First order optimality conditions for constrained optimization

杨俊锋

南京大学数学系

December 15, 2014

Outline

Outline

Problems with convex feasible set

Feasible set and active set

(Examples

Different feasible directions

First order necessary optimality conditions

- A geometric necessary condition
- KKT optimality conditions
- Constraint qualifications

First order sufficient conditions

Problems with convex feasible set

Let $f : R^n \to R$ be continuously differentiable and $\Omega \subset R^n$ be convex. Consider generic optimization of the form

$$\min\{f(x) : s.t.\ x \in \Omega\}.$$

Problems with convex feasible set

Let $f : R^n \to R$ be continuously differentiable and $\Omega \subset R^n$ be convex. Consider generic optimization of the form

$$\min\{f(x) : s.t.\ x \in \Omega\}.$$

Define the set of descent directions at $x \in \Omega$ by

$$\mathcal{D}_{\text{descent}}(x) := \{d \in R^n \mid \nabla f(x)^T d < 0\},$$

Problems with convex feasible set

Let $f : R^n \to R$ be continuously differentiable and $\Omega \subset R^n$ be convex. Consider generic optimization of the form

$$\min\{f(x) : s.t.\ x \in \Omega\}.$$

Define the set of descent directions at $x \in \Omega$ by

$$\mathcal{D}_{\text{descent}}(x) := \{d \in R^n \mid \nabla f(x)^T d < 0\},$$

and the set of feasible directions at $x \in \Omega$ by

$$\mathcal{D}_{\text{feasible}}(x) := \{d \in R^n \mid \ \exists \delta > 0 \text{ such that } x + \alpha d \in \Omega,\ \forall \alpha \in [0, \delta)\}.$$

Problems with convex feasible set

Let $f : R^n \to R$ be continuously differentiable and $\Omega \subset R^n$ be convex. Consider generic optimization of the form

$$\min\{f(x) : s.t.\ x \in \Omega\}.$$

Define the set of descent directions at $x \in \Omega$ by

$$\mathcal{D}_{\text{descent}}(x) := \{d \in R^n \mid \nabla f(x)^T d < 0\},$$

and the set of feasible directions at $x \in \Omega$ by

$$\mathcal{D}_{\text{feasible}}(x) := \{d \in R^n \mid \ \exists \delta > 0 \text{ such that } x + \alpha d \in \Omega,\ \forall \alpha \in [0, \delta)\}.$$

It can be shown that

$$\mathcal{D}_{\text{feasible}}(x) := \{\beta(y - x) \mid y \in \Omega, \beta > 0\}.$$

A point x^* is optimal if and only if

$$x^* \in \Omega \quad \text{and} \quad \mathcal{D}_{\text{feasible}}(x^*) \cap \mathcal{D}_{\text{descent}}(x^*) = \emptyset,$$

A point x^* is optimal if and only if

$$x^* \in \Omega \quad \text{and} \quad \mathcal{D}_{\text{feasible}}(x^*) \cap \mathcal{D}_{\text{descent}}(x^*) = \emptyset,$$

or, equivalently,

$$x^* \in \Omega, \quad (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall x \in \Omega.$$

A point x^* is optimal if and only if

$$x^* \in \Omega \quad \text{and} \quad \mathcal{D}_{\text{feasible}}(x^*) \cap \mathcal{D}_{\text{descent}}(x^*) = \emptyset,$$

or, equivalently,

$$x^* \in \Omega, \quad (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall x \in \Omega.$$

It can be shown that x^* satisfies

$$x^* = \text{Proj}_\Omega \left(x^* - \nabla f(x^*)\right).$$

Therefore, solving the underlying optimization problem reduces to solving nonlinear equations

$$F(x) := x - \text{Proj}_\Omega \left(x - \nabla f(x)\right) = 0.$$

A point x^* is optimal if and only if

$$x^* \in \Omega \quad \text{and} \quad \mathcal{D}_{\text{feasible}}(x^*) \cap \mathcal{D}_{\text{descent}}(x^*) = \emptyset,$$

or, equivalently,

$$x^* \in \Omega, \quad (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall x \in \Omega.$$

It can be shown that x^* satisfies

$$x^* = \text{Proj}_{\Omega}\left(x^* - \nabla f(x^*)\right).$$

Therefore, solving the underlying optimization problem reduces to solving nonlinear equations

$$F(x) := x - \text{Proj}_{\Omega}\left(x - \nabla f(x)\right) = 0.$$

More general problem is variational inequality:

$$\text{find } x^* \in \Omega \quad \text{such that} \quad (x - x^*)^T F(x^*) \geq 0, \quad \forall x \in \Omega,$$

where $F : R^n \to R^n$ and $\Omega \subset R^n$.

If Ω is nonconvex, the notion of feasible directions is not trivially generalized.

If Ω is nonconvex, the notion of feasible directions is not trivially generalized.

Though a geometric viewpoint for

$$\min\{f(x) : s.t.\ x \in \Omega\}$$

is possible, we first derive optimality conditions for the concrete problem

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & c_i(x) = 0, \qquad i \in \mathcal{E}, \\ & c_i(x) \geq 0, \qquad i \in \mathcal{I}, \end{aligned}$$

where f and c_i ($i \in \mathcal{E} \cup \mathcal{I}$) are all smooth, real-valued functions defined on R^n, and $\mathcal{E}$ and $\mathcal{I}$ are finite sets of indices.

Outline

Problems with convex feasible set

Feasible set and active set

(Examples

Different feasible directions

First order necessary optimality conditions
- A geometric necessary condition
- KKT optimality conditions
- Constraint qualifications

First order sufficient conditions

Feasible set and active set

Definition (feasible set)

$$\mathcal{F} := \{x \in R^n : \; c_i(x) = 0, i \in \mathcal{E}; c_i(x) \geq 0, i \in \mathcal{I}\}$$

Feasible set and active set

Definition (feasible set)

$$\mathcal{F} := \{x \in R^n : \; c_i(x) = 0, i \in \mathcal{E}; c_i(x) \geq 0, i \in \mathcal{I}\}$$

Definition (active set)

Let $x \in \mathcal{F}$ be a feasible point. An inequality constraint $c_i(x) \geq 0$ is said to be active at x if $c_i(x) = 0$ and inactive at x if $c_i(x) > 0$; all equality constraints are said to be active at x. The active set at $x \in \mathcal{F}$ is

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

Outline

Constraints can make problem much complicated

- We have seen already that global solutions are difficult to find even when there are no constraints.

Constraints can make problem much complicated

- We have seen already that global solutions are difficult to find even when there are no constraints.
- The situation may be improved when we add constraints, since the feasible set might exclude many of the local minima and it may be comparatively easy to pick the global minimum from those that remain.

Constraints can make problem much complicated

- We have seen already that global solutions are difficult to find even when there are no constraints.
- The situation may be improved when we add constraints, since the feasible set might exclude many of the local minima and it may be comparatively easy to pick the global minimum from those that remain.
- However, constraints can also make things more difficult.

Constraints can make problem much complicated

- We have seen already that global solutions are difficult to find even when there are no constraints.
- The situation may be improved when we add constraints, since the feasible set might exclude many of the local minima and it may be comparatively easy to pick the global minimum from those that remain.
- However, constraints can also make things more difficult. Example

$$\begin{array}{ll} \min & 0.01x_1^2 + (x_2 + 100)^2 \\ s.t. & x_2 - \cos x_1 \geq 0. \end{array}$$

Constraints can make problem much complicated

- We have seen already that global solutions are difficult to find even when there are no constraints.
- The situation may be improved when we add constraints, since the feasible set might exclude many of the local minima and it may be comparatively easy to pick the global minimum from those that remain.
- However, constraints can also make things more difficult. Example

$$\begin{array}{ll} \min & 0.01x_1^2 + (x_2 + 100)^2 \\ s.t. & x_2 - \cos x_1 \geq 0. \end{array}$$

With the constraint, there are local solutions near

$$(x_1, x_2) = (k\pi, -1), \quad k = \pm 1, \pm 3, \pm 5, \ldots$$

Smoothness

- Smoothness of objective functions and constraints is an important issue in characterizing solutions, just as in the unconstrained case.

Smoothness

- Smoothness of objective functions and constraints is an important issue in characterizing solutions, just as in the unconstrained case.
- It ensures that the objective function and the constraints all behave in a reasonably predictable way and therefore allows algorithms to make good choices for search directions.

Smoothness

- Smoothness of objective functions and constraints is an important issue in characterizing solutions, just as in the unconstrained case.
- It ensures that the objective function and the constraints all behave in a reasonably predictable way and therefore allows algorithms to make good choices for search directions.
- Graphs of nonsmooth functions contain “kinks" or “jumps" where the smoothness breaks down.

Smoothness

- Smoothness of objective functions and constraints is an important issue in characterizing solutions, just as in the unconstrained case.
- It ensures that the objective function and the constraints all behave in a reasonably predictable way and therefore allows algorithms to make good choices for search directions.
- Graphs of nonsmooth functions contain “kinks" or “jumps" where the smoothness breaks down.
- The feasible region for constrained optimization problem may contain many kinks and sharp edges, and this, in general, does not mean that the constraint functions are nonsmooth.

- For example, the set

$$\Omega := \{x = (x_1, x_2) \mid \|x\|_1 \leq 1\}$$

can also be described by

$$\Omega := \{x = (x_1, x_2) \mid x_1 \pm x_2 \leq 1, -x_1 \pm x_2 \leq 1\}.$$

- For example, the set

$$\Omega := \{x = (x_1, x_2) \mid \|x\|_1 \leq 1\}$$

can also be described by

$$\Omega := \{x = (x_1, x_2) \mid x_1 \pm x_2 \leq 1, -x_1 \pm x_2 \leq 1\}.$$

- On the other hand, nonsmooth unconstrained problems can sometimes be reformulated as smooth constrained problems, e.g.,

$$\min_{x \in \Omega} \{f(x) := \max(x^2, x)\}$$

is equivalent to

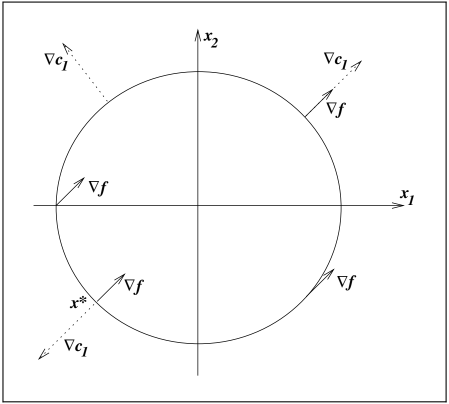
$$\min_{x,t} \{t : \ s.t. \ t \geq x, t \geq x^2, x \in \Omega\}.$$

Example: A single equality constraint

$$\min\{x_1 + x_2 : \ s.t. \ x_1^2 + x_2^2 - 2 = 0\},$$

i.e., $n = 2$, $\mathcal{I} = \emptyset$, $\mathcal{E} = \{1\}$,

$$f(x) = x_1 + x_2 \quad \text{and} \quad c_1(x) = x_1^2 + x_2^2 - 2.$$



- The solution is $x^* = (-1, -1)^T$;
- From other points on the circle, can find a way to move that stays feasible while decreasing f;
- It holds that $\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*)$, where $\lambda_1^* = -\frac{1}{2}$;

- The solution is $x^* = (-1, -1)^T$;
- From other points on the circle, can find a way to move that stays feasible while decreasing f;
- It holds that $\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*)$, where $\lambda_1^* = -\frac{1}{2}$;

Suppose that $c_1(x) = 0$. We require $c_1(x + d) = 0$, i.e.,

$$0 = c_1(x + d) \approx c_1(x) + \nabla c_1(x)^T d = \nabla c_1(x)^T d.$$

Hence, the direction d retains feasibility w.r.t. c_1, to first order, when it satisfies $\nabla c_1(x)^T d = 0$.

- The solution is $x^* = (-1, -1)^T$;
- From other points on the circle, can find a way to move that stays feasible while decreasing f;
- It holds that $\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*)$, where $\lambda_1^* = -\frac{1}{2}$;

Suppose that $c_1(x) = 0$. We require $c_1(x + d) = 0$, i.e.,

$$0 = c_1(x + d) \approx c_1(x) + \nabla c_1(x)^T d = \nabla c_1(x)^T d.$$

Hence, the direction d retains feasibility w.r.t. c_1, to first order, when it satisfies $\nabla c_1(x)^T d = 0$.

If there is no direction d such that

$$\nabla c_1(x)^T d = 0 \quad \text{and} \quad \nabla f(x)^T d < 0,$$

then is it likely that x is a local minimizer. Easy to check that such direction d does not exist only when $\nabla f(x)$ and $\nabla c_1(x)$ are parallel, i.e., $\nabla f(x) = \lambda_1 \nabla c_1(x)$ for some scalar λ_1.

Define Lagrangian function

$$\mathcal{L}(x, \lambda) := f(x) - \lambda_1 c_1(x).$$

For this example, x^* is optimal and there exists $\lambda_1^* \in R$ such that

$$\begin{aligned} \nabla_x \mathcal{L}(x^*, \lambda_1^*) &= 0, \\ c_1(x^*) &= 0. \end{aligned}$$

(will be made precise later.)

Define Lagrangian function

$$\mathcal{L}(x, \lambda) := f(x) - \lambda_1 c_1(x).$$

For this example, x^* is optimal and there exists $\lambda_1^* \in R$ such that

$$\begin{aligned} \nabla_x \mathcal{L}(x^*, \lambda_1^*) &= 0, \\ c_1(x^*) &= 0. \end{aligned}$$

(will be made precise later.)

- Similar condition holds at $(1, 1)$, which indicates that this condition is not sufficient to be a minimum;

Define Lagrangian function

$$\mathcal{L}(x, \lambda) := f(x) - \lambda_1 c_1(x).$$

For this example, x^* is optimal and there exists $\lambda_1^* \in R$ such that

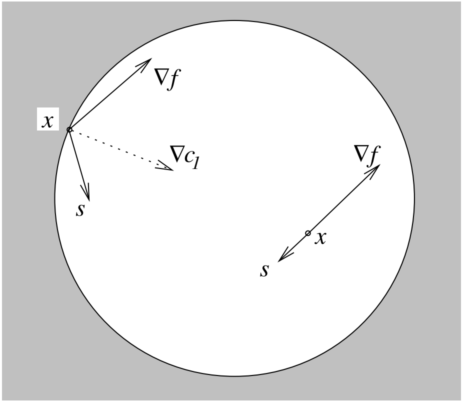
$$\begin{aligned} \nabla_x \mathcal{L}(x^*, \lambda_1^*) &= 0, \\ c_1(x^*) &= 0. \end{aligned}$$

(will be made precise later.)

- Similar condition holds at $(1, 1)$, which indicates that this condition is not sufficient to be a minimum;
- By replacing c_1 by $-c_1$, we see that the sign of λ_1^* is not essential (in this case, λ_1^* changes from $-1/2$ to $1/2$).

Example: A single inequality constraint

$$\min\{x_1 + x_2 : \ s.t.\ 2 - x_1^2 - x_2^2 \geq 0\}.$$



- The solution is still $x^* = (-1, -1)^T$;
- It holds that $\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*)$ with $\lambda_1^* = \frac{1}{2}$.
- The sign of λ_1^* cannot be changed.

- The solution is still $x^* = (-1, -1)^T$;
- It holds that $\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*)$ with $\lambda_1^* = \frac{1}{2}$.
- The sign of λ_1^* cannot be changed.

Suppose that $c_1(x) \geq 0$. We require $c_1(x + d) \geq 0$, i.e.,

$$0 \leq c_1(x + d) \approx c_1(x) + \nabla c_1(x)^T d.$$

Hence, the direction d retains feasibility w.r.t. c_1, to first order, when it satisfies

$$c_1(x) + \nabla c_1(x)^T d \geq 0.$$

- The solution is still $x^* = (-1, -1)^T$;
- It holds that $\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*)$ with $\lambda_1^* = \frac{1}{2}$.
- The sign of λ_1^* cannot be changed.

Suppose that $c_1(x) \geq 0$. We require $c_1(x+d) \geq 0$, i.e.,

$$0 \leq c_1(x+d) \approx c_1(x) + \nabla c_1(x)^T d.$$

Hence, the direction d retains feasibility w.r.t. c_1, to first order, when it satisfies

$$c_1(x) + \nabla c_1(x)^T d \geq 0.$$

- If $c_1(x) > 0$, then can find d such that

  $$c_1(x) + \nabla c_1(x)^T d \geq 0 \quad \text{and} \quad \nabla f(x)^T d < 0,$$

  unless $\nabla f(x) = 0$. Thus, if $c_1(x) > 0$ and $\nabla f(x) = 0$, then x is likely a local minimizer.

- Suppose $c_1(x) = 0$. The conditions

$$\nabla c_1(x)^T d \geq 0 \quad \text{and} \quad \nabla f(x)^T d < 0$$

cannot be simultaneously satisfied by some $d \in R^n$ only when $\nabla f(x)$ and $\nabla c_1(x)$ point in the same direction, i.e.,

$$\nabla f(x) = \lambda_1 \nabla c_1(x), \quad \text{for some } \lambda_1 \geq 0.$$

In this case, x is likely a local minimizer.

- Suppose $c_1(x) = 0$. The conditions

$$\nabla c_1(x)^T d \geq 0 \quad \text{and} \quad \nabla f(x)^T d < 0$$

cannot be simultaneously satisfied by some $d \in R^n$ only when $\nabla f(x)$ and $\nabla c_1(x)$ point in the same direction, i.e.,

$$\nabla f(x) = \lambda_1 \nabla c_1(x), \quad \text{for some } \lambda_1 \geq 0.$$

In this case, x is likely a local minimizer.

Define Lagrangian function

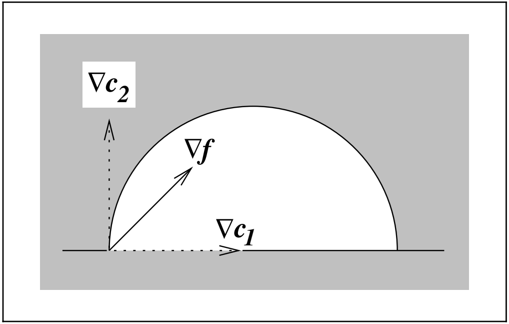
$$\mathcal{L}(x, \lambda) := f(x) - \lambda_1 c_1(x).$$

For this example, x^* is optimal and there exists $\lambda_1^* \in R$ such that

$$\begin{aligned} \nabla_x \mathcal{L}(x^*, \lambda_1^*) &= 0, \\ \lambda_1^* c_1(x^*) &= 0, \\ \lambda_1^* &\geq 0. \end{aligned}$$

Example: Two inequality constraints

$$\min\{x_1 + x_2 : \ s.t.\ 2 - x_1^2 - x_2^2 \geq 0, x_2 \geq 0\}.$$



- The solution is $x^* = (-\sqrt{2}, 0)^T$, a point at which both constraints are active. If both constraints are active at a feasible point x, then at this point a direction d is a feasible descent direction, to first-order, if it satisfies

$$\nabla c_i(x)^T d \geq 0,\ i \in \mathcal{I} = \{1, 2\}, \quad \nabla f(x)^T d < 0.$$

  Easy to see that at $(-\sqrt{2}, 0)^T$ no such direction exists.

- The solution is $x^* = (-\sqrt{2}, 0)^T$, a point at which both constraints are active. If both constraints are active at a feasible point x, then at this point a direction d is a feasible descent direction, to first-order, if it satisfies

$$\nabla c_i(x)^T d \geq 0,\ i \in \mathcal{I} = \{1, 2\}, \quad \nabla f(x)^T d < 0.$$

  Easy to see that at $(-\sqrt{2}, 0)^T$ no such direction exists.
- Define Lagrangian function by

$$\mathcal{L}(x, \lambda) = f(x) - \lambda_1 c_1(x) - \lambda_2 c_2(x).$$

  For this example, $x^* = (-\sqrt{2}, 0)^T$ is optimal and there exists λ^* such that

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

  In fact, $\lambda^* = (\sqrt{2}/4, 1)^T$.

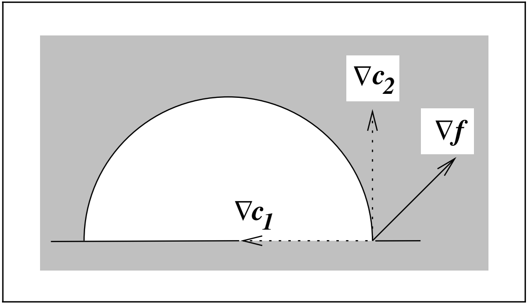
- At $x = (\sqrt{2}, 0)^T$, where both constraints are active, it is easy to find d such that

$$\nabla c_i(x)^T d \geq 0,\ i \in \mathcal{I} = \{1, 2\}, \quad \nabla f(x)^T d < 0.$$

Thus, $x = (\sqrt{2}, 0)^T$ is not a solution. Easy to show that at this point

$$\nabla_x \mathcal{L}(x, \lambda) = 0$$

holds only when $\lambda = (-\sqrt{2}/4, 1)^T$. Thus the three conditions given in the last slide cannot be satisfied simultaneously.



- At $x = (1, 0)^T$ only the constraint $c_2(x) = x_2 \geq 0$ is active. Since $c_1(x) > 0$ at this point, stepping forward in any direction d will maintain $c_1(x + d) > 0$ as long as d is sufficiently small. Thus, no need to take care of the first constraint. At $(1, 0)$, a direction d is feasible and descent, to first order, if it satisfies

$$\nabla c_2(x)^T d \geq 0, \quad \nabla f(x)^T d < 0.$$

Clearly such d exists, and thus $x = (1, 0)^T$ is not a solution. In this case, easy to show that the following conditions cannot be satisfied simultaneously:

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

- The previous examples show that it might be possible to derive necessary optimality conditions for $x^* \in \mathcal{F}$ to be a local optimal solution by using

$$\{\nabla f(x), \nabla c_i(x),\ i \in \mathcal{I}\}.$$

- The previous examples show that it might be possible to derive necessary optimality conditions for $x^* \in \mathcal{F}$ to be a local optimal solution by using

$$\{\nabla f(x), \nabla c_i(x),\ i \in \mathcal{I}\}.$$

- However, this is not trivial since a geometric set can be described in different ways, and this could cause problem.

- The previous examples show that it might be possible to derive necessary optimality conditions for $x^* \in \mathcal{F}$ to be a local optimal solution by using

$$\{\nabla f(x), \nabla c_i(x),\ i \in \mathcal{I}\}.$$

- However, this is not trivial since a geometric set can be described in different ways, and this could cause problem. For example, consider

$$\min\{f(x) := x_1 + x_2 :\ s.t.\ c_1(x) := (x_1^2 + x_2^2 - 2)^2 = 0\}.$$

  In this case, at the solution $x^* = (-1, -1)^T$, the following conditions cannot be fulfilled anymore:

$$\begin{aligned} \nabla_x \mathcal{L}(x^*, \lambda_1^*) &= 0 \quad \text{for some } \lambda_1^* \in R, \\ c_1(x^*) &= 0. \end{aligned}$$

  (can show that $\nabla c_1(x^*)^T d = 0$ for all $d \in R^n$.)

Outline

Three types of feasible directions

Recall that the feasible set is defined by

$$\mathcal{F} := \{x \in R^n : \; c_i(x) = 0, i \in \mathcal{E}; c_i(x) \geq 0, i \in \mathcal{I}\}.$$

Three types of feasible directions

Recall that the feasible set is defined by

$$\mathcal{F} := \{x \in R^n : \ c_i(x) = 0, i \in \mathcal{E}; c_i(x) \geq 0, i \in \mathcal{I}\}.$$

Definition (feasible directions)

Let $x \in \mathcal{F}$. A direction $d \in R^n$ is said to be a feasible direction at $x \in \mathcal{F}$ if there exists $\delta > 0$ such that

$$x + \alpha d \in \mathcal{F}, \quad \forall \alpha \in [0, \delta).$$

The set of all feasible directions at a feasible point $x \in \mathcal{F}$ is denoted by $FD(x, \mathcal{F})$.

Definition (linearized feasible directions)

Let $x \in \mathcal{F}$. Suppose that all functions c_i, $i \in \mathcal{I}$, are differentiable. A direction $d \in R^n$ is said to be a linearized feasible direction if it satisfies

$$\nabla c_i(x)^T d = 0, \quad i \in \mathcal{E}; \quad \nabla c_i(x)^T d \geq 0, \quad i \in \mathcal{I} \cap \mathcal{A}(x).$$

The set of all linearized feasible directions at a feasible point $x \in \mathcal{F}$ is denoted by $LFD(x, \mathcal{F})$.

Definition (linearized feasible directions)

Let $x \in \mathcal{F}$. Suppose that all functions c_i, $i \in \mathcal{I}$, are differentiable. A direction $d \in R^n$ is said to be a linearized feasible direction if it satisfies

$$\nabla c_i(x)^T d = 0, \quad i \in \mathcal{E}; \quad \nabla c_i(x)^T d \geq 0, \quad i \in \mathcal{I} \cap \mathcal{A}(x).$$

The set of all linearized feasible directions at a feasible point $x \in \mathcal{F}$ is denoted by $LFD(x, \mathcal{F})$.

Definition (sequential feasible directions)

Let $x \in \mathcal{F}$. A direction $d \in R^n$ is said to be a sequential feasible direction at $x \in \mathcal{F}$ if there exists $\{d_k \in R^n : k = 1, 2, \ldots\}$ and $\{\delta_k > 0 : k = 1, 2, \ldots\}$ such that

$$d_k \to d, \quad \delta_k \to 0 \quad \text{and} \quad x + \delta_k d_k \in \mathcal{F}, \quad \forall\, k.$$

The set of all sequential feasible directions at a feasible point $x \in \mathcal{F}$ is denoted by $SFD(x, \mathcal{F})$.

Remarks

- *For any $x \in \mathcal{F}$, the sets $FD(x, \mathcal{F})$, $LFD(x, \mathcal{F})$ and $SFD(x, \mathcal{F})$ are all cones.*

Remarks

- *For any $x \in \mathcal{F}$, the sets $FD(x, \mathcal{F})$, $LFD(x, \mathcal{F})$ and $SFD(x, \mathcal{F})$ are all cones.*
- *$FD(x, \mathcal{F})$ is simple, but mainly used when $\mathcal{F}$ is convex.*

Remarks

- *For any* $x \in \mathcal{F}$*, the sets* $FD(x, \mathcal{F})$*,* $LFD(x, \mathcal{F})$ *and* $SFD(x, \mathcal{F})$ *are all cones.*
- $FD(x, \mathcal{F})$ *is simple, but mainly used when* $\mathcal{F}$ *is convex.*
- $SFD(x, \mathcal{F})$ *does not depend on how* $\mathcal{F}$ *is described and is completely determined by the geometry of* $\mathcal{F}$*. However, it is not easy to characterize and thus not easy to use. This set is also called the tangent cone at* $x \in \mathcal{F}$ *(elements in which are called tangent directions).*

Remarks

- *For any $x \in \mathcal{F}$, the sets $FD(x, \mathcal{F})$, $LFD(x, \mathcal{F})$ and $SFD(x, \mathcal{F})$ are all cones.*
- *$FD(x, \mathcal{F})$ is simple, but mainly used when $\mathcal{F}$ is convex.*
- *$SFD(x, \mathcal{F})$ does not depend on how $\mathcal{F}$ is described and is completely determined by the geometry of $\mathcal{F}$. However, it is not easy to characterize and thus not easy to use. This set is also called the tangent cone at $x \in \mathcal{F}$ (elements in which are called tangent directions).*
- *$LFD(x, \mathcal{F})$ is easy to use, but it depends on how $\mathcal{F}$ is described.*

Outline

Problems with convex feasible set

Feasible set and active set

(Examples

Different feasible directions

First order necessary optimality conditions

- A geometric necessary condition
- KKT optimality conditions
- Constraint qualifications

First order sufficient conditions

A geometric necessary condition

Theorem (Geometric necessary optimality condition)

Let x^* be a local optimal solution. Suppose f and c_i, $i \in \mathcal{I}$, are all differentiable. Then, it holds that

$$\nabla f(x^*)^T d \geq 0, \quad \forall d \in SFD(x^*, \mathcal{F}).$$

A geometric necessary condition

Theorem (Geometric necessary optimality condition)

Let x^* be a local optimal solution. Suppose f and c_i, $i \in \mathcal{I}$, are all differentiable. Then, it holds that

$$\nabla f(x^*)^T d \geq 0, \quad \forall d \in SFD(x^*, \mathcal{F}).$$

The converse of this theorem is not true. Consider, e.g.,

$$\min\{x_2 : \ s.t. \ x_2 \geq -x_1^2\}.$$

A geometric necessary condition

Theorem (Geometric necessary optimality condition)

Let x^* be a local optimal solution. Suppose f and c_i, $i \in \mathcal{I}$, are all differentiable. Then, it holds that

$$\nabla f(x^*)^T d \geq 0, \quad \forall d \in SFD(x^*, \mathcal{F}).$$

The converse of this theorem is not true. Consider, e.g.,

$$\min\{x_2 : \ s.t.\ x_2 \geq -x_1^2\}.$$

Theorem

Let $x \in \mathcal{F}$. Suppose all c_i's are differentiable at x. It holds that

$$FD(x, \mathcal{F}) \subseteqq SFD(x, \mathcal{F}) \subseteqq LFD(x, \mathcal{F}).$$

Constraint qualifications

Note that $SFD(x, \mathcal{F}) \subsetneqq LFD(x, \mathcal{F})$ can happen. E.g., the singleton $\mathcal{F} = \{x^* = (0,0)^T\}$ can be expressed as

$$\mathcal{F} = \{x \in R^2 : c_1(x) \geq 0, c_2(x) \geq 0\},$$

where

$$c_1(x) = 1 - x_1^2 - (x_2 - 1)^2 \quad \text{and} \quad c_2(x) = -x_2.$$

Easy to verify that $SFD(x^*, \mathcal{F}) = \{(0,0)^T\}$ and

$$LFD(x^*, \mathcal{F}) = \{(d_1, 0)^T \mid d_1 \in R\}.$$

Constraint qualifications

Note that $SFD(x, \mathcal{F}) \subsetneqq LFD(x, \mathcal{F})$ can happen. E.g., the singleton $\mathcal{F} = \{x^* = (0, 0)^T\}$ can be expressed as

$$\mathcal{F} = \{x \in R^2 : c_1(x) \geq 0, c_2(x) \geq 0\},$$

where

$$c_1(x) = 1 - x_1^2 - (x_2 - 1)^2 \quad \text{and} \quad c_2(x) = -x_2.$$

Easy to verify that $SFD(x^*, \mathcal{F}) = \{(0, 0)^T\}$ and

$$LFD(x^*, \mathcal{F}) = \{(d_1, 0)^T \mid d_1 \in R\}.$$

Constraint qualifications (约束规范) are conditions under which there holds

$$SFD(x, \mathcal{F}) = LFD(x, \mathcal{F}).$$

Lemma (Farkas lemma)

Let p, q be any nonnegative integers. Suppose

$$\{a_i : i = 1, 2, \dots, p\},\ \{b_j : j = 1, 2, \dots, q\} \text{ and } v$$

are vectors in R^n. Then, the system

$$\begin{array}{rcll} a_i^T d & = & 0, & i = 1, 2, \dots, p, \\ b_j^T d & \geq & 0, & j = 1, 2, \dots, q, \\ v^T d & < & 0, & \end{array}$$

has no solution if and only if there exist $\lambda_i \in R$, $i = 1, 2, \dots, p$, and $\lambda_j \geq 0$, $j = 1, 2, \dots, q$, such that

$$v = \sum_{i=1}^{p} \lambda_i a_i + \sum_{j=1}^{q} \lambda_j b_j.$$

Theorem (KKT optimality conditions)

Suppose that all functions defining the problem are C^1 and that x^* is a local minimizer. If $SFD(x^*, \mathcal{F}) = LFD(x^*, \mathcal{F})$, then there exist $\lambda_i^* \in R$, $i \in \mathcal{E} \cup \mathcal{I}$, such that the following conditions are satisfied:

$$\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= \nabla f(x^*) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* \nabla c_i(x^*) = 0, \\
c_i(x^*) &= 0, \quad i \in \mathcal{E}, \\
c_i(x^*) &\geq 0, \quad i \in \mathcal{I}, \\
\lambda_i^* &\geq 0, \quad i \in \mathcal{I}, \\
\lambda_i^* c_i(x^*) &= 0, \quad i \in \mathcal{I},
\end{aligned}$$

where $\mathcal{L}(x, \lambda)$ is called the Lagrange function and is defined by

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

The vector λ is usually referred to as Lagrange multipliers.

- The first order necessary optimality conditions presented in the last slide is implied by the Farkas lemma.

[1] H. W. Kuhn and A. W. Tucker, Nonlinear programming, in: J. Neyman, ed., Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (University of California Press, Berkeley, California, 1951) 481-492.

[2] W. Karush, Minima of functions of several variables with inequalities as side conditions, Master's thesis, University of Chicago, Chicago, Illinois, (1939).

- The first order necessary optimality conditions presented in the last slide is implied by the Farkas lemma.
- The theorem was given by Kuhn and Tucker in 1951. [1]

[1] H. W. Kuhn and A. W. Tucker, Nonlinear programming, in: J. Neyman, ed., Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (University of California Press, Berkeley, California, 1951) 481-492.

[2] W. Karush, Minima of functions of several variables with inequalities as side conditions, Master's thesis, University of Chicago, Chicago, Illinois, (1939).

- The first order necessary optimality conditions presented in the last slide is implied by the Farkas lemma.
- The theorem was given by Kuhn and Tucker in 1951. [1]
- Similar conditions were considered by Karush in 1939. [2]

---

[1] H. W. Kuhn and A. W. Tucker, Nonlinear programming, in: J. Neyman, ed., Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (University of California Press, Berkeley, California, 1951) 481-492.

[2] W. Karush, Minima of functions of several variables with inequalities as side conditions, Master's thesis, University of Chicago, Chicago, Illinois, (1939).

- The first order necessary optimality conditions presented in the last slide is implied by the Farkas lemma.
- The theorem was given by Kuhn and Tucker in 1951. [1]
- Similar conditions were considered by Karush in 1939. [2]
- Now it was widely referred to as the Karush-Kuhn-Tucker conditions, or KKT conditions for short.

[1] H. W. Kuhn and A. W. Tucker, Nonlinear programming, in: J. Neyman, ed., Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (University of California Press, Berkeley, California, 1951) 481-492.

[2] W. Karush, Minima of functions of several variables with inequalities as side conditions, Master's thesis, University of Chicago, Chicago, Illinois, (1939).

- The first order necessary optimality conditions presented in the last slide is implied by the Farkas lemma.
- The theorem was given by Kuhn and Tucker in 1951. [1]
- Similar conditions were considered by Karush in 1939. [2]
- Now it was widely referred to as the Karush-Kuhn-Tucker conditions, or KKT conditions for short.
- A point that satisfies the conditions is referred to as a KKT point. Without convexity assumptions, a KKT point is the best we can expect to achieve in most cases.

---

[1] H. W. Kuhn and A. W. Tucker, Nonlinear programming, in: J. Neyman, ed., Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (University of California Press, Berkeley, California, 1951) 481-492.

[2] W. Karush, Minima of functions of several variables with inequalities as side conditions, Master's thesis, University of Chicago, Chicago, Illinois, (1939).

- The first order necessary optimality conditions presented in the last slide is implied by the Farkas lemma.
- The theorem was given by Kuhn and Tucker in 1951. [1]
- Similar conditions were considered by Karush in 1939. [2]
- Now it was widely referred to as the Karush-Kuhn-Tucker conditions, or KKT conditions for short.
- A point that satisfies the conditions is referred to as a KKT point. Without convexity assumptions, a KKT point is the best we can expect to achieve in most cases.
- The function $\mathcal{L}(x, \lambda)$ can be traced back to Lagrange (1760) and thus is referred to as the Lagrange function.

---

[1] H. W. Kuhn and A. W. Tucker, Nonlinear programming, in: J. Neyman, ed., Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (University of California Press, Berkeley, California, 1951) 481-492.

[2] W. Karush, Minima of functions of several variables with inequalities as side conditions, Master's thesis, University of Chicago, Chicago, Illinois, (1939).

LP example

Consider standard form LP and its dual:

$$\min\{c^T x : \ s.t. Ax = b, x \geq 0\} \quad \text{and} \quad \max\{b^T y : \ s.t. \ A^T y \leq c\}.$$

The Lagrange function: $\mathcal{L}(x, y, s) = c^T x - y^T(Ax - b) - s^T x$.
The KKT conditions:

$$\begin{aligned} \nabla_x \mathcal{L}(x, y, s) = c - A^T y - s = 0, \\ Ax = b, \quad x \geq 0, \\ s_i \geq 0, \ s_i x_i = 0, \ \forall i, \end{aligned}$$

are equivalent to

$$\begin{aligned} A^T y + s = c, \\ Ax = b, \quad x \geq 0, \\ s \geq 0, \ s^T x = 0. \end{aligned}$$

For LP, we already know that these conditions are also sufficient for x (resp. (y, s)) to be primal (resp. dual) optimal.

If $SFD(x^*, \mathcal{F}) \subsetneq LFD(x^*, \mathcal{F})$, a local minimizer can fail to be a KKT point

$$\begin{aligned} \min_{x \in R^2} \quad & x_1 \\ s.t. \quad & x_1^3 - x_2 \geq 0, \\ & x_2 \geq 0. \end{aligned}$$

- Global solution $x^* = (0, 0)^T$ is not a KKT point.
- Note that

$$\begin{aligned} SFD(x^*, \mathcal{F}) &= \left\{ d \in R^2 \mid d = (d_1, 0)^T, d_1 \geq 0 \right\}, \\ LFD(x^*, \mathcal{F}) &= \left\{ d \in R^2 \mid d = (d_1, 0)^T, d_1 \in R \right\}. \end{aligned}$$

The condition $SFD(x^*, \mathcal{F}) = LFD(x^*, \mathcal{F})$ is sufficient but not necessary

$$\begin{aligned} \min_{x \in R^2} \quad & x_2 \\ s.t. \quad & x_1^2 + (x_2 - 1)^2 - 1 = 0, \\ & x_1^2 + (x_2 + 1)^2 - 1 = 0. \end{aligned}$$

- Global solution $x^* = (0, 0)^T$ is a KKT point.
- Note that

$$\begin{aligned} SFD(x^*, \mathcal{F}) &= \left\{ d \in R^2 \mid d = (0, 0)^T \right\}, \\ LFD(x^*, \mathcal{F}) &= \left\{ d \in R^2 \mid d = (d_1, 0)^T, d_1 \in R \right\}. \end{aligned}$$

Other constraint qualifications: LFCQ

Definition (LFCQ)

Let x^* be a local minimizer. If all active constraint functions

$$\{c_i(x) : \ i \in \mathcal{A}(x^*)\}$$

are linear functions, we say that linear function constraint qualification (LFCQ) holds at x^*.

Other constraint qualifications: LFCQ

Definition (LFCQ)

Let x^* be a local minimizer. If all active constraint functions

$$\{c_i(x) : \ i \in \mathcal{A}(x^*)\}$$

are linear functions, we say that linear function constraint qualification (LFCQ) holds at x^*.

Theorem

Let x_0 be a feasible point. If all active constraint functions

$$\{c_i(x) : \ i \in \mathcal{A}(x_0)\}$$

are linear functions, then it holds that $SFD(x_0, \mathcal{F}) = LFD(x_0, \mathcal{F})$.

Other constraint qualifications: LFCQ

Definition (LFCQ)

Let x^* be a local minimizer. If all active constraint functions

$$\{c_i(x): \; i \in \mathcal{A}(x^*)\}$$

are linear functions, we say that linear function constraint qualification (LFCQ) holds at x^*.

Theorem

Let x_0 be a feasible point. If all active constraint functions

$$\{c_i(x): \; i \in \mathcal{A}(x_0)\}$$

are linear functions, then it holds that $SFD(x_0, \mathcal{F}) = LFD(x_0, \mathcal{F})$.

Corollary

If LFCQ holds at a local minimizer x^*, then x^* is a KKT point.

Other constraint qualifications: LFCQ

Definition (LFCQ)

Let x^* be a local minimizer. If all active constraint functions

$$\{c_i(x) : \; i \in \mathcal{A}(x^*)\}$$

are linear functions, we say that linear function constraint qualification (LFCQ) holds at x^*.

Theorem

Let x_0 be a feasible point. If all active constraint functions

$$\{c_i(x) : \; i \in \mathcal{A}(x_0)\}$$

are linear functions, then it holds that $SFD(x_0, \mathcal{F}) = LFD(x_0, \mathcal{F})$.

Corollary

If LFCQ holds at a local minimizer x^*, then x^* is a KKT point.

Corollary

If all constraint functions $\{c_i(x) : \; i \in \mathcal{E} \cup \mathcal{I}\}$ are linear functions, then any local solution x^* is a KKT point.

Other constraint qualifications: LICQ

Definition (LICQ)

Let x^* be a local minimizer. If all active constraint gradients

$$\{\nabla c_i(x^*) : \; i \in \mathcal{A}(x^*)\}$$

are linearly independent, we say that the linear independence constraint qualification (LICQ) holds at x^*.

Other constraint qualifications: LICQ

Definition (LICQ)

Let x^* be a local minimizer. If all active constraint gradients

$$\{\nabla c_i(x^*) : \; i \in \mathcal{A}(x^*)\}$$

are linearly independent, we say that the linear independence constraint qualification (LICQ) holds at x^*.

Theorem

Let $x_0 \in \mathcal{F}$ be a feasible point. If $\{\nabla c_i(x_0) : \; i \in \mathcal{A}(x_0)\}$ are linearly independent, then $SFD(x_0, \mathcal{F}) = LFD(x_0, \mathcal{F})$.

Other constraint qualifications: LICQ

Definition (LICQ)

Let x^* be a local minimizer. If all active constraint gradients

$$\{\nabla c_i(x^*) : \; i \in \mathcal{A}(x^*)\}$$

are linearly independent, we say that the linear independence constraint qualification (LICQ) holds at x^*.

Theorem

Let $x_0 \in \mathcal{F}$ be a feasible point. If $\{\nabla c_i(x_0) : \; i \in \mathcal{A}(x_0)\}$ are linearly independent, then $SFD(x_0, \mathcal{F}) = LFD(x_0, \mathcal{F})$.

Corollary

Let x^* be a local minimizer. If LICQ holds at x^*, then x^* is a KKT point.

Other constraint qualifications: MFCQ

Definition (MFCQ)

Let x^* be a local minimizer. If $\{\nabla c_i(x^*) : \ i \in \mathcal{E}\}$ are linearly independent, and

$$\{d \mid \nabla c_i(x^*)^T d = 0, i \in \mathcal{E}; \nabla c_i(x^*)^T d > 0, i \in \mathcal{A}(x^*) \cap \mathcal{I}\} \neq \emptyset,$$

then we say that Mangasarian-Fromowitz CQ holds at x^*.

Other constraint qualifications: MFCQ

Definition (MFCQ)

Let x^* be a local minimizer. If $\{\nabla c_i(x^*) : \ i \in \mathcal{E}\}$ are linearly independent, and

$$\{d \mid \nabla c_i(x^*)^T d = 0, i \in \mathcal{E}; \nabla c_i(x^*)^T d > 0, i \in \mathcal{A}(x^*) \cap \mathcal{I}\} \neq \emptyset,$$

then we say that Mangasarian-Fromowitz CQ holds at x^*.

It can be shown that if MFCQ holds at a feasible point x then $SFD(x, \mathcal{F}) = LFD(x, \mathcal{F})$. Also, LICQ implies MFCQ.

Other constraint qualifications: MFCQ

Definition (MFCQ)

Let x^* be a local minimizer. If $\{\nabla c_i(x^*) : \ i \in \mathcal{E}\}$ are linearly independent, and

$$\{d \mid \nabla c_i(x^*)^T d = 0, i \in \mathcal{E}; \nabla c_i(x^*)^T d > 0, i \in \mathcal{A}(x^*) \cap \mathcal{I}\} \neq \emptyset,$$

then we say that Mangasarian-Fromowitz CQ holds at x^*.

It can be shown that if MFCQ holds at a feasible point x then $SFD(x, \mathcal{F}) = LFD(x, \mathcal{F})$. Also, LICQ implies MFCQ.

Theorem

Let x^* be a local minimizer. If MFCQ holds at x^*, then x^* is a KKT point.

Outline

Theorem (A first order sufficient condition)

Suppose all functions are C^1. Let x^* be a feasible point. If

$$\nabla f(x^*)^T d > 0, \forall 0 \neq d \in SFD(x^*, \mathcal{F}),$$

then x^* is a strict local minimizer.

Proof.

Refer to Theorem 8.2.16 in the book by Yuan-Sun (Chinese version). $\square$

Theorem

Consider convex optimization problem of the form

$$\begin{aligned} \min_{x \in R^n} \quad & f(x) \\ s.t. \quad & Ax = b, \\ & c_i(x) \leq 0, \\ & i = 1, 2, \ldots, q, \end{aligned}$$

where $f \in C^1(R^n)$ *and* $c_i \in C^1(R^n)$, $i = 1, 2, \ldots, q$, *are all convex functions,* $A \in R^{p \times n}$ *and* $b \in R^p$. *If* x^* *is a KKT point, then it is a global optimal solution.*

Theorem

Consider convex optimization problem of the form

$$\begin{aligned} \min_{x \in R^n} \quad & f(x) \\ s.t. \quad & Ax = b, \\ & c_i(x) \leq 0, \\ & i = 1, 2, \ldots, q, \end{aligned}$$

where $f \in C^1(R^n)$ and $c_i \in C^1(R^n)$, $i = 1, 2, \ldots, q$, are all convex functions, $A \in R^{p \times n}$ and $b \in R^p$. If x^* is a KKT point, then it is a global optimal solution.

Define the Lagrange function $\mathcal{L} : R^n \times R^p \times R^q \to R$ by

$$\mathcal{L}(x, \lambda, \eta) := f(x) - \lambda^T (Ax - b) + \sum_{i=1}^{q} \eta_i c_i(x).$$

Clearly, for any fixed λ and η, $\mathcal{L}$ is convex in x.

Proof

Suppose x^* is a KKT point, i.e., there exist $\lambda^* \in R^p$ and $\eta^* \in R^q$ such that

$$\begin{aligned} \nabla f(x^*) - A^T\lambda^* + \sum_{i=1}^{q} \eta_i^* \nabla c_i(x^*) &= 0, \\ Ax^* &= b, \\ c_i(x^*) \leq 0, \quad \forall i = 1, 2, \ldots, q&, \\ \eta^* &\geq 0, \\ \eta_i^* c_i(x^*) = 0, \quad \forall i = 1, 2, \ldots, q&. \end{aligned}$$

Proof

Suppose x^* is a KKT point, i.e., there exist $\lambda^* \in R^p$ and $\eta^* \in R^q$ such that

$$\begin{aligned}
\nabla f(x^*) - A^T\lambda^* + \sum_{i=1}^{q} \eta_i^* \nabla c_i(x^*) &= 0, \\
Ax^* &= b, \\
c_i(x^*) \leq 0, \quad \forall i = 1, 2, \ldots, q&, \\
\eta^* &\geq 0, \\
\eta_i^* c_i(x^*) = 0, \quad \forall i = 1, 2, \ldots, q&.
\end{aligned}$$

Since $\mathcal{L}$ is convex and differentiable in x, for any feasible x, it holds that

$$\mathcal{L}(x, \lambda^*, \eta^*) \geq \mathcal{L}(x^*, \lambda^*, \eta^*) + \langle \nabla_x \mathcal{L}(x^*, \lambda^*, \eta^*), x - x^* \rangle.$$

Proof continued

Or, equivalently,

$$\begin{aligned} & f(x) + \sum_{i=1}^{q} \eta_i^* c_i(x) \\ \geq \quad & f(x^*) + \sum_{i=1}^{q} \eta_i^* c_i(x^*) + \langle \nabla_x \mathcal{L}(x^*, \lambda^*, \eta^*), x - x^* \rangle \\ = \quad & f(x^*). \end{aligned}$$

Proof continued

Or, equivalently,

$$
\begin{aligned}
& f(x) + \sum_{i=1}^{q} \eta_i^* c_i(x) \\
\geq \quad & f(x^*) + \sum_{i=1}^{q} \eta_i^* c_i(x^*) + \langle \nabla_x \mathcal{L}(x^*, \lambda^*, \eta^*), x - x^* \rangle \\
= \quad & f(x^*).
\end{aligned}
$$

As a result, for all feasible x, it holds that

$$
f(x) \geq f(x^*) - \sum_{i=1}^{q} \eta_i^* c_i(x) \geq f(x^*).
$$