

Lagrange Multipliers and the Karush-Kuhn-Tucker conditions

March 20, 2012

Goal:

Want to find the maximum or minimum of a function subject to some constraints.

Formal Statement of Problem:

Given functions f, g_1, \dots, g_m and h_1, \dots, h_l defined on some domain $\Omega \subset \mathbf{R}^n$ the optimization problem has the form

$$\min_{\mathbf{x} \in \Omega} f(\mathbf{x}) \text{ subject to } g_i(\mathbf{x}) \leq 0 \quad \forall i \text{ and } h_j(\mathbf{x}) = 0 \quad \forall j$$

We will derive/state sufficient and necessary for (local) optimality when there are

- ① no constraints,
- ② only equality constraints,
- ③ only inequality constraints,
- ④ equality and inequality constraints.

Unconstrained Optimization

Assume:

Let $f : \Omega \rightarrow \mathbb{R}$ be a continuously differentiable function.

Necessary and sufficient conditions for a local minimum:

\mathbf{x}^* is a local minimum of $f(\mathbf{x})$ if and only if

- 1 f has zero gradient at \mathbf{x}^* :

$$\nabla_{\mathbf{x}} f(\mathbf{x}^*) = \mathbf{0}$$

- 2 and the Hessian of f at \mathbf{w}^* is positive semi-definite:

$$\mathbf{v}^t (\nabla^2 f(\mathbf{x}^*)) \mathbf{v} \geq 0, \forall \mathbf{v} \in \mathbb{R}^n$$

where

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{pmatrix}$$

Assume:

Let $f : \Omega \rightarrow \mathbb{R}$ be a continuously differentiable function.

Necessary and sufficient conditions for local maximum:

\mathbf{x}^* is a local maximum of $f(\mathbf{x})$ if and only if

- 1 f has zero gradient at \mathbf{x}^* :

$$\nabla f(\mathbf{x}^*) = \mathbf{0}$$

- 2 and the Hessian of f at \mathbf{x}^* is negative semi-definite:

$$\mathbf{v}^t (\nabla^2 f(\mathbf{x}^*)) \mathbf{v} \leq \mathbf{0}, \forall \mathbf{v} \in \mathbb{R}^n$$

where

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{pmatrix}$$

Constrained Optimization: Equality Constraints

Problem:

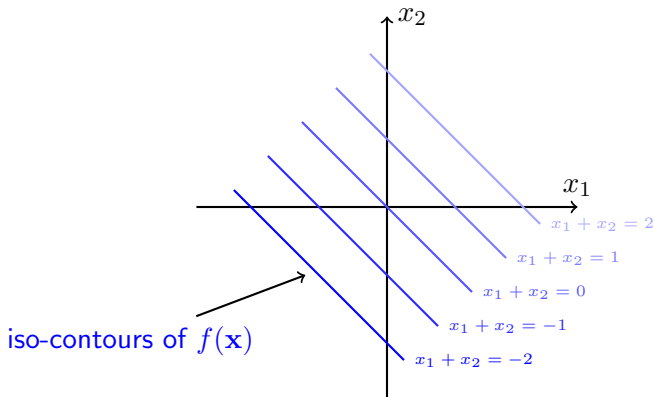
This is the constrained optimization problem we want to solve

$$\min_{\mathbf{x} \in \mathbb{R}^2} f(\mathbf{x}) \text{ subject to } h(\mathbf{x}) = 0$$

where

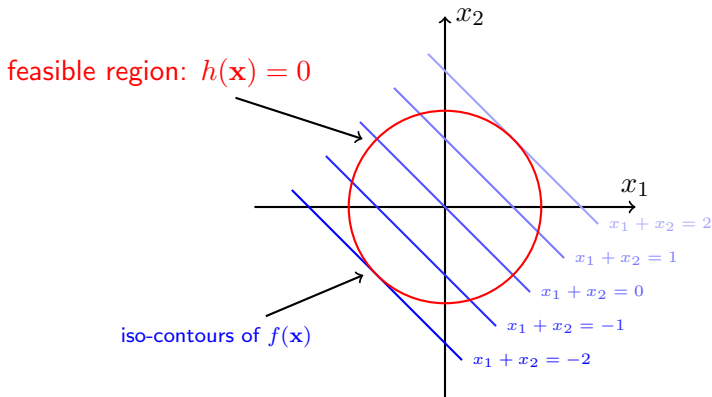
$$f(\mathbf{x}) = x_1 + x_2 \text{ and } h(\mathbf{x}) = x_1^2 + x_2^2 - 2$$

Tutorial example - Cost function



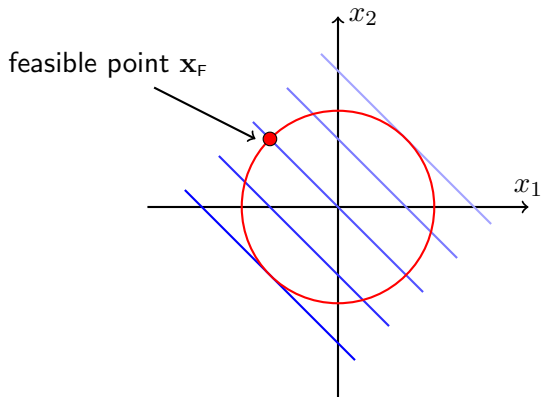
$$f(\mathbf{x}) = x_1 + x_2$$

Tutorial example - Feasible region

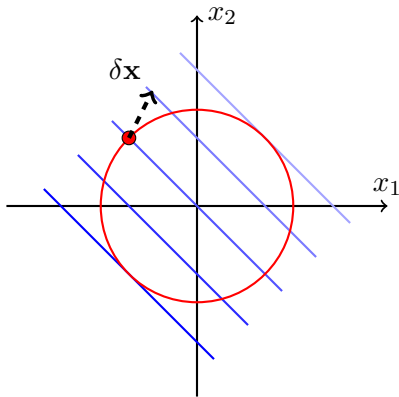


$$h(\mathbf{x}) = x_1^2 + x_2^2 - 2$$

Given a point \mathbf{x}_F on the constraint surface

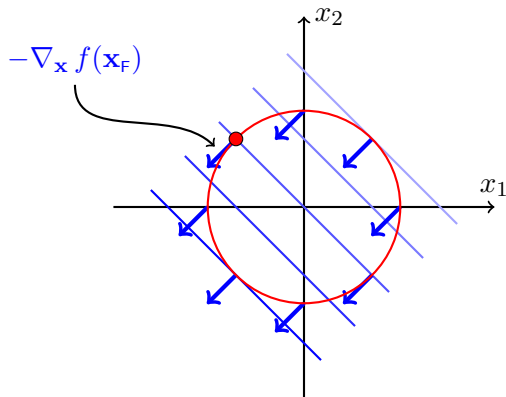


Given a point \mathbf{x}_F on the constraint surface



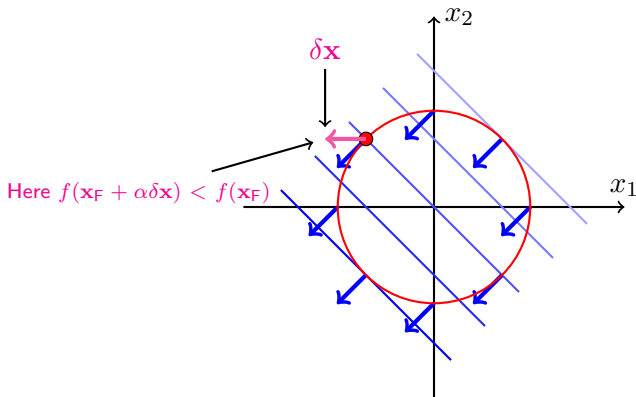
Find $\delta \mathbf{x}$ s.t. $h(\mathbf{x}_F + \alpha \delta \mathbf{x}) = 0$ and $f(\mathbf{x}_F + \alpha \delta \mathbf{x}) < f(\mathbf{x}_F)$?

Condition to decrease the cost function



At any point $\tilde{\mathbf{x}}$ the direction of steepest descent of the cost function $f(\mathbf{x})$ is given by $-\nabla_{\mathbf{x}} f(\tilde{\mathbf{x}})$.

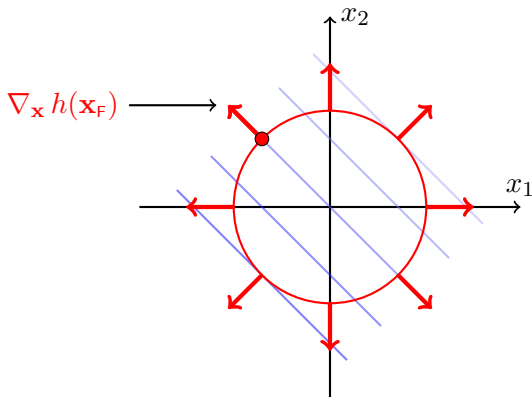
Condition to decrease the cost function



To move $\delta \mathbf{x}$ from \mathbf{x} such that $f(\mathbf{x} + \delta \mathbf{x}) < f(\mathbf{x})$ must have

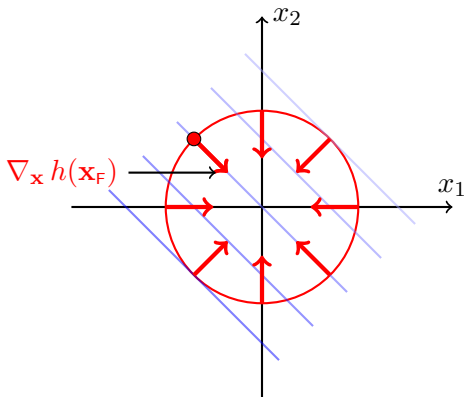
$$\delta \mathbf{x} \cdot (-\nabla_{\mathbf{x}} f(\mathbf{x})) > 0$$

Condition to remain on the constraint surface



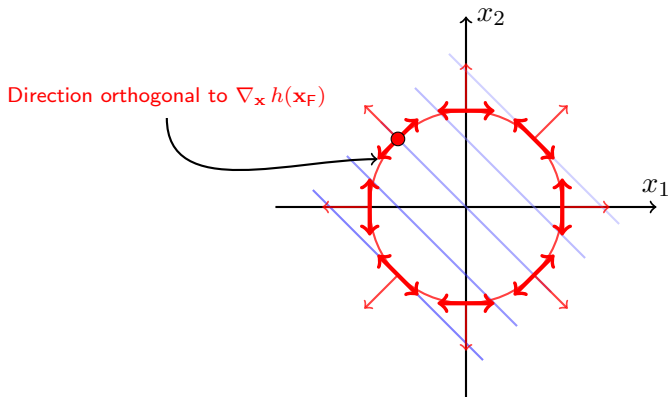
Normals to the constraint surface are given by $\nabla_{\mathbf{x}} h(\mathbf{x})$

Condition to remain on the constraint surface



Note the direction of the normal is arbitrary as the constraint be imposed as either $h(\mathbf{x}) = 0$ or $-h(\mathbf{x}) = 0$

Condition to remain on the constraint surface



To move a small $\delta \mathbf{x}$ from \mathbf{x} and remain on the constraint surface we have to move in a direction orthogonal to $\nabla_{\mathbf{x}} h(\mathbf{x})$.

If \mathbf{x}_F lies on the constraint surface:

- setting $\delta\mathbf{x}$ orthogonal to $\nabla_{\mathbf{x}} h(\mathbf{x}_F)$ ensures $h(\mathbf{x}_F + \delta\mathbf{x}) = 0$.
- And $f(\mathbf{x}_F + \delta\mathbf{x}) < f(\mathbf{x}_F)$ only if

$$\delta\mathbf{x} \cdot (-\nabla_{\mathbf{x}} f(\mathbf{x}_F)) > 0$$

Consider the case when

$$\nabla_{\mathbf{x}} f(\mathbf{x}_F) = \mu \nabla_{\mathbf{x}} h(\mathbf{x}_F)$$

where μ is a scalar.

When this occurs

- If $\delta \mathbf{x}$ is orthogonal to $\nabla_{\mathbf{x}} h(\mathbf{x}_F)$ then

$$\delta \mathbf{x} \cdot (-\nabla_{\mathbf{x}_F} f(\mathbf{x})) = -\delta \mathbf{x} \cdot \mu \nabla_{\mathbf{x}} h(\mathbf{x}_F) = 0$$

- Cannot move from \mathbf{x}_F to **remain on the constraint surface** and **decrease (or increase) the cost function.**

This case corresponds to a constrained local optimum!

Consider the case when

$$\nabla_{\mathbf{x}} f(\mathbf{x}_F) = \mu \nabla_{\mathbf{x}} h(\mathbf{x}_F)$$

where μ is a scalar.

When this occurs

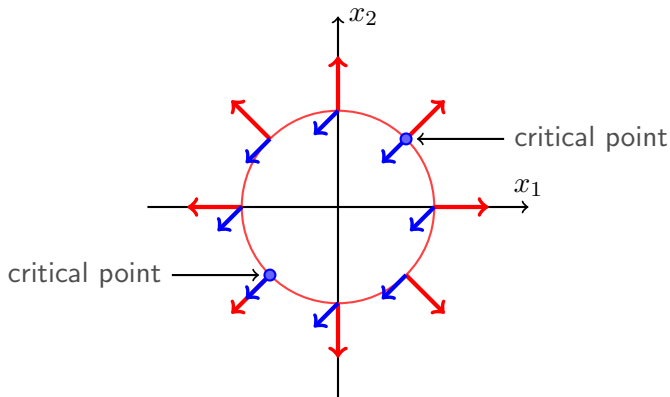
- If $\delta \mathbf{x}$ is orthogonal to $\nabla_{\mathbf{x}} h(\mathbf{x}_F)$ then

$$\delta \mathbf{x} \cdot (-\nabla_{\mathbf{x}_F} f(\mathbf{x})) = -\delta \mathbf{x} \cdot \mu \nabla_{\mathbf{x}} h(\mathbf{x}_F) = 0$$

- Cannot move from \mathbf{x}_F to **remain on the constraint surface** and **decrease (or increase) the cost function**.

This case corresponds to a constrained local optimum!

Condition for a local optimum



A constrained local optimum occurs at \mathbf{x}^* when $\nabla_{\mathbf{x}} f(\mathbf{x}^*)$ and $\nabla_{\mathbf{x}} h(\mathbf{x}^*)$ are parallel that is

$$\nabla_{\mathbf{x}} f(\mathbf{x}^*) = \mu \nabla_{\mathbf{x}} h(\mathbf{x}^*)$$

From this fact Lagrange Multipliers make sense

Remember our constrained optimization problem is

$$\min_{\mathbf{x} \in \mathbb{R}^2} f(\mathbf{x}) \quad \text{subject to} \quad h(\mathbf{x}) = 0$$

Define the **Lagrangian** as

$$\mathcal{L}(\mathbf{x}, \mu) = f(\mathbf{x}) + \mu h(\mathbf{x})$$

Then \mathbf{x}^* a local minimum \iff there exists a unique μ^* s.t.

- ① $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \mu^*) = \mathbf{0}$
- ② $\nabla_{\mu} \mathcal{L}(\mathbf{x}^*, \mu^*) = 0$
- ③ $\mathbf{y}^t (\nabla_{\mathbf{xx}}^2 \mathcal{L}(\mathbf{x}^*, \mu^*)) \mathbf{y} \geq 0 \quad \forall \mathbf{y} \text{ s.t. } \nabla_{\mathbf{x}} h(\mathbf{x}^*)^t \mathbf{y} = 0$

The case of multiple equality constraints

The constrained optimization problem is

$$\min_{\mathbf{x} \in \mathbb{R}^2} f(\mathbf{x}) \quad \text{subject to} \quad h_i(\mathbf{x}) = 0 \quad \text{for } i = 1, \dots, l$$

Construct the **Lagrangian** (introduce a multiplier for each constraint)

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^l \mu_i h_i(\mathbf{x}) = f(\mathbf{x}) + \boldsymbol{\mu}^t \mathbf{h}(\mathbf{x})$$

Then \mathbf{x}^* a local minimum \iff there exists a unique $\boldsymbol{\mu}^*$ s.t.

- ① $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*) = \mathbf{0}$
- ② $\nabla_{\boldsymbol{\mu}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*) = \mathbf{0}$
- ③ $\mathbf{y}^t (\nabla_{\mathbf{x}\mathbf{x}}^2 \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*)) \mathbf{y} \geq 0 \quad \forall \mathbf{y} \text{ s.t. } \nabla_{\mathbf{x}} h(\mathbf{x}^*)^t \mathbf{y} = 0$

Constrained Optimization: Inequality Constraints

Problem:

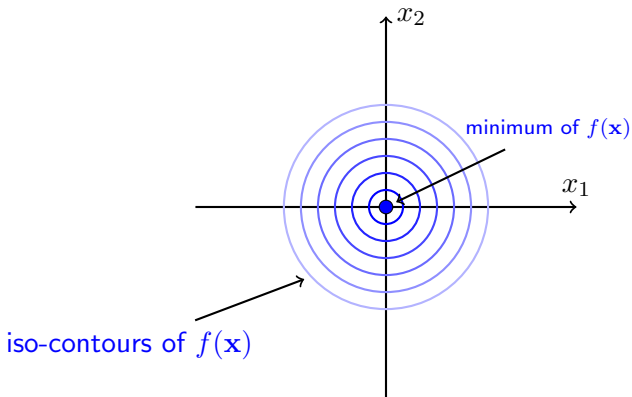
Consider this constrained optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^2} f(\mathbf{x}) \text{ subject to } g(\mathbf{x}) \leq 0$$

where

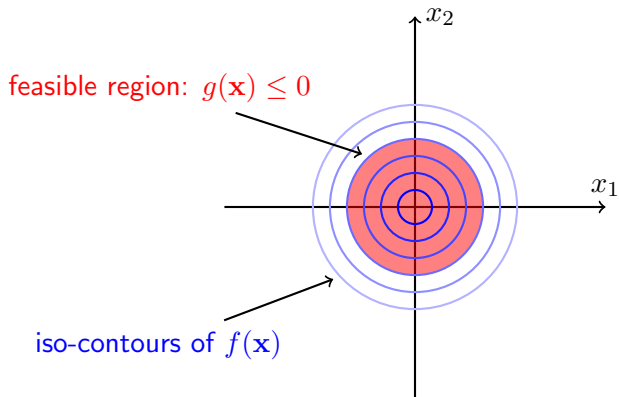
$$f(\mathbf{x}) = x_1^2 + x_2^2 \text{ and } g(\mathbf{x}) = x_1^2 + x_2^2 - 1$$

Tutorial example - Cost function



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

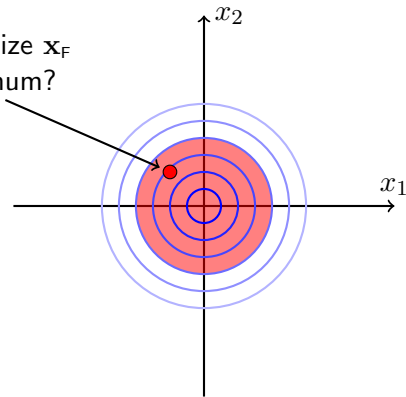
Tutorial example - Feasible region



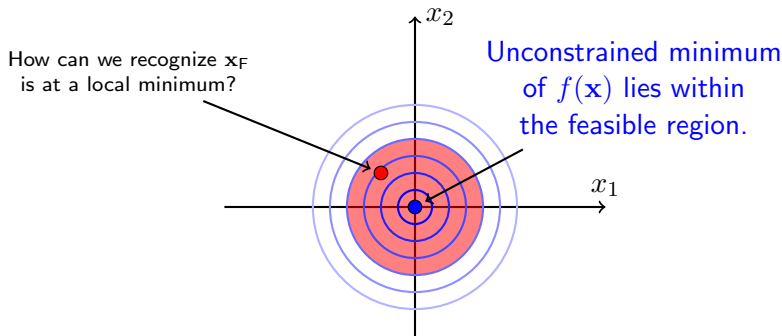
$$g(\mathbf{x}) = x_1^2 + x_2^2 - 1$$

How do we recognize if \mathbf{x}_F is at a local optimum?

How can we recognize \mathbf{x}_F is at a local minimum?



Remember \mathbf{x}_F denotes a feasible point.



当 $f(\mathbf{x})$ 在可行域之内时, $f(\mathbf{x})$ 本身的极小值, 在经过 $g(\mathbf{x})$ 约束之后, 还是原来的那个值, 也就是说, 这 $g(\mathbf{x})$ 这个约束条件没有起到作用

\therefore Necessary and sufficient conditions for a constrained local minimum are the same as for an unconstrained local minimum.

$$\nabla_{\mathbf{x}} f(\mathbf{x}_F) = \mathbf{0} \quad \text{and} \quad \nabla_{\mathbf{xx}} f(\mathbf{x}_F) \text{ is positive definite}$$

This Tutorial Example has an inactive constraint

Problem:

Our constrained optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^2} f(\mathbf{x}) \text{ subject to } g(\mathbf{x}) \leq 0$$

where

$$f(\mathbf{x}) = x_1^2 + x_2^2 \text{ and } g(\mathbf{x}) = x_1^2 + x_2^2 - 1$$

Constraint is not active at the local minimum ($g(\mathbf{x}^*) < 0$):

Therefore the local minimum is identified by the same conditions as in the unconstrained case.

Problem:

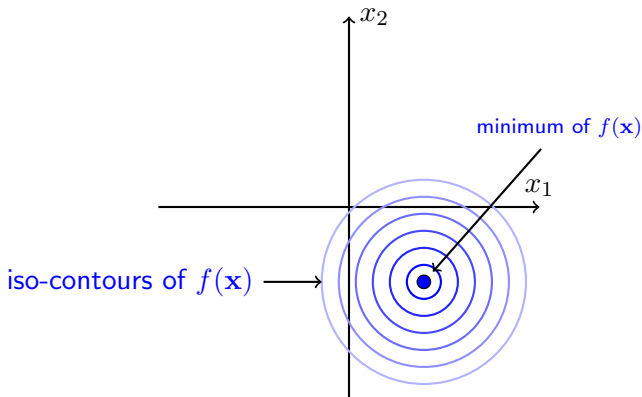
This is the constrained optimization problem we want to solve

$$\min_{\mathbf{x} \in \mathbb{R}^2} f(\mathbf{x}) \text{ subject to } g(\mathbf{x}) \leq 0$$

where

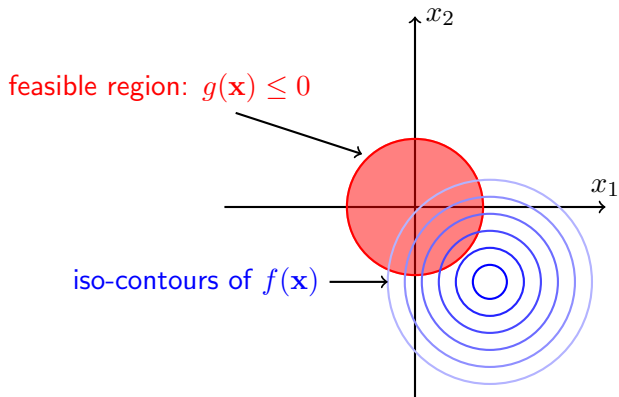
$$f(\mathbf{x}) = (x_1 - 1.1)^2 + (x_2 - 1.1)^2 \text{ and } g(\mathbf{x}) = x_1^2 + x_2^2 - 1$$

Tutorial example - Cost function



$$f(\mathbf{x}) = (x_1 - 1.1)^2 + (x_2 - 1.1)^2$$

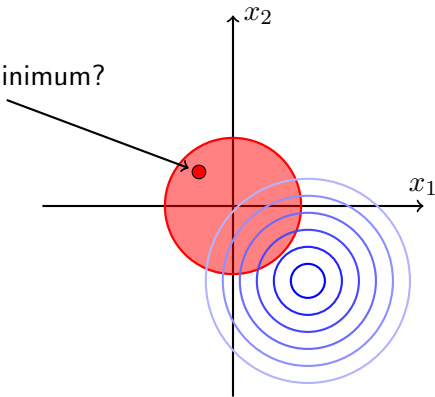
Tutorial example - Feasible region



$$g(\mathbf{x}) = x_1^2 + x_2^2 - 1$$

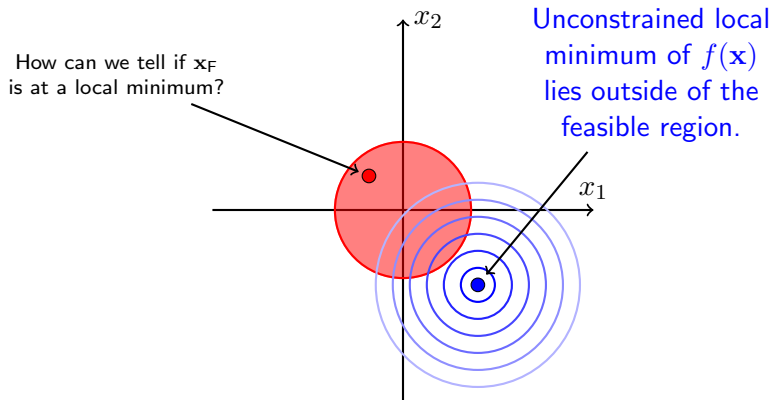
How do we recognize if \mathbf{x}_F is at a local optimum?

Is \mathbf{x}_F at a local minimum?



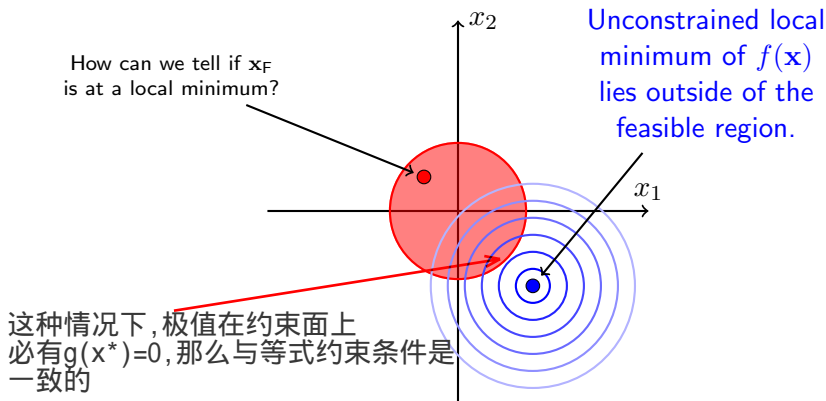
Remember \mathbf{x}_F denotes a feasible point.

How do we recognize if \mathbf{x}_F is at a local optimum?



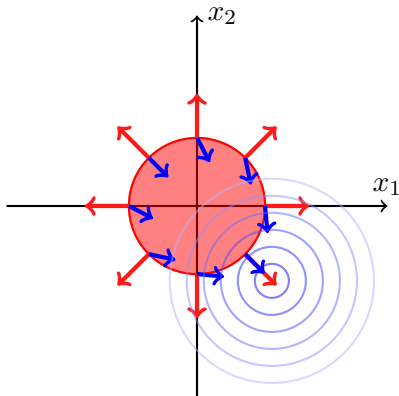
\therefore the constrained local minimum occurs on the surface of the constraint surface.

How do we recognize if \mathbf{x}_F is at a local optimum?



\therefore Effectively have an optimization problem with an **equality constraint**: $g(\mathbf{x}) = 0$.

Given an equality constraint

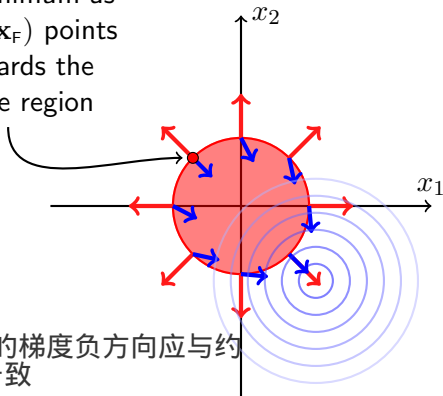


A local optimum occurs when $\nabla_{\mathbf{x}} f(\mathbf{x})$ and $\nabla_{\mathbf{x}} g(\mathbf{x})$ are parallel:

$$-\nabla_{\mathbf{x}} f(\mathbf{x}) = \lambda \nabla_{\mathbf{x}} g(\mathbf{x})$$

Want a constrained local minimum...

X Not a constrained local minimum as $-\nabla_{\mathbf{x}} f(\mathbf{x}_F)$ points in towards the feasible region

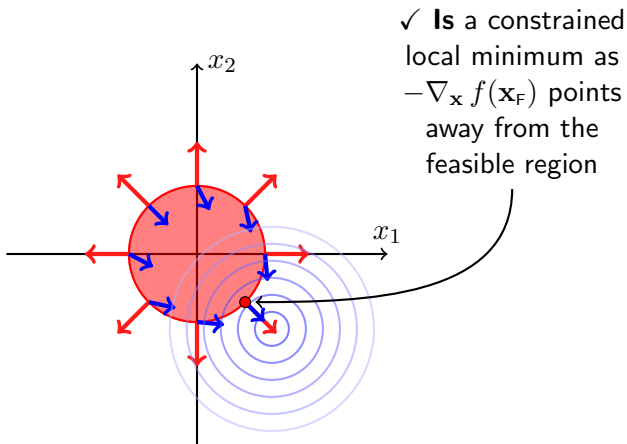


极小值时, $f(\mathbf{x})$ 的梯度负方向应与约束条件梯度方向一致

\therefore Constrained local minimum occurs when $-\nabla_{\mathbf{x}} f(\mathbf{x})$ and $\nabla_{\mathbf{x}} g(\mathbf{x})$ point in the same direction:

$$-\nabla_{\mathbf{x}} f(\mathbf{x}) = \lambda \nabla_{\mathbf{x}} g(\mathbf{x}) \quad \text{and} \quad \lambda > 0$$

Want a constrained local minimum...



∴ Constrained local minimum occurs when $-\nabla_{\mathbf{x}} f(\mathbf{x})$ and $\nabla_{\mathbf{x}} g(\mathbf{x})$ point in the same direction:

$$-\nabla_{\mathbf{x}} f(\mathbf{x}) = \lambda \nabla_{\mathbf{x}} g(\mathbf{x}) \quad \text{and} \quad \lambda > 0$$

Summary of optimization with one inequality constraint

Given

$$\min_{\mathbf{x} \in \mathbb{R}^2} f(\mathbf{x}) \quad \text{subject to} \quad g(\mathbf{x}) \leq 0$$

第二种, 极值在约束面上, 且此时 $f(\mathbf{x})$ 梯度负方向与约束面梯度方向一

If \mathbf{x}^* corresponds to a constrained local minimum then

Case 1:

Unconstrained local minimum occurs **in** the feasible region.

- ① $g(\mathbf{x}^*) < 0$
- ② $\nabla_{\mathbf{x}} f(\mathbf{x}^*) = 0$
- ③ $\nabla_{\mathbf{x}\mathbf{x}} f(\mathbf{x}^*)$ is a positive semi-definite matrix.

Case 2:

Unconstrained local minimum lies **outside** the feasible region.

- ① $g(\mathbf{x}^*) = 0$
- ② $-\nabla_{\mathbf{x}} f(\mathbf{x}^*) = \lambda \nabla_{\mathbf{x}} g(\mathbf{x}^*)$
with $\lambda > 0$
- ③ $\mathbf{y}^t \nabla_{\mathbf{x}\mathbf{x}} L(\mathbf{x}^*) \mathbf{y} \geq 0$ for all \mathbf{y} orthogonal to $\nabla_{\mathbf{x}} g(\mathbf{x}^*)$.

第一种情况, 退化为直接求 $f(\mathbf{x})$ 的极小值, 等价将 λ 置为 0

Karush-Kuhn-Tucker conditions encode these conditions

Given the optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^2} f(\mathbf{x}) \text{ subject to } g(\mathbf{x}) \leq 0$$

Define the **Lagrangian** as

$$f + \lambda g = L$$

满足下方条件时, 即有f的
负梯度方向与约束面梯度方向一致

$$\mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda g(\mathbf{x})$$

Then \mathbf{x}^* a local minimum \iff there exists a unique λ^* s.t.

① $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \lambda^*) = 0$

② $\lambda^* \geq 0$

③ $\lambda^* g(\mathbf{x}^*) = 0$

④ $g(\mathbf{x}^*) \leq 0$

⑤ Plus positive definite constraints on $\nabla_{\mathbf{xx}} \mathcal{L}(\mathbf{x}^*, \lambda^*)$.

case1: 看做lambda为0, 所以有lambda*g=0

case2: 极值只能发生在约束面上, 所以必有g=0

These are the **KKT conditions**.

Let's check what the KKT conditions imply

Case 1 - Inactive constraint:

- When $\lambda^* = 0$ then have $\mathcal{L}(\mathbf{x}^*, \lambda^*) = f(\mathbf{x}^*)$.
- Condition KKT 1 $\implies \nabla_{\mathbf{x}} f(\mathbf{x}^*) = \mathbf{0}$.
- Condition KKT 4 $\implies \mathbf{x}^*$ is a feasible point.

Case 2 - Active constraint:

- When $\lambda^* > 0$ then have $\mathcal{L}(\mathbf{x}^*, \lambda^*) = f(\mathbf{x}^*) + \lambda^* g(\mathbf{x}^*)$.
- Condition KKT 1 $\implies \nabla_{\mathbf{x}} f(\mathbf{x}^*) = -\lambda^* \nabla_{\mathbf{x}} g(\mathbf{x}^*)$.
- Condition KKT 3 $\implies g(\mathbf{x}^*) = 0$.
- Condition KKT 3 also $\implies \mathcal{L}(\mathbf{x}^*, \lambda^*) = f(\mathbf{x}^*)$.

KKT conditions for multiple inequality constraints

Given the optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^2} f(\mathbf{x}) \quad \text{subject to} \quad g_j(\mathbf{x}) \leq 0 \quad \text{for } j = 1, \dots, m$$

Define the **Lagrangian** as

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{j=1}^m \lambda_j g_j(\mathbf{x}) = f(\mathbf{x}) + \boldsymbol{\lambda}^t \mathbf{g}(\mathbf{x})$$

Then \mathbf{x}^* a local minimum \iff there exists a unique $\boldsymbol{\lambda}^*$ s.t.

- 1 $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}$
- 2 $\lambda_j^* \geq 0$ for $j = 1, \dots, m$
- 3 $\lambda_j^* g(\mathbf{x}^*) = 0$ for $j = 1, \dots, m$
- 4 $g_j(\mathbf{x}^*) \leq 0$ for $j = 1, \dots, m$
- 5 Plus positive definite constraints on $\nabla_{\mathbf{x}\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*)$.

KKT for multiple equality & inequality constraints

Given the constrained optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^2} f(\mathbf{x})$$

subject to

$$h_i(\mathbf{x}) = 0 \text{ for } i = 1, \dots, l \text{ and } g_j(\mathbf{x}) \leq 0 \text{ for } j = 1, \dots, m$$

Define the **Lagrangian** as

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\mu}^t \mathbf{h}(\mathbf{x}) + \boldsymbol{\lambda}^t \mathbf{g}(\mathbf{x})$$

Then \mathbf{x}^* a local minimum \iff there exists a unique $\boldsymbol{\lambda}^*$ s.t.

- ① $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) = \mathbf{0}$
- ② $\lambda_j^* \geq 0$ for $j = 1, \dots, m$
- ③ $\lambda_j^* g_j(\mathbf{x}^*) = 0$ for $j = 1, \dots, m$
- ④ $g_j(\mathbf{x}^*) \leq 0$ for $j = 1, \dots, m$
- ⑤ $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$
- ⑥ Plus positive definite constraints on $\nabla_{\mathbf{x}\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*)$.