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

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

In these notes..

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 $f : \Omega \to \mathbb{R}$ 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 $\mathbf{x}^*$:

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

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

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

where

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

Assume:

Let $f : \Omega \to \mathbb{R}$ 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 $\mathbf{x}^*$:

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

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

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

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:

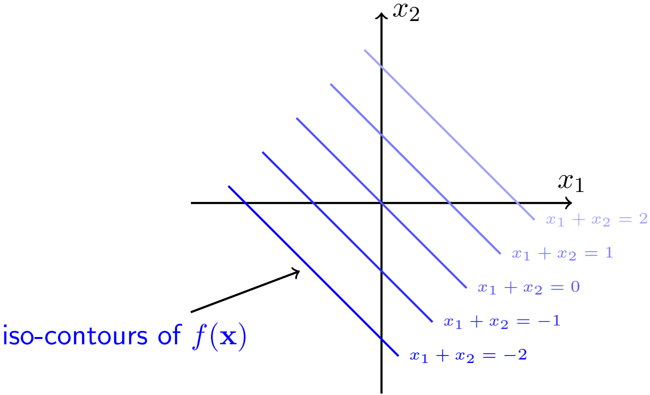
This is the constrained optimization problem we want to solve

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

where

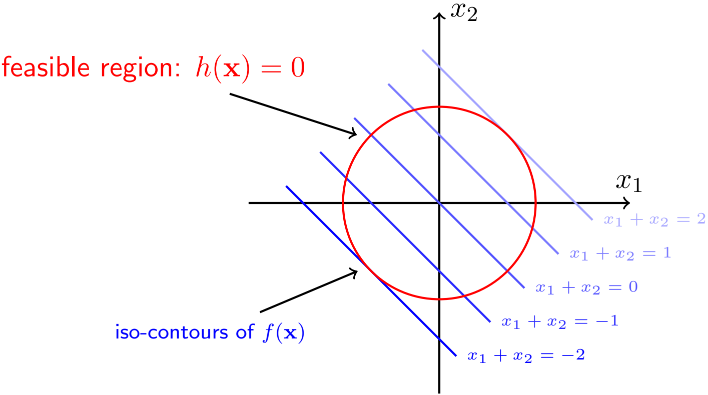
$$\boxed{f(\mathbf{x}) = x_1 + x_2 \textbf{ and } h(\mathbf{x}) = x_1^2 + x_2^2 - 2}$$

Tutorial example - Cost function



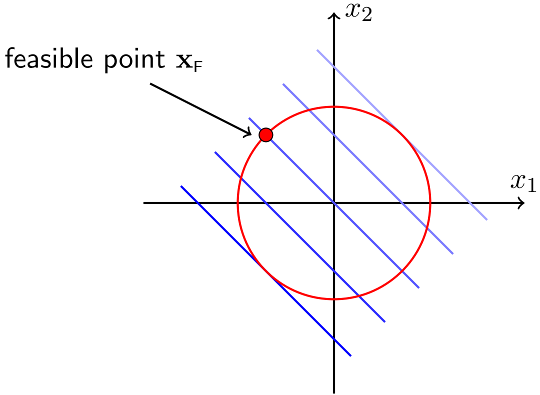
$$\boxed{f(\mathbf{x}) = x_1 + x_2}$$

Tutorial example - Feasible region

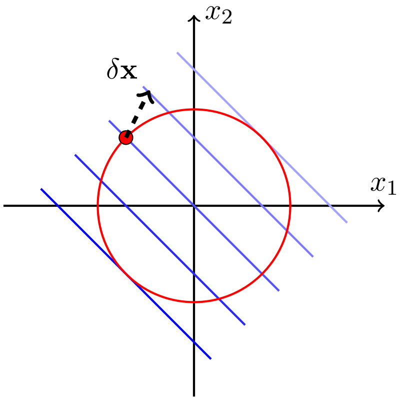


$$\boxed{h(\mathbf{x}) = x_1^2 + x_2^2 - 2}$$

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

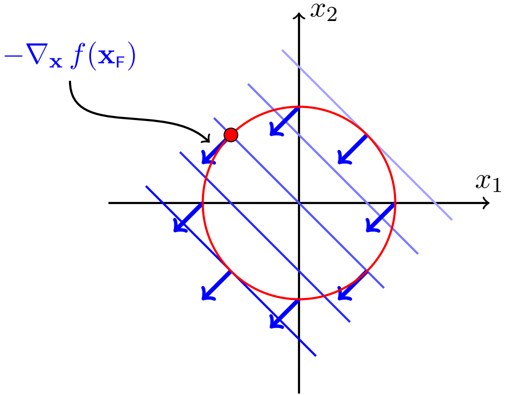


Given a point $\mathbf{x}_{\text{F}}$ on the constraint surface

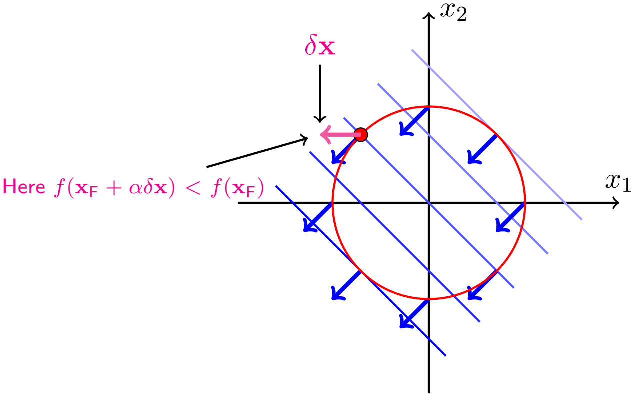


Find $\delta\mathbf{x}$ s.t. $h(\mathbf{x}_{\text{F}} + \alpha\,\delta\mathbf{x}) = 0$ **and** $f(\mathbf{x}_{\text{F}} + \alpha\,\delta\mathbf{x}) < f(\mathbf{x}_{\text{F}})$?

Condition to decrease the cost function



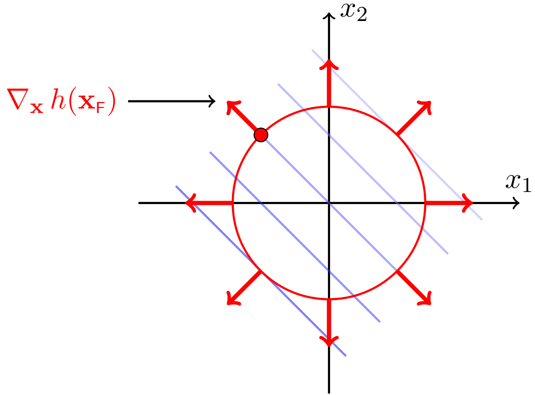
At any point $\tilde{\mathbf{x}}$ the direction of steepest descent of the cost function $f(\mathbf{x})$ is given by $-\nabla_{\mathbf{x}} f(\tilde{\mathbf{x}})$.



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

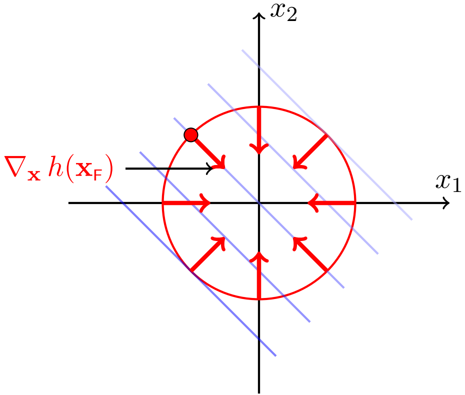
$$\boxed{\delta\mathbf{x} \cdot (-\nabla_{\mathbf{x}}\, f(\mathbf{x})) > 0}$$

Condition to remain on the constraint surface



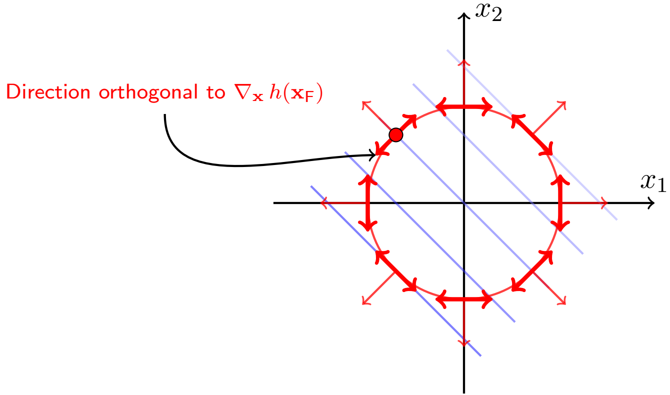
Normals to the constraint surface are given by $\nabla_{\mathbf{x}}\, h(\mathbf{x})$

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 $\mathbf{x}_\mathsf{F}$ lies on the constraint surface:

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

$$\delta\mathbf{x} \cdot (-\nabla_\mathbf{x}\, f(\mathbf{x}_\mathsf{F})) > 0$$

Consider the case when

$$\nabla_{\mathbf{x}} f(\mathbf{x}_{\mathsf{F}}) = \mu \nabla_{\mathbf{x}} h(\mathbf{x}_{\mathsf{F}})$$

where μ is a scalar.

When this occurs

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

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

- Cannot move from $\mathbf{x}_{\mathsf{F}}$ 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

$$\nabla_{\mathbf{x}} f(\mathbf{x}_{\mathsf{F}}) = \mu \nabla_{\mathbf{x}} h(\mathbf{x}_{\mathsf{F}})$$

where μ is a scalar.

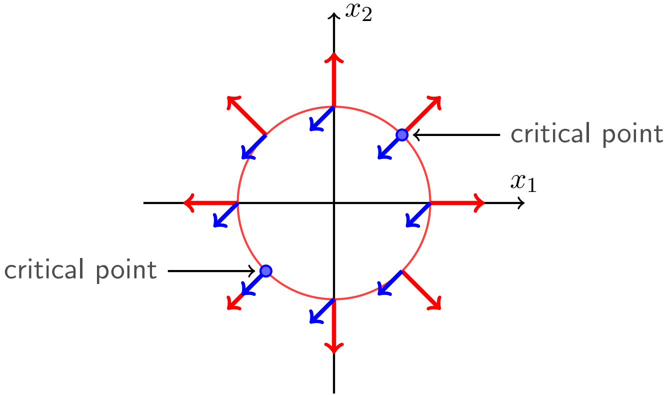
When this occurs

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

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

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

This case corresponds to a constrained 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

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

From this fact Lagrange Multipliers make sense

Remember our constrained optimization problem is

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

Define the **Lagrangian** as

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

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

From this fact Lagrange Multipliers make sense

Remember our constrained optimization problem is

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

Define the **Lagrangian** as note $\mathcal{L}(\mathbf{x}^*, \mu^*) = f(\mathbf{x}^*)$
$\downarrow$

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

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

1. $\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*, \mu^*) = \mathbf{0}$ $\leftarrow$ encodes $\nabla_{\mathbf{x}} f(\mathbf{x}^*) = \mu^* \nabla_{\mathbf{x}} h(\mathbf{x}^*)$
2. $\nabla_{\mu}\mathcal{L}(\mathbf{x}^*, \mu^*) = 0$ $\leftarrow$ encodes the equality constraint $h(\mathbf{x}^*) = 0$
3. $\mathbf{y}^t(\nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}^*, \mu^*))\mathbf{y} \geq 0 \quad \forall \mathbf{y}$ s.t. $\nabla_{\mathbf{x}} h(\mathbf{x}^*)^t \mathbf{y} = 0$

$\uparrow$
Positive definite Hessian tells us we have a local minimum

The case of multiple equality constraints

The constrained optimization problem is

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

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

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

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

Constrained Optimization:
Inequality Constraints

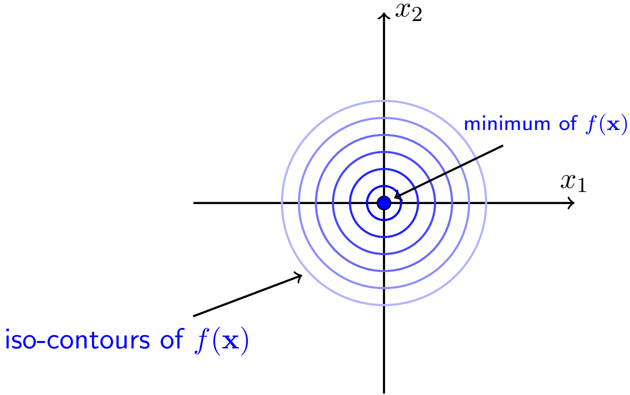
Problem:

Consider this constrained optimization problem

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

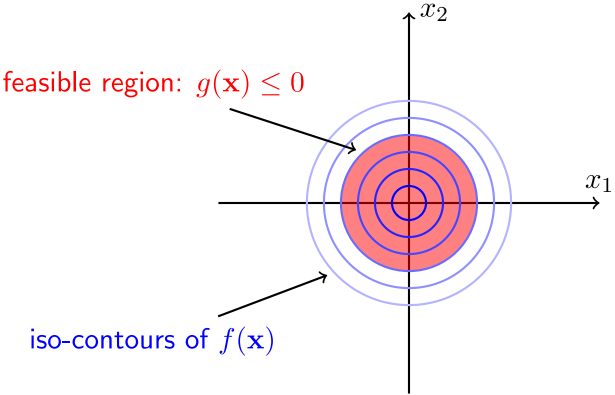
where

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



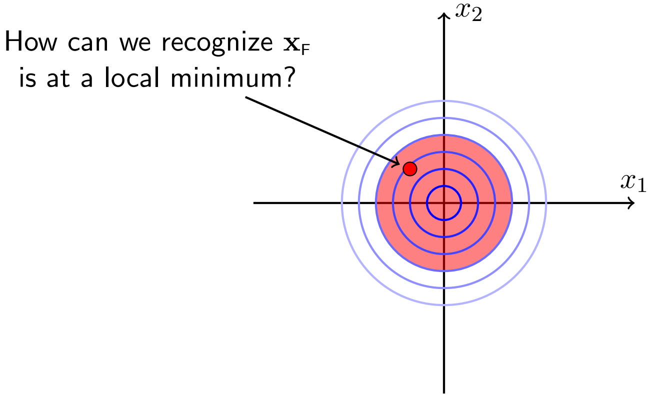
$$\boxed{f(\mathbf{x}) = x_1^2 + x_2^2}$$

Tutorial example - Feasible region

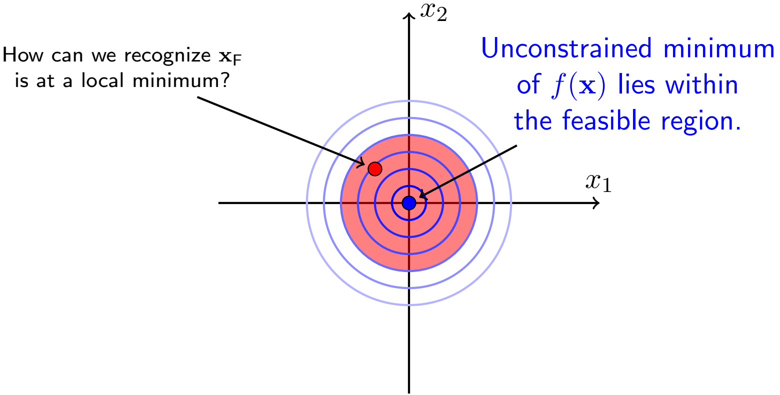


$$\boxed{g(\mathbf{x}) = x_1^2 + x_2^2 - 1}$$

How do we recognize if $\mathbf{x}_{\mathsf{F}}$ is at a local optimum?



Remember $\mathbf{x}_{\mathsf{F}}$ denotes a feasible point.



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

$$\boxed{\nabla_{\mathbf{x}} f(\mathbf{x}_\mathsf{F}) = \mathbf{0} \quad \text{and} \quad \nabla_{\mathbf{xx}} f(\mathbf{x}_\mathsf{F}) \text{ is positive definite}}$$

This Tutorial Example has an inactive constraint

Problem:

Our constrained optimization problem

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

where

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

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.

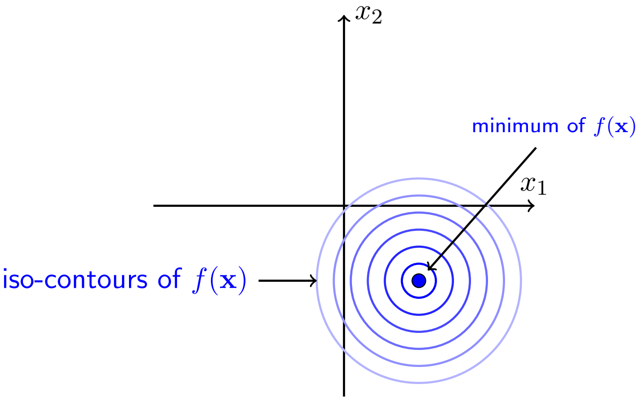
Problem:

This is the constrained optimization problem we want to solve

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

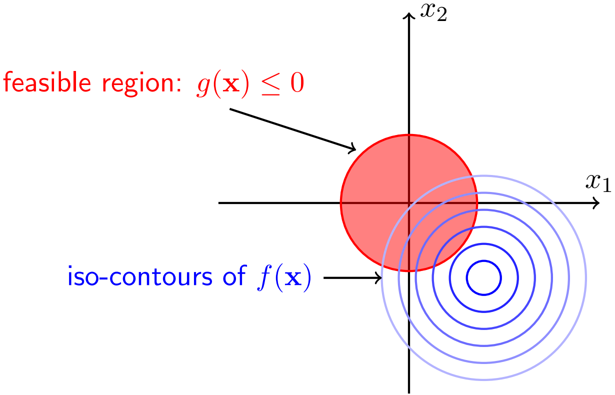
where

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



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

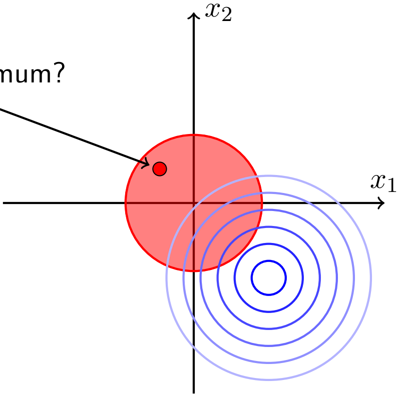
Tutorial example - Feasible region



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

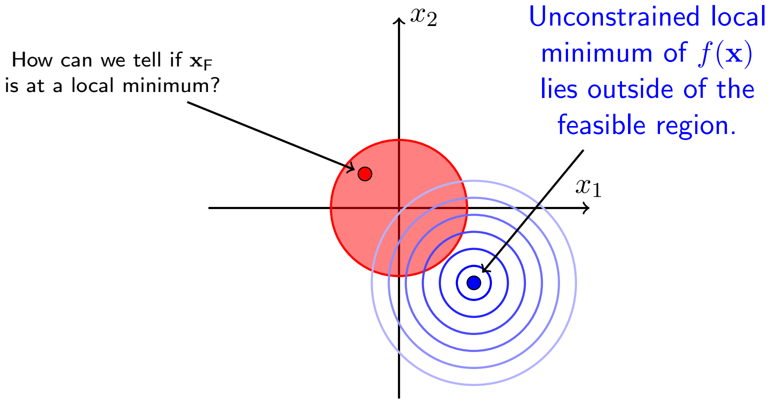
How do we recognize if $\mathbf{x}_{\mathsf{F}}$ is at a local optimum?

Is $\mathbf{x}_{\mathsf{F}}$ at a local minimum?



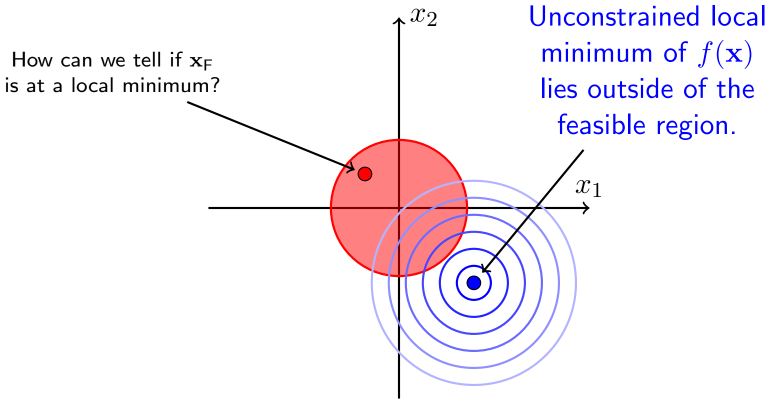
Remember $\mathbf{x}_{\mathsf{F}}$ denotes a feasible point.

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

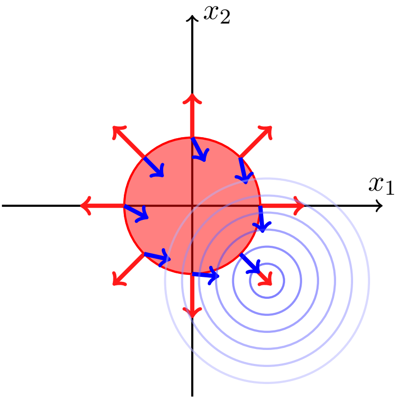


$\therefore$ the constrained local minimum occurs on the surface of the constraint surface.

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



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

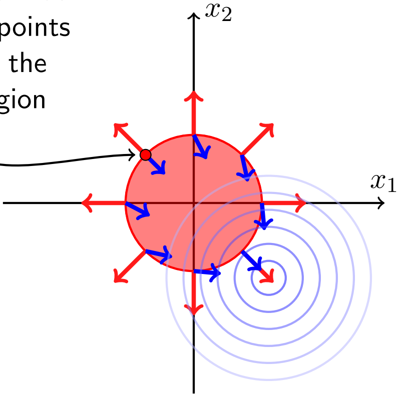


A local optimum occurs when $\nabla_{\mathbf{x}} f(\mathbf{x})$ and $\nabla_{\mathbf{x}} g(\mathbf{x})$ are parallel:

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

Want a constrained local minimum...

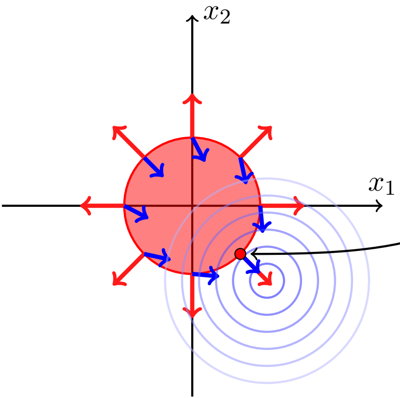
✗ **Not** a constrained local minimum as $-\nabla_{\mathbf{x}} f(\mathbf{x}_{\mathsf{F}})$ points in towards the feasible region



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

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

Want a constrained local minimum...



✓ **Is** a constrained local minimum as $-\nabla_{\mathbf{x}}\, f(\mathbf{x}_{\mathsf{F}})$ points away from the feasible region

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

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

Summary of optimization with one inequality constraint

Given

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

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. $\nabla_{\mathbf{x}} f(\mathbf{x}^*) = \mathbf{0}$
3. $\nabla_{\mathbf{xx}} f(\mathbf{x}^*)$ 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. $\mathbf{y}^t \nabla_{\mathbf{xx}} L(\mathbf{x}^*) \mathbf{y} \geq 0$ for all $\mathbf{y}$ orthogonal to $\nabla_{\mathbf{x}} g(\mathbf{x}^*)$.

Karush-Kuhn-Tucker conditions encode these conditions

Given the optimization problem

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

Define the **Lagrangian** as

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

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

1. $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \lambda^*) = \mathbf{0}$
2. $\lambda^* \geq 0$
3. $\lambda^*\, g(\mathbf{x}^*) = 0$
4. $g(\mathbf{x}^*) \leq 0$
5. Plus positive definite constraints on $\nabla_{\mathbf{xx}}\, \mathcal{L}(\mathbf{x}^*, \lambda^*)$.

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

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

Define the **Lagrangian** as

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

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 $j = 1, \ldots, m$
3. $\lambda_j^* \, g(\mathbf{x}^*) = 0$ for $j = 1, \ldots, m$
4. $g_j(\mathbf{x}^*) \leq 0$ for $j = 1, \ldots, m$
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

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

subject to

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

Define the **Lagrangian** as

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

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