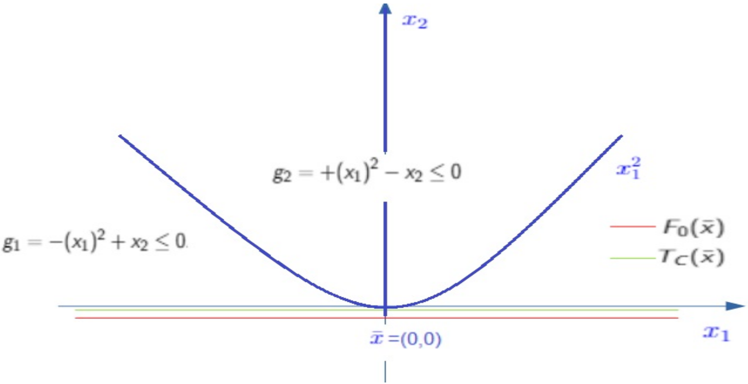
- **Example 4.** Consider the problem

$$\begin{aligned} \text{minimize} \quad & x_2 \\ \text{subject to} \quad & g_1 = -(x_1)^2 + x_2 \leq 0, \\ & g_2 = +(x_1)^2 - x_2 \leq 0 \end{aligned}$$

  The local optimal is $\bar{x} = (0,0)^T$ where both the constraints are active and $\nabla g_1(\bar{x}) = (0,1)^T$ and $\nabla g_1(\bar{x}) = (0,-1)^T$ while $\nabla f(\bar{x}) = (0,1)^T$.

- Clearly LICQ fails. But $T_C(\bar{x}) = F_0(\bar{x}) = \{(x_1, x_2),\ x_2 = 0, x_1 \in R\}$ Thus ACQ holds and characterizes the local optimal as a KKT optimal.

- An infinite number of solutions exist for the stationarity condition $-\nabla f(\bar{x}) = u_1 \nabla g_1(\bar{x}) + u_2 \nabla g_2(\bar{x})$. Moreover complimentarity exists as both the constraints are active.


x_2
$g_2 = +(x_1)^2 - x_2 \leq 0$
x_1^2
$g_1 = -(x_1)^2 + x_2 \leq 0$
$F_0(\bar{x})$
$T_C(\bar{x})$
$\bar{x}$ =(0,0)
x_1

GCQ as the weakest constraint qualification

Theorem

$T_C(\bar{x})^o = F_0(\bar{x})^o$ if and only if for every differentiable function f with local minimizer at $\bar{x}$ the KKT conditions are satisfied.

- GCQ introduced by M. Guinard in 1969.
- The theorem is due to Gould and Tolle (1971) who showed that GCQ is the weakest constraint qualification.
- **What if GCQ fails ? Then for the same constraint set, constraint functions and local minimizer $\bar{x}$, there must exist a function for which KKT conditions fail to satisfy.**

We observed this in example 5 which follows next.

An Example where GCQ fails

- The following is an example commonly found in optimization literature and this particular form was taken from [6]. However, the observation being talked about was made by us and details have been worked out in the write up [19].

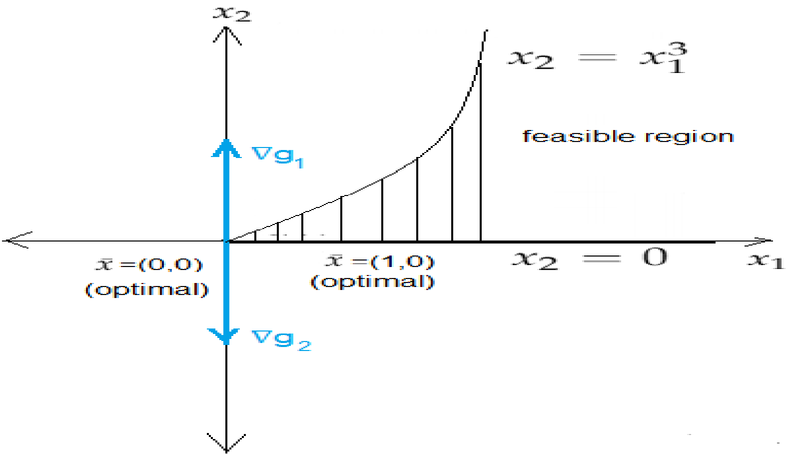
- **Example 5.** Consider the problem

$$\begin{aligned} \text{minimize} \quad & f_1 = x_2 \\ \text{subject to} \quad & g_1 = x_2 - (x_1)^3 \le 0, \\ & g_2 = -x_2 \le 0 \end{aligned}$$

- Every point of the form $(x_1, 0)^T$, $x_1 \ge 0$ is a local optimal with objective function value 0.

- We classify $(0,0)^T$ and some other local minimum along the x_1 axis say $\bar{x} = (1,0)^T$.

An Example



x_2
$x_2 = x_1^3$
feasible region
∇g_1
$\bar{x} = (0,0)$
(optimal)
$\bar{x} = (1,0)$
(optimal)
$x_2 = 0$
x_1
∇g_2

$\bar{x} = (1, 0)^T$

- Both inequality constraints g_1 and g_2 are active at $\bar{x} = (0, 0)^T$. Only g_2 active at $\bar{x} = (1, 0)^T$.
- Local optimal can be observed to be at $\bar{x} = (1, 0)^T$, with optimal objective function value 0.
- Gradients of active constraints at $\bar{x}$, $\nabla g_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$ and $\nabla f_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.
- LICQ satisfied $\bar{x} = (1, 0)^T$.
- $u_1 = 1$ uniquely satisfies the KKT equations classifying $\bar{x} = (1, 0)^T$ as a KKT point.

$\bar{x} = (0,0)^T$

- Local Optimal $\bar{x} = (0,0)^T$, with optimal objective function value 0.

- Both inequality constraint g_1 and g_2 active at $\bar{x} = (0,0)^T$.

- Gradients of active constraints at $\bar{x}$, $\nabla g_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $\nabla g_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$ and $\nabla f_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

- LICQ fails at $\bar{x}$.

- Infinite number of solutions to the KKT conditions.$\bar{x} = (0,0)^T$ is a KKT point for the problem.

Change the objective function

- Consider the problem

$$\begin{aligned} \text{minimize} \quad & f_2 = x_1 \\ \text{subject to} \quad & x_2 - (x_1)^3 \leq 0, \\ & -x_2 \leq 0 \end{aligned}$$

- Local Optimal $\bar{x} = (0,0)^T$, with optimal objective function value 0.

- Both inequality constraint g_1 and g_2 active at $\bar{x} = (0,0)^T$.

- Gradients of active constraints at $\bar{x}$, $\nabla g_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $\nabla g_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$ and $\nabla f_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

- No solution to the KKT stationarity condition.

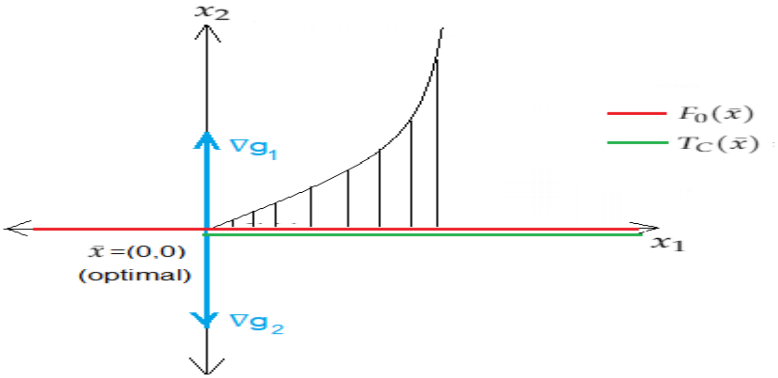
So Far

- Constraint set, constraint functions and local optimal $\bar{x} = (0,0)^T$ stay the same.

- **For f_1, $\bar{x} = (0,0)^T$ is a KKT point but for f_2, $\bar{x} = (0,0)^T$ fails to be a KKT point.**

- **This suggests the failure of GCQ to hold at $\bar{x} = (0,0)^T$.**

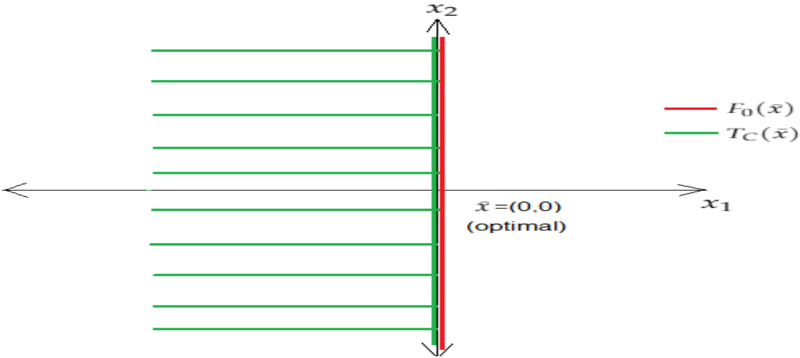
ACQ fails



$$F_0(\bar{x}) = \{d = (d_1, 0)^T : d_1 \in \mathbb{R}\}.$$

$$T_C(\bar{x}) = \{(d_1, 0) : d_1 \geq 0\}$$

GCQ fails



$$F_0(\bar{x})^o = \{(d_1, d_2)^T,\ d_1 = 0, d_2 \in \mathbb{R}\}$$

$$T_C(\bar{x})^o = \{(d_1, d_2)^T,\ d_1 \leq 0,\ d_2 \in \mathbb{R}\}.$$

$T_C(\bar{x})^o \neq F_0(\bar{x})^o$ and hence GCQ fails to hold at $\bar{x}$

KKT Conditions are necessary and not sufficient

- KKT conditions are first order necessary conditions.
- Second order information needed to classify the stationary point as local minimum or maximum.
- When are the KKT conditions sufficient ?
- **Question** More precisely, **for what type of problems and constraint qualifications are the KKT conditions sufficient ?**
- Before answering this question we look at another crucial notion in the theory of constrained optimization and raise another important question.

The Dual Problem

- Associated with the optimization problem P is a closely related problem D called the dual of the original problem (called the primal problem in this context).
- Then for any feasible solution $x \in C$

$$f(x) \geq f(x) + u^T g(x) + v^T h(x) = L(x, u, v)$$

for $u \geq 0$.

- Take infimums and obtain

$$f^* \geq inf_{x \in X} L(x, u, v) = \theta(u, v)$$

- The dual problem $\mathcal{D}$ can now be stated as

$$\boxed{\begin{array}{ll} \text{maximize} & \theta(u, v) \\ \text{subject to} & u \geq 0, \\ \text{where} & \theta(u, v) = \text{argmin}_x L(x, u, v), \qquad (\mathcal{D}) \end{array}}$$

The Dual Problem- Example

- Consider the problem,

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

  where $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$.

- Write the greater than type inequality as a less than type inequality constraint and form the Lagrangian as $L(x, u, v) = c^T x - u^T x + v^T(b - Ax) = -b^T v + (c - u - A^T v)^T x$, $u \geq 0$.

- The dual objective function can now be given as $g(u, v) = \inf_x L(x, u, v) = b^T v + \inf_x (c - u - A^T v)^T x$

The Dual Problem- Example

- Note that we must have $c - u - A^T v = 0$ in order for $(c - u - A^T v)^T x$ to attain an infimum (being a linear function). Thus $A^T v + u = c$ and thus we write the dual problem as

$$\begin{aligned} \text{maximize} \quad & b^T v \\ \text{subject to} \quad & A^T v + u = c \\ & u \geq 0 \qquad\qquad (\mathcal{D}) \end{aligned}$$

- The original or primal problem $\mathcal{P}$ is actually a linear programming problem in the standard form and we have now found its dual. This will be used later.

The Duality Gap

Theorem

Weak Duality Theorem - Let x be primal feasible and u, v be dual feasible, that is $u \geq 0$, then

$$f(x) \geq \theta(u, v)$$

- **Duality Gap-** The difference between the primal optimal objective function value and dual optimal objective function value.
- **Strong Duality-** When the duality gap between the primal optimal objective function value and dual optimal objective function value is zero.
- **If the duality gap is zero, it can be used as a stopping criterion for algorithms to solve constrained optimization problem.**
- **Question** When is the duality gap zero ?

Two Questions

- **1.** What type of problems and constraint qualifications are the KKT conditions sufficient for the KKT point to be the local optimal (in fact the global optimal) ?

- **2.** What type of problems and constraint qualification can we be assured of strong duality ?

- **Answer- Convex Optimization Problem + Slater's constraint Qualification.**

The requisite theory of convex analysis has been covered in the write up [20].

Convex Optimization Problem

- Consider the problem

$$\boxed{\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & g_j(x) \le 0,\ \ j = 1, 2, ..., l & \\ & Ax = b,\ \ i = 1, 2, ..., m & (\mathcal{CP}) \\ & x \in \mathbb{R}^n & \end{array}}$$

- We have the following assumptions.
  1. The functions f, g_j are convex functions in C^1.
  2. The convex functions h_i associated with equality constraints are affine. That is

$$h_i(x) = a_i^T x - b_i,\ \text{for all } i = 1 \ldots m$$

Convex Optimization Problem

- The constraint set for $\mathcal{CP}$ is given by

$$C = \{x \in \mathbb{R}^n,\ g(x) \leq 0,\ Ax = b\}$$

  The constraint set C is an intersection of convex sets and hence is a convex set.

- **For a convex optimization problem, the local minimum is the global minimum.**

- In fact for the convex optimization problem $\mathcal{CP}$ it can be shown that the **KKT point is the global optimizer for the problem**. A detailed proof for this has been included in the write ups [19,20].

Slater's Constraint Qualification

- **Slater's constraint qualification.** These qualifications states that there should exist $x \in \mathbb{R}^n$ which satisfies the inequality constraints with strict inequality, that is $g_i(x) < 0$ for all i.
- SCQ ensures MFCQ holds true at every point in the constraint set.
- **As a result the local (hence global optimal) is a KKT point.**
- We thus obtain the following chain of implications for a convex set.

$$\boxed{\text{SCQ} \Rightarrow \text{MFCQ} \Rightarrow \text{ACQ} \Rightarrow \text{GCQ}}$$

Convex Optimization Problem and KKT Multipliers

- For the Convex optimization problem *CP* with Slaters constraint qualification, strong duality holds and

$$(\bar{x}, \bar{u}, \bar{v}) \text{ satisfy KKT conditions } \iff \begin{pmatrix} \bar{x} \text{ primal solution} \\ (\bar{u}, \bar{v}) \text{ dual solution} \\ f(\bar{x}) = \theta(\bar{u}, \bar{v}) \end{pmatrix}$$

- As a result, **solving the KKT equations amounts to solving the problem.**

- There exist counterexamples which show that in the absence of Slaters constraint qualification, strong duality may not exist and that the global optimal may not be a KKT point. These examples have been included in the write up [20].

- Duality gap can be used as a stopping criterion.

- We can write the KKT conditions for the convex optimization problem $\mathcal{CP}$ as

$$\boxed{\begin{aligned} \nabla f(x) + \sum_{j=1}^{l} u_j \nabla g_j(x) + A^T v = 0 \\ u_j g_j(x) = 0,\ j = 1, \dots, l, \\ Ax = b,\ g_j(x) \leq 0,\ j = 1, \dots, l, \\ u \geq 0 \end{aligned}}$$

All the results in the previous three slides have been studied and explained in detail in the write ups [19,20].

- A particular class of convex optimization problem we will focus on is the linear programming problem.

- The linear programming problem in the standard form

$$\boxed{\begin{array}{lll} \text{minimize} & c^T x & \\ \text{subject to} & Ax = b & \\ & x \geq 0 & \qquad\qquad \text{(LP- Standard Form)} \end{array}}$$

  where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.

- **KKT Conditions (LP)-**

$$\boxed{\begin{array}{c} A^T v + u = c \\ Ax = b \\ x_j u_j = 0, \; j = 1, \ldots, n \\ (x, u) \geq 0 \end{array}}$$

Complimentarity/Duality Gap

- $\bar{x}$ and $(\bar{u}, \bar{v})$ are primal and dual solutions for the given linear programming problem iff they satisfy KKT equations.

- LP - Primal

$$\begin{aligned} \min \quad & c^T x \\ \text{sub} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

LP - Dual

$$\begin{aligned} \max \quad & b^T v \\ \text{sub} \quad & A^T v + u = c \\ & u \geq 0 \end{aligned}$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.

- A simple calculation gives the duality gap-

$$c^T x - b^T v = c^T x - v^T Ax = c^T x - x^T A^T v = c^T x - x^T (c - u) = x^T u$$

- **Thus complimentarity (referred to as complimentary slackness in LP literature !) measures the duality gap.**

Summary So Far

- We have so far seen the classical KKT theory and the role of constraint qualifications.
- We looked at the dual problem associated with a constrained optimization problem and defined the duality gap.
- We saw that for a convex optimization problem, solving the KKT conditions for $(\bar{x}, \bar{u}, \bar{v})$ amounts to solving the primal and dual problems with $\bar{x}$ being the primal optimal and $(\bar{u}, \bar{v})$ being the dual optimal with zero duality gap.
- We saw that the for a particular type of convex optimization problem, the linear programming problem, the complimentarity condition allows us to measure the duality gap.

What Lies Ahead- A historical perspective first

- We shall now look at an important class of algorithms to solve the convex optimization problems, called the barrier method which is a part of a bigger class of algorithms called the sequential unconstrained minimization techniques(SUMT, Fiacco and Mccormick 1969).

- Linear programming(LP) has played a crucial role in the history of mathematical optimization. The classical simplex method to solve the LP is a landmark algorithm in the field of computational mathematics across disciplines.

- In 1972 it was shown by Klee and Minty that LP has exponential complexity in the worst case. They showed this using a rather pathological example of what came to be known as the Klee and Minty cube. However, for all practical problems, the simplex was the go to method for its remarkable efficiency in practice.

What Lies Ahead- A historical perspective first

- Karmarkar's algorithm (1984)- polynomial worst case complexity + practically comparable to the simplex method for large problems (a controversial claim !).
- But connections between Karmarkar's algorithm and the barrier method were found in (1988) which rekindled interest in the barrier method (SUMT).
- The barrier method give way to the primal dual interior point methods.
- Primal dual interior point methods - polynomial worst case complexity + practically comparable to the Simplex for large problems (no controversy here !).

What Lies Ahead

- We shall concern ourselves with the barrier method to solve the convex optimization problem and check for its correctness for a linear programming problem.
- We will motivate the primal dual interior point methods using the barrier method. Some important methods in this class include the short step algorithm, the long step algorithm, Mehrotra's predictor corrector algorithm etc.
- We shall focus on the short step algorithm to solve the linear programming problem.
- We will point out the salient features of this method based on theoretical details that have been fully explored in the write up [21].
- We will check the correctness of the method and confirm the theoretical foundations of the method with computational experiments. We first start with a discussion of the perturbed KKT system.

Perturbed KKT System

- Solving the KKT system as it is is an aggressive strategy.
- Instead, perturb the complimentarity condition and obtain the perturbed KKT system.

$$\boxed{\begin{aligned} \nabla f(x) + \sum_{j=1}^{l} u_j \nabla g_j(x) + A^T v = 0 \\ u_j g_j(x) = -\mu,\ j = 1, \ldots, l, \\ Ax = b,\ g_j(x) \leq 0,\ j = 1, \ldots, l, \\ u \geq 0 \end{aligned}}$$

- Employ Newton's method to solve a family of a system of non linear equations parametrized by μ, where $\mu \to 0^+$.
- The perturbed system approximates the original KKT system better as $\mu \to 0^+$.

Instead Solve the Perturbed KKT System

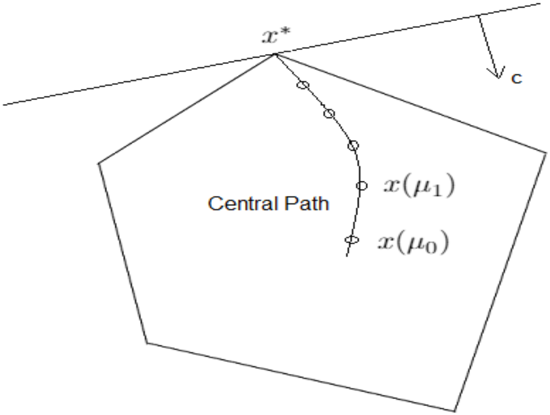
- Assuming we are able to satisfy the feasibility conditions, we have a non linear system of equations to solve.

- A very deep convergence theory exists which confirms that as $\mu \to 0^+$, the sequence of solutions to the perturbed KKT system $(\bar{x}(\mu), \bar{u}(\mu), \bar{v}(\mu)$ converges to the solution $(\bar{x}, \bar{u}, \bar{v})$ of the original problem.

- **Primal Dual Central Path-** The locus of points $(\bar{x}(\mu), \bar{u}(\mu), \bar{v}(\mu))$ as $\mu \to 0^+$ carves out a differentiable path called the primal dual central path $\mathcal{C}$.

- The projection of primal dual central path on the primal space is called the **primal central path**.

The Central Path



Connections with the Logarithmic Barrier Function

- **Goal-** To make the inequality constraints implicit in the objective function

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \leq 0 \end{array}$$

where $g : \mathbb{R}^n \to \mathbb{R}^m$ and $C = \{x \in \mathbb{R}^n;\ g(x) \leq 0\}$.

- Define indicator function $I_C(x)$

$$I_C(x) = \begin{cases} 0, & x \in C \\ +\infty, & x \notin C \end{cases}$$

$$\Rightarrow f(x) + I_C(x) = \begin{cases} f(x), & x \in C \\ +\infty, & x \notin C \end{cases}$$

- Indicator function is discontinuous, makes the calculus difficult.

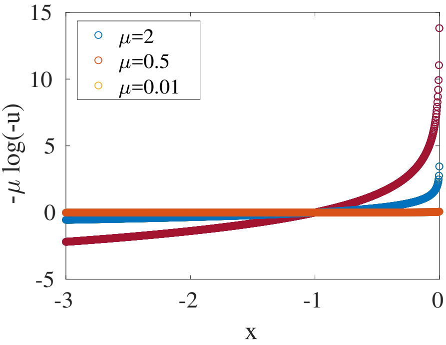
Logarithmic Barrier Function

- Instead use Frisch's logarithmic barrier function defined as

$$I_{log}(x) = -\mu \sum_{j=1}^{l} log(-g_j(x))$$

  $I_{log}(x)$ is twice continuously differentiable.

- The function $-\mu log(-u)$ gives a better approximation of the indicator function as $\mu \rightarrow 0^+$



Convex Optimization problem with Log Barrier

- The convex optimization problem $(\mathcal{CP})$ can be written with inequality constraints implicit into the objective function using the Log barrier as,

$$
\begin{aligned}
\text{minimize} \quad & f(x) - \mu \sum_{j=1}^{l} log(-g_j(x)) \\
\text{subject to} \quad & Ax = b \qquad\qquad (\mathcal{CLB}) \\
& x \in \mathbb{R}^n
\end{aligned}
$$

- The KKT conditions for this problem are given as

$$\nabla f(x) - \mu \sum_{j=1}^{l} \frac{1}{g_j(x)} \nabla g_j(x) + A^T v = 0$$

$$Ax = b$$

Rewrite and obtain perturbed KKT

- Make the substitution $u_j = -\frac{\mu}{g_j(x)}$ for all j and obtain

$$\boxed{\begin{aligned} \nabla f(x) + \sum_{j=1}^{l} u_j \nabla g_j(x) + A^T v = 0 \\ u_j g_j(x) = -\mu,\ j = 1, \ldots, l, \\ Ax = b,\ \ g_j(x) \leq 0,\ j = 1, \ldots, l, \\ u \geq 0 \end{aligned}}$$

- These are the perturbed KKT conditions.
- A one to one correspondence between the perturbed KKT system and the logarithmic barrier problem.
- **Minimizing the barrier problem $\iff$ solving the perturbed KKT conditions**

Dual Points from the KKT Conditions

- From the perturbed KKT conditions, the dual feasible points can be read off.

- Suppose $\bar{x}(\mu)$, $\bar{u}(\mu)$, $\bar{v}(\mu) = \mu w$ solves the perturbed KKT system corresponding to μ. We have then

$$\nabla f(\bar{x}(\mu)) + \sum_{j=1}^{l} \bar{u}_j(\mu) \nabla g_j(\bar{x}(\mu)) + A^T \bar{v}(\mu) = 0$$

$$\bar{u}_j(\mu) g_j(\bar{x}(\mu)) = -\mu, \; j = 1, \dots, l$$

$$A\bar{x}(\mu) = b, \; g_j(\bar{x}(\mu)) \leq 0, \; j = 1, \dots, l$$

$$\bar{u}(\mu) \geq 0$$

- For each μ, from the corresponding KKT system, we can isolate the value of u_j from the complimentarity condition as $\bar{u}_j(\mu) = -\frac{\mu}{g_j(\bar{x}(\mu))}$.

Dual Points from the KKT Conditions

- Claim: $\bar{u}_j(\mu) = -\frac{\mu}{g_j(\bar{x}(\mu))}$ and $\bar{v}(\mu) = \mu w$ are dual feasible points for the original problem.

- The Lagrangian $L(x, u, v)$ for the particular choice of $u = \bar{u}(\mu)$ and $v = \bar{v}(\mu)$ can be written as

$$L(x, \bar{u}, \bar{v}) = f(x) + \sum_{j=1}^{l} \bar{u}_j g_j(x) + \bar{v}^T(Ax - b)$$

- $\bar{x}(\mu)$ is the minimizer for $L(x, \bar{u}, \bar{v})$ since $\nabla L(\bar{x}(\mu), \bar{u}, \bar{v}) = 0$ (stationarity).

- The dual objective function for the original problem is

$$g(u, v) = inf_x\, L(x, u, v),\; u \geq 0$$

$$g(\bar{u}, \bar{v}) = inf_x\, L(x, \bar{u}, \bar{v})$$

Dual Points from the KKT Conditions

- Since $\bar{x}(\mu)$ minimizes $L(x, \bar{u}, \bar{v})$, we have $g(\bar{u}, \bar{v}) > -\infty$, $(\bar{u}, \bar{v})$ lies in the domain of the Lagrangian dual objective function for the original problem.
- Notice $\bar{x}(\mu)$ is primal feasible because it is the solution to the barrier problem which is defined only on the feasible set.
- This allows to bound the duality gap having calculated the solution to the barrier problem. We have

$$g(\bar{u}, \bar{v}) = inf_x \, L(x, \bar{u}, \bar{v})$$

$$g(\bar{u}, \bar{v}) = f(\bar{x}(\mu)) + \sum_{j=1}^{l} \bar{u}_j g_j(\bar{x}(\mu)) + \bar{v}^T(A\bar{x}(\mu) - b)$$

$$g(\bar{u}, \bar{v}) = f(\bar{x}(\mu)) - l\mu$$

Duality Gap from solution to the barrier problem

- We thus have

$$f(\bar{x}(\mu)) - g(\bar{u}, \bar{v}) = l\mu$$

Let $\bar{f}$ be the primal optimal objective function value for original problem $\mathcal{CP}$. Since $\bar{u}(\mu), \bar{v}(\mu)$ is dual feasible for the original problem, we have $\bar{f} \geq g(\bar{u}, \bar{v})$ thus

$$-g(\bar{u}, \bar{v}) \geq -\bar{f}$$

$$f(\bar{x}(\mu)) - g(\bar{u}, \bar{v}) \geq f(\bar{x}(\mu)) - \bar{f}$$

This finally gives

$$\boxed{f(\bar{x}(\mu)) - \bar{f} \leq l\mu}$$

- Thus the difference between the optimal objective function values of the barrier problem and the original problem is bounded above by $l\mu$ where l is the number of inequality constraints and $\mu \to 0^+$.

The Barrier Method - Basic Idea

- Solve the perturbed KKT system for a given μ using Newton's method.

- Assuming the non negativity conditions are met, we first isolate the value of u as $-\frac{\mu.}{g(x)}$ from the complimentarity and substitute into the stationarity to obtain following system

$$r(x, v) = \begin{pmatrix} \nabla f(x) + \mu \sum_{j=1}^{l} -\frac{1}{g_j(x)} \nabla g_j(x) + A^T v \\ Ax - b \end{pmatrix} = 0$$

- Newton's update is now given as

$$\begin{pmatrix} H(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta v \end{pmatrix} = -r(x, v)$$

The Barrier Method - Basic Idea

- where $H(x)$ is given as

$$H(x) = \nabla^2 f(x) + \sum_{j=1}^{l} \frac{\mu}{g_j^2(x)} \nabla g_j(x) \nabla g_j^T(x) + \sum_{j=1}^{l} -\frac{\mu}{g_j(x)} \nabla^2 g_j(x)$$

- The non negativity conditions can be met using proper initialization and backtracking.

- Update μ and use the solution $x^*(\mu)$ as the initial solution for the next iteration.

- Thus there are inner and outer iterations. Inner iterations solve the perturbed KKT for given value of μ while outer iterations update μ.

Algorithm 1 The Barrier Method

Given strictly feasible $x = x^0$, $\mu = \mu^0$, $\beta < 1$, tolerance $= \epsilon$

while $\mu l > \epsilon$ **do**

Compute $\bar{x}(\mu)$ by minimizing $f(x) - \mu \sum_{j=1}^{l} log(g_j(x))$ subject $Ax = b$ starting at x^0.
Equivalently solve the perturbed KKT conditions using the Newton's method.

Update $x = \bar{x}(\mu)$

Decrease $\mu = \beta\mu$.

end output - optimal $\bar{x}(\mu)$

Correctness of the method

- We generate an instance of linear programming problem in such a way that we have a starting point on the central path.
- Matrix $A \in \mathbb{R}^{5\times 10}$, $x^0 \in \mathbb{R}^{10}$, $x^0 > 0$ and $v^0 \in \mathbb{R}^5$ using the rand command on MATLAB.
- Choose an initial value of $\mu^0 > 0$. We used $\mu^0 = 1$ and was reduced by a factor of $\frac{1}{50}$.
- Then $u^0 = \frac{1.}{x^0}$. This ensures $x_j^0 u_j^0 = 1, \forall j$.
- Construct b as $b = Ax^0$ and c as $c = A^T v + u$
- This ensures we have a linear programming instance with starting point on the central path.

Correctness of the method

- The solution obtained was compared with the GNU Linear Programming Kit(GLPK) to test the correctness. The GLPK was set to its default settings to use Simplex method.

$\bar{x}$ – Barrier Method	$\bar{x}$ – GNU-LPK
0.0000313587	0.00000
0.0000018036	0.00000
0.9735660219	0.97357
0.0000018292	0.00000
1.4714308889	1.47143
0.0000072748	0.00000
0.0000027989	0.00000
0.7606919232	0.76070
0.9242006268	0.92419
2.6958763375	2.69590

Table: Optimal Solution to a LP of size 5×10

Primal Dual Methods for LP - Motivation

- As pointed out earlier, barrier method gives way to a more sophisticated class of algorithms called the primal dual interior point algorithms.
- These primal dual interior point methods have been shown to compete with the simplex method.
- The method we shall focus is one of the first methods in this class called the short step algorithm. This method per se is not competitive enough but provides a theoretical foundation for methods that do actually compete with the simplex method.
- **Key idea**- Staying on the central path is not a priority. The priority is to reach the optimal solution, even if it is by staying close to the central path and not on the central path itself.

Primal Dual Methods for LP - Motivation

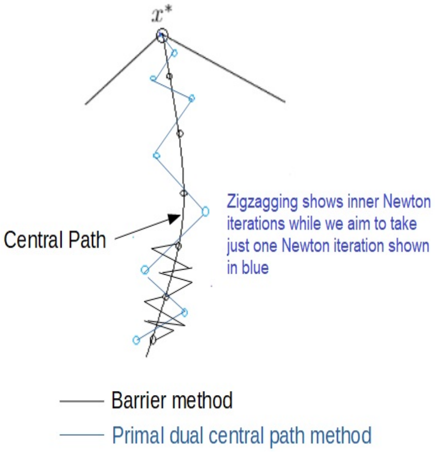
- Staying on the central path amounts to solving the perturbed KKT system in the inner loop for a particular value of μ till a solution is obtained. This is equivalent to solving the corresponding barrier optimization problem till optimality. Several Newton iterations are required.

- We can instead choose not to completely solve the KKT system. **Instead we take only one Newton step and move on.**

- **Since we do not solve the problem completely we do not stay on the central path.** This is because central path is the locus of optimal solutions to the family of barrier problems (or a solution to a family of KKT equations) parametrized by μ.

A Question Arises

- If not the central path, then how to we guide the iterates to stay on their course towards reaching the optimal solution.
- Answer- An obvious answer is to stay close to the central path if not the central path.
- The zigzagging shows the inner Newton's iterations while we aim to take just one Newton step shown in blue
- So we define a neighborhoods of the central path to make the idea of staying close to the central path precise. But an important discussion first.



Averages - An important discussion !!

- Look at the perturbed KKT conditions again where $\sigma \in (0,1)$

$$
\begin{aligned}
A^T v + u &= c && (1)\\
Ax &= b && (2)\\
x_j u_j &= \sigma\mu, \; j = 1, \ldots, n && (3)\\
(x, u) &\geq 0 && (4)
\end{aligned}
$$

- The solution $x_{\sigma\mu}, u_{\sigma\mu}, v_{\sigma\mu}$ to the KKT equations corresponding to the parameter $\sigma\mu$ would be such that the average $\frac{x^T u}{n}$ equals $\sigma\mu$ and each individual term $x_j u_j$ equals the average.

- It is possible that conditions $1, 2$ and 4 hold and the average $\frac{x^T u}{n}$ equals $\sigma\mu$. Yet we do not have a solution to the perturned KKT conditions if individual values of terms $x_j u_j$ do not equal the average.

Neighborhood of the Central Path

- Several neighborhoods of the central path are defined in literature, like the $N_{-\infty}(\gamma)$ nbh or the $N_2(\theta)$ nbh.

- We define the $N_2(\theta)$ neighborhood of the central path as

$$N_2(\theta) = \{(x, u, v) \in \mathcal{F}^0;\ ||XU - \mu e||_2 \leq \theta\mu\}$$

where

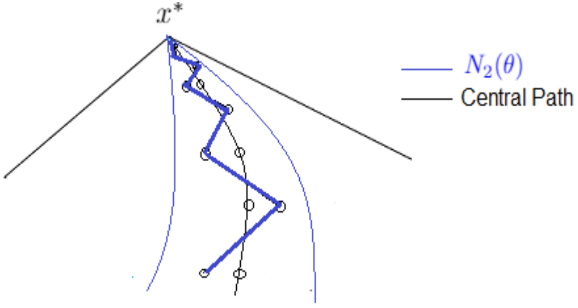
$$\mathcal{F}^\circ = \{(x, v, u) \,|\, Ax = b,\ A^T v + u = c,\ (x, u) > 0\}$$

and

$$X = diag(x_1, \dots, x_n),\ X = diag(u_1, \dots, u_n)$$

- The neighborhood really is a collection of all points that satisfy the KKT conditions $1, 2$ and 4. But they deviate from the central path $\mathcal{C}$ only because the pairwise products $x_j u_j$ are not identical. The $N_2(\theta)$ nbh ensures that the deviation is bounded.

- The following picture helps picturize the $N_2(\theta)$ nbh.



The $N_2(\theta)$ neighborhood

- Iterates are forced to stay in the neighborhood as they traverse through the strictly feasible primal dual space to the optimal solution.

Short Step Algorithm - Basic Idea

- The linearized KKT system is given as

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ U & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta v \\ \Delta u \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -XUe + \sigma \mu e \end{pmatrix} \tag{1}$$

- If we were to actually solve the system using Newton updates, then the solution $(x^{k+1}, v^{k+1}, u^{k+1})$ would lie on the central path with a reduction in the duality gap by the factor σ_k and each term $x_j^{k+1} u_j^{k+1}$ being equal to $\sigma_k \mu_k$.

- We do not solve the KKT system till we actually obtain a solution, instead we take only one Newton step by solving the system only once. Thus, our solution does not lie on the central path as expected.

Short Step Algorithm - Basic Idea

- However with clever choices of $\theta \in (0,1)$ and the parameter $\sigma_k \in (0,1)$ we can ensure that $(x^{k+1}, v^{k+1}, u^{k+1}) = (x^k, v^k, u^k) + (\Delta x, \Delta v, \Delta u)$ where $(\Delta x, \Delta v, \Delta u)$ is obtained by solving (1) for $\sigma_k \mu_k$, lies in $N_2(\theta)$.

- This means that the $(x^{k+1}, v^{k+1}, u^{k+1}) \in \mathcal{F}^0$ and the average values of $x_j^{k+1} u_j^{k+1}$ are at the most equal to $\sigma_k \mu_k$ and that individual $x_j^{k+1} u_j^{k+1}$ are not necessarily equal to $\sigma_k \mu_k$.

- However as $\sigma_k \mu_k$ approaches 0, the $N_2(\theta)$ neighborhood becomes narrower forcing the corresponding iterates to stay closer and closer to the central path and finally tapering into the solution.

- The short step algorithm is due to Monteiro and Adler (1989).

Algorithm 2 The Short Step Algorithm

Given $\theta, \delta \in (0,1)$ such that $\frac{\theta^2+\delta^2}{2^{\frac{3}{2}}(1-\theta)} \leq \left(1 - \frac{\delta}{\sqrt{n}}\right)\theta$

$\theta, \delta = 0.4$ satisfy the relationship.

Set $\sigma = \left(1 - \frac{\delta}{\sqrt{n}}\right)\theta$ and choose $\omega^0 = (x^0, u^0, v^0) \in N_2(\theta)$

avg $= \frac{{x^0}^T u^0}{n}$, set eps=tol

while avg>eps **do**

solve 1 with $\omega = \omega^k$ for $\Delta\omega = (\Delta x, \Delta v, \Delta u)$

$\omega^{k+1} = \omega^k + \Delta\omega$

avg $= \frac{{x^{k+1}}^T u^{k+1}}{n}$

end output - optimal $(\bar{x}, \bar{v}, \bar{u})$

Theorem

Given the choices of θ and σ as in the short step algorithm and starting with a $\omega^0 = (x^0, u^0, v^0) \in N_2(\theta)$, we have the following,

1. $\omega^{k+1} = (x^{k+1}, v^{k+1}, u^{k+1}) \in N_2(\theta)$ *for all* k *and*

2. $\mu(\omega^{k+1}) = \frac{x^{k+1\,T} v^{k+1}}{n} = \sigma \mu(\omega^k)$

3. $\mu(\omega^k) = \left(1 - \frac{\delta}{\sqrt{n}}\right)^n \mu(\omega_0)$

- A complete proof of this theorem was studied and has been included in the write up [21].

- Conclusion (1)- states that starting with a point in the neighborhood $N_2(\theta)$ with the choices of θ and σ as stated in the algorithm, the successive iterates stay in the neighborhood $N_2(\theta)$.

- Conclusion (2)- Average drop obtained by taking only one step $(\Delta x, \Delta v, \Delta u)$ from (x^k, v^k, u^k) by solving 1 causes the average to reduce by the same factor.

- But since $(x^{k+1}, v^{k+1}, u^{k+1}) \in N_2(\theta)$, it means conditions 1, 2 and 4 of the perturbed KKT system (pg. 91/125) hold true. Moreover the average has reduced by exactly the same factor σ as is expected of the solution to the the perturbed KKT. **So do we have a solution ?**

- **Does it mean that just one Newton iteration yields solution to the KKT system ?**

Averages Again !!

- The answer is **NO.**
- The reduction in the average going from ω^k to ω^{k+1} is equal to σ but the perturbed complimentarity condition $x_j u_j = \sigma\mu$ is still not satisfied.
- Perturbed complimentarity condition is satisfied when the $x_j^{k+1} u_j^{k+1} = \sigma\mu^k$ for every j.
- The theorem only concludes that the average $\frac{{x^{k+1}}^T u^{k+1}}{n}$ equals $\sigma\mu^k$.
- Individual terms $x_j^{k+1} u_j^{k+1}$ may however differ from $\sigma\mu^k$.

Conclusion (3) yields number of operations

- Denote $\mu(\omega^k)$ as μ^k and $\mu(\omega^0)$ as μ^0.
- Conclusion 3 of the theorem states then $\mu^k = \left(1 - \frac{\delta}{\sqrt{n}}\right)^k \mu^0$ or equivalently $\frac{\mu^k}{\mu^0} = \left(1 - \frac{\delta}{\sqrt{n}}\right)^k$
- Now $\left(1 - \frac{\delta}{\sqrt{n}}\right)^k \leq \epsilon \iff k log\left(1 - \frac{\delta}{\sqrt{n}}\right) \leq log \epsilon$ Using the inequality $log(1+x) \leq x$, $x \geq -1$, we have $-k\frac{\delta}{\sqrt{n}} \leq log\epsilon \Rightarrow klog\left(1 - \frac{\delta}{\sqrt{n}}\right) \leq log\epsilon$. Finally we have $-k\frac{\delta}{\sqrt{n}} \leq log\epsilon \iff k \geq \frac{\sqrt{n}}{\delta}|log\epsilon|$. Thus,

$$\boxed{\frac{\mu^k}{\mu^0} \leq \epsilon \text{ in } \mathcal{O}(\sqrt{n}|log\epsilon|) \text{ Newton's iterations}}$$

Correctness of the method

- Generate instances of linear programming problem in such a way that we have a starting point on the central path.
- The algorithm requires the starting point to be in the $N_2(\theta)$ nbh. A point on the central path is in the nbh trivially.
- Matrix $A \in \mathbb{R}^{m \times n}$, $x^0 \in \mathbb{R}^n$, $x^0 > 0$ and $v^0 \in \mathbb{R}^m$ using the rand command on MATLAB.
- Choose an initial value of $\mu^0 > 0$. We used $\mu^0 = 20$.
- Then $u^0 = \frac{20.}{x^0}$. This ensures $x_j^0 u_j^0 = 20, \forall j$.
- Construct b as $b = Ax^0$ and c as $c = A^T v + u$
- This ensures we have a linear programming instance with starting point in the $N_2(\theta)$ neighborhood.

- The correctness of the algorithm implemented can be seen in the table where we compare the result with GNU-LPK

$\bar{x}$ – Short Step Algorithm	$\bar{x}$ – GNU-LPK
1.3171891711810	1.31719
0.3508849735971	0.35088
0.0000000039733	0.00000
0.0000000445164	0.00000
0.5075037554320	0.50750
0.3898336310555	0.38983
0.0000000119323	0.00000
0.0000000117004	0.00000
0.0000000167616	0.00000
0.8692352323639	0.86924

Table: Optimal Solution for a 5×10 LP obtained by Barrier method vs GNU-LPK

Computational proof of concept

- The table shows the values of $\sigma = 1 - \frac{0.4}{\sqrt{n}}$.

number of variables (n)	σ
10	0.87351
30	0.92697
40	0.93675
50	0.94343
60	0.94836
70	0.95219
80	0.95528

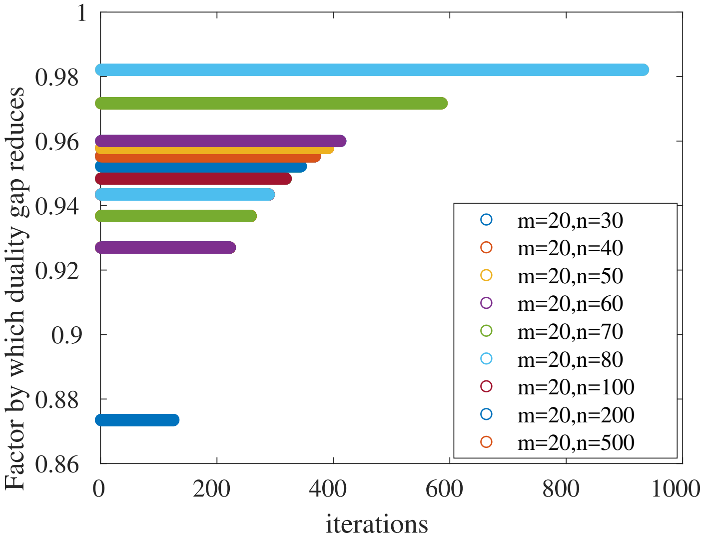
Table: n vs σ

- This is the factor which we would expect our iterates to have their averages $\frac{1}{n} x^{k+1\,T} u^{k+1}$ reduced from the average $\frac{1}{n} x^{k\,T} u^k$. We validate this theoretical assertion now.

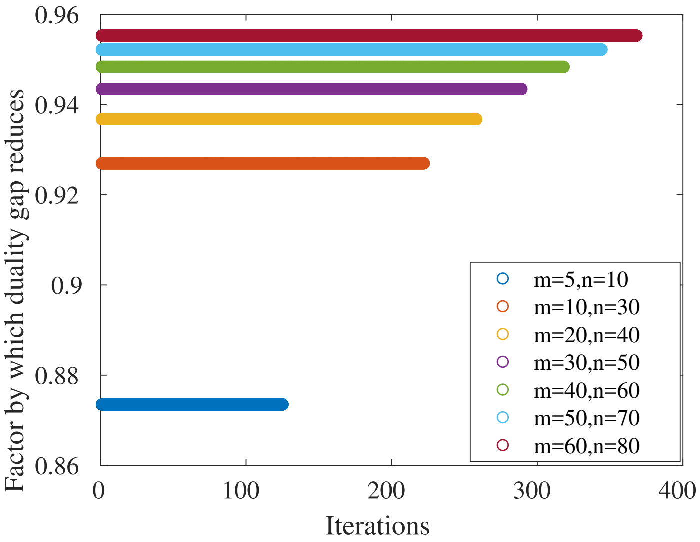
Computational proof of concept

- We solve LP instances for constant value of $m = 20$ and increasing values of $n = 30, \ldots, 500$.
- We then solve instances of LP with increasing values of $m = 5, \ldots, 60$ and increasing values of $n = 10, \ldots, 80$.
- As the value of n increases, the number of iterations go up as seen on the $x-$ axis.
- On the $y-$ axis we see that for given n we have constant value of reduction in the average, equal to the expected reduction as shown in the table earlier.

Computational proof of concept



Computational proof of concept



Computational proof of concept

- We start on the central path, but since we do not completely solve the perturbed system the theoretical assertion is that we do not stay on the central path.

- We have shown that the average drop observed computationally matches the theory.

- **We now show that even though the average drop is exactly as what we would expect had we stayed on the central path, the iterates are such that** $x_j^{k+1} u_j^{k+1} = \sigma \mu^k$ **does not hold true for all** j. Instead it is $\frac{1}{n} x_j^{k+1\,T} u_j^{k+1}$ **which equals** $\sigma \mu^k$.

- And thus the assertion that we do not stay on the central path but stay 'close' to it has been shown in the following example.

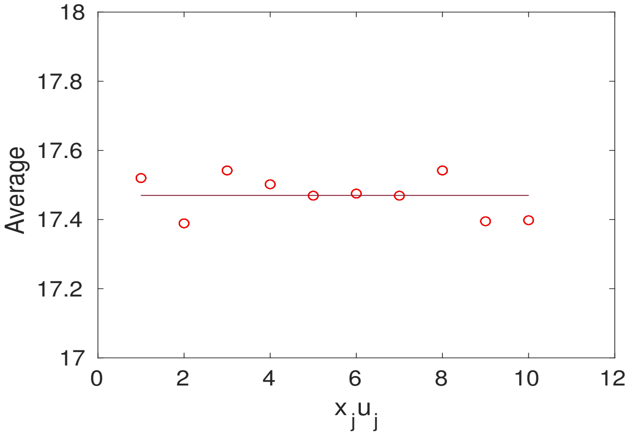
Computational proof of concept

- We consider a randomly generated instance of LP with $m = 5$ and $n = 10$ such that the starting point is on the central path. Start with an average $\mu^0 = 20$ and consider one iteration.

$\mu(\omega^0) = 20$	$x_j^0 u_j^0$	$x_j^1 u_j^1$	$\mu(\omega^1)$	σ
20	20	17.520	17.470	0.87351
	20	17.389	17.470	
	20	17.542	17.470	
	20	17.502	17.470	
	20	17.469	17.470	
	20	17.475	17.470	
	20	17.469	17.470	
	20	17.542	17.470	
	20	17.395	17.470	
	20	17.398	17.470	

Table: Iterates not on the central path

Computational proof of concept



Simplex Method vs Interior Point Methods - The difference in approach

- **KKT Conditions-**

$$\boxed{\begin{array}{c} A^T v + u = c \\ Ax = b \\ x_j u_j = 0, \; j = 1, \ldots, n \\ (x, u) \geq 0 \end{array}}$$

- Simplex algorithm (G. Dantzig, 1948) is one of the most popular algorithms in the history of computational mathematics across various disciplines.
- Interior point methods maintain the KKT conditions 1, 2 and 4 and try to satisfy the condition 3 in the limit.
- Simplex method maintains conditions 1, 2 and 3 throughout the process while trying to find the optimal vertex in the correct orthant.

Constrained optimization problem on smooth manifold

- A constrained optimization problem on a smooth manifold is given as

$$\begin{aligned} &\text{minimize} && f(p) \\ &\text{subject to} && g_j(p) \leq 0, \;\; j = 1, 2, ..., l \\ & && h_i(p) = 0, \;\; i = 1, 2, ..., m \\ & && p \in \mathcal{M} \end{aligned} \qquad (\mathcal{P}_M)$$

where $\mathcal{M}$ is a smooth manifold and the functions f, g and h are assumed to be in C^1.

Translation from R^n to a Smooth Manifold $\mathcal{M}$

- The discussion in subsequent slides is based on Bergmann and Herzog (2019).
- Tangent Space $R^n \Longrightarrow$ Tangent Space $T_{\mathcal{M}}(p)$
- Tangent Cone $T_C(\bar{x}) \Longrightarrow$ Tangent Cone $T_{\mathcal{M}}(\Omega, p)$
- Cone of linearized feasible directions $F_0(\bar{x}) \Longrightarrow$ Linearizing Cone $T_{\mathcal{M}}^{\text{lin}}(\Omega, p)$
- Cotangent/Dual Space R_n(row vector) $\Longrightarrow$ Dual Space $T_{\mathcal{M}}^*(p)$
- Derivative $\nabla f \in R_n \Longrightarrow$ Differential $(df)(p) \in T_{\mathcal{M}}^*(p)$

- **Tangent vector to the constraint set Ω.** $[\dot{\gamma}(0)] \in T_{\mathcal{M}}(p)$ such that for all C^1 functions f defined near p,

$$[\dot{\gamma}(0)](f) = \lim_{k \to \infty} \frac{f(p_k) - f(p)}{t_k}, \ (p_k) \subseteq \Omega, \ t_k \to 0$$

- **Active constraints**

$$\mathcal{A}(p) = \{j : g_j(p) = 0\}.$$

Denote indices corresponding to equality constraints by $\mathcal{E}$.

- **Bouligand Tangent Cone.**

$$T_{\mathcal{M}}(\Omega, p) = \left\{[\dot\gamma(0)] \in T_{\mathcal{M}}(p) : [\dot\gamma(0)] \text{ tangent vector to } \Omega \text{ at } p\right\}$$

- **Cone of linearized feasible directions.**

$$T^{\text{lin}}_{\mathcal{M}}(\Omega, p) = \left\{[\dot\gamma(0)] \in T_{\mathcal{M}}(p) : [\dot\gamma(0)](g^j) \leq 0 \; [\dot\gamma(0)](h^i)(p) = 0\right\}$$

$\forall j \in \mathcal{A}(p)$ and $i \in \mathcal{E}$

- **Abadie Constraint Qualification.**

$$T^{\text{lin}}_{\mathcal{M}}(\Omega, p) = T_{\mathcal{M}}(\Omega, p)$$

- The original paper uses the Guinard constraint qualification to derive the KKT conditions. The proof was modified to use ACQ instead of GCQ.
- A detailed derivation has been included in the write up [22].

Necessary Condition for Optimality

Theorem

First order necessary optimality condition. Suppose $p^* \in \Omega$ is a local minimizer of $\mathcal{P}_1$. Then we have

$$[\dot{\gamma}(0)](f) \geq 0$$

for all tangent vectors $[\dot{\gamma}(0)] \in T_{\mathcal{M}}(\Omega, p)$

Proof.

Suppose $[\dot{\gamma}(0)] \in T_{\mathcal{M}}(\Omega, p^*)$ and (p_k, t_k) is the associated tangential sequence. Then since p^* is a local minimum, we have for sufficiently large $k \in \mathbb{N}$

$$\frac{f(p_k) - f(p^*)}{t_k} \geq 0$$

which gives $[\dot{\gamma}(0)](f) \geq 0$. □

Farkas' Lemma

Lemma

For a finite dimensional vector space V*, let* $A \in \mathcal{L}(V, R^q)$ *and* $b \in V^*$*. Also denote adjoint of* A *as* $A^* \in \mathcal{L}(V, \mathbb{R}^*)$ *Then the following are equivalent,*

1. *The system* $A^* y = b$ *has a solution* $y \in \mathbb{R}_{||}$*, which satisfies* $y \geq 0$

2. *For any* $d \in V$*,* $Ad \geq 0$ *implies* $bd \geq 0$*.*

The KKT Conditions

Theorem

Karush-Kuhn-Tucker Necessary conditions. For the problem, let p be a local optimal solution. Assume the Abadie's constraint qualification hold true. Then there exist vectors $\mu \in \mathbb{R}^l$ and $\lambda \in \mathbb{R}^m$

$$g_j(p) \leq 0 \text{ and } h_i(p) = 0 \ \forall i, j \ \textit{(Feasibility)}$$

$$(df)(p) + \sum_{j=1}^{l} \mu_j (dg_j)(p) + \sum_{i=1}^{m} \lambda_i (dh_i)(p) = 0 \ \textit{(Stationarity)}$$

$$\mu_j g_j(p) = 0, \ j = 1, \dots, l \ , \mu \geq 0 \ \textit{(Complimentarity)}$$

- **Linear Independence Constraint Qualification (LICQ)**

$$\boxed{\begin{array}{l}\text{The set } \{(dh^i)(p)\} \cup \{(dg^j)(p)\} i \in \mathcal{E} \text{ and } j \in \mathcal{A}(p) \text{ is a} \\ \text{linearly independent set in the cotangent space } T^*_{\mathcal{M}}(\Omega, p)\end{array}}$$

- **Mangasarian Fromovitz constraint qualifications (MFCQ)**

$$\boxed{\begin{array}{l}\text{The set } \{(dh^i)(p)\} \text{ is linearly independent and there} \\ \text{exists a } [\dot\gamma(0)] \in T_M(p) \text{ such that} (dg^j)(p)[\dot\gamma(0)] \leq 0 \\ \text{for all } j \in \mathcal{A}(p) \text{ and } (dh^i)(p)[\dot\gamma(0)] = 0 \text{ for all } i \in \mathcal{E}\end{array}}$$

- **Guinard's constraint qualification (GCQ)**

$$\boxed{T^{\text{lin}}_{\mathcal{M}}(\Omega, p)^\circ = T_{\mathcal{M}}(\Omega, p)^\circ}$$

- And finally the chain of implications also holds true. If LICQ holds at $p \in \mathcal{M}$, then

$$\boxed{\text{LICQ} \Rightarrow \text{MFCQ} \Rightarrow \text{ACQ} \Rightarrow \text{GCQ}}$$

Back to Research Objectives

- The following is a summary of the current status of constrained optimization theory on manifolds.

Constrained Opt (Euclidean)	Constrained Opt (Riemannian)
KKT Theory	Bergmann and Herzog (2019)
Projected Gradient Method	Bergman and Herzog (2019)
Penalty Methods	Boumal and Liu (2019)
Augmented Lagrangian Method	Boumal and Liu (2019)
Barrier / Interior Point Methods	??
Duality Theory	??
Convex Analysis (Fenchel Conjugate)	Bergmann and Herzog (2019)

Back to Research Objectives

- Barrier and interior point methods exist for general non-linear programming problems as well (although, not studied during the course of the candidacy), their generalization to optimization on manifolds remains to be seen.
- However, barrier and interior point methods have been most prominently used for solving convex optimization problems. Primal Dual interior point methods were developed for linear programming problems.
- Convexity however is more challenging on general smooth manifolds as compared with $\mathbb{R}^n$. Several notions of convexity have been proposed and one of them is retraction convexity.
- We are about embark on a comparative study of retraction vs geodesic convexity.

References

1. Mokhtar S. Bazaraa, Hanif D. Sherali, C.M.Shetty, *Nonlinear Programming: Theory and Algorithms 3rd Edition*. John Wiley and Sons, Inc, 2006.

2. Jorge Noecedal, Stephen J. Wright. *Numerical Optimization, second edition*. Springer.

3. Rodrigo Eustaquio, Elizabeth Karas and Ademir Alves Ribeiro
`Constraint qualifications for nonlinear programming`

4. F. J. Gould and Jon W. Tolle
`A necessary and sufficient qualification for constrained optimization`, SIAM Journal on Applied Mathematics, 20:164-172, 1971

References

5. H. KUHN AND A. TUCKER, Nonlinear programming,Proc. Second Berkeley Symposium on Mathematical Statistics and Probability, J. Neyman, ed., University of California Press,Berkeley, 1951, pp. 481-492.

6. Yoram Louzoun.: Optimization course Class 7. https://u.cs.biu.ac.il/ louzouy/courses/opt.html

7. Stephen J. Wright Primal Dual Interior Point Methods SIAM

8. Stephen Boyd and Lieven Vanderberghe Convex Optimization Cambridge University Press

9. Anders Forsgren, Philip E. Gill, Margaret H. Wright Interior point methods for nonlinear optimization,SIAM Review Vol 44, No 4, pp 525-597.

References

10. Ryan Tibshirani
http://www.stat.cmu.edu/ ryantibs/convexopt/ lectures/barr-method.pdf.

11. Ryan Tibshirani
https://www.youtube.com/watch?v=s8-O28F-YE4.

12. Ryan Tibshirani
http://www.stat.cmu.edu/ ryantibs/convexopt/ lectures/primal-dual.pdf

13. Ryan Tibshirani
https://www.youtube.com/watch?v=UfPvb$_K78ik$

14. Nicholas Boumal.
An Introduction to Optimization on Smooth Manifolds.

References

15. Ronny Bergmann and Roland Herzog.
`Intrinsic Formulation of KKT Conditions and Constraint Qualifications On Smooth Manifolds.`

16. Changshuo Liu and Nicholas Boumal.
`Simple Algorithms for Optimization on Riemannian Manifolds with Constraints`

17. Ronny Bergmann and Roland Herzog.
`Fenchel Duality Theory and a Primal Dual Algorithm on Riemannian Manifold.`

18. Nicolas Andreasson, Anton Evgrafov and Michael Patriksson
`An Introduction to Continuous Optimization: Foundations and Fundamental Algorithms.`

- 19. Classical Theory of Constrained Optimization
https://github.com/TejasNatuOpt/Candidacy-Material.
- 20. Convex Analysis and Optimization
https://github.com/TejasNatuOpt/Candidacy-Material
- 21. Penalty and Interior Point Methods
https://github.com/TejasNatuOpt/Candidacy-Material.
- 22. Constrained Optimization on Smooth Manifolds
https://github.com/TejasNatuOpt/Candidacy-Material.
- 23. Shirish Shevade, NPTEL Indian Institute of Science, Bangalore
https://www.youtube.com/watch?v=iePvYbZJm8E.

Questions