

# Numerical Optimization

## Lecture Notes #28 Constrained Optimization

Peter Blomgren,  
([blomgren.peter@gmail.com](mailto:blomgren.peter@gmail.com))

Department of Mathematics and Statistics  
Dynamical Systems Group  
Computational Sciences Research Center  
San Diego State University  
San Diego, CA 92182-7720  
<http://terminus.sdsu.edu/>

Fall 2018



# Outline

- 1 Constrained Optimization
  - KKT First Order Necessary Conditions
  - Second Order Conditions
- 2 Some Approaches
  - Problems and Algorithms

## Introduction

We have spent a lots of effort on the Unconstrained Optimization problem, we now take a very quick look at the fundamentals of Constrained Optimization — we will quickly realize that things get quite “interesting!”

Problem 0: Constrained Optimization

$$\min_{\vec{x} \in \mathbb{R}^n} f(\vec{x}) \quad \text{subject to} \quad \begin{cases} c_i(\vec{x}) = 0, & i \in \mathcal{E} \\ c_i(\vec{x}) \geq 0, & i \in \mathcal{I} \end{cases}$$

where  $i \in \mathcal{E}$  are the equality constraints, and  $i \in \mathcal{I}$  the inequality constraints.

The smoothness (or lack thereof) for the objective  $f(\vec{x})$  and the constraint functions  $c_i(\vec{x})$  will impact the difficulty of solving the problem.

## The Feasible Set

With the following definition of all allowable points:

Definition (The Feasible Set)

Let

$$\Omega = \{\vec{x} \in \mathbb{R}^n : c_i(\vec{x}) = 0 \forall i \in \mathcal{E}, \text{ and } c_i(\vec{x}) \geq 0 \forall i \in \mathcal{I}\}$$

We can rewrite the problem more compactly as

Problem 1: Constrained Optimization

$$\min_{\vec{x} \in \Omega} f(\vec{x}).$$

Our goal is to state necessary and sufficient conditions for optimality.

## Local Solution

### Definition (Local Solution)

A point  $\vec{x}^*$  is a local solution of Problem 1 if  $\vec{x}^* \in \Omega$  and there is a neighborhood  $\mathcal{N}$  of  $\vec{x}^*$  such that  $f(\vec{x}) \geq f(\vec{x}^*) \forall \vec{x} \in \mathcal{N} \cap \Omega$ .

### Definition (Strict Local Solution)

A point  $\vec{x}^*$  is a strict local solution of Problem 1 if  $\vec{x}^* \in \Omega$  and there is a neighborhood  $\mathcal{N}$  of  $\vec{x}^*$  such that  $f(\vec{x}) > f(\vec{x}^*) \forall \vec{x} \in \mathcal{N} \cap \Omega$ .

### Definition (Isolated Local Solution)

A point  $\vec{x}^*$  is an isolated local solution of Problem 1 if  $\vec{x}^* \in \Omega$  and there is a neighborhood  $\mathcal{N}$  of  $\vec{x}^*$  such that  $\vec{x}^*$  is the only local solution in  $\mathcal{N} \cap \Omega$ .

# Smoothness

It is usually (always?) advantageous to express constraints and objectives in as smooth a way as possible; e.g. we can replace single non-smooth conditions, like

$$\text{ns } \|\vec{x}\|_1 = |x_1| + |x_2| \leq 1$$

with several smooth constraints

$$\text{s\#1 } x_1 + x_2 \leq 1$$

$$\text{s\#2 } x_1 - x_2 \leq 1$$

$$\text{s\#3 } -x_1 + x_2 \leq 1$$

$$\text{s\#4 } -x_1 - x_2 \leq 1$$

## The Active Set

### Definition (Active Set)

The active set  $\mathcal{A}(\vec{x})$  at any feasible  $\vec{x}$  consists of the equality constraint indices from  $\mathcal{E}$  and the indices of the inequality constraint indices from  $\mathcal{I}$  for which  $c_i(\vec{x}) = 0$ , *i.e.*

$$\mathcal{A}(\vec{x}) = \mathcal{E} \cup \{i \in \mathcal{I} : c_i(\vec{x}) = 0\}.$$

At a feasible point  $\vec{x}$ , the inequality constraint  $i \in \mathcal{I}$  is said to be active if  $c_i(\vec{x}) = 0$  and inactive if  $c_i(\vec{x}) > 0$ .

## Linear Independence Constraint Qualification

Definition (LICQ: Linear Independence Constraint Qualification)

Given a point  $\vec{x}$  and the active set  $\mathcal{A}(\vec{x})$ , we say that the linear independence constraint qualification (LICQ) holds if the set of active constraint gradients

$$\{\nabla c_i(\vec{x}), i \in \mathcal{A}(\vec{x})\}$$

is linearly independent.



## The Lagrangian Function

Our final building block before stating the first order conditions for optimality is the:

Definition (The Lagrangian Function,  $\mathcal{L}(\vec{x}, \vec{\lambda})$ )

$$\mathcal{L}(\vec{x}, \vec{\lambda}) = f(\vec{x}) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(\vec{x})$$

The Lagrange multipliers,  $\lambda_i$ , are used to “pull” the solution back to the feasible set.

## KKT First Order Necessary Conditions

Theorem (KKT:FONC — First Order Necessary Conditions)

*Suppose that  $\vec{x}^*$  is a local solution to Problem 1, that the functions  $f$  and  $c_i$  are continuously differentiable, and that the LICQ holds at  $\vec{x}^*$ . Then there is a Lagrange multiplier vector  $\vec{\lambda}^*$ , with components  $\lambda_i(\vec{x}^*)$ ,  $i \in \mathcal{E} \cup \mathcal{I}$ , such that the following conditions are satisfied at  $(\vec{x}^*, \vec{\lambda}^*)$ :*

$$\begin{aligned}\nabla_{\vec{x}} \mathcal{L}(\vec{x}, \vec{\lambda}) &= 0, \\ c_i(\vec{x}^*) &= 0, \quad \forall i \in \mathcal{E} \\ c_i(\vec{x}^*) &\geq 0, \quad \forall i \in \mathcal{I} \\ \lambda_i^* &\geq 0, \quad \forall i \in \mathcal{I} \\ \lambda_i^* c_i(\vec{x}^*) &= 0, \quad i \in \mathcal{I} \cup \mathcal{E}.\end{aligned}$$

The Karush–Kuhn–Tucker conditions.

## KKT First Order Necessary Conditions (compact form)

Theorem (KKT:FONC — Compact Form)

*Suppose that  $\vec{x}^*$  is a local solution to Problem 1, that the functions  $f$  and  $c_i$  are continuously differentiable, and that the LICQ holds at  $\vec{x}^*$ . Then there is a Lagrange multiplier vector  $\vec{\lambda}^*$ , with components  $\lambda_i(\vec{x}^*)$ ,  $i \in \mathcal{E} \cup \mathcal{I}$ , such that the following conditions are satisfied at  $(\vec{x}^*, \vec{\lambda}^*)$ :*

$$0 = \nabla_{\vec{x}} \mathcal{L}(\vec{x}, \vec{\lambda}) = \nabla f(\vec{x}^*) - \sum_{i \in \mathcal{A}(\vec{x}^*)} \lambda_i^* \nabla c_i(\vec{x}^*).$$

## Strict Complementarity

### Definition (Strict Complementarity)

Given a local solution  $\vec{x}^*$  of Problem 1, and a vector  $\vec{\lambda}^*$  satisfying the KKT:FONC, we say that the strict complementarity condition holds if exactly one of  $\lambda_i^*$  or  $c_i(\vec{x}^*)$  is zero for each index  $i \in \mathcal{I}$ . In other words, we have  $\lambda_i^* > 0 \forall i \in \mathcal{I} \cap \mathcal{A}(\vec{x}^*)$ .

We sweep the proof of KKT:FONC under our infinitely stretchable rug. Not because it is not important (it is!), but we are somewhat short on time.

## Linearized Feasible Directions

Definition (Set of Linearized Feasible Directions)

Given a feasible point  $\vec{x}$  and the active constrain set  $\mathcal{A}(\vec{x})$ , the set of linearized feasible directions is

$$\mathcal{F}(\vec{x}) = \left\{ \vec{d} \text{ such that } \begin{array}{l} \vec{d}^T \nabla c_i(\vec{x}) = 0, \quad \forall i \in \mathcal{E} \\ \vec{d}^T \nabla c_i(\vec{x}) \geq 0, \quad \forall i \in \mathcal{A}(\vec{x}) \cap \mathcal{I} \end{array} \right\}$$

## Second Order Conditions

Second order conditions will help determine the impact of directions  $\vec{w} \in \mathcal{F}(\vec{x}^*)$  for which  $\vec{w}^T \nabla f(\vec{x}^*) = 0$ , *i.e.* directions which are “locally flat.”

From this point on we need the functions  $f$  and  $c_i$  to be twice continuously differentiable.

## Critical Cone

Given  $\mathcal{F}(\vec{x}^*)$ , and some Lagrange multiplier vector  $\lambda^*$  satisfying KKT:FONC, we define:

Definition (Critical Cone)

$$C(\vec{x}^*, \vec{\lambda}^*) = \{ \vec{w} \in \mathcal{F}(\vec{x}^*) : \nabla c_i(\vec{x}^*)^T \vec{w} = 0, \forall i \in \mathcal{A}(\vec{x}^*) \cap \mathcal{I} \text{ with } \lambda_i^* = 0 \}.$$

Or equivalently

$$\vec{w} \in C(\vec{x}^*, \vec{\lambda}^*) \Leftrightarrow \begin{cases} \nabla c_i(\vec{x}^*)^T \vec{w} = 0, & \forall i \in \mathcal{E} \\ \nabla c_i(\vec{x}^*)^T \vec{w} = 0, & \forall i \in \mathcal{A}(\vec{x}^*) \cap \mathcal{I} \text{ with } \lambda_i^* > 0 \\ \nabla c_i(\vec{x}^*)^T \vec{w} \geq 0, & \forall i \in \mathcal{A}(\vec{x}^*) \cap \mathcal{I} \text{ with } \lambda_i^* = 0 \end{cases}$$

The critical cone  $C(\vec{x}^*, \vec{\lambda}^*)$  contains the directions from  $\mathcal{F}(\vec{x}^*)$  for which it is not clear from first derivative information whether  $f$  will increase or decrease.

## Second Order Necessary Conditions

## Theorem (Second Order Necessary Conditions)

*Suppose that  $\vec{x}^*$  is a local solution of Problem 1, and that the LICQ condition is satisfied. Let  $\vec{\lambda}^*$  be the Lagrange multiplier vector for which the KKT:FONC are satisfied. Then*

$$\vec{w}^T \nabla_{\vec{x}\vec{x}}^2 \mathcal{L}(\vec{x}, \vec{\lambda}) \vec{w} \geq 0, \quad \forall \vec{w} \in C(\vec{x}^*, \vec{\lambda}^*).$$

Interpretation: The Hessian of the Lagrangian has non-negative curvature along critical directions.



## Second Order Sufficient Conditions

Theorem (Second Order Sufficient Conditions)

*Suppose that for some feasible point  $\vec{x}^* \in \mathbb{R}^n$  there is a Lagrange multiplier vector  $\vec{\lambda}^*$  such that KKT:FONC are satisfied. Suppose also that*

$$\vec{w}^T \nabla_{\vec{x}\vec{x}}^2 \mathcal{L}(\vec{x}, \vec{\lambda}) \vec{w} > 0, \quad \forall \vec{w} \in C(\vec{x}^*, \vec{\lambda}^*), \quad \vec{w} \neq \vec{0}.$$

*Then  $\vec{x}^*$  is a strict local solution for Problem 1.*

Much remains to be said; however, everything grows out of these fundamental definitions and theorems; leveraging special cases, weakening and strengthening conditions, and looking for alternatives.

# Some Approaches

- **Linear Programming — The Simplex Method**
  - $f$  and  $c_i$  linear functions
  - Leonid Kantorovich, 1939 — Linear Programming.
  - George Datzig, 1947 — The Simplex Method.
  - John von Neumann, 1947 — Theory of Duality.
  - The *worst case complexity* for The Simplex Method is exponential, but it is remarkably efficient in practice.

# Some Approaches

## ● Interior Point Methods, Primal-Dual Methods

- $c_i$  are *strict* inequalities.
- Better theoretical behavior than The Simplex Method.
- Leonid Khachiyan, 1979 — The Ellipsoid Method (polynomial runtime,  $\mathcal{O}(n^6L)$ )
- Narendra Karmarkar, 1984 — Projective Algorithm,  $\mathcal{O}(n^{3.5}L^2 \cdot \log L \cdot \log \log L)$ , where  $n$  is the number of variables and  $L$  is the number of bits of input to the algorithm.

# Some Approaches

- **Quadratic Programming**

- Special case: Exploit the structure of the problem:
  - Quadratic Objective, Linear Constraints
  - Active Set methods
  - Interior Point methods
  - Gradient Projection methods
- Appears as sub-problems for: Sequential Quadratic Programming, Penalty/Augmented Lagrangian Methods, and Interior Point Methods.

# Some Approaches

- **Penalty / Augmented Lagrangian Methods**

- Constraints are represented by additions to the objective
- *Quadratic Penalty Terms* — add the square of the constraint discrepancies: intuitive, fairly simple to implement
- *Non-smooth Penalty Terms* —  $\ell_1$  and  $\ell_0$  penalty functions
- *Method of Multipliers* — estimated for the Lagrange multipliers are used.

# Index