Multiple Constraints and Non-regular Solution in DDN

Suikei Wang
u6624985

11 August 2020

Table of Contents

Table of Contents

Optimality Conditions for Unconstrained Problems

- We consider the unconstrained minimization problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

where f is a given function on $\mathbb{R}^n$

- In order to determine the minimizer, it is important to understand what can happen at a minimizer, and at what condition a point must be a minimizer.

Theorem 1

Let $f : \Omega \to \mathbb{R}$ be a function defined on a set $\Omega \subset \mathbb{R}^n$
(i) If f is differentiable at x^*, then x^* is a critical point of f, i.e. $\nabla f(x^*) = 0$
(ii) If f is twice continuous differentiable on Ω, then the Hessian (x^*) is positive semidefinite.

Optimality Conditions for Unconstrained Problems

Definition 1

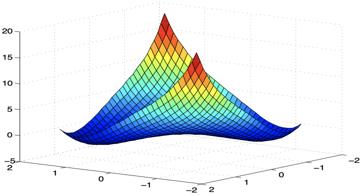
Any point x^* satisfying $\nabla f(x^*) = 0$ is called a critical point of f.

Theorem 1 says that any minimizer of f is a critical point of f .

Examples

Determine the minimizers of the function:

$$f(x, y) = x^4 - 4xy + y^4$$



Theorem 2

Let $f : \Omega \to \mathbb{R}$ be a twice continuous differentiable function on an open set $\Omega \subset \mathbb{R}^n$ and let x^* be an interior point of Ω. If

$$\nabla f(x^*) = 0 \quad \text{and} \quad \nabla^2 f(x^*) > 0,$$

then x^* is a strict local minimizer of f on Ω.

The result in Theorem 2 is best possible for general functions. More can be said if we consider convex functions.

Theorem 3

Let f be a function defined on a convex set $\Omega \subset \mathbb{R}^n$.
(a) If f is convex, any local minimizer of f is also a global minimizer.
(b) If f is strictly convex, it has at most one global minimizer.

Theorem 4

Let f be a convex function with continuous first partial derivatives. Then any critical point of f is a global minimizer.

Proof. Let x^* be a critical point of f , i.e. $\nabla^2 f\left(x^*\right) = 0$. Since f is convex, for any x we have:

$$f(x) \geq f\left(x^*\right) + \left(x - x^*\right)^T \nabla f\left(x^*\right) = f\left(x^*\right)$$

Therefore x^* is a global minimizer of f.

Unconstrained Nodes in DDN

Gradient of unconstrained cases

Consider a function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$. Let

$$y(x) \in \underset{u \in \mathbb{R}^m}{\arg\min} f(x, u)$$

Assume $y(x)$ exists and that f is second-order differentiable in the neighborhood of the point $(x, y(x))$. Set $H = D^2_{YY} f(x, y(x)) \in \mathbb{R}^{m \times m}$ and $B = D^2_{XY} f(x, y(x)) \in \mathbb{R}^{m \times n}$. Then for H non-singular the derivative of y with respect to x is

$$Dy(x) = -H^{-1} B$$

Proof. First-order optimality condition: $D_Y f(x, y(x)) = 0_{1 \times m}$, where y is the optimal point. Transposing and differentiating both sides:

$$0_{m \times n} = D(D_Y f(x, y))^T = D^2_{XY} f(x, y) + D^2_{YY} f(x, y) Dy(x)$$

Optimality Conditions for Constrained Problems

A general formulation for constrained optimization problems is as follows:

$$
\begin{array}{ll}
\text{minimize } f(x) \\
\text{subject to } \begin{cases} c_i(x) = 0 & \text{for } i = 1, \cdots, m_e, \\ c_i(x) \leq 0 & \text{for } i = m_e + 1, \cdots, m \end{cases}
\end{array}
$$

where f and c_i are smooth real-valued functions on $\mathbb{R}^n$, and m_e and m are nonnegative integers with $m_e \leq m$. We set

$$E := \{1, \cdots, m_e\} \quad \text{and} \quad I := \{m_e + 1, \cdots, m\}$$

as index sets of equality constraints and inequality constraints, respectively. We call

- f the objective function;
- $c_i, i \in E$ the equality constraints; and
- $c_i, i \in I$ the inequality constraints

Definition 2

The feasible set of the constrained optimization problem is defined to be

$$F := \{x \in \mathbb{R}^n : c_i(x) = 0 \text{ for } i \in E \text{ and } c_i(x) \leq 0 \text{ for } i \in I\}$$

Any point $x \in F$ is called a feasible point of this constrained optimization problem. We call this problem infeasible if $F = 0$

Definition 3

- A feasible point $x^* \in F$ is called a local minimizer if there is a neighbourhood $B(x^*, \delta) := \{x \in \mathbb{R}^n : \|x - x^*\| < \delta\}$ of x^* with $\delta > 0$ such that

$$f(x) \geq f(x^*) \qquad \text{for all } x \in F \cap B(x^*, \delta)$$

- x^* is called a strict local minimizer if there is a neighbourhood $B(x^*, \delta) := \{x \in \mathbb{R}^n : \|x - x^*\| < \delta\}$ of x^* with $\delta > 0$ such that

$$f(x) > f(x^*) \qquad \text{for all } x \in F \cap B(x^*, \delta)$$

Definition 4

- A feasible point $x^* \in F$ is called a global minimizer if

$$f(x) \geq f(x^*) \quad \text{for all } x \in F$$

- x^* is called a strict global minimizer if

$$f(x) > f(x^*) \quad \text{for all } x \in F$$

Let x^* be a local minimizer. If there is an index $i \in I$ such that $c_i(x^*) < 0$, then, x^* is still the local minimizer of the problem obtained by deleting i-th constraint. In this situation, we say that the i-th constraint is inactive at x^*

Definition 5

At a feasible point $x \in F$, the index $i \in I$ is said to be active if $c_i(x^*) = 0$ and inactive if $c_i(x^*) < 0$.

Definition 6

For $x \in F$, its active set $A(x)$ is defined as $A(x) = E \cup \{i \in I : c_i(x) = 0\}$.

- We will give necessary and sufficient conditions for a feasible point x to be a local minimizer.
- These conditions will be derived by considering the change of f on the feasible set along certain directions. This leads to introduce the concept of tangent cones.

Definition 7

Given $x \in F$, a vector $d \in \mathbb{R}^n$ is said to be tangent to F at x if there is a sequence $\{z_k\} \subset F$ with $z_k \to x$ and a sequence of positive numbers $\{t_k\}$ with $t_k \to 0$ such that

$$d = \lim_{k \to \infty} \frac{z_k - x}{t_k}$$

We use $T_X F$ to denote the set of all vectors tangent to F at x and call it the tangent cone of F at x.

Theorem 5

Given $x \in F$, we define

$$LFD(x) :=$$

$$\left\{d \in \mathbb{R}^n : d^T \nabla c_i(x) = 0 \text{ for } i \in E; d^T \nabla c_i(x) \leq 0 \text{ for } i \in I \cap A(x)\right\}$$

and call it the set of linearized feasible directions of F at x.

Theorem 6 (Karush-Kuhn-Tucker Theorem)

Let $x^* \in F$ be a local minimizer. If

$$T_{x^*} F = LFD\left(x^*\right),$$

then there exists $\lambda^* = (\lambda_1^*, \cdots, \lambda_m^*)^T \in \mathbb{R}^m$ such that

$$\nabla f\left(x^*\right) + \sum_{i \in E \cup I} \lambda_i^* \nabla c_i\left(x^*\right) = 0, \qquad \text{(Lagrangian stationary)}$$

Cont'd

$$\left.\begin{array}{ll} c_i\left(x^*\right) = 0 & \text{for all } i \in E, \\ c_i\left(x^*\right) \leq 0 & \text{for all } i \in I, \end{array}\right\} \quad \text{(primal feasibility)}$$

The equations above are called the Karush-Kuhn-Tucker (KKT) conditions, a point x is called a KKT point if there exists λ such that (x^*, λ^*) satisfies the KKT conditions.

Constrained Nodes in DDN

Gradient of equality constrained cases

Consider a function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $h : \mathbb{R}^m \to \mathbb{R}^p = [h_1, \ldots, h_p]^T$. Let

$$y(x) \in \underset{u \in \mathbb{R}^m}{\arg\min} f(x, u)$$

$$\text{subject to } h_i(u) = 0, i = 1, \ldots, p$$

Assume that $y(x)$ exists, that f and $h = [h_1, \ldots, h_p]^T$ are second-order differentiable in the neighborhood of $x(, y(x))$, and that $\text{rank}(D_Y h(x, y)) = p$. Then for H non-singular:

$$Dy(x) = H^{-1}A^T \left(AH^{-1}A^T\right)^{-1} \left(AH^{-1}B - C\right) - H^{-1}B$$

Cont'd

where

$$A = D_Y h(x,y) \in \mathbb{R}^{p\times m}, B = D^2_{XY} f(x,y) - \sum_{i=1}^{p} \lambda_i D^2_{XY} h_i(x,y) \in \mathbb{R}^{m\times n}$$

$$C = D_X h(x,y) \in \mathbb{R}^{p\times n}, H = D^2_{YY} f(x,y) - \sum_{i=1}^{p} \lambda_i D^2_{YY} h_i(x,y) \in \mathbb{R}^{m\times m}$$

$$\lambda^T A = D_Y f(x,y) \to A^T \lambda = D_Y f(x,y)^T$$

Proof. According to the method of Lagrange multipliers, we get the Lagrange function:

$$\mathcal{L}(x,y,\lambda) = f(x,y) - \sum_{i=1}^{p} \lambda_i h_i(x,y)$$

Cont'd

Since we assume y is the optimal point for the fixed input x, and $D_Y h(x, y)$ is full rank, y is a regular point. $(D_Y f(x, y) = 0_{1\times m})$ Therefore, Lagrange multiplier λ is existed at the point (y, λ):

$$\begin{bmatrix} \left(\mathrm{D}_Y f(x, y) - \sum_{i=1}^{p} \lambda_i \mathrm{D}_Y h_i(x, y)\right)^T \\ h(x, y) \end{bmatrix} = 0_{m+p}$$

where first m rows are from differentiating $\mathcal{L}$ w.r.t y. Last p rows are from differentiating $\mathcal{L}$ w.r.t λ.

- The optimal point of the unconstrained problem automatically satisfies the constraints $(D_Y f(x, y) = 0_{1\times m})$
- $D_Y f(x, y)$ is non-zero and orthogonal to the constraint surface defined by $h(x, y) = 0$

Cont'd

From the second case we have:

$$D_Y f(x,y) = \sum_{i=1}^{p} \lambda_i D_Y h_i(x,y) = \lambda^T A$$

Then differentiating the gradient of the Lagrangian with x:

$$\begin{bmatrix} \mathrm{D}^2_{YY} f - \sum_{i=1}^{p} \lambda_i \mathrm{D}^2_{YY} h_i & -\mathrm{D}_Y h^T \\ \mathrm{D}_Y h & 0_{p\times p} \end{bmatrix} \begin{bmatrix} \mathrm{D}y \\ \mathrm{D}\lambda \end{bmatrix}$$

$$+ \begin{bmatrix} \mathrm{D}^2_{XY} f - \sum_{i=1}^{p} \lambda_i \mathrm{D}^2_{XY} h_i \\ \mathrm{D}_X h \end{bmatrix} = 0_{(m+p)\times n}$$

Gradient of inequality constrained cases

Consider a function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ and $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^q$. Let:

$$\begin{array}{ll} y \in \arg\min_{u \in \mathbb{R}^m} & f(x, u) \\ \text{subject to} & h(x, u) = 0, i = 1, \dots p \\ \text{subject to} & g(x, u) \leq 0, i = 1, \dots, q \end{array}$$

Assume that $y(x)$ exists, that f, h and g are second-order differentiable in the neighborhood of $(x, y(x))$, and that all inequality constraints are active at $y(x)$. Let $\tilde{h} = [h_1, \dots, h_p, g_1, \dots, g_q]$ and assume $\text{rank}(D_Y \tilde{h}(x, y) = p + q)$. Then for H non-singular:

$$Dy(x) = H^{-1}A^T(AH^{-1}A^T)^{-1}(AH^{-1}B - C) - H^{-1}B$$

where

$$A = D_Y \tilde{h}(x, y) \in \mathbb{R}^{(p+q) \times m}$$

$$B = \mathrm{D}^2_{XY} f(x, y) - \sum^{p+q} \lambda_i \mathrm{D}^2_{XY} \tilde{h}_i(x, y) \in \mathbb{R}^{m \times n}$$

Cont'd

$$C = \mathrm{D}_X \tilde{h}(x, y) \in \mathbb{R}^{(p+q)\times n}$$

$$H = \mathrm{D}^2_{YY} f(x, y) - \sum_{i=1}^{p+q} \lambda_i \mathrm{D}^2_{YY} \tilde{h}_i(x, y) \in \mathbb{R}^{m\times m}$$

Inactive and active constraints

- If constraint is inactive at solution y, $g(x, y) < 0$, then we must have $D_Y f(x, y) = 0$ and can take $Dy(x)$ to be the same as for the unconstrained case.
- If the constraints is active at the solution, $g(x, y) = 0$ but $D_Y f(x, y) \neq 0$, $\lambda \neq 0$, the gradient $Dy(x)$ is the same as replacing the inequality constraints with the equality.
- If the constraints is active at y and $D_Y f(x, y) = 0$ then $Dy(x)$ is undefined. Choose either the constrained or constrained gradient.

Examples

Illustration of different scenarios for the solution to inequality constrained nodes.

- y_1: the local minimum strictly satisfying the constraints
- y_2: on the boundary of the constraint set with the negative gradient of the objective pointing outside of the set
- y_3: on the boundary of the constraint set and is also a local minimum

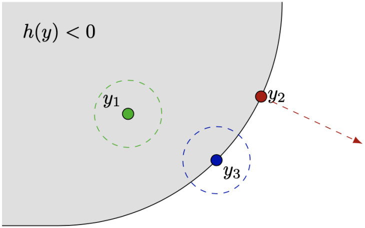


Table of Contents

Non-regular Solution

The solution of Lagrange multipliers may not always regular. The existence of the Lagrange multipliers λ is guaranteed by assumption that $y(x)$ is a regular point.

Definition 8

A feasible point u is said to be regular if the equality constraint gradients $D_U h_i$ and the active inequality constraint gradients $D_U g_i$ are linearly independent, or there are no equality constraints and the inequality constraints are all inactive at u.

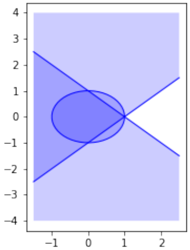
- The first-order derivatives of the equality constraints and active inequality constraints are linearly dependent.
- The number of active constraints exceeds the dimensionality of the output.

Overdetermined System

A system of equations is considered overdetermined if there are more equations than unknowns.

Examples

$$
\begin{aligned}
y \in \quad & \text{argmin}_u && \tfrac{1}{2}\|u - x\|^2 \\
& \text{subject to} && u_1^2 + u_2^2 - 1 \leq 0 \\
& && u_1 - u_2 - 1 \leq 0 \\
& && u_1 + u_2 - 1 \leq 0
\end{aligned}
$$



The linear system of equations $Ax = b$ where A is an m x n matrix with $m > n$, i.e., there are more equations than unknowns, usually does not have solutions. ($Ax \neq b$ for all x). When this is the case, we want to find an x such that the residual vector $r = b - Ax$ is, in some sense, as small as possible.

Least Square Method

The solution x given be the least squares method minimizes $\|r\|_2 = \|b - Ax\|_2$, i.e., the square sum of errors:

$$\|r\|_2 = \sum_{i=1}^{m} \left[b_i - \sum_{j=1}^{n} a_{ij} x_j \right]^2 = f(x_1, \ldots, x_n)$$

The min-max solution x minimizes $\|r\|_\infty$, i.e., the component of the residual vector: $\max(|r_1|, \ldots, |r_n|)$
The minimum value is obtained when $(x_1, \ldots, x_n)$ satisfies
$\nabla f(x_1, \ldots, x_n) = 0 \quad <=> A^T A x = A^T b \quad <=> x = (A^T A)^{-1} A^T b$

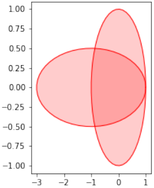
Rank Deficient

The rank deficiency of a matrix is one of the columns is fully explained by the others, in the sense that it is a linear combination of the others. A trivial example is when a column is duplicated.

Examples

$$
\begin{aligned}
y \in \quad & \operatorname{argmin}_u && \tfrac{1}{2}\|u - x\|^2 & \\
& \text{subject to} && u_1^2 + u_2^2 - 1 \leq 0 & (h_1) \\
& && \tfrac{1}{4}(u_1 + 1)^2 + 4u_2^2 - 1 \leq 0 & (h_2)
\end{aligned}
$$

At $x_2 = 0$ both constraints are active. This results in A being rank deficient.



Non-Convex Case

Also a results in rank deficiency. The constraint set is defined as the area which is a non-convex optimization problem.

Examples

$$\begin{array}{rlll} y \in & \text{argmin}_u & \frac{1}{2}\|u - x\|^2 & \\ & \text{subject to} & u_1^2 + u_2^2 - 1 \leq 0 & (h_1) \\ & & \frac{1}{4}(u_1 + 1)^2 + 4u_2^2 - 1 \geq 0 & (h_2) \end{array}$$

At $x = (0.75, 0)$ we get $y = (1, 0)$ and both constraints are active. The results in A being rank deficient.