Numerical Optimization

Lecture 15: Constrained Optimization：CQ and Optimality Conditions

王浩
信息科学与技术学院
Email: wanghao1@shanghaitech.edu.cn

Problem statement

We often want only to optimize f within a feasible set Ω:

$$\begin{aligned} &\min_x \ f(x) \\ &\text{subject to/such that (s.t.) } x \in \Omega. \end{aligned} \tag{P^*}$$

The set Ω may represent:

- the (effective) domain of f, or, more generally,
- restrictions on the variables in x.

Commonly, we pose this type of constrained optimization problem as

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

However, there are big differences between (P) and (P^*). (See next set of notes.)

Solution types in unconstrained optimization

Recall:

- $x_* \in \mathbb{R}^n$ is a global solution if

$$f(x_*) \leq f(x) \quad \forall x \in \mathbb{R}^n.$$

- $x_* \in \mathbb{R}^n$ is a local solution if there is a neighborhood $\mathcal{N}$ of x_* s.t.

$$f(x_*) \leq f(x) \quad \forall x \in \mathcal{N} \cap \mathbb{R}^n.$$

- $x_* \in \mathbb{R}^n$ is a strict/strong local solution if there is a neighborhood $\mathcal{N}$ of x_* s.t.

$$f(x_*) < f(x) \quad \forall x \in (\mathcal{N} \cap \mathbb{R}^n) \backslash x_*.$$

Solution types in constrained optimization

The statements do not change much for constrained optimization:

- $x_* \in \Omega$ is a global solution if

$$f(x_*) \leq f(x) \quad \forall x \in \Omega.$$

- $x_* \in \Omega$ is a local solution if there is a neighborhood $\mathcal{N}$ of x_* s.t.

$$f(x_*) \leq f(x) \quad \forall x \in \mathcal{N} \cap \Omega.$$

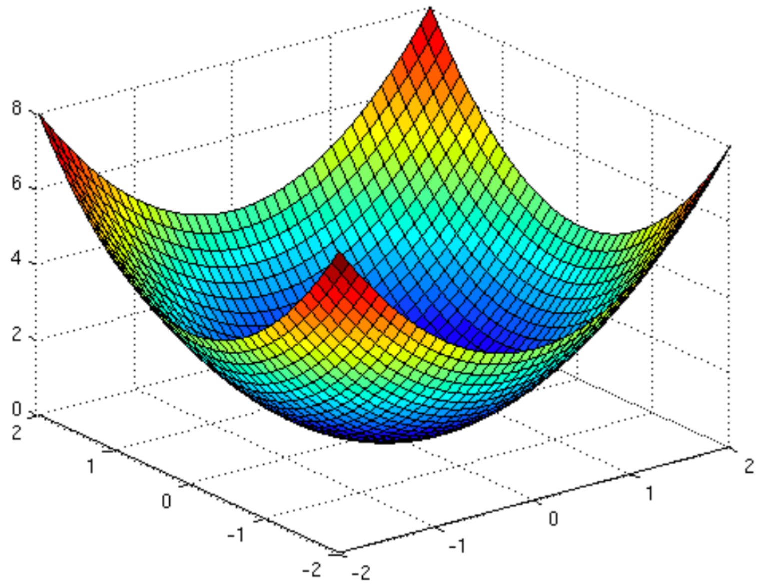
- $x_* \in \Omega$ is a strict/strong local solution if there is a neighborhood $\mathcal{N}$ of x_* s.t.

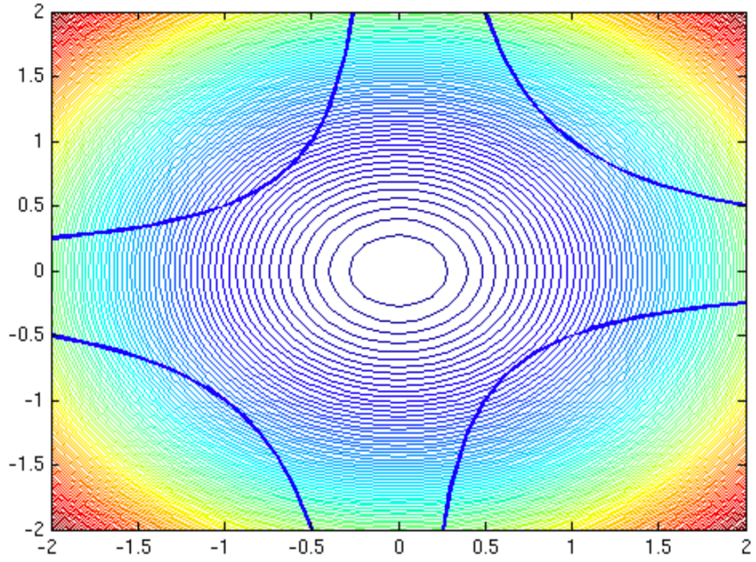
$$f(x_*) < f(x) \quad \forall x \in (\mathcal{N} \cap \Omega) \backslash x_*.$$

Constraints can lead to difficulty

$$\min_x f(x) = x_1^2 + x_2^2$$

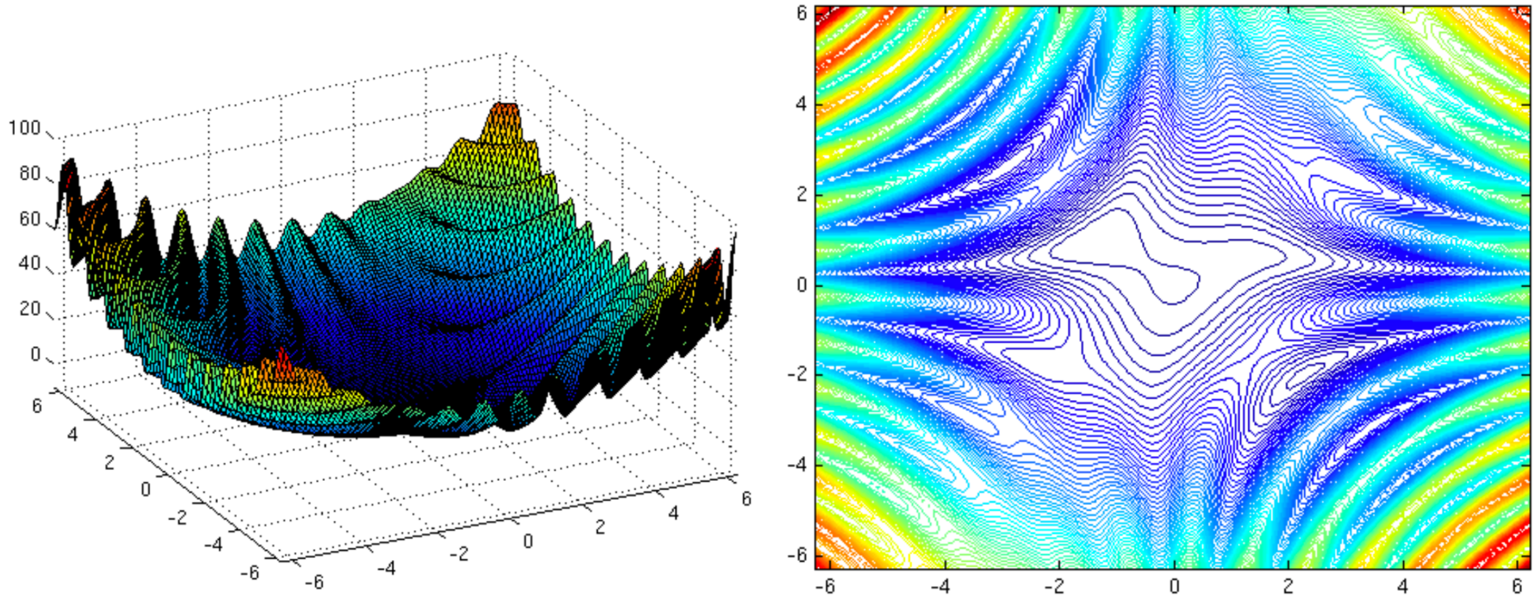
$$\text{s.t. } c(x) = \left\{ \begin{array}{r} 1 - x_1x_2, \text{ if } x_1x_2 \geq 0 \\ 2x_1x_2 + 1, \text{ if } x_1x_2 < 0 \end{array} \right\} \leq 0$$





Constraints can make a problem easier

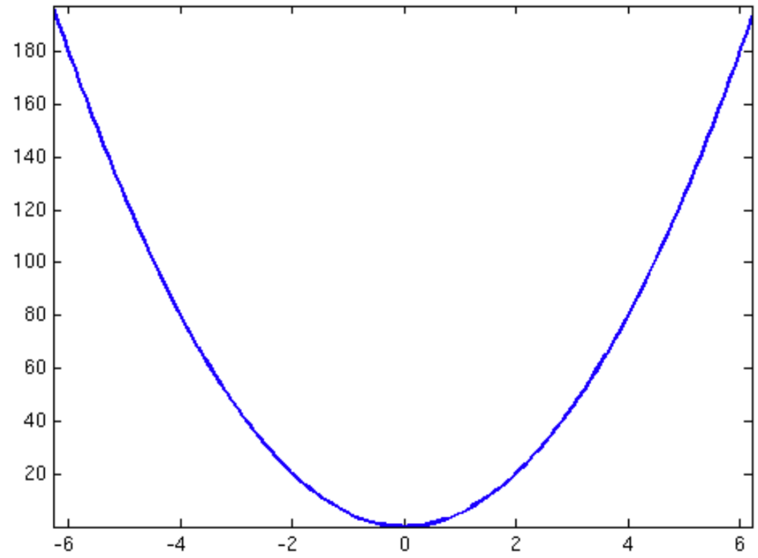
$$\min_{x}\ f(x) = (x_2 - 2x_1)\sin(x_1 x_2) + x_1^2 + x_2^2$$
$$\text{s.t. } c(x) = 2x_1 - x_2 = 0$$



$(f \text{ in } \mathbb{R}^2)$

Constraints can make a problem easier

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



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

Optimality conditions

We aim to derive optimality conditions for constrained problems.

- Necessary conditions are those that must be satisfied by any solution.
- Sufficient conditions at x_* are those that imply x_* is a solution.

Recall that for unconstrained optimization, the necessary conditions were

$$x_* \text{ is a minimizer} \implies \begin{cases} \nabla f(x_*) = 0 \\ \nabla^2 f(x_*) \succeq 0, \end{cases}$$

and sufficient conditions were

$$\left.\begin{aligned} \nabla f(x_*) &= 0 \\ \nabla^2 f(x_*) &\succ 0 \end{aligned}\right\} \implies x_* \text{ is a strict minimizer.}$$

Convex optimization

A constrained problem is convex if and only if Ω is convex; i.e., we need

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

Thus, a constrained problem is nonconvex if we have

- a nonconvex objective function f;
- a nonaffine equality constraint function c^i;
- a nonconvex inequality constraint function c^i.

Notice that this includes

- $c^i(x) = 0$ where c^i is nonlinear;
- $c^i(x) \geq 0$ where $c^i(x)$ is convex!

Convexity implies that local minimizers are global minimizers.

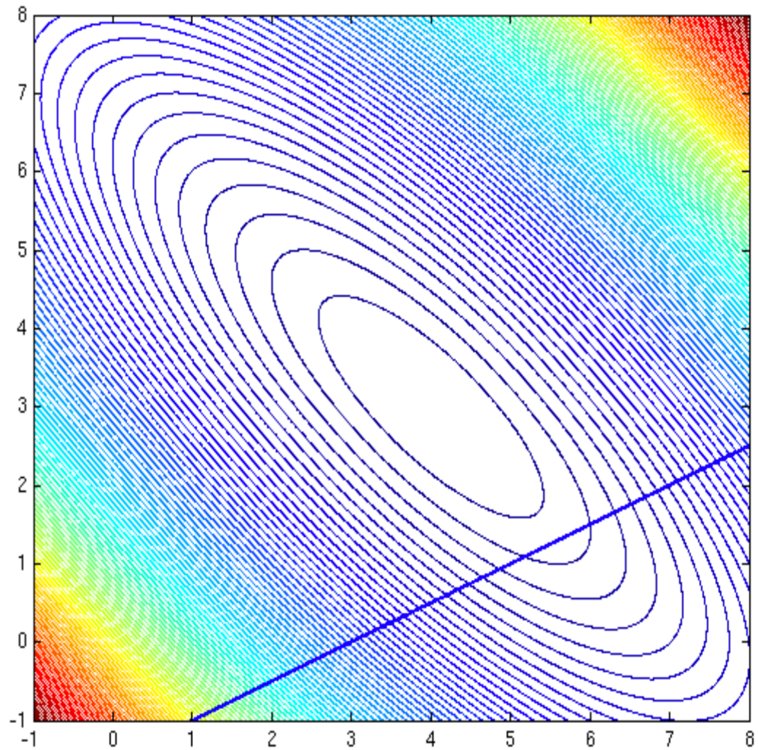
Single equality constraint

Consider the following problem with a single equality constraint:

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

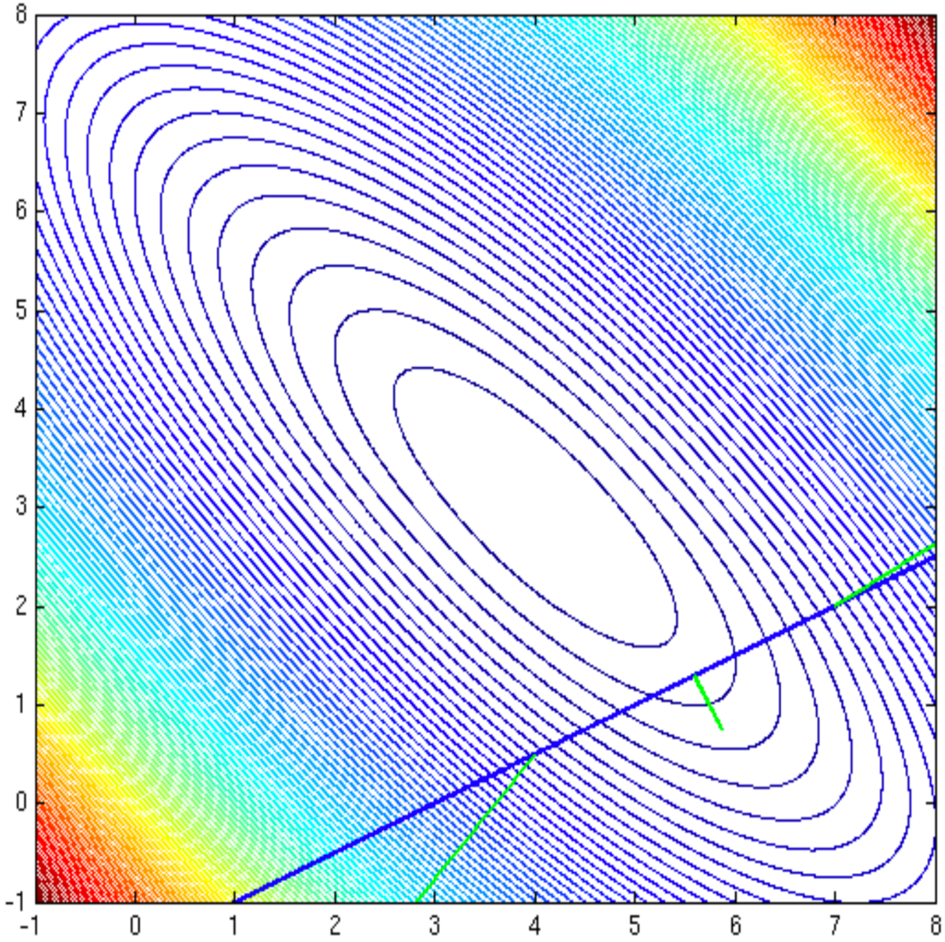
$$\text{s.t. } c(x) = -x_1 + 2x_2 + 3 = 0.$$

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



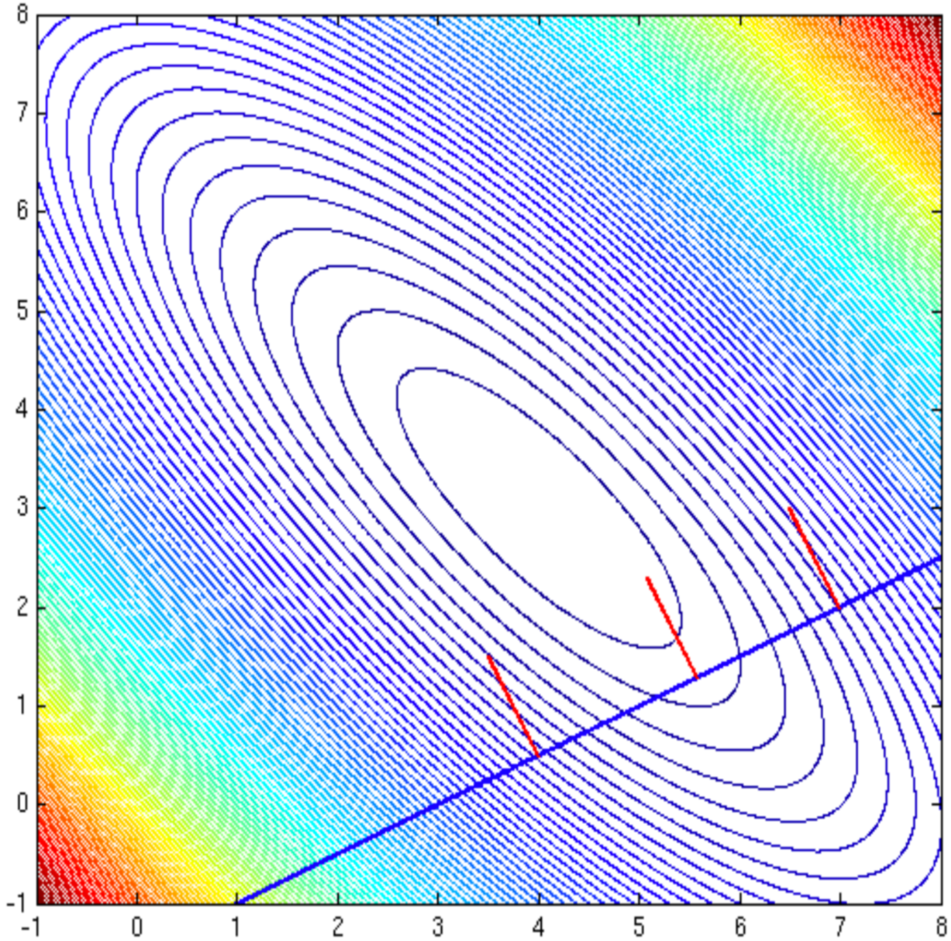
Objective gradients on the constraint

Clearly, for this problem, $\nabla f(x) \neq 0$ at all x satisfying $c(x) = 0$:



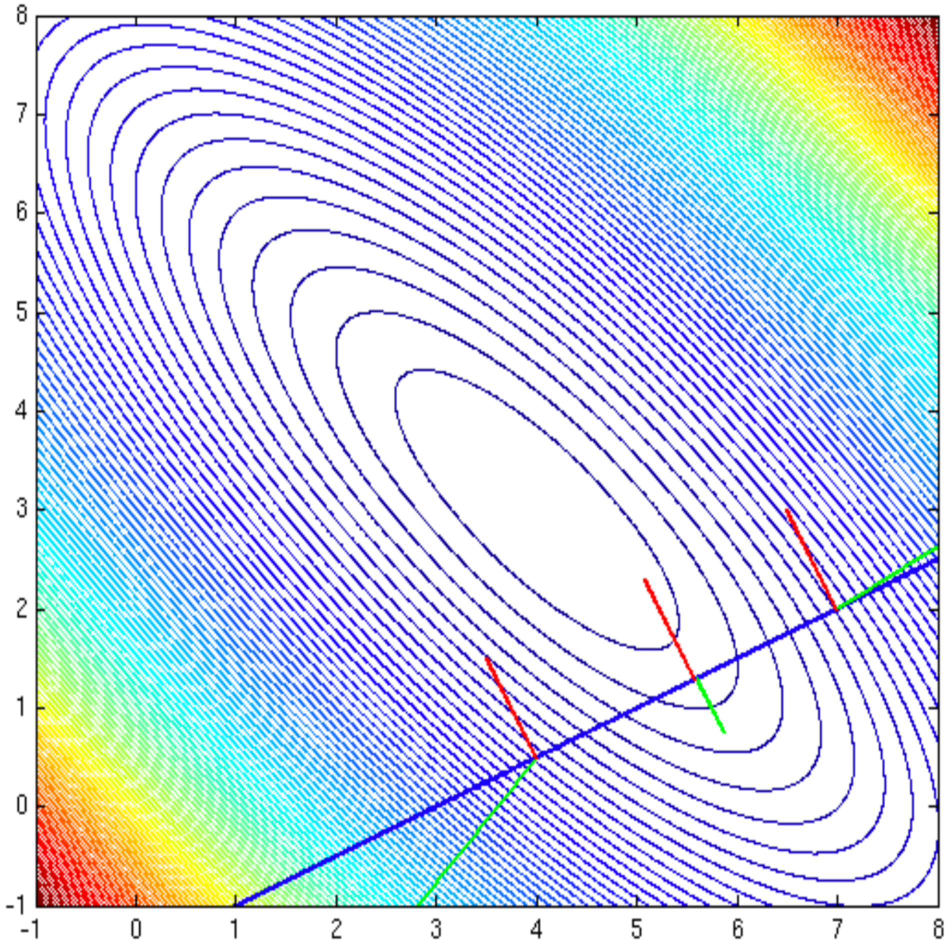
Constraint gradients on the constraint

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



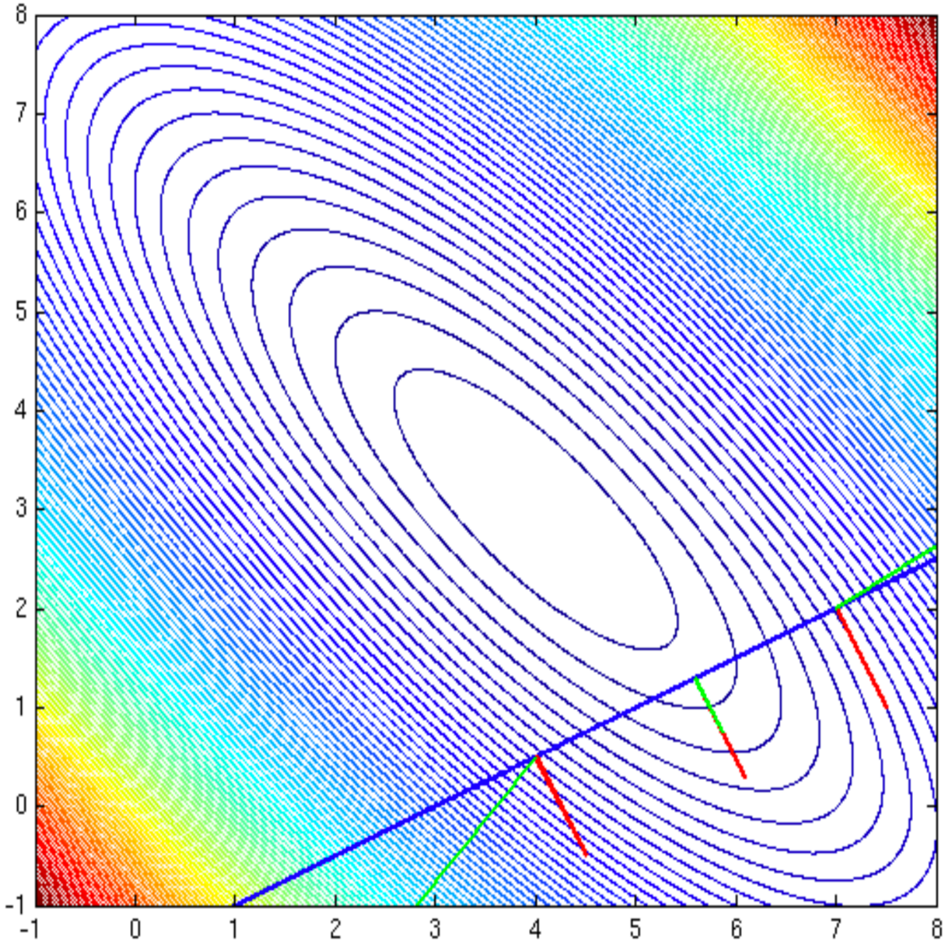
Objective and constraint gradients on the constraint

Notice that the gradients “line up” at the solution x_*:



Objective and constraint gradients on the constraint

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



Optimality conditions

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 $\lambda \in \mathbb{R}^n$

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

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

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

Single nonlinear equality constraint

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

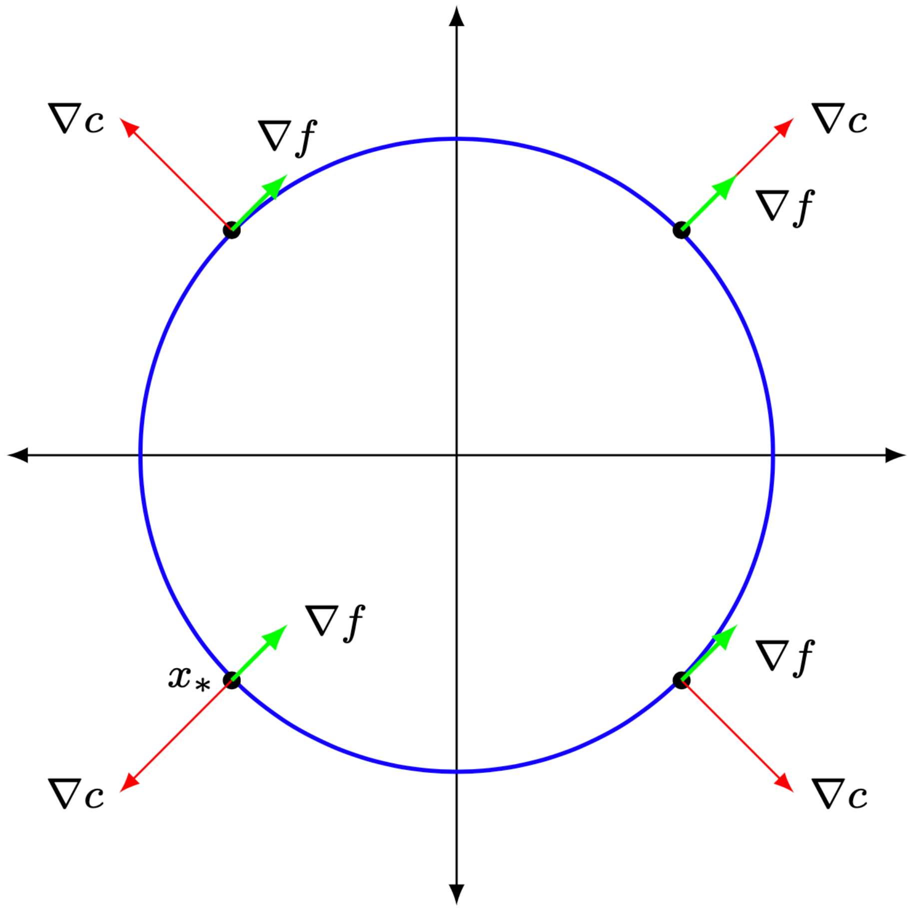
$$\begin{aligned} \nabla_x L(x, \lambda) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \lambda \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} =& \ 0 \\ x_1^2 + x_2^2 - 2 =& \ 0. \end{aligned}$$

There are two solutions!

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

Objective and constraint gradients　条件未必充分

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



A troubling example 条件未必会成立

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!

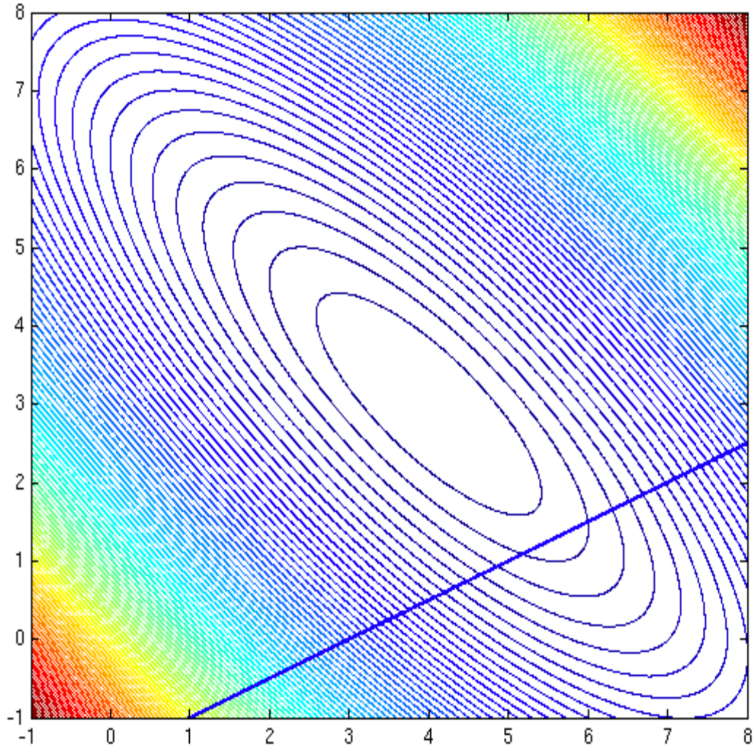
Single inequality constraint

Consider the following problem with a single inequality constraint:

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

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

Now, the solution is clearly at $x_* = (4, 3)$, i.e., where $\nabla f(x) = 0$ and $c(x) < 0$.



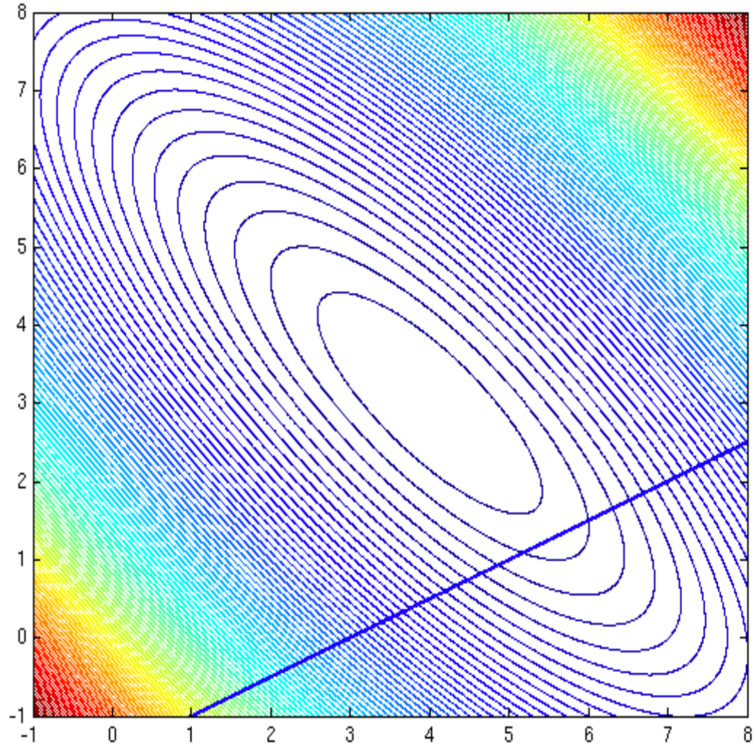
Single inequality constraint

Consider the following problem with a single inequality constraint:

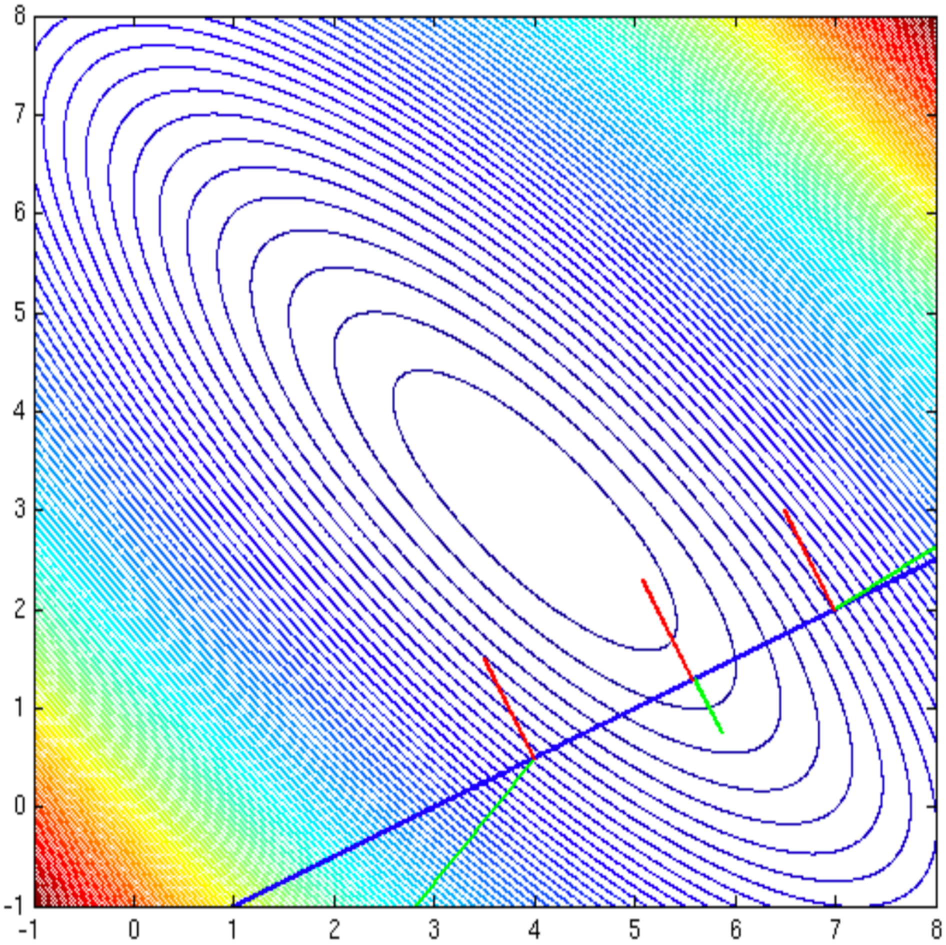
$$\min_x \; f(x) = -\begin{bmatrix} 32 \\ 31 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

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

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



Objective and constraint gradients



Optimality conditions

If the optimal solution x_* is 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

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

Single nonlinear inequality constraint

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

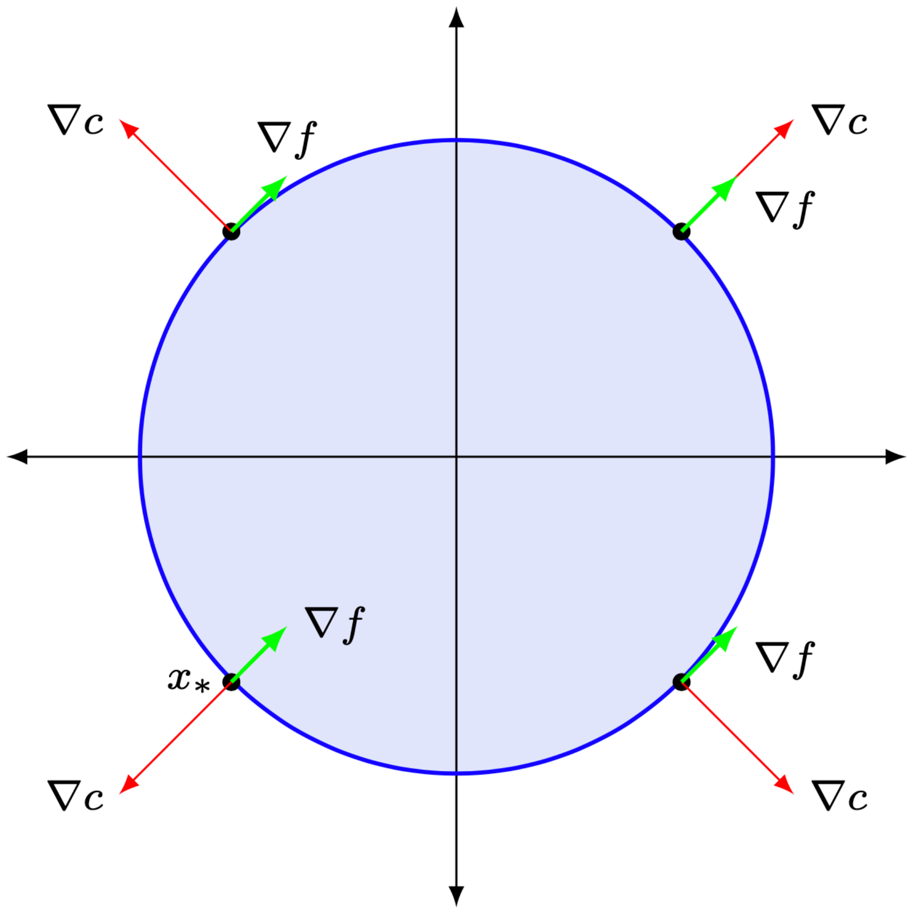
$$L(x, \lambda) = f(x) + \lambda c(x) = x_1 + x_2 + \lambda(x_1^2 + x_2^2 - 2),$$

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)?

Objective and constraint gradients



Discussion of problem

- If the constraint is inactive, then the optimality conditions require $\nabla f(x) = 0$ to be zero. However, $\nabla f(x) = (1,1)^T$ at all x, so there are no solutions to the optimality conditions with $c(x) < 0$. Alternatively, we can note that at any interior point, there exists $d = -\gamma \nabla f(x)$ with $\gamma > 0$ sufficient small such that $c(x+d) \leq 0$ and $\nabla f(x)^T d < 0$.
- If the constraint is active, then the optimality conditions require

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

for some $\lambda \geq 0$. Clearly, this can only occur at $x = (-1, -1)$ with $\lambda = \frac{1}{2}$.

Another troubling example

Consider the following problem:

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

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!

Karush-Kuhn-Tucker Conditions

Now let's turn our attention back to generally constrained problems:

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

Define the Lagrangian

$$L(x, \lambda) = f(x) + \lambda^T c(x).$$

Theorem (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:

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

Definition (LICQ)

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

$$\{\nabla c^i(x) : i \in \mathcal{E} \cup \mathcal{A}(x)\}$$

is linearly independent at x.

在势力场中具有理想约束的质点系的平衡条件是势能对于每个广义坐标的偏导数分别等于零。

对于一个自由度系统，系统具有一个广义坐标q，因此系统势能可以表示为q的一元函数 即$V=V(q)$
当系统平衡时，在平衡位置处有

$$\frac{\mathrm{d}V}{\mathrm{d}q}=0$$

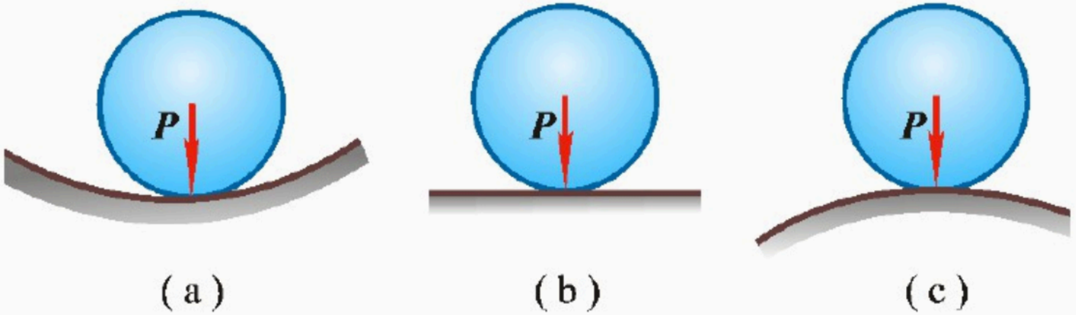
——带拉格朗日乘子的质点系动力学方程
第一类拉格朗日方程

——第二类拉格朗日方程

稳定平衡：在平衡位置处系统势能具有极小值。

不稳定平衡：在平衡位置上系统势能具有极大值。

随遇平衡：系统在某位置附近其势能是不变的。



如果系统处于稳定平衡状态，则在平衡位置处系统势能具有极小值。

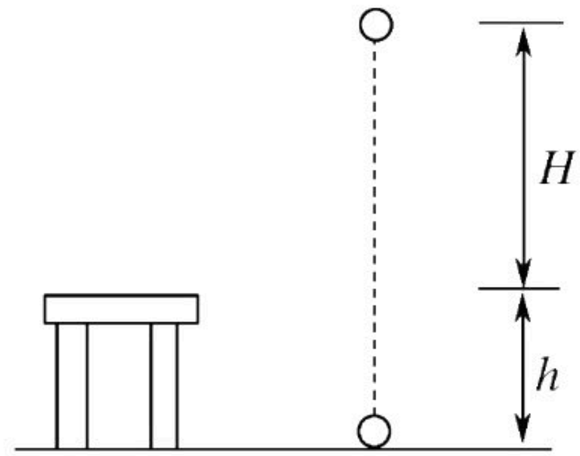
即系统势能对广义坐标的二阶导数大于零

$$\frac{d^2V}{dq^2} > 0$$

——一个自由度系统平衡的稳定性判据

$$\min_{H} \quad G(H) = mgH$$

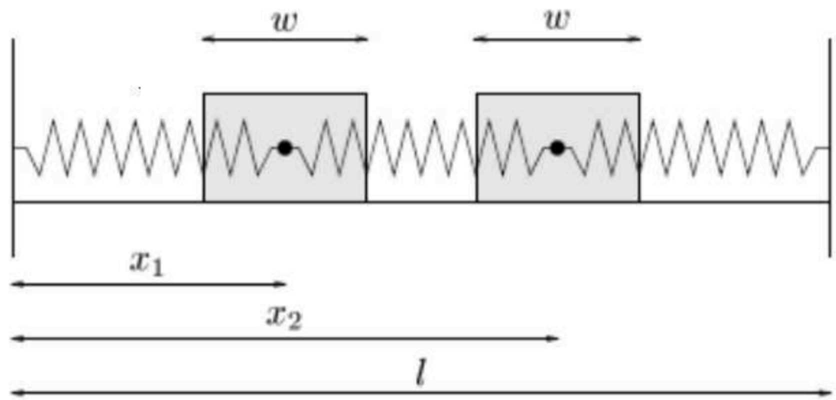
$$\text{s.t.} \quad H \geq h$$



$$\nabla G(H) = mg, \quad \nabla (H - h) = 1$$

重力： $g = -\nabla G(H) = -mg$，支撑力： $\lambda \cdot 1$

$$\boxed{\begin{array}{c} -mg + \lambda = 0 \\ H \geq h \quad \lambda \geq 0 \\ \lambda (H - h) = 0 \end{array}}$$



两个滑块以及左面和右面的墙壁用弹簧连在一起，滑块的宽为 $w > 0$ ，滑块不可能穿入墙内或者另一个滑块内.

弹簧的弹性势能可以表示成滑块位置的函数，即

$$f(x_1, x_2) = \frac{1}{2}k_1 x_1^2 + \frac{1}{2}k_2(x_2 - x_1)^2 + \frac{1}{2}k_3(l - x_2)^2$$

其中 $k_i > 0$ 分别是三个弹簧的弹性系数，那么势能函数满足的约束不等式为

$$x_1 - \frac{w}{2} \geq 0, \quad -x_1 + x_2 - w \geq 0, \quad -x_2 - \frac{w}{2} + l \geq 0.$$

系统的平衡位置 x^* 即为势能函数在上述约束条件下的最小值点.

$$\begin{aligned} \textbf{minimize} \quad & \tfrac{1}{2}[k_1x_1^2 + k_2(x_2 - x_1)^2 + k_3(l - x_2)^2] \\ \textbf{subject to} \quad & x_1 - \tfrac{w}{2} \geq 0 \\ & -x_1 + x_2 - w \geq 0 \\ & -x_2 - \tfrac{w}{2} + l \geq 0 \end{aligned}$$

$$\lambda_1(x_1 - w/2) = 0, \quad \lambda_2(-x_1 + x_2 - w) = 0, \quad \lambda_3(-x_2 + w/2 - l) = 0$$

$$\begin{bmatrix} k_1x_1 - k_2(x_2 - x_1) \\ k_2(x_2 - x_1) - k_3(l - x_2) \end{bmatrix} - \lambda_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \lambda_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} - \lambda_3 \begin{bmatrix} 0 \\ -1 \end{bmatrix} = 0.$$

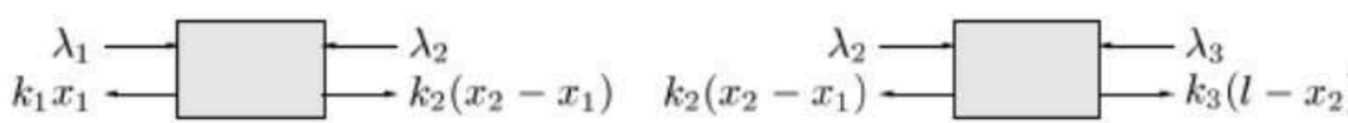


图 5.12 系统的受力分析.

First-order optimality conditions

Abstract constraint sets

Consider a general constrained optimization problem over a closed set $\mathcal{X}$:

$$\begin{aligned} \min_x \ & f(x) \\ \text{s.t. } & x \in \mathcal{X}. \end{aligned}$$

This representation is abstract, but frees us from particular representations of $\mathcal{X}$ that may be problematic when defining optimality conditions. In particular, the normal and tangent cones can be defined for a given $x \in \mathcal{X}$ without having any algebraic description of $\mathcal{X}$...

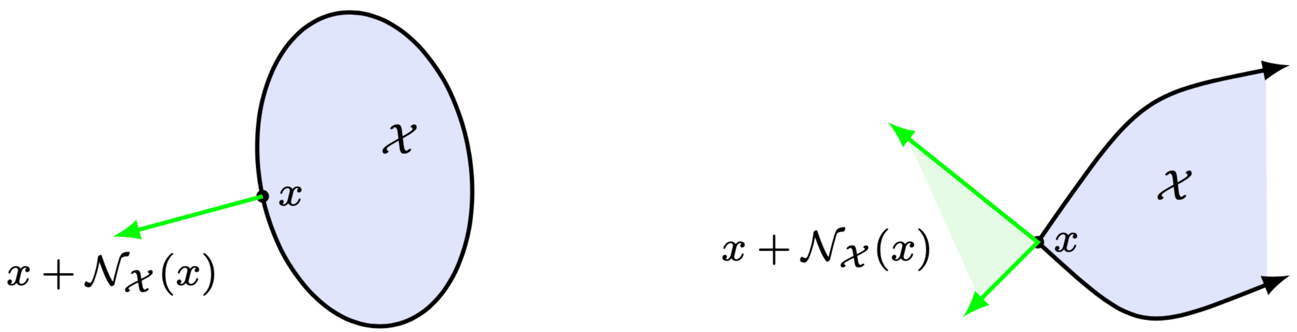
Normal cone

Definition (Normal Cone)

Given a nonempty convex $\mathcal{X} \subseteq \mathbb{R}^n$ and $x \in \mathcal{X}$, the normal cone of $\mathcal{X}$ at x is

$$\mathcal{N}_{\mathcal{X}}(x) := \{g : g^T(\bar{x} - x) \leq 0 \text{ for all } \bar{x} \in \mathcal{X}\}.$$

If $x \in \text{int}(\mathcal{X})$, then clearly $\mathcal{N}_{\mathcal{X}}(x) = \{0\}$, but for $x \notin \text{int}(\mathcal{X})$ the normal cone contains at least one halfline.



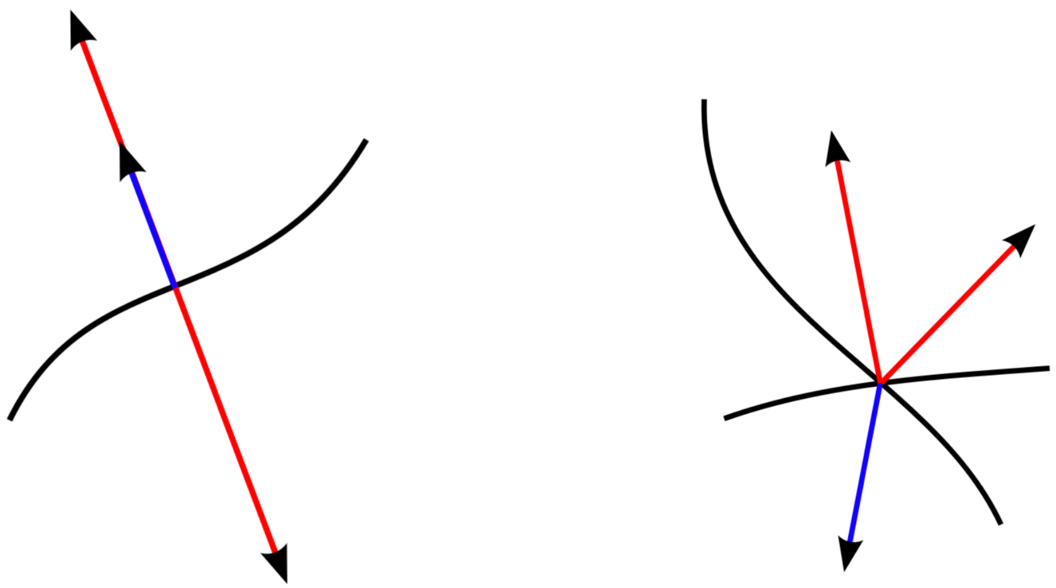
Fundamental necessary condition (using the normal cone) for convex $\mathcal{X}$

Theorem

If x_* is a minimizer of f in (convex) $\mathcal{X}$, then

$$-\nabla f(x_*) \in \mathcal{N}_{\mathcal{X}}(x_*).$$

That is, if x_* is a minimizer, then the steepest descent direction for f at x_* is in the normal cone of $\mathcal{X}$ at x_*. (Blue vector denotes $\nabla f(x_*)$.)



In the unconstrained case, $\mathcal{X} = \mathbb{R}^n$ and $\mathcal{N}_{\mathcal{X}}(x_*) = \{0\}$.

Tangent cone

Definition (Tangent direction)

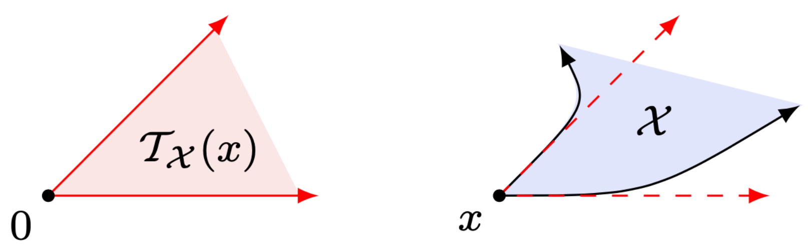
A direction $d \in \mathbb{R}^n$ is tangent to $\mathcal{X} \subseteq \mathbb{R}^n$ at a point $x \in \mathcal{X}$ if there exists a sequence of points $\{x_k\} \in \mathcal{X}$ and positive scalars $\{\tau_k\}$ such that

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

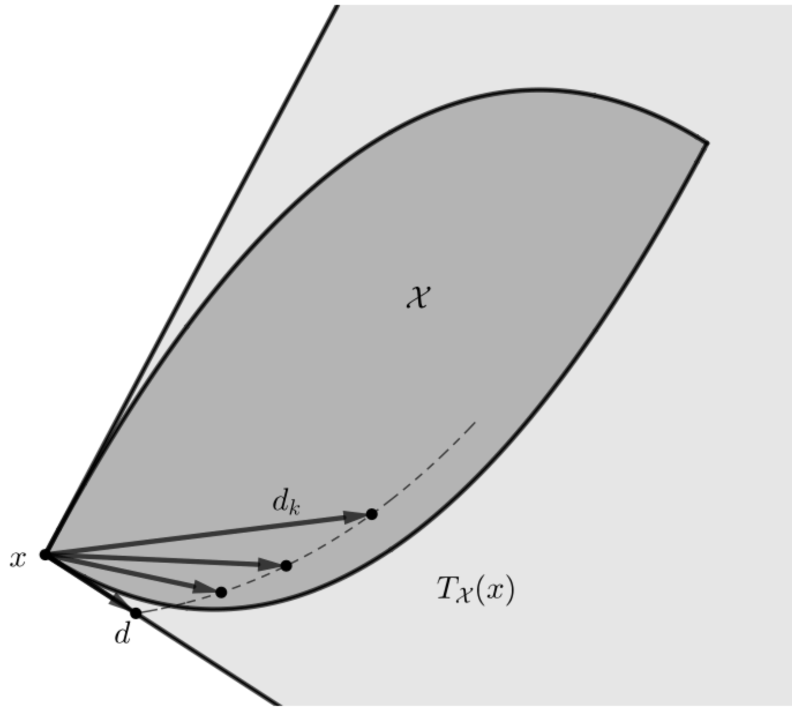
Definition (Tangent cone)

The tangent cone corresponding to a set $\mathcal{X} \subseteq \mathbb{R}^n$ at $x \in \mathcal{X}$ is

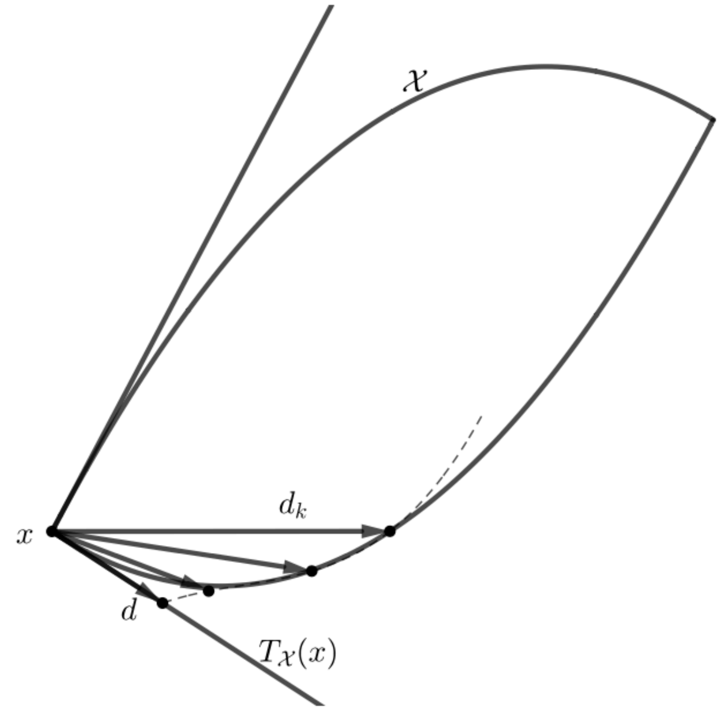
$$\mathcal{T}_{\mathcal{X}}(x) := \{d : d \text{ is tangent to } \mathcal{X} \text{ at } x\}.$$



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



(a) 不等式约束



(b) 等式约束

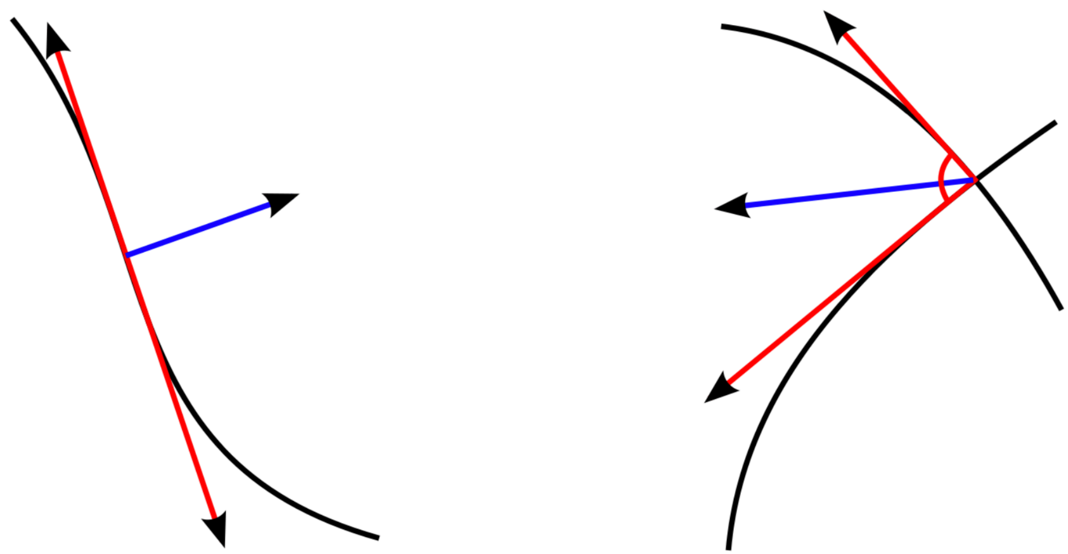
Fundamental necessary condition (using the tangent cone)

Theorem 几何最优性条件

If x_* is a minimizer of f in $\mathcal{X}$, then

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

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 $\mathcal{X}$ at x_*. (Blue vector denotes $\nabla f(x_*)$.)



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

等价表述：切锥和下降方向集合，交集为空。

Feasible directions

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

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

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

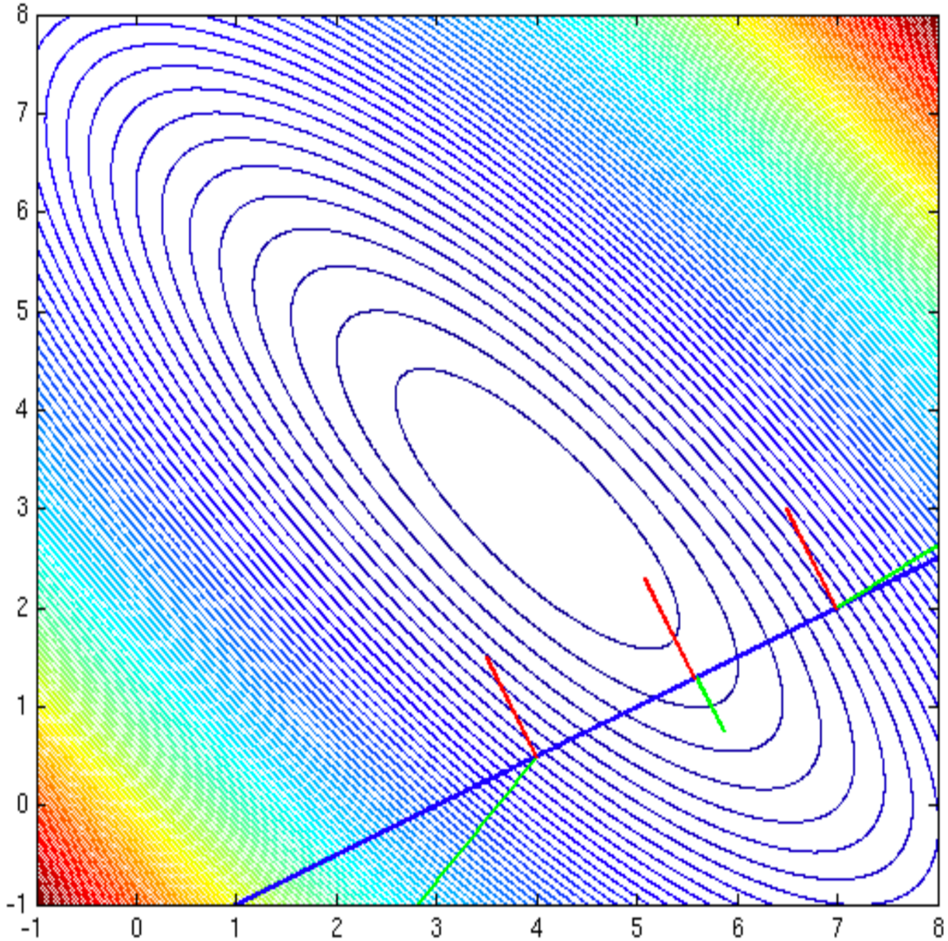
- If $c(x) < 0$, then any small displacement means we remain feasible.
- If $c(x) = 0$, then $\nabla c(x)^T d \leq 0$ (and Taylor's theorem) means we stay feasible.

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

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

Feasible directions illustration

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

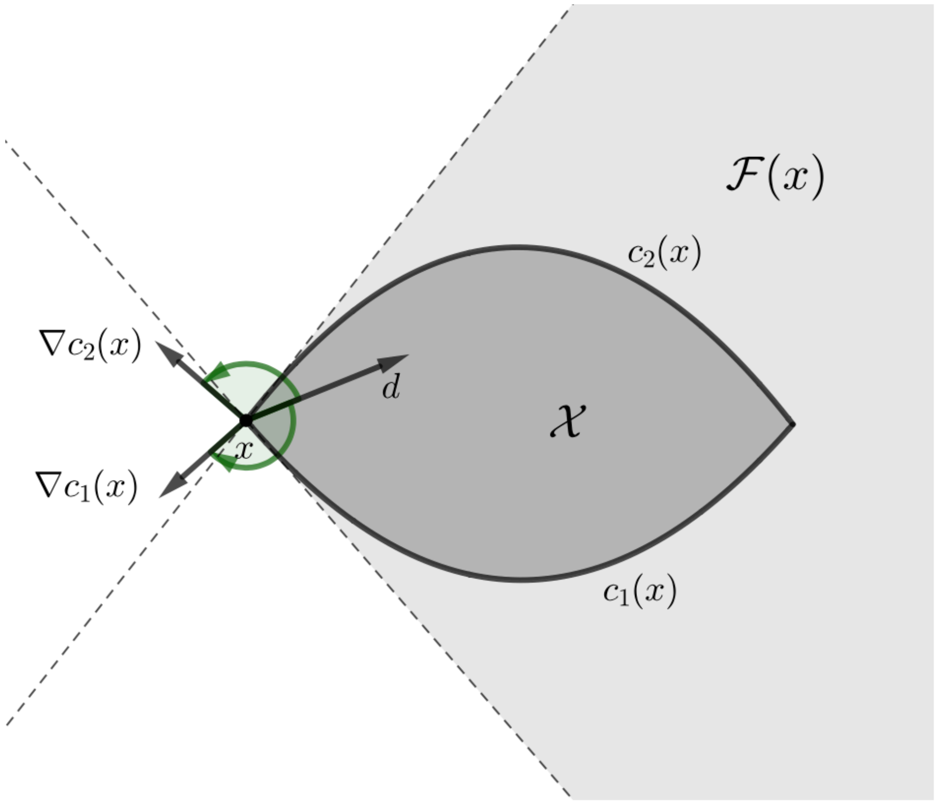


Central to the proof of the KKT theorem is Farkas's theorem.

- At a feasible point x, let

$$\mathcal{F}(x) := \{d : \nabla c^i(x)^T d = 0,\ i \in \mathcal{E};\ \nabla c^i(x)^T d \le 0,\ i \in \mathcal{A}(x)\}$$

be the set of linearized feasible directions at x.



Farkas's theorem

Central to the proof of the KKT theorem is Farkas's theorem.

- At a feasible point x, let

$$\mathcal{F}(x) := \{d : \nabla c^i(x)^T d = 0,\ i \in \mathcal{E};\ \nabla c^i(x)^T d \leq 0,\ i \in \mathcal{A}(x)\}$$

  be the set of linearized feasible directions at x.
- Farkas's theorem implies that only one of the following has a solution:

  System 1: $\nabla f(x)^T d < 0$ and $d \in \mathcal{F}(x)$ for some d;

  System 2: $\nabla f(x) + \nabla c(x)\lambda = 0$ and $\lambda^i \geq 0,\ i \in \mathcal{A}(x)$ for some λ.

FARKAS' LEMMA

(Textbook： Numerical Optimization)

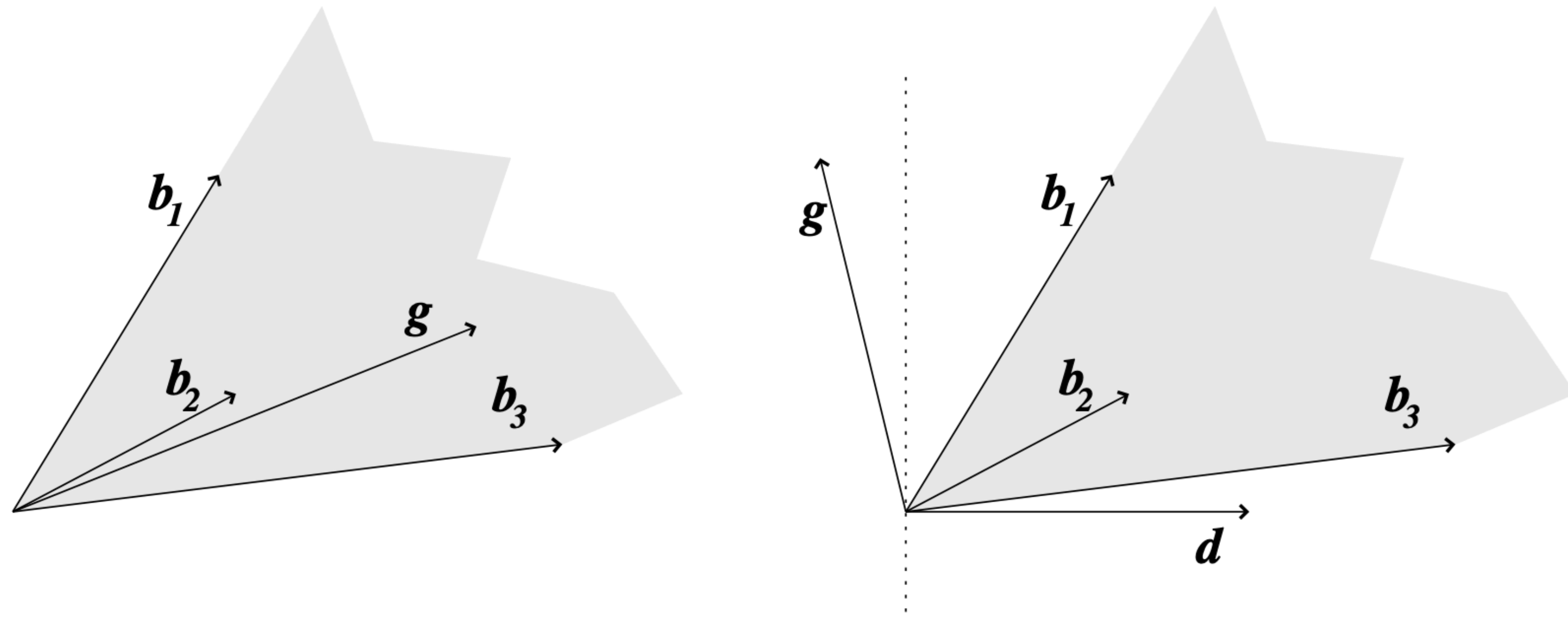
The most important step in proving Theorem 12.1 is a classical theorem of the alternative known as *Farkas' Lemma*. This lemma considers a cone K defined as follows:

$$K = \{By + Cw \mid y \geq 0\}, \tag{12.45}$$

where B and C are matrices of dimension $n \times m$ and $n \times p$, respectively, and y and w are vectors of appropriate dimensions. Given a vector $g \in \mathbb{R}^n$, Farkas' Lemma states that one (and only one) of two alternatives is true. Either $g \in K$, or else there is a vector $d \in \mathbb{R}^n$ such that

$$g^T d < 0, \quad B^T d \geq 0, \quad C^T d = 0. \tag{12.46}$$

$$K = \{By + Cw \mid y \geq 0\}$$



$$g \in \mathbb{R}^n$$

$$g^T d < 0, \quad B^T d \geq 0, \quad C^T d = 0$$

Farkas Lemma

> $Ax \leq \vec{0}$ and $c^\top x > 0$ 或者 $A^\top y = c$ and $y \geq 0$ 之中有且仅有一个是可解的。

证明：

构造一个线性规划问题如下：

$$\begin{gathered}\text{minimize } 0^\top y \\ s.t.\ A^\top y = c \\ y \geq 0\end{gathered}$$

以及它的对偶形式：

$$\begin{gathered}\text{maximize } 0^\top x \\ \text{s.t. } Ax \leq b\end{gathered}$$

Tangent cone and linearized feasible directions

Recall the cone of linearized feasible directions

$$\mathcal{F}(x_*) := \{d : \nabla c^i(x_*)^T d = 0,\ i \in \mathcal{E};\ \nabla c^i(x_*)^T d \leq 0,\ i \in \mathcal{A}(x_*)\}.$$

- Constraint qualifications often guarantee a useful relationship between $\mathcal{F}(x_*)$ and the cone of tangent directions $\mathcal{T}_{\mathcal{X}}(x_*)$.
- For example, we have in general that

  $$\mathcal{T}_{\mathcal{X}}(x_*) \subseteq \mathcal{F}(x_*),$$

  but if the constraints are affine or if the LICQ holds, then

  $$\mathcal{T}_{\mathcal{X}}(x_*) = \mathcal{F}(x_*).$$

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

  System 1: $\nabla f(x)^T d < 0$ and $d \in \mathcal{F}(x)$ for some d;

  System 2: $\nabla f(x) + \nabla c(x)\lambda = 0$ and $\lambda^i \geq 0,\ i \in \mathcal{A}(x)$ for some λ.

一般CQ下的KKT条件

Suppose x_* *is a local solution at which* $\boxed{\mathcal{T}_{\mathcal{X}}(x_*) = \mathcal{F}(x_*)}$ *Then, there exists Lagrange multipliers* λ_* *such that* (x_*, λ_*) *satisfies the following:*

稳定性条件 $\nabla_x L(x, \lambda) = \nabla f(x) + \nabla c(x)\lambda = 0$

原始可行性条件 $c^i(x) = 0, \quad i \in \mathcal{E}$

原始可行性条件 $c^i(x) \le 0, \quad i \in \mathcal{I}$

对偶可行性条件 $\lambda^i \ge 0, \quad i \in \mathcal{I}$

互补可行性条件 $\lambda^i c^i(x) = 0, \quad i \in \mathcal{I}.$

约束规范
(Constraint Qualification)

LICQ

Definition (LICQ)

The linear independence constraint qualification (LICQ) holds at x_* if the set of constraint gradients

$$\nabla c^i(x_*),\ i \in \mathcal{E} \cup \mathcal{A}(x_*)$$

is linearly independent.

- Naturally, this fails if any of the active constraint gradients is zero.
- It also fails for the following sets of constraints:

$$\begin{array}{rcrcrcr} c^1(x) & = & x_1 & + & x_2 & \leq & 2 \\ c^2(x) & = & -x_1 & - & x_2 & \leq & -2 \end{array} \quad \text{(linear!)}$$

  and (assuming $x_* = (1,1)$)

$$\begin{array}{rcrcrcrcr} c^1(x) & = & (x_1 - \frac{1}{2})^2 & + & (x_1 - \frac{1}{2})(x_2 - \frac{3}{2}) & + & (x_2 - \frac{3}{2})^2 & \leq & \frac{1}{4} \\ c^2(x) & = & \frac{3}{2}(x_1 - \frac{1}{2})^2 & + & (x_1 - \frac{1}{2})(x_2 - \frac{3}{2}) & + & \frac{3}{2}(x_2 - \frac{3}{2})^2 & \leq & \frac{1}{2} \end{array}$$

  and anytime there are $n+1$ active constraints!
- The LICQ implies a unique set of multipliers.

Affine constraints

If all constraints that are active at a solution are affine, then no other qualification is needed and x_* is necessarily a KKT point.

- Thus, Lagrange multipliers always exist for linear optimization problems.
- They also always exist for quadratic optimization problems, i.e.,

$$\begin{aligned} &\min\ c^T x + \tfrac{1}{2} x^T Q x \\ &\text{s.t. } Ax = b,\ x \geq 0. \end{aligned}$$

- Note, obviously, that the LICQ does not imply that the constraints are affine.
- Note also that affine constraints do not imply the LICQ (as we saw).

Slater conditions

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

Definition (Weak Slater Condition)

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

$$c^i(x) = 0,\ i \in \mathcal{E};\quad c^i(x) \le 0,\ i \in \mathcal{I}\ (\text{affine});\quad c^i(x) < 0,\ i \in \mathcal{I}\ (\text{non-affine}).$$

Definition (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., $\exists x$ with

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

- 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.

MFCQ

Definition

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

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

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, compact set Λ_* such that for all $\lambda_* \in \Lambda_*$, the point (x_*, λ_*) satisfies the KKT conditions.
- The LICQ $\Longrightarrow$ the MFCQ, but not vice versa; e.g., the MFCQ holds for

$$\begin{array}{ccccccccc} c^1(x) & = & (x_1 - \frac{1}{2})^2 & + & (x_1 - \frac{1}{2})(x_2 - \frac{3}{2}) & + & (x_2 - \frac{3}{2})^2 & \le & \frac{1}{4} \\ c^2(x) & = & \frac{3}{2}(x_1 - \frac{1}{2})^2 & + & (x_1 - \frac{1}{2})(x_2 - \frac{3}{2}) & + & \frac{3}{2}(x_2 - \frac{3}{2})^2 & \le & \frac{1}{2} \end{array}$$

MFCQ

Definition

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

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

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
- It implies the existen
  $\lambda_* \in \Lambda_*$, the point (
- The LICQ $\Longrightarrow$ the

$$\begin{aligned} c^1(x) &= (x_1 - \\ c^2(x) &= \tfrac{3}{2}(x_1 - \end{aligned}$$

Olvi L. Mangasarian	
Born	12 January 1934 (age 86) Baghdad, Iraq
Died	15 March 2020 (aged 86) Madison, Wisconsin
Alma mater	Harvard University
Scientific career	
Fields	Mathematics
Institutions	University of Wisconsin–Madison
Doctoral advisor	Bernard Budiansky

Other constraint qualifications

- Constant rank
- Constant positive linear dependence
- Quasi-normality
- Quasi-regularity
- Pseudo-normality
- Abadie
- Guignard
- Second-order

二阶条件
（Second-order Conditions）

Critical cones

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

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

That is,

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

Note that if we have strict complementarity, i.e., if we have

$$\lambda^i + c^i(x) > 0,\ i \in \mathcal{A}(x),$$

then the critical cone is simply the set of all d such that

$$\nabla c^i(x)^T d = 0,\ i \in \mathcal{E} \cup \mathcal{A}(x).$$

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

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

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

Second-order necessary conditions

Theorem

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:

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

Second-order sufficient conditions

Theorem

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

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

Then, x_* is a strict local solution.