SI152: Numerical Optimization

Lecture 13: KKT and CQ

Hao Wang
Email: haw309@gmail.com

ShanghaiTech University

November 11, 2025

Outline

Recall the following optimality for convex set constraint.

$$\min_x \quad f(x) \quad \text{s.t. } x \in \Omega,$$

Theorem 1

If x^* is a minimizer of $f \in \mathcal{C}$ in Ω, then

$$-\nabla f(x^*) \in \mathcal{N}_\Omega(x^*).$$

If f is convex, then it is also sufficient condition.

That is,

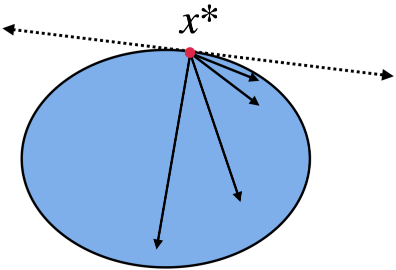
$$\nabla f(x^*)^T (y - x^*) \geq 0, \quad \forall y \in \Omega.$$

What are those directions $d = y - x^*$?

They are called the tangent directions which forms the Tangent Cone: $\mathcal{T}_\Omega(x^*)$.

That is, the optimality condition becomes:

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



What about nonconvex sets?

Definition 2 (Tangent direction)

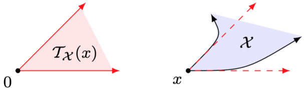
A direction $d \in \mathbb{R}^n$ is tangent to $\Omega \subset \mathbb{R}^n$ at a point $x \in \Omega$ if there exists a sequence of points $\{x_k\} \in \Omega$ and positive scalars $\{\tau_k\}$ such that

$$0 = \lim_{k\to\infty} \tau_k \quad \text{and} \quad d = \lim_{k\to\infty} \frac{1}{\tau_k}(x_k - x).$$

Definition 3 (Tangent cone)

The tangent cone corresponding to a set $\Omega \subset \mathbb{R}^n$ at $x \in \Omega$ is

$$\mathcal{T}_\Omega(x) := \{d \mid d \text{ is tangent to } \Omega \text{ at } x\}.$$



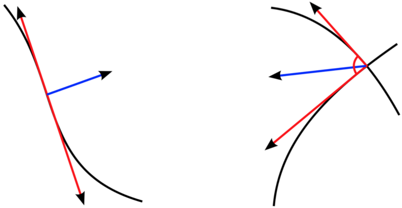
One can verify that for any $\Omega \subset \mathbb{R}^n$ and $x \in \Omega$, the set $\mathcal{T}_\Omega(x)$ is a closed cone.

Theorem 4 (Geometric Necessary Optimality Condition)

If x^* is a minimizer of $f \in \mathcal{C}$ in Ω, then

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

That is, if x_* is a minimizer, then there is no d that is both a descent direction for f at x_* and tangent to Ω at x_*. (Blue vector denotes $\nabla f(x_*)$.)



In the unconstrained case, $\Omega = \mathbb{R}^n$ and $\mathcal{T}_\Omega(x_*) = \mathbb{R}^n$.

Proof.

We proceed by contradiction. Suppose there exists a direction d in the tangent cone such that $\nabla f(x^*)^T d < 0$. Let $\{z^k\}$ and $\{t_k\}$ be the sequences referred to in the definition of the tangent cone. Then we have

$$\begin{aligned} f(z^k) &= f(x^*) + (z^k - x^*)^T \nabla f(x^*) + o(\|z^k - x^*\|) \\ &= f(x^*) + t_k \nabla f(x^*)^T d + o(t_k). \end{aligned}$$

Since $\nabla f(x^*)^T d < 0$, the first-order term eventually dominates the remainder term, so for sufficiently small t_k,

$$f(z^k) < f(x^*) + \frac{1}{2} t_k \nabla f(x^*)^T d, \quad \forall \text{ sufficiently large } k.$$

Thus, in any neighborhood of x^*, we can find feasible points where the objective function value is strictly less than $f(x^*)$, contradicting the assumption that x^* is a local minimizer. $\square$

Outline

Constrained optimization problem as

$$\begin{aligned} \min_{x} \ & f(x) \\ \text{s.t.} \ & c_i(x) = 0, i \in \mathcal{E}, \\ & c_i(x) \leq 0, i \in \mathcal{I}. \end{aligned}$$

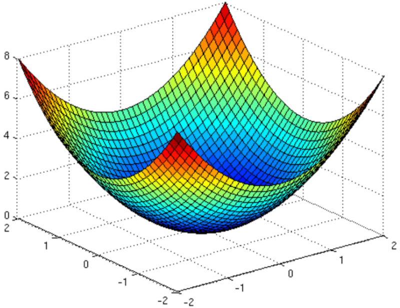
- It is difficult to describe $\mathcal{T}_\Omega$ with

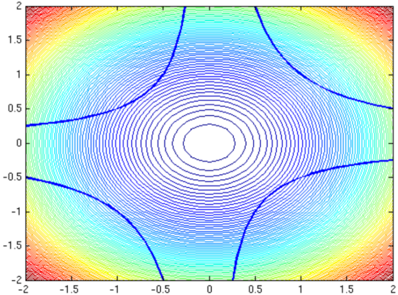
$$\Omega := \{x \in \mathbb{R}^n \mid c_i(x) = 0, i \in \mathcal{E}; c_i(x) \leq 0, i \in \mathcal{I}\}.$$

Constraints can lead to difficulty

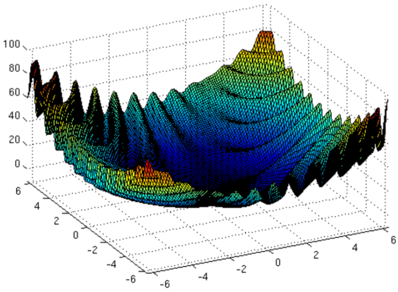
$$\min_{x} \; f(x) = x_1^2 + x_2^2$$

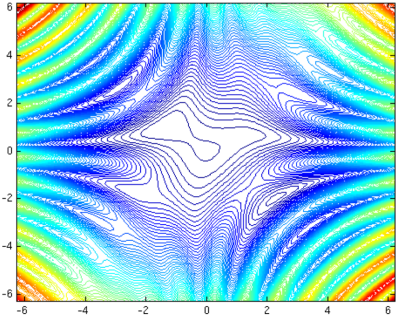
$$\text{s.t. } c(x) = \begin{Bmatrix} 1 - x_1x_2, \text{ if } x_1x_2 \geq 0 \\ 2x_1x_2 + 1, \text{ if } x_1x_2 < 0 \end{Bmatrix} \leq 0$$



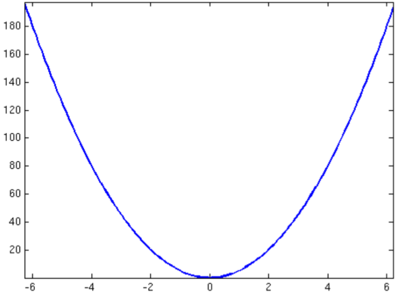


$$\min_x \ f(x) = (x_2 - 2x_1)\sin(x_1 x_2) + x_1^2 + x_2^2$$
$$\text{s.t. } c(x) = 2x_1 - x_2 = 0$$





$$\min_{x} \; f(x) = (x_2 - 2x_1)\sin(x_1 x_2) + x_1^2 + x_2^2$$
$$\text{s.t. } c(x) = 2x_1 - x_2 = 0$$

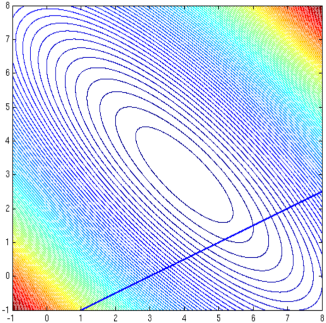


(f on the line $2x_1 = x_2$)

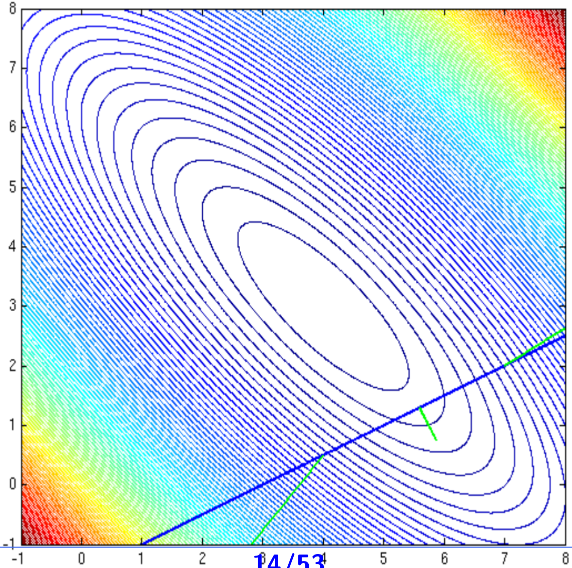
Consider the following problem with a single equality constraint:

$$\min_x f(x) = -\begin{bmatrix}32\\31\end{bmatrix}^T \begin{bmatrix}x_1\\x_2\end{bmatrix} + \frac{1}{2}\begin{bmatrix}32\\31\end{bmatrix}^T \begin{bmatrix}x_1\\x_2\end{bmatrix}^T \begin{bmatrix}5 & 4\\4 & 5\end{bmatrix}\begin{bmatrix}32\\31\end{bmatrix}$$
$$\text{s.t. } c(x) = -x_1 + 2x_2 + 3 = 0.$$

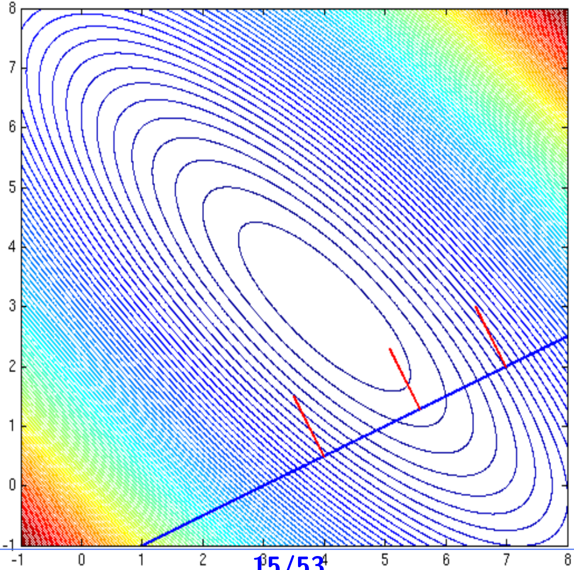
The solution is clearly at $x_* \approx (5.6, 1.3)$



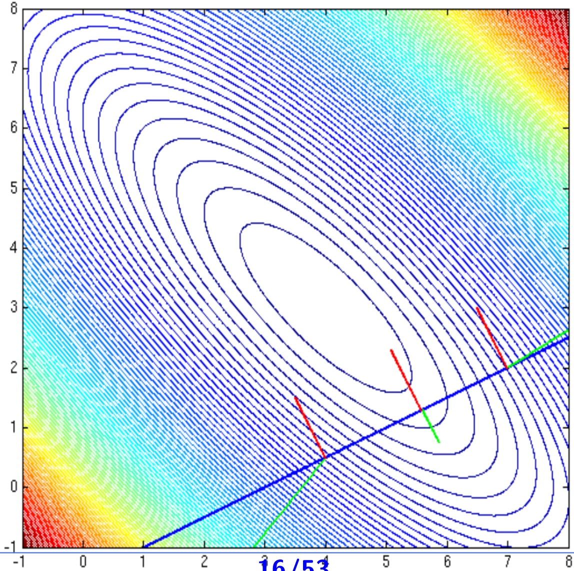
Clearly, for this problem, $\nabla f(x) \neq 0$ at all feasible x



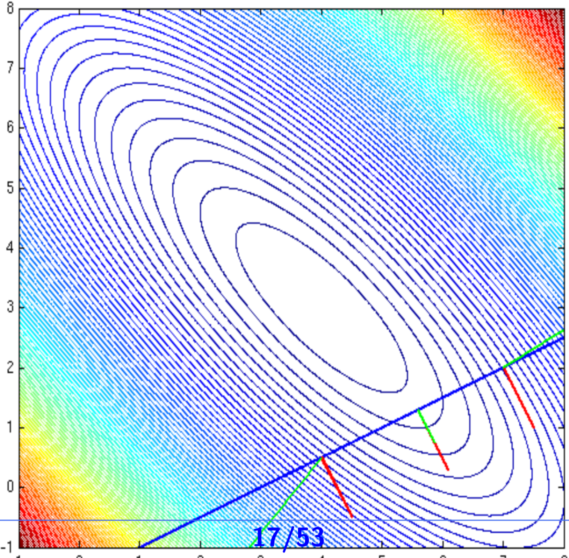
Observe gradients of the constraint at x satisfying $c(x) = 0$



Notice that the gradients "line up" at the solution x_*:



The picture is similar if the constraint is (equivalently) written as $x_1 - 2x_2 - 3 = 0$:



The optimal solution x_* exists at a feasible point where, for some $\lambda \in \mathbb{R}$,

$$\nabla f(x) + \nabla c(x)\lambda = 0.$$

Thus, defining the Lagrangian with Lagrange multiplier $\lambda \in \mathbb{R}$

$$L(x,\lambda) := f(x) + \lambda c(x),$$

the optimality conditions for our problem are

$$\begin{aligned} \nabla_x L(x,\lambda) = \nabla f(x) + \nabla c(x)\lambda &= 0 \\ c(x) &= 0. \end{aligned}$$

In particular, for our problem, this means

$$\begin{bmatrix} 5 & 4 & -1 \\ 4 & 5 & 2 \\ -1 & 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \lambda \end{bmatrix} = - \begin{bmatrix} -32 \\ -31 \\ 3 \end{bmatrix}$$

Suppose we are at a feasible point x and consider a small displacement along d:

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

Taylor's theorem suggests that we remain feasible if

$$\nabla c(x)^T d = 0.$$

(In fact, for our example, we do indeed remain feasible along such directions.) An alternative way to state that we are at a solution is if there is no d such that

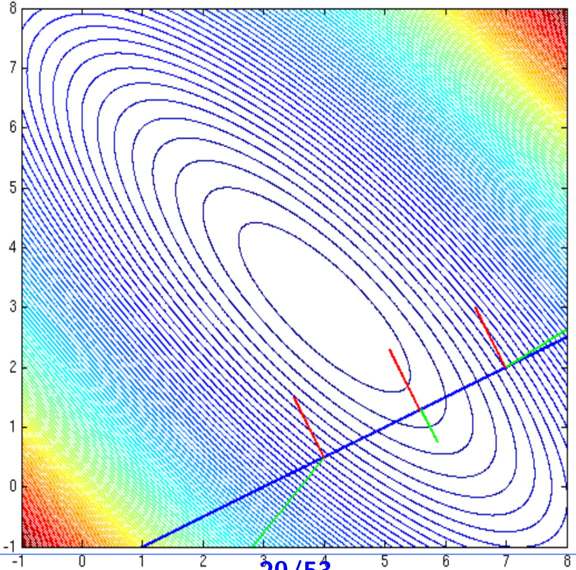
$$\nabla c(x)^T d = 0 \quad \text{and} \quad \nabla f(x)^T d < 0.$$

Indeed, Farkas's theorem implies that only one of the following has a solution:

System 1: $\nabla f(x) + \nabla c(x)\lambda = 0$ for some $\lambda \in \mathbb{R}$;

System 2: $\nabla c(x)^T d = 0$ and $\nabla f(x)^T d < 0$ for some $d \in \mathbb{R}^n$.

Only at x_* does $\nabla c(x)^T d = 0$ imply $\nabla f(x)^T d = 0$:



Consider the following problem:

$$\begin{aligned}\min_x \; & f(x) = x_1 + x_2 \\ \text{s.t. } & c(x) = x_1^2 + x_2^2 - 2 = 0.\end{aligned}$$

The Lagrangian is

$$L(x,\lambda) = f(x) + \lambda c(x) = x_1 + x_2 + \lambda(x_1^2 + x_2^2 - 2),$$

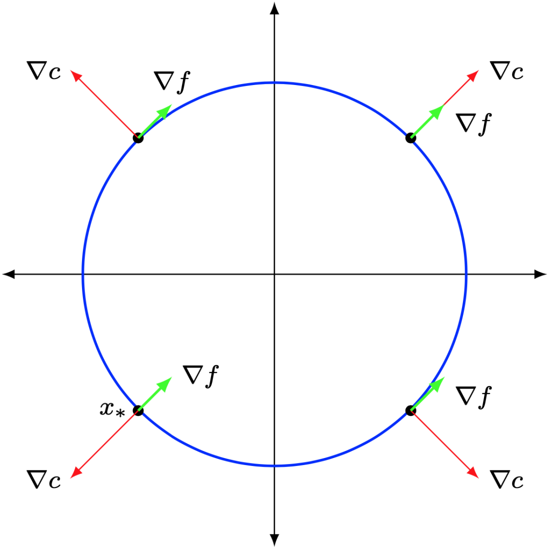
and so the first-order optimality conditions are

$$\begin{aligned}\nabla_x L(x,\lambda) = \begin{bmatrix}1\\1\end{bmatrix} + \lambda \begin{bmatrix}2x_1\\2x_2\end{bmatrix} &= 0\\ x_1^2 + x_2^2 - 2 &= 0\end{aligned}$$

There are two solutions!

$$x = (1,1), \lambda = -\frac{1}{2} \text{ and } x = (-1,-1), \lambda = -\frac{1}{2}.$$

Why only first-order? $x = (1, 1)$ is a maximizer!



Consider the following problem:

$$\min_x \; x$$
$$\text{s.t. } x^2 = 0.$$

The solution is obviously $x_* = 0$. However, defining the Lagrangian

$$L(x, \lambda) = x + \lambda x^2.$$

the "optimality conditions"

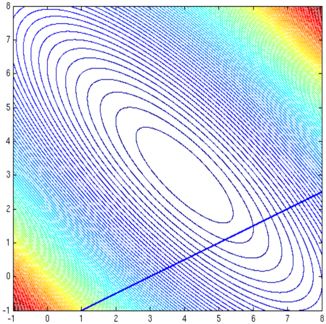
$$\begin{aligned} 1 + 2\lambda x &= 0 \\ x^2 &= 0 \end{aligned}$$

have no solution!

Consider the following problem with a single inequality constraint:

$$\min_x \; f(x) = -\begin{bmatrix}32\\31\end{bmatrix}^T \begin{bmatrix}x_1\\x_2\end{bmatrix} + \frac{1}{2}\begin{bmatrix}32\\31\end{bmatrix}^T \begin{bmatrix}x_1\\x_2\end{bmatrix}^T \begin{bmatrix}5 & 4\\4 & 5\end{bmatrix}\begin{bmatrix}32\\31\end{bmatrix}$$
$$\text{s.t. } c(x) = x_1 - 2x_2 - 3 \le 0.$$

The solution is clearly at $x_* \approx (4, 3)$ where $\nabla f(x) = 0$ and $c(x) < 0$.

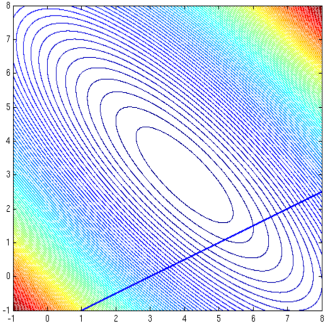


Consider the following problem with a single inequality constraint:

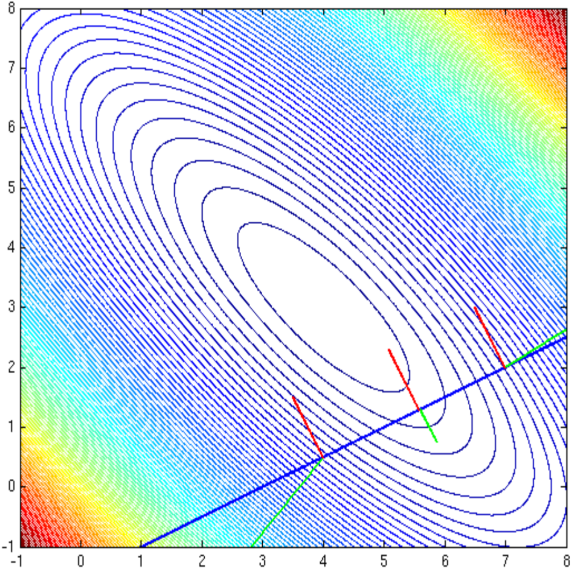
$$\min_x \; f(x) = -\begin{bmatrix} 32 \\ 31 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 32 \\ 31 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} 32 \\ 31 \end{bmatrix}$$

$$\text{s.t. } c(x) = -x_1 + 2x_2 + 3 \leq 0.$$

The solution is clearly at $x_* \approx (5.6, 1.3)$.



Objective gradients on the constraint



Optimality conditions

The optimal solution x_* exists at a strictly feasible, then, necessarily, x_* yields

$$\nabla f(x) = 0.$$

If the constraint is active at x_*, then for some $\lambda \geq 0$,

$$\nabla_x L(x, \lambda) = \nabla f(x) + \nabla c(x)\lambda = 0.$$

Considering both cases together, the optimality conditions are

$$\begin{aligned} \nabla_x L(x, \lambda) = \nabla f(x) + \nabla c(x)\lambda &= 0 \\ c(x) &\leq 0 \\ \lambda &\geq 0 \\ \lambda c(x) &= 0. \end{aligned}$$

In particular, for our problem, this means

$$\begin{aligned} \begin{bmatrix} -32 \\ -31 \end{bmatrix} + \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \lambda \begin{bmatrix} -1 \\ 2 \end{bmatrix} &= 0 \\ -x_1 + 2x_2 + 3 &\leq 0 \\ \lambda &\geq 0 \end{aligned}$$

Suppose we are at a feasible point x and consider a small displacement along d.

- If $c(x) < 0$, then any small displacement means we remain feasible.
- If $c(x) = 0$, then $\nabla c(x)^T d \leq 0$ (and Taylor's theorem) means we stay feasible.

An alternative way to state that we are at a solution is if there is no d such that

$$\begin{cases} \nabla f(x)^T d < 0 & \text{if } c(x) < 0; \\ \nabla f(x)^T d < 0 \text{ and } \nabla c(x)^T d \leq 0 & \text{if } c(x) = 0. \end{cases}$$

Indeed, Farkas's theorem implies that only one of the following has a solution:

System 1: $\nabla f(x) + \nabla c(x)\lambda = 0$ for some $\lambda \geq 0$;

System 2: $\nabla c(x)^T d \leq 0$ and $\nabla f(x)^T d < 0$ for some $d \in \mathbb{R}^n$.

Consider the following problem:

$$\min_{x} \; f(x) = x_1 + x_2$$
$$\text{s.t. } c(x) = x_1^2 + x_2^2 - 2 \le 0.$$

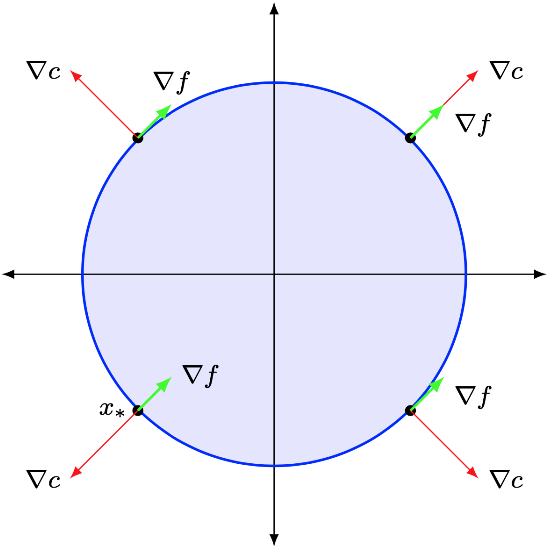
The Lagrangian is

$$L(x, \lambda) = f(x) + \lambda c(x) = x_1 + x_2 + \lambda(x_1^2 + x_2^2 - 2),$$

and so the first-order optimality conditions are

$$\begin{aligned} \nabla_x L(x, \lambda) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \lambda \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} &= 0 \\ x_1^2 + x_2^2 - 2 &\le 0 \\ \lambda &\ge 0 \\ \lambda(x_1^2 + x_2^2 - 2) &= 0 \end{aligned}$$

What is(are) the solution(s)!


∇c
∇f
∇c
∇f
∇f
x_*
∇f
∇c
∇c

Consider the following problem:

$$\min_x x \quad \text{s.t. } x^2 \leq 0.$$

The solution is obviously $x_* = 0$. However, defining the Lagrangian

$$L(x, \lambda) = x + \lambda x^2.$$

the "optimality conditions"

$$\begin{aligned} 1 + 2\lambda x &= 0 \\ x^2 &\leq 0 \\ \lambda &\geq 0 \\ \lambda x^2 &= 0 \end{aligned}$$

have no solution!

Constrained optimization problem as

$$\begin{aligned} \min_x \ & f(x) \\ \text{s.t. } & c_i(x) = 0, i \in \mathcal{E}, \\ & c_i(x) \leq 0, i \in \mathcal{I}. \end{aligned}$$

It looks under some conditions (remember the counter-examples), an optimal solution satisfy the Karush-Kuhn-Tucker conditions (KKT conditions):

$$\begin{aligned} \nabla_x L(x, \lambda) = \nabla f(x) + \sum_{i \in \mathcal{E} \cup \mathcal{I}} \nabla c_i(x) \lambda = 0 \\ c_i(x) = 0, i \in \mathcal{E} \\ c_i(x) \leq 0, i \in \mathcal{I} \\ \lambda_i \geq 0, i \in \mathcal{I} \\ \lambda_i c_i(x) = 0, i \in \mathcal{I}. \end{aligned}$$

Equivalently, under some conditions (remember the counter-examples), an optimal solution is feasible and satisfy

$$\nabla_x L(x, \lambda) = \nabla f(x) + \sum_{i \in \mathcal{E} \cup \mathcal{A}(x)} \nabla c_i(x) \lambda = 0$$

$$\lambda_i \geq 0, i \in \mathcal{A}(x)$$

Farkas's Theorem implies that there is no solution d in

$$\mathcal{F}_\Omega(x) := \{d \in \mathbb{R}^n \mid \nabla c_i(x)^T d \leq 0, i \in \mathcal{A}(x); \nabla c_i(x)^T d = 0, i \in \mathcal{E}\}$$

Equivalently, under some conditions, an optimal solution is feasible and satisfy

$$\nabla f(x)^T d \geq 0 \quad \forall d \in \mathcal{F}_\Omega(x)$$

Equivalently, under some conditions, an optimal solution is feasible and satisfy

$$\nabla f(x)^T d \geq 0 \quad \forall d \in \mathcal{F}_\Omega(x)$$

In other words, $\mathcal{F}_\Omega(x)$ has no descent direction!
Previously, we know that an optimal solution is feasible and satisfy

$$\nabla f(x)^T d \geq 0 \quad \forall d \in \mathcal{T}_\Omega(x)$$

In other words, $\mathcal{T}_\Omega(x)$ has no descent direction!

- $\mathcal{T}_\Omega(x)$ is good, but hard to describe.
- $\mathcal{F}_\Omega(x)$ has simpler description, we don't need to verify all $d \in \mathcal{F}_\Omega(x)$, we only need to verify its alternatively equivalent form, the KKT conditions!
- Generally, which one of the following is correct? (Lemma 12.2, textbook)

$$\mathcal{T}_\Omega(x) \subset \mathcal{F}_\Omega(x) \quad \text{or} \quad \mathcal{F}_\Omega(x) \subset \mathcal{T}_\Omega(x)$$

- It is ideal if we have a problem with $\mathcal{T}_\Omega(x) = \mathcal{F}_\Omega(x)$

$$\begin{aligned} \min_x \ & f(x) \\ \text{s.t. } & c_i(x) = 0, i \in \mathcal{E}, \\ & c_i(x) \le 0, i \in \mathcal{I}. \end{aligned}$$

Conditions that guarantee $\mathcal{T}_\Omega(x) = \mathcal{F}_\Omega(x)$ are called constraint qualification (CQ). Without a CQ, we may even don't have the multipliers.

Definition 5 (LICQ)

The linear independence constraint qualification (LICQ) is said to hold at x if the set of active constraint gradients

$$\{\nabla c_i(x) \mid i \in \mathcal{E} \cup \mathcal{A}(x)\}$$

is linearly independent at x.

What is the geometrical intuition of LICQ?

Theorem 6 (KKT points under the LICQ)

Suppose x_* is a local solution at which the LICQ holds. Then, there exists Lagrange multipliers λ_* such that (x_*, λ_*) satisfies the following:

$$\nabla_x L(x, \lambda) = \nabla f(x) + \sum_{i \in \mathcal{E} \cup \mathcal{I}} \nabla c_i(x) \lambda = 0$$

$$c_i(x) = 0, i \in \mathcal{E}$$

$$c_i(x) \leq 0, i \in \mathcal{I}$$

$$\lambda_i \geq 0, i \in \mathcal{I}$$

$$\lambda_i c_i(x) = 0, i \in \mathcal{I}.$$

LICQ implies the uniqueness of λ_*!

- If all constraints that are active at a solution are affine, then no other qualification is needed and x_* is necessarily a KKT point. (recall the optimality conditions for LP)
- Therefore, Lagrange multipliers always exist for linear optimization problems
- Quadratic problems are also safe, i.e.,

$$\min\ c^T x + \frac{1}{2} x^T Q x$$
$$\text{s.t. } Ax = b, x \geq 0$$

- LICQ does not imply all constraints are affine. Affine constraints do not imply the LICQ

Suppose our problem is convex (i.e., affine $c_i, i \in \mathcal{E}$ and convex $c_i, i \in \mathcal{I}$)

Definition 7 (Weak Slater condition)

The weak Slater condition is satisfied if there exists a feasible point strictly satisfying all non-affine inequalities, i.e., $\exists x$ such that

$$c_i(x) = 0, i \in \mathcal{E}; c_i(x) \leq 0, i \in \mathcal{I} \text{ (affine)}; c_i(x) < 0, i \in \mathcal{I} \text{ (non-affine)}.$$

Definition 8 (Strong Slater condition)

The strong Slater condition is satisfied if

$$\nabla c_i(x), i \in \mathcal{E} \text{ are linearly independent}$$

and there exists a feasible point strictly satisfying all inequalities, i.e., $\exists x$ such that

$$c_i(x) = 0, i \in \mathcal{E}; c_i(x) < 0, i \in \mathcal{I}.$$

- The weak condition implies the existence of a nonempty, closed, convex set Λ_* such that for all $\lambda_* \in \Lambda_*$, the point (x_*, λ_*) satisfies the KKT conditions.
- The strong condition implies the existence of such a Λ_* that is bounded.

Definition 9

The Mangasarian-Fromovitz constraint qualification (MFCQ) holds at x_* if

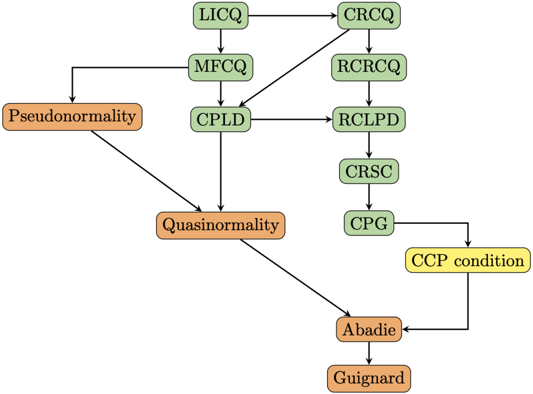
$$\nabla c_i(x_*), \quad i \in \mathcal{E} \text{ are linearly independent}$$

and there exists d such that

$$\begin{aligned} \nabla c_i(x_*)^T d &= 0, i \in \mathcal{E} \\ \nabla c_i(x_*)^T d &< 0, i \in \mathcal{A}(x_*). \end{aligned}$$

- This is an extension of the strong Slater condition to nonconvex problems
- It implies the existence of a nonempty, closed set Λ_* such that for all $\lambda_* \in \Lambda_*$, the point (x_*, λ_*) satisfies the KKT conditions.
- The LICQ $\implies$ the MFCQ.

Other constraint qualifications



Outline

Constrained optimization problem as

$$\begin{aligned} \min_x \ & f(x) \\ \text{s.t. } & c_i(x) = 0, i \in \mathcal{E}, \\ & c_i(x) \leq 0, i \in \mathcal{I}. \end{aligned}$$

Let $\Lambda = \{\lambda : \lambda_i \geq 0, i \in \mathcal{I}\}$. The primal function as

$$L_P(x) := \sup_{\lambda \in \Lambda} L(x, \lambda)$$

and the dual function as

$$L_D(\lambda) := \inf_x L(x, \lambda)$$

The primal problem is to find

$$\min_x L_P(x) = \min_x \sup_{\lambda \in \Lambda} L(x, \lambda)$$

The same argument as in LP case, we have

$$L_P(x) = \begin{cases} f(x) & \text{if } x \text{ is feasible} \\ \infty & \text{otherwise} \end{cases}$$

The dual problem is to find

$$\max_{\lambda \in \Lambda} L_D(\lambda) := \max_{\lambda \in \Lambda} \inf_x L(x, \lambda) = f(x) + \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x)$$

The dual function $L_D(\lambda)$ is always concave (why?), so that the dual problem is a convex problem!

Weak duality still holds for nonlinear cases.

Theorem 10 (Weak duality)

For every x and $\lambda \in \Lambda$, we have

$$L_D(\lambda) \le L_P(x).$$

There is generally no guarantee for strong duality!

Definition 11 (Saddle points)

A point (x_*, λ_*) with $\lambda_* \in \Lambda$ is called a saddle point of the Lagrangian if for all $x \in \mathbb{R}^n$ and $\lambda \in \Lambda$ we have

$$L(x_*, \lambda) \leq L(x_*, \lambda_*) \leq L(x, \lambda_*)$$

Theorem 12

If the Lagrangian has a saddle point (x_*, λ_*), then x_* is a solution of the primal problem, $\lambda_* \in \Lambda$ is a solution of the dual problem, and the following duality holds

$$\min_x L_P(x) = \max_{\lambda \in \Lambda} L_D(\lambda).$$

But, generally,

saddle points $\not\Leftrightarrow$ KKT points

For nonconvex cases,

- A saddle of the Lagrangian doesn't mean the exists of KKT points (CQ needed).
- KKT points may not be a saddle (may be local optimal, could be a saddle of the primal, or a maximum)

For convex cases,

Theorem 13

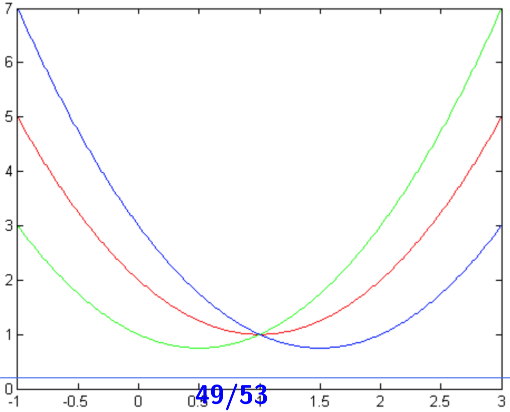
If f is convex, $c_i, i \in \mathcal{E}$ are affine, and $c_i, i \in \mathcal{I}$ are convex, then a point x_* satisfies the KKT conditions with $\lambda_* \in \Lambda$ if and only if (x_*, λ_*) is a saddle point of the Lagrangian.

A constraint qualification guaranteeing the existence of Lagrange multipliers for the KKT conditions guarantees a duality relationship as well.

$$\min_x x^2 \quad \text{s.t. } 1 - x \leq 0$$

$$L(x, \lambda) = x^2 + \lambda(1 - x).$$

Plots of $L(x, \lambda)$ for $\lambda = 1, 2, 3$.



Outline

At a KKT point (x, λ), definite the critical cone

$$\mathcal{C}(x,\lambda) = \{d \in \mathcal{F}(x) \mid \nabla c_i(x)^T d = 0, i \in \mathcal{A}(x) \ \text{ with } \lambda_i > 0\}.$$

That is

$$d \in \mathcal{C}(x,\lambda) \iff \begin{cases} \nabla c_i(x)^T d = 0, & i \in \mathcal{E}, \\ \nabla c_i(x)^T d = 0, & i \in \mathcal{A}(x) \text{ with } \lambda_i > 0 \\ \nabla c_i(x)^T d \leq 0, & i \in \mathcal{A}(x) \text{ with } \lambda_i = 0. \end{cases}$$

It is clear that from the KKT conditions that for $d \in \mathcal{C}(x,\lambda)$ we have

$$d^T \nabla f(x) + d^T \nabla c(x) \lambda = 0 \implies d^T \nabla f(x) = 0,$$

so first-order information tells us nothing about how f changes along d.

Theorem 14

Suppose x_* is a local solution at which the LICQ holds, and let λ_* be the corresponding multipliers such that (x_*, λ_*) is a KKT point. Then,

$$d^T \nabla^2_{xx} L(x_*, \lambda_*) d \geq 0 \text{ for all } d \in \mathcal{C}(x_*, \lambda_*).$$

Here, $\nabla^2_{xx} L(x, \lambda)$ is the Hessian of the Lagrangian:

$$\nabla^2_{xx} L(x, \lambda) = \nabla^2 f(x) + \sum \lambda_i \nabla^2 c_i(x).$$

Theorem 15

Suppose x_* is a feasible point for which there is a Lagrange multiplier vector such that the KKT conditions hold. Suppose also that

$$d^T \nabla^2_{xx} L(x_*, \lambda_*) d > 0 \text{ for all } d \in \mathcal{C}(x_*, \lambda_*) \setminus \{0\}.$$

Then, x_* is a strict local solution.