

Lecture 3 - Optimality Conditions

Dinesh Krishnamoorthy

Department of Engineering Cybernetics, NTNU

Spring 2026

Unconstrained optimization

Consider the unconstrained optimization problem

$$\min_{\mathbf{x}} f(\mathbf{x})$$

Local Optimum

A point $\mathbf{x}^*$ is a local optimum if it yields the lowest cost of its feasible neighborhood.

$$f(\mathbf{x}^*) \leq f(\mathbf{x}), \forall \mathbf{x} \in N_\delta(\mathbf{x}^*) \cap \mathcal{F}$$

i.e., cost does not improve along any direction $\mathbf{d}$,

$$f(\mathbf{x}^*) \leq f(\mathbf{x}^* + \mathbf{d})$$

Unconstrained Optimization

Consider the unconstrained optimization problem

$$\min_{\mathbf{x}} f(\mathbf{x})$$

- ▶ 1st Order Taylor series expansion about $\mathbf{x}^*$:

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \sum_{i=1}^n \frac{\partial f(\mathbf{x}^*)}{\partial x_i} (x_i - x_i^*) + O(\|\mathbf{x} - \mathbf{x}^*\|)$$

- ▶ Vector notation:

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top \mathbf{d} + O(\|\mathbf{d}\|)$$

Necessary Condition for a Local Minimum

- Change in f :

$$\delta f(\mathbf{x}^*) = f(\mathbf{x}) - f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)^\top \mathbf{d} + O(\|\mathbf{d}\|)$$

- If $\mathbf{x}^*$ is a local minimizer of f , then $\delta f(\mathbf{x}^*) \geq 0$ for all $\mathbf{d} \in \mathbb{R}^n$:

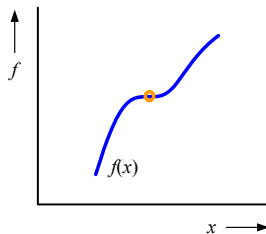
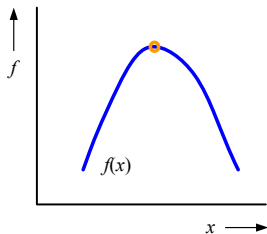
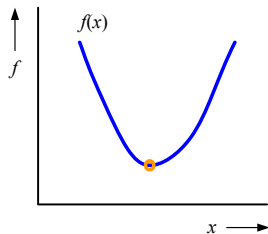
$$\rightarrow \quad \nabla f(\mathbf{x}^*) = 0$$

- If $\nabla f(\mathbf{x}^*) = 0$, then $\mathbf{x}^*$ is a **stationary point** of f .

Is this condition sufficient for local optimality?

Stationary Points of a Univariate Function

- Univariate function: $\frac{df}{dx} = 0$



- $\nabla f(x^*) = 0$ not sufficient!

→ Curvature information required!

Interior Minimum - Unconstrained

- Minimization problem:

$$\min_{\mathbf{x}} f(\mathbf{x})$$

- 2nd Order Taylor series expansion about $\mathbf{x}^*$:

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \sum_{i=1}^n \frac{\partial f(\mathbf{x}^*)}{\partial x_i} (x_i - x_i^*) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_i \partial x_j} (x_i - x_i^*)^2 + O(\|\mathbf{x} - \mathbf{x}^*\|^2)$$

- Vector notation:

$$\delta f(\mathbf{x}^*) = f(\mathbf{x}) - f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)^\top \mathbf{d} + \frac{1}{2} \mathbf{d}^\top \mathbf{H}(\mathbf{x}^*) \mathbf{d} + O(\|\mathbf{d}\|^2)$$

Sufficient Conditions for a Local Minimum

- Change in f with curvature information:

$$\delta f(\mathbf{x}^*) = f(\mathbf{x}) - f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)^\top \mathbf{d} + \frac{1}{2} \mathbf{d}^\top H(\mathbf{x}^*) \mathbf{d} + O(\|\mathbf{d}\|^2)$$

Sufficient Conditions for a Local Minimum

- Change in f with curvature information:

$$\delta f(\mathbf{x}^*) = f(\mathbf{x}) - f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)^\top \mathbf{d} + \frac{1}{2} \mathbf{d}^\top H(\mathbf{x}^*) \mathbf{d} + O(\|\mathbf{d}\|^2)$$

- $\mathbf{x}^*$ is a local minimizer of $f \implies \delta f(\mathbf{x}^*) \geq 0$ for all $\mathbf{d} \in \mathbb{R}^n$

Sufficient Conditions for a Local Minimum

- Change in f with curvature information:

$$\delta f(\mathbf{x}^*) = f(\mathbf{x}) - f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)^\top \mathbf{d} + \frac{1}{2} \mathbf{d}^\top H(\mathbf{x}^*) \mathbf{d} + O(\|\mathbf{d}\|^2)$$

- $\mathbf{x}^*$ is a local minimizer of $f \implies \delta f(\mathbf{x}^*) \geq 0$ for all $\mathbf{d} \in \mathbb{R}^n$
- $\delta f(\mathbf{x}^*) > 0$ for all $\mathbf{d} \in \mathbb{R}^n \implies \mathbf{x}^*$ is a local minimizer of f

Sufficient Conditions for a Local Minimum

- Change in f with curvature information:

$$\delta f(\mathbf{x}^*) = f(\mathbf{x}) - f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)^\top \mathbf{d} + \frac{1}{2} \mathbf{d}^\top H(\mathbf{x}^*) \mathbf{d} + O(\|\mathbf{d}\|^2)$$

- $\mathbf{x}^*$ is a local minimizer of $f \implies \delta f(\mathbf{x}^*) \geq 0$ for all $\mathbf{d} \in \mathbb{R}^n$
- $\delta f(\mathbf{x}^*) > 0$ for all $\mathbf{d} \in \mathbb{R}^n \implies \mathbf{x}^*$ is a local minimizer of f

$\begin{aligned} \nabla f(\mathbf{x}^*) &= 0 \\ \nabla f(\mathbf{x}^*) &= 0, \mathbf{d}^\top H(\mathbf{x}^*) \mathbf{d} > 0 \end{aligned}$	$\begin{aligned} & \text{(necessary)} \\ & \text{(sufficient)} \end{aligned}$
--	---

i.e., if at the stationary point the **Hessian** $H(\mathbf{x}^*)$ is **positive definite**, then for sure we have a **local** minimum.

What about convex functions?

Sufficient Conditions for a Local Minimum

- Change in f with curvature information:

$$\delta f(\mathbf{x}^*) = f(\mathbf{x}) - f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)^\top \mathbf{d} + \frac{1}{2} \mathbf{d}^\top H(\mathbf{x}^*) \mathbf{d} + O(\|\mathbf{d}\|^2)$$

- $\mathbf{x}^*$ is a local minimizer of $f \implies \delta f(\mathbf{x}^*) \geq 0$ for all $\mathbf{d} \in \mathbb{R}^n$
- $\delta f(\mathbf{x}^*) > 0$ for all $\mathbf{d} \in \mathbb{R}^n \implies \mathbf{x}^*$ is a local minimizer of f

$\begin{aligned} \nabla f(\mathbf{x}^*) &= 0 \\ \nabla f(\mathbf{x}^*) &= 0, \mathbf{d}^\top H(\mathbf{x}^*) \mathbf{d} > 0 \end{aligned}$	$\begin{aligned} & \text{(necessary)} \\ & \text{(sufficient)} \end{aligned}$
--	---

i.e., if at the stationary point the **Hessian** $H(\mathbf{x}^*)$ is **positive definite**, then for sure we have a **local** minimum.

What about convex functions?

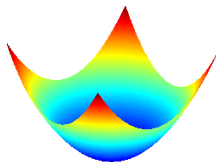
They have a PSD Hessian **for any** $\mathbf{x}$ in their domain, so that the stationarity condition is already necessary and sufficient.

Stationary Points of a Quadratic 2-D Function

► 2-D function: $f(\mathbf{x}) = \mathbf{x}^\top H \mathbf{x}$

$$H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

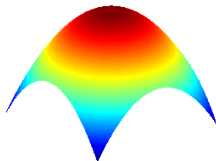
positive definite



minimum

$$H = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

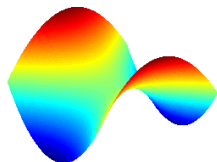
negative definite



maximum

$$H = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

indefinite



saddle point

Constrained Optima

- Minimization problem:

$$\begin{array}{ll}\min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{g}(\mathbf{x}) \leq 0 \\ & \mathbf{h}(\mathbf{x}) = 0\end{array}$$

- Necessary condition for $\mathbf{x}^*$ to be a minimizer for f :

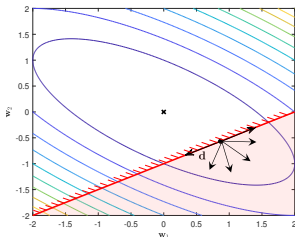
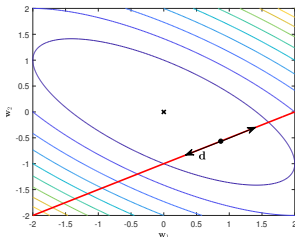
$$\delta f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)^\top \mathbf{d} + O(\|\mathbf{d}\|) \geq 0 \quad \forall \mathbf{d} \in \mathcal{F}$$

- $\mathbf{d} \in \mathcal{F}$: feasible directions

→ How to find $\mathbf{d} \in \mathcal{F}$?

Feasible directions

- $\mathbf{d}$ is a feasible direction for $h(\mathbf{x}) = 0$ and $g(\mathbf{x}) \leq 0$ in $\mathbf{x}^*$ if $\nabla h(\mathbf{x}^*)^\top \mathbf{d} = 0$, and $\nabla g(\mathbf{x}^*)^\top \mathbf{d} \leq 0$, respectively



- **Intuition:** No *admissible* direction in the feasible region $\mathbf{d} \in \mathcal{F}$, can improve the cost locally.
- Therefore, conditions that guarantee that a point has no feasible descent directions $\mathbf{d}$, are necessary conditions for optimality

Constrained Optima - Equality Constrained

- Conceptual idea: derive optimality conditions by reformulating into unconstrained problem
- Let's first consider only equality constraints:

$$\begin{array}{ll}\min_x & f(x) \\ \text{s.t.} & h(x) = 0\end{array}$$

- Lagrangian function:

$$\min_x f(x) + \lambda^\top h(x)$$

$\begin{array}{l} h(x^*) = 0 \\ \nabla f(x^*) + \nabla h(x^*)\lambda = 0 \end{array}$

(necessary conditions)

But what is a Lagrange multiplier? Why is this the necessary condition?

Reduced Gradient

► If $\mathbf{x}^* \in \mathbb{R}^n$ is a minimizer, then $\forall \mathbf{d} \in \mathcal{F} \subseteq \mathbb{R}^n$:

$$\begin{array}{lll} h(\mathbf{x}^*) & = & 0 \quad \text{(feasibility)} \\ \delta f(\mathbf{x}^*) & = & \nabla f(\mathbf{x}^*)^\top \mathbf{d} \geq 0 \quad \text{(optimality)} \\ \delta h_j(\mathbf{x}^*) & = & \nabla h_j(\mathbf{x}^*)^\top \mathbf{d} = 0 \quad j = 1, \dots, m \quad \text{(admissibility)} \end{array}$$

→ use $\delta h_j(\mathbf{x}^*) = 0$ to eliminate $m < n$ variables from $\mathbf{x}$

Reduced Gradient

- Let $p = n - m > 0$, take $\delta \mathbf{r} \in \mathbb{R}^p$ and $\delta \mathbf{s} \in \mathbb{R}^m$:

$$\mathbf{d} = \begin{bmatrix} \delta \mathbf{r} \\ \delta \mathbf{s} \end{bmatrix} \quad \begin{array}{ll} p & \text{decision variables (free)} \\ m \text{ (size of } h) & \text{state variables (dependent)} \end{array}$$

- Also:

$$\nabla f^\top = \left[\frac{\partial f}{\partial \mathbf{r}_1}, \dots, \frac{\partial f}{\partial \mathbf{r}_p}, \frac{\partial f}{\partial \mathbf{s}_1}, \dots, \frac{\partial f}{\partial \mathbf{s}_m} \right] \triangleq \left[\frac{\partial f}{\partial \mathbf{r}}^\top, \frac{\partial f}{\partial \mathbf{s}}^\top \right]$$

$$\nabla h_j^\top = \left[\frac{\partial h_j}{\partial \mathbf{r}_1}, \dots, \frac{\partial h_j}{\partial \mathbf{r}_p}, \frac{\partial h_j}{\partial \mathbf{s}_1}, \dots, \frac{\partial h_j}{\partial \mathbf{s}_m} \right] \triangleq \left[\frac{\partial h_j}{\partial \mathbf{r}}^\top, \frac{\partial h_j}{\partial \mathbf{s}}^\top \right]$$

Reduced Gradient

- Substitute in optimality conditions:

$$\delta f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)^\top \mathbf{d} = \frac{\partial f}{\partial \mathbf{r}}^\top \delta \mathbf{r} + \frac{\partial f}{\partial \mathbf{s}}^\top \delta \mathbf{s} \geq 0$$

$$\delta h_j(\mathbf{x}^*) = \nabla h_j(\mathbf{x}^*)^\top \mathbf{d} = \frac{\partial h_j}{\partial \mathbf{r}}^\top \delta \mathbf{r} + \frac{\partial h_j}{\partial \mathbf{s}}^\top \delta \mathbf{s} = 0 \quad j = 1, \dots, m$$

- Constraint equations in matrix form:

$$\delta \mathbf{h} = \frac{\partial \mathbf{h}}{\partial \mathbf{r}}^\top \delta \mathbf{r} + \frac{\partial \mathbf{h}}{\partial \mathbf{s}}^\top \delta \mathbf{s}$$

$$\frac{\partial \mathbf{h}}{\partial \mathbf{r}} = \begin{bmatrix} \frac{\partial h_1}{\partial r_1} & \dots & \frac{\partial h_m}{\partial r_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_1}{\partial r_p} & \dots & \frac{\partial h_m}{\partial r_p} \end{bmatrix}$$

$$\frac{\partial \mathbf{h}}{\partial \mathbf{s}} = \begin{bmatrix} \frac{\partial h_1}{\partial s_1} & \dots & \frac{\partial h_m}{\partial s_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_1}{\partial s_m} & \dots & \frac{\partial h_m}{\partial s_m} \end{bmatrix}$$

Reduced Gradient

- Optimality and admissibility conditions:

$$\begin{aligned}\delta f(\mathbf{x}^*) &= \frac{\partial f}{\partial \mathbf{r}}^\top \delta \mathbf{r} + \frac{\partial f}{\partial \mathbf{s}}^\top \delta \mathbf{s} \geq 0 \\ \delta h(\mathbf{x}^*) &= \frac{\partial h}{\partial \mathbf{r}}^\top \delta \mathbf{r} + \frac{\partial h}{\partial \mathbf{s}}^\top \delta \mathbf{s} = 0\end{aligned}$$

Reduced Gradient

- Optimality and admissibility conditions:

$$\begin{aligned}\delta f(x^*) &= \frac{\partial f}{\partial \mathbf{r}}^\top \delta \mathbf{r} + \frac{\partial f}{\partial \mathbf{s}}^\top \delta \mathbf{s} \geq 0 \\ \delta h(x^*) &= \frac{\partial h}{\partial \mathbf{r}}^\top \delta \mathbf{r} + \frac{\partial h}{\partial \mathbf{s}}^\top \delta \mathbf{s} = 0\end{aligned}$$

- Use bottom equation to eliminate $\delta \mathbf{s}$:

$$\delta \mathbf{s} = - \left[\frac{\partial h}{\partial \mathbf{s}} \right]^{-\top} \frac{\partial h}{\partial \mathbf{r}}^\top \delta \mathbf{r}$$

Reduced Gradient

- Optimality and admissibility conditions:

$$\begin{aligned}\delta f(x^*) &= \frac{\partial f}{\partial \mathbf{r}}^\top \delta \mathbf{r} + \frac{\partial f}{\partial \mathbf{s}}^\top \delta \mathbf{s} \geq 0 \\ \delta h(x^*) &= \frac{\partial h}{\partial \mathbf{r}}^\top \delta \mathbf{r} + \frac{\partial h}{\partial \mathbf{s}}^\top \delta \mathbf{s} = 0\end{aligned}$$

- Use bottom equation to eliminate $\delta \mathbf{s}$:

$$\delta \mathbf{s} = - \left[\frac{\partial h}{\partial \mathbf{s}} \right]^{-\top} \frac{\partial h}{\partial \mathbf{r}}^\top \delta \mathbf{r}$$

- Substitute in top expression:

$$\delta f(x^*) = \underbrace{\left\{ \frac{\partial f}{\partial \mathbf{r}}^\top - \frac{\partial f}{\partial \mathbf{s}}^\top \left[\frac{\partial h}{\partial \mathbf{s}} \right]^{-\top} \frac{\partial h}{\partial \mathbf{r}}^\top \right\}}_{\text{reduced gradient}} \delta \mathbf{r} \geq 0$$

Reduced Gradient

► If $\mathbf{x}^* \in \mathbb{R}^n$ is a minimizer, then $\forall \delta \mathbf{r} \in \mathbb{R}^p$:

$$\mathbf{h}(\mathbf{x}^*) = 0 \quad (\text{feasibility})$$

$$\frac{\partial f(\mathbf{x}^*)}{\partial \mathbf{r}} - \frac{\partial \mathbf{h}(\mathbf{x}^*)}{\partial \mathbf{r}} \left[\frac{\partial \mathbf{h}(\mathbf{x}^*)}{\partial \mathbf{s}} \right]^{-1} \frac{\partial f(\mathbf{x}^*)}{\partial \mathbf{s}} = 0 \quad (\text{optimality and admissibility})$$

Reduced Gradient

- If $\mathbf{x}^* \in \mathbb{R}^n$ is a minimizer, then $\forall \delta \mathbf{r} \in \mathbb{R}^p$:

$$\mathbf{h}(\mathbf{x}^*) = 0 \quad (\text{feasibility})$$

$$\frac{\partial f(\mathbf{x}^*)}{\partial \mathbf{r}} - \frac{\partial \mathbf{h}(\mathbf{x}^*)}{\partial \mathbf{r}} \left[\frac{\partial \mathbf{h}(\mathbf{x}^*)}{\partial \mathbf{s}} \right]^{-1} \frac{\partial f(\mathbf{x}^*)}{\partial \mathbf{s}} = 0 \quad (\text{optimality and admissibility})$$

- To reformulate in terms of $\nabla f(\mathbf{x}^*)$ and $\nabla h_j(\mathbf{x}^*)$, define:

$$\boldsymbol{\lambda} := - \left[\frac{\partial \mathbf{h}(\mathbf{x}^*)}{\partial \mathbf{s}} \right]^{-1} \frac{\partial f(\mathbf{x}^*)}{\partial \mathbf{s}} \in \mathbb{R}^m \quad (\text{Lagrange multipliers})$$

Reduced Gradient - Necessary Conditions

- With $\lambda = - \left[\frac{\partial h(x^*)}{\partial s} \right]^{-1} \frac{\partial f(x^*)}{\partial s}$, we have:

$$\begin{aligned} \frac{\partial f(x^*)}{\partial r} + \frac{\partial h(x^*)}{\partial r} \lambda &= 0 \\ \frac{\partial f(x^*)}{\partial s} + \frac{\partial h(x^*)}{\partial s} \lambda &= 0 \end{aligned}$$

- and $x = [r^\top, s^\top]^\top$ gives:

$$\frac{\partial f}{\partial x} + \frac{\partial h}{\partial x} \lambda = 0$$

- Now: if $x^* \in \mathbb{R}^n$ is a minimizer, then:

$\begin{aligned} h(x^*) &= 0 \\ \nabla f(x^*) + \nabla h(x^*) \lambda &= 0 \end{aligned}$	(necessary conditions)
---	------------------------

- Intuition: Objective gradient balanced by the constraint gradient

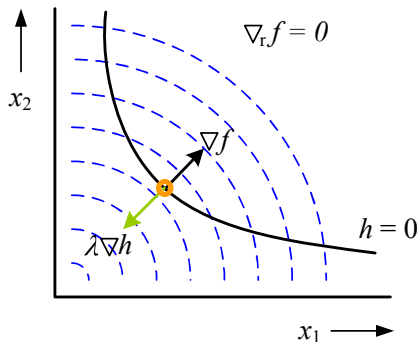
$$\nabla f(x^*) = -\nabla h(x^*) \lambda$$

Necessary Conditions - Illustration

Necessary conditions:

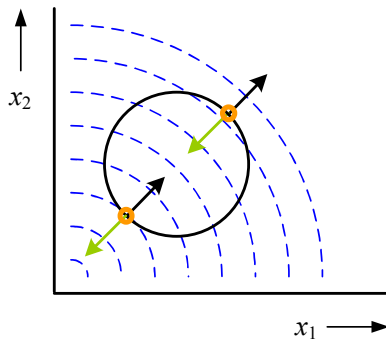
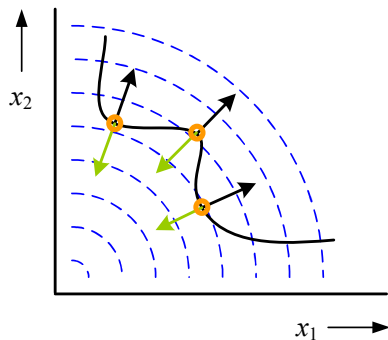
1. Feasible $h(x^*) = 0$
2. Objective gradient balanced by constraint gradient

$$\nabla f(x^*) = -\nabla h(x^*)\lambda$$



Also sufficient?

Necessary Conditions - Illustration



Stationary points both minimizers and maximizers

→ Not sufficient!

A “Trick” - the Lagrangian Function

- Define the **Lagrangian function** $L(x, \lambda)$:

$$L(x, \lambda) = f(x) + \lambda^\top h(x)$$

- Necessary conditions (stationary points) with $\lambda^* \neq 0$:

$$\begin{array}{l} \frac{\partial L(x^*, \lambda^*)}{\partial x} = 0 \rightarrow \\ \frac{\partial L(x^*, \lambda^*)}{\partial \lambda} = 0 \rightarrow \end{array} \quad \boxed{\begin{array}{l} \nabla f(x^*) + \nabla h(x^*) \lambda^* = 0 \\ h(x^*) = 0 \end{array}}$$

- Sufficient condition for x^* to be a minimizer:

A “Trick” - the Lagrangian Function

- Define the **Lagrangian function** $L(x, \lambda)$:

$$L(x, \lambda) = f(x) + \lambda^\top h(x)$$

- Necessary conditions (stationary points) with $\lambda^* \neq 0$:

$$\begin{array}{l} \frac{\partial L(x^*, \lambda^*)}{\partial x} = 0 \rightarrow \\ \frac{\partial L(x^*, \lambda^*)}{\partial \lambda} = 0 \rightarrow \end{array} \quad \boxed{\begin{array}{l} \nabla f(x^*) + \nabla h(x^*) \lambda^* = 0 \\ h(x^*) = 0 \end{array}}$$

- Sufficient condition for x^* to be a minimizer:

Hessian of Lagrange function with respect to x , $L_{xx}(x^*, \lambda^*)$, is **positive definite** on the subspace tangent to $h(x)$ at x^*

Lagrangian Function for Inequality Constraints

- ▶ Define the **Lagrangian function** $L(x, \mu)$:

$$L(x, \mu) = f(x) + \mu^\top g(x)$$

- ▶ Only active inequality constraints contribute
- ▶ Necessary conditions (stationary points) with $\mu^* \geq 0$:

$$\nabla f(x^*) + \nabla g(x^*)\mu^* = 0$$

$$g(x^*) \leq 0$$

$$g(x^*) \circ \mu^* = 0$$

$$\mu^* \geq 0$$

(Lagrange gradient = 0)

(feasibility)

(complementary slackness)

(dual feasibility)

- ▶ Sufficient condition for x^* to be a minimizer:

Lagrangian Function for Inequality Constraints

- Define the **Lagrangian function** $L(x, \mu)$:

$$L(x, \mu) = f(x) + \mu^\top g(x)$$

- Only active inequality constraints contribute
- Necessary conditions (stationary points) with $\mu^* \geq 0$:

$$\nabla f(x^*) + \nabla g(x^*)\mu^* = 0$$

$$g(x^*) \leq 0$$

$$g(x^*) \circ \mu^* = 0$$

$$\mu^* \geq 0$$

(Lagrange gradient = 0)

(feasibility)

(complementary slackness)

(dual feasibility)

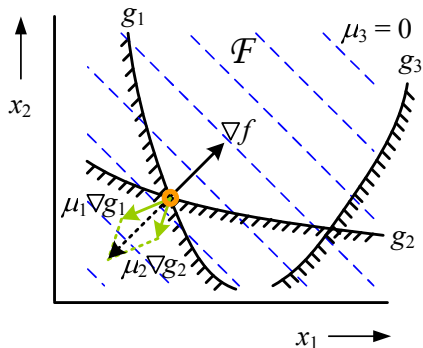
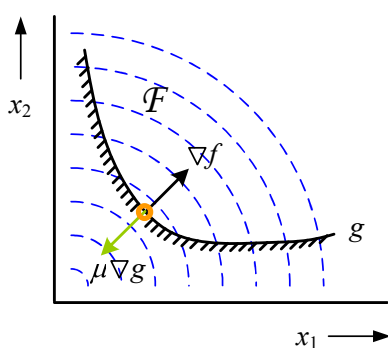
- Sufficient condition for x^* to be a minimizer:

Hessian of Lagrange function $L_{xx}(x^*, \mu^*)$ is **positive definite** on the subspace tangent to the **active** inequality constraints

Inequality Constraints - Illustration

Necessary conditions:

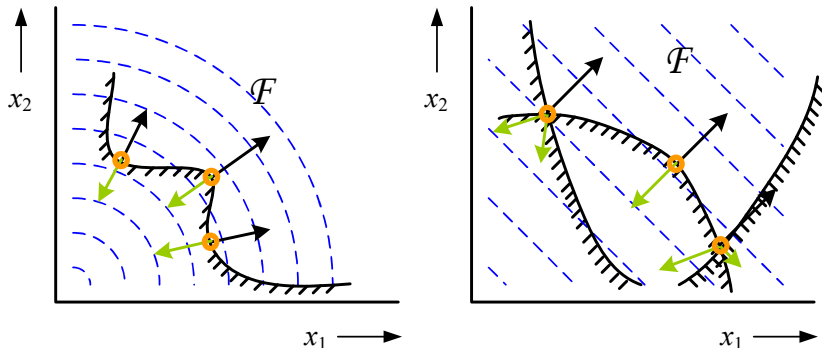
1. feasible
2. $\mu^* \geq 0$
3. objective balanced by **active** constraints
4. complementary slackness



Inequality Constraints - Illustration

Sufficient condition:

Hessian of Lagrange function $L_{xx}(x^*, \mu^*)$ is **positive definite** on the subspace tangent to the active inequality constraints



For case $m = n$ (no tangent subspace), KKT condition is sufficient

Intuition of the Lagrange multipliers

Physical interpretation

- ▶ Force exerted by the constraint
- ▶ $\mu \geq 0$ for inequality constraints implies that force acts only in one direction

$$\lambda > 0$$



$$\lambda < 0$$



Economic interpretation of Lagrange multipliers

- ▶ Marginal improvement in the cost by relaxing the constraint by one unit

$$\begin{aligned}\nabla f(x^*) + \nabla h(x^*)\lambda^* &= 0 \\ \Rightarrow \lambda^* &= -\nabla h(x^*)^{-1}\nabla f(x^*)\end{aligned}$$

i.e., rate of change in optimal value w.r.t. changes in constraint enforcement

- ▶ the sensitivity of the objective value $f(x^*)$ w.r.t the constraints
- ▶ Shadow price of the constraint
- ▶ $\mu \geq 0$ for inequality constraints implies that prices cannot be negative

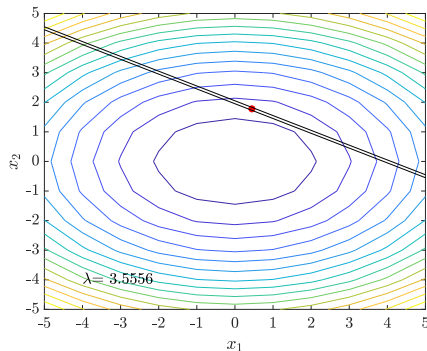
Intuitions on KKT point

Consider a ball moved down the hill...

$$c = 2$$

$$\min_{x_1, x_2} 2x_1^2 + x_2^2$$

$$\text{s.t. } -0.5x_1 - x_2 + c = 0$$



Intuition: λ tells us how hard the constraint is pushing or pulling the solution against the slope of the cost.

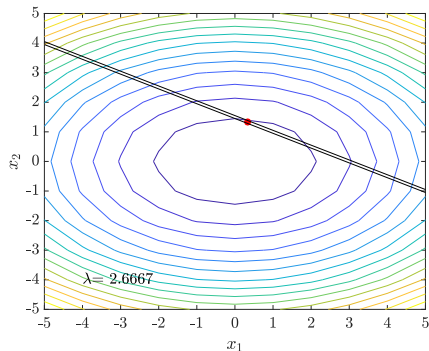
Intuitions on KKT point

Consider a ball moved down the hill...

$$c = 1.5$$

$$\min_{x_1, x_2} 2x_1^2 + x_2^2$$

$$\text{s.t. } -0.5x_1 - x_2 + c = 0$$



Intuition: λ tells us how hard the constraint is pushing or pulling the solution against the slope of the cost.

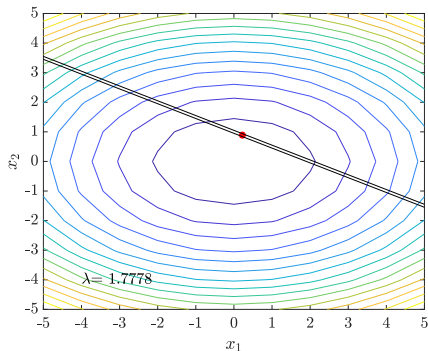
Intuitions on KKT point

Consider a ball moved down the hill...

$$c = 1$$

$$\min_{x_1, x_2} 2x_1^2 + x_2^2$$

$$\text{s.t. } -0.5x_1 - x_2 + c = 0$$



Intuition: λ tells us how hard the constraint is pushing or pulling the solution against the slope of the cost.

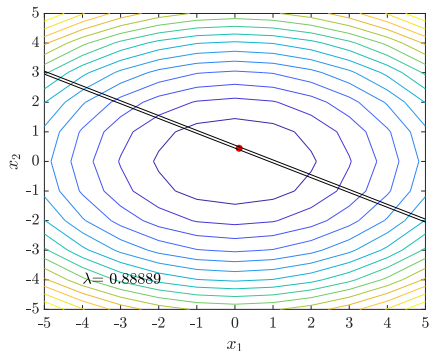
Intuitions on KKT point

Consider a ball moved down the hill...

$$c = 0.5$$

$$\min_{x_1, x_2} 2x_1^2 + x_2^2$$

$$\text{s.t. } -0.5x_1 - x_2 + c = 0$$



Intuition: λ tells us how hard the constraint is pushing or pulling the solution against the slope of the cost.

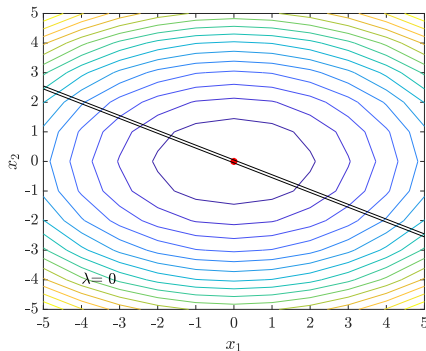
Intuitions on KKT point

Consider a ball moved down the hill...

$$c = 0$$

$$\min_{x_1, x_2} 2x_1^2 + x_2^2$$

$$\text{s.t. } -0.5x_1 - x_2 + c = 0$$



Intuition: λ tells us how hard the constraint is pushing or pulling the solution against the slope of the cost.

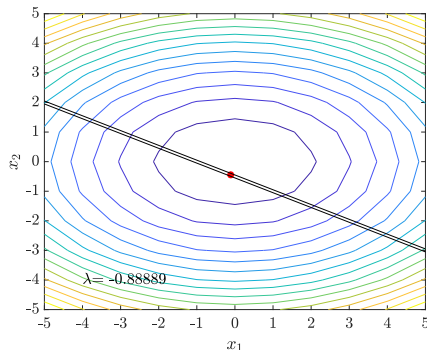
Intuitions on KKT point

Consider a ball moved down the hill...

$$c = -0.5$$

$$\min_{x_1, x_2} 2x_1^2 + x_2^2$$

$$\text{s.t. } -0.5x_1 - x_2 + c = 0$$



Intuition: λ tells us how hard the constraint is pushing or pulling the solution against the slope of the cost.

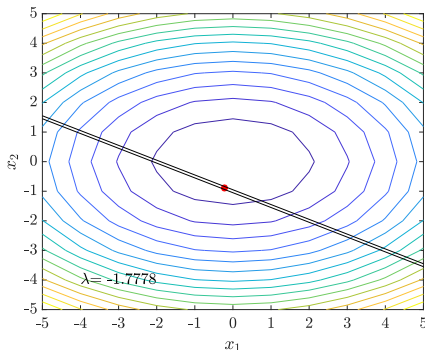
Intuitions on KKT point

Consider a ball moved down the hill...

$$c = -1$$

$$\min_{x_1, x_2} 2x_1^2 + x_2^2$$

$$\text{s.t. } -0.5x_1 - x_2 + c = 0$$



Intuition: λ tells us how hard the constraint is pushing or pulling the solution against the slope of the cost.

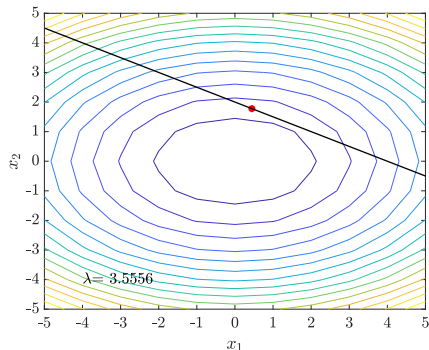
Intuitions on KKT point

Consider a Ball rolling down a hill...

$$c = 2$$

$$\min_{x_1, x_2} 2x_1^2 + x_2^2$$

$$\text{s.t. } -0.5x_1 - x_2 + c \leq 0$$



Intuition: λ tells us how hard the solution is pushing against constraint.

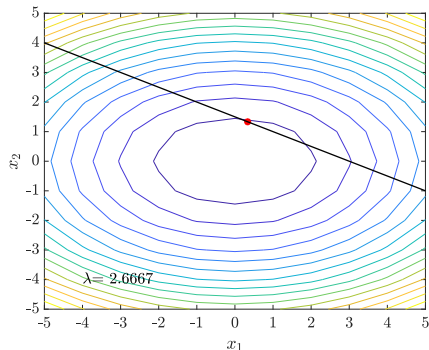
Intuitions on KKT point

Consider a Ball rolling down a hill...

$$c = 1.5$$

$$\min_{x_1, x_2} 2x_1^2 + x_2^2$$

$$\text{s.t. } -0.5x_1 - x_2 + c \leq 0$$



Intuition: λ tells us how hard the solution is pushing against constraint.

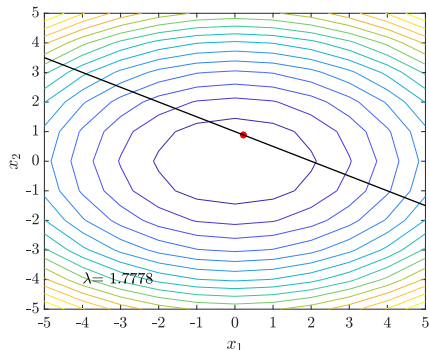
Intuitions on KKT point

Consider a Ball rolling down a hill...

$$c = 1$$

$$\min_{x_1, x_2} 2x_1^2 + x_2^2$$

$$\text{s.t. } -0.5x_1 - x_2 + c \leq 0$$



Intuition: λ tells us how hard the solution is pushing against constraint.

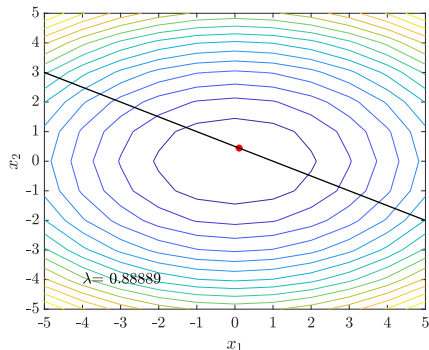
Intuitions on KKT point

Consider a Ball rolling down a hill...

$$c = 0.5$$

$$\min_{x_1, x_2} 2x_1^2 + x_2^2$$

$$\text{s.t. } -0.5x_1 - x_2 + c \leq 0$$



Intuition: λ tells us how hard the solution is pushing against constraint.

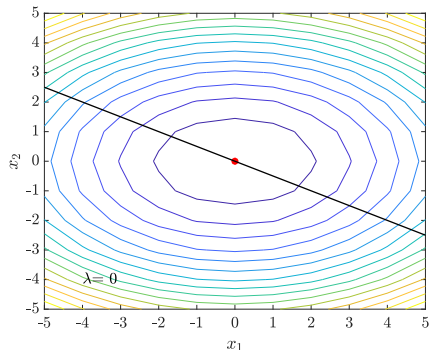
Intuitions on KKT point

Consider a Ball rolling down a hill...

$$c = 0$$

$$\min_{x_1, x_2} 2x_1^2 + x_2^2$$

$$\text{s.t. } -0.5x_1 - x_2 + c \leq 0$$



Intuition: λ tells us how hard the solution is pushing against constraint.

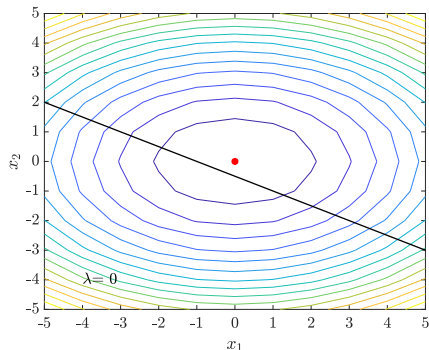
Intuitions on KKT point

Consider a Ball rolling down a hill...

$$c = -0.5$$

$$\min_{x_1, x_2} 2x_1^2 + x_2^2$$

$$\text{s.t. } -0.5x_1 - x_2 + c \leq 0$$



Intuition: λ tells us how hard the solution is pushing against constraint.

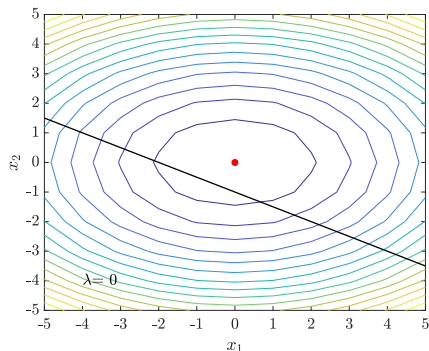
Intuitions on KKT point

Consider a Ball rolling down a hill...

$$c = -1$$

$$\min_{x_1, x_2} 2x_1^2 + x_2^2$$

$$\text{s.t. } -0.5x_1 - x_2 + c \leq 0$$



Intuition: λ tells us how hard the solution is pushing against constraint.

Intuitions on KKT point

For equality constraints, multipliers may have both signs, $\lambda_i \in \mathbb{R}$

For inequality constraints, multipliers must be non-negative $\mu_i \in \mathbb{R}_{\geq 0}$

- ▶ $\mu_i = 0, g_i(x) > 0 \Rightarrow$ inactive constraint
- ▶ $\mu_i > 0, g_i(x) = 0 \Rightarrow$ strictly active constraint
- ▶ $\mu_i = 0, g_i(x) = 0 \Rightarrow$ weakly active constraint

Strict Complementarity: No weakly active constraint

Active constraint set

For a point $x^* \in \mathcal{F}$, the index set i for which $g_i(x^*) = 0$ is called an **active set**, denoted by $\mathbb{A}$.

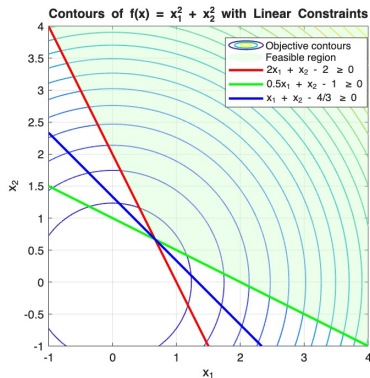
Inequality Constraints - Example

$$\min_{x_1, x_2} x_1^2 + x_2^2$$

$$\text{s.t. } 2x_1 + x_2 - 2 \geq 0$$

$$0.5x_1 + x_2 - 1 \geq 0$$

$$x_1 + x_2 - 4/3 \geq 0$$

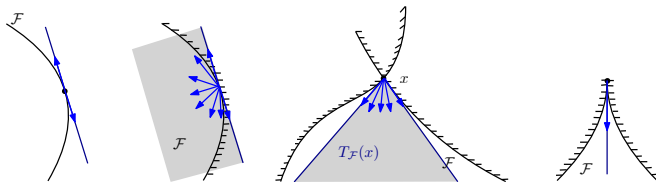


Is the KKT condition satisfied?

Geometric description of Feasible directions

Tangent cone

The tangent cone to a feasible set $\mathcal{F}$ at a point $x \in \mathcal{F}$, denoted by $T_{\mathcal{F}}(x)$, is a geometric description of the set of linearized feasible directions.



Note: The definition of $T_{\mathcal{F}}(x)$ depends on the geometry of the feasible set $\mathcal{F}$.

Geometric Optimality condition

Descent direction

Set of directions that improve the cost function

$$\mathcal{D}_f = \{\mathbf{d} \in \mathbb{R}^n \mid \nabla f(\mathbf{x})^\top \mathbf{d} < 0\}$$

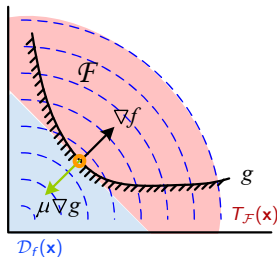
Taylor series expansion around $f(\mathbf{x})$: $f(\mathbf{x} + \mathbf{d}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^\top \mathbf{d}$

Geometric optimality condition

If $\mathbf{x}^*$ is a local optimum, then

$$\mathcal{D}_f(\mathbf{x}^*) \cap T_{\mathcal{F}}(\mathbf{x}^*) = \emptyset$$

i.e., the cost cannot improve along any feasible direction



Algebraic description of Feasible directions

Linearized feasible directions

Given a feasible point x and the active constraint set $\mathbb{A}(x)$, the set of linearized feasible directions $\mathcal{D}_{\mathcal{F}}(x)$ is

$$\mathcal{D}_{\mathcal{F}} = \left\{ \mathbf{d} \mid \begin{array}{l} \mathbf{d}^{\top} \nabla h(x) = 0 \\ \mathbf{d}^{\top} \nabla g_{\mathbb{A}}(x) = 0 \end{array} \right\}$$

- ▶ Constraint qualifications are needed to ensure that geometric ($\mathcal{T}_{\mathcal{F}}$) and algebraic descriptions ($\mathcal{D}_{\mathcal{F}}$) are equivalent.
- ▶ These conditions ensure that $\mathcal{D}_{\mathcal{F}}(x)$ constructed by linearizing the algebraic description of $\mathcal{F}$ captures the essential geometric features of $\mathcal{F}$ in the vicinity of x , as represented by $\mathcal{T}_{\mathcal{F}}(x)$.

The most used constraint qualification is LICQ.

Regularity

- ▶ A point $x^* \in \mathcal{F}$ is said to be **regular** if the gradient of the equality constraints $\nabla h(x^*)$ and the active inequality constraints $\nabla g_{\mathbb{A}}(x^*)$ have **full row rank**.

$$\text{rank} \left(\begin{bmatrix} \nabla h(x^*) \\ \nabla g_{\mathbb{A}}(x^*) \end{bmatrix} \right) = m_h + \text{card}(\mathbb{A})$$

- ▶ Point $x^* \in \mathcal{F}$ is also then said to respect Linear Independence Constraint Qualification (LICQ)

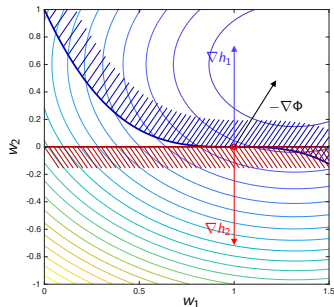
The first order necessary condition of optimality can only be formulated if a **constraint qualification** is satisfied.

LICQ

Linear Independence Constraint Qualification (LICQ) holds at $x^* \in \mathcal{F}$ iff $\nabla g(x^*)$ and $\nabla h_{\mathbb{A}}(x^*)$ are linearly independent.

- ▶ **Consequence of violating LICQ**: Inability to solve the system of equations arising from the KKT conditions.

Example 1: Failure of LICQ



Optimum $x^* = [1, 0]^T$

$$\min_x f(x) = (x_1 - 1.35)^2 + (x_2 - 0.6)^2$$

$$\text{s.t. } g_1(x) = x_2 - (1 - x_1)^3 \leq 0$$

$$g_2(x) = -x_2 \leq 0$$

$$\nabla f(x^*) = \begin{bmatrix} 2(x_1^* - 1.35) \\ 2(x_2^* - 0.6) \end{bmatrix} = \begin{bmatrix} -0.7 \\ -1.2 \end{bmatrix}$$

$$\nabla g_1(x^*) = \begin{bmatrix} 3(1 - x_1^*)^2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\nabla g_2(x^*) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

$$\nabla \mathcal{L} = \nabla f(x^*) + \mu_1 \nabla g_1(x^*) + \mu_2 \nabla g_2(x^*)$$

$$= \begin{bmatrix} -0.7 \\ \mu_1 - \mu_2 - 1.2 \end{bmatrix} \neq 0$$

KKT condition not satisfied!

Intuition of the Lagrange multipliers

Physical interpretation

- ▶ Force exerted by the constraint
- ▶ $\mu \geq 0$ for inequality constraints implies that force acts only in one direction

$$\lambda > 0$$



$$\lambda < 0$$



Intuition of the Lagrange multipliers

Physical interpretation

- ▶ Force exerted by the constraint
- ▶ $\mu \geq 0$ for inequality constraints implies that force acts only in one direction

$$\lambda > 0$$



$$\lambda < 0$$



What happens if constraint qualifications do not hold?

- ▶ Forces undefined/non-unique (i.e. forces exerted by the constraints cannot balance the slope of the cost function)

$$\nabla g(x^*)\mu \neq -\nabla f(x^*)$$

Example 2

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

$$\text{s.t. } h_1(x) = x_1 + x_2 - 1 = 0$$

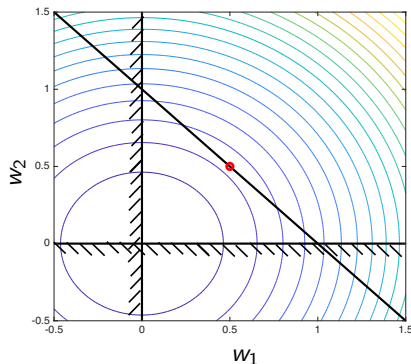
$$g_1(x) = -x_1 \leq 0$$

$$g_2(x) = -x_2 \leq 0$$

$$\nabla f = [2x_1, 2x_2]^T = [1, 1]^T$$

$$\nabla h_1 = [1, 1]^T,$$

$$\nabla g_1 = [-1, 0]^T, \nabla g_2 = [0, -1]^T$$



$$\text{Optimum } x^* = [0.5, 0.5]^T$$

Does LICQ hold?

KKT conditions

If x^* is a local minimizer of

$$\begin{aligned} \min_x & f(x) \\ \text{s.t. } & h(x) = 0 \\ & g(x) \leq 0 \end{aligned}$$

and LICQ holds at x^* ,

For two vectors of the same size $a, b \in \mathbb{R}^n$, $a \circ b$ denotes the Hadamard product (i.e., element-wise product).

KKT conditions

If x^* is a local minimizer of

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & h(x) = 0 \\ & g(x) \leq 0 \end{aligned}$$

and LICQ holds at x^* , then there exists a unique vector $\lambda^* \in \mathbb{R}^{n_h}$ and $\mu^* \in \mathbb{R}^{n_g}$ such that

$$\nabla f(x^*) + \nabla h(x^*)\lambda^* + \nabla g(x^*)\mu^* = 0 \quad (\text{Stationarity condition})$$

$$h(x^*) = 0 \quad (\text{Primal feasibility})$$

$$g(x^*) \leq 0 \quad (\text{Primal feasibility})$$

$$\mu^* \geq 0 \quad (\text{Dual feasibility})$$

$$\mu^* \circ g(x^*) = 0 \quad (\text{Complementary slackness})$$

For two vectors of the same size $a, b \in \mathbb{R}^n$, $a \circ b$ denotes the Hadamard product (i.e., element-wise product).

KKT conditions

If x^* is a local minimizer of

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & h(x) = 0 \\ & g(x) \leq 0 \end{aligned}$$

and LICQ holds at x^* , then there exists a unique vector $\lambda^* \in \mathbb{R}^{n_h}$ and $\mu^* \in \mathbb{R}^{n_g}$ such that

$$\nabla f(x^*) + \nabla h(x^*)\lambda^* + \nabla g(x^*)\mu^* = 0 \quad (\text{Stationarity condition})$$

$$h(x^*) = 0 \quad (\text{Primal feasibility})$$

$$g(x^*) \leq 0 \quad (\text{Primal feasibility})$$

$$\mu^* \geq 0 \quad (\text{Dual feasibility})$$

$$\mu^* \circ g(x^*) = 0 \quad (\text{Complementary slackness})$$

and $[x^*, \lambda^*, \mu^*]^T$ is called a **KKT point**.

For two vectors of the same size $a, b \in \mathbb{R}^n$, $a \circ b$ denotes the Hadamard product (i.e., element-wise product).

Convex problems

Slater Constraint Qualification (strictly feasible)

$$\exists x \quad \text{s.t.} \quad h(x) = 0 \text{ and } g(x) < 0$$

KKT conditions for convex problems

If $f(x)$ and $g(x)$ are convex, and $h(x)$ is affine, satisfying the Slater's condition, then the KKT conditions are necessary and sufficient for optimality.

Feasible Directions and the Critical Cone

- **Linearized feasible directions** at a feasible point x^* :

$$\mathcal{D}_{\mathcal{F}}(x^*) := \left\{ d \mid \begin{array}{l} \nabla h(x^*)d = 0, \\ \nabla g_{\mathbb{A}}(x^*)d \leq 0 \end{array} \right\}$$

- **Critical cone** at a KKT point (x^*, λ^*, μ^*) :

$$\mathcal{C}(x^*, \lambda^*, \mu^*) := \left\{ d \in \mathcal{D}_{\mathcal{F}}(x^*) \mid \begin{array}{ll} \nabla h(x^*)d = 0, & \\ \nabla g_{\mathbb{A}_+}(x^*)d = 0, & \mu_{\mathbb{A}_+} > 0 \\ \nabla g_{\mathbb{A}_0}(x^*)d = 0, & \mu_{\mathbb{A}_0} = 0 \end{array} \right\}$$

Second-Order Sufficient Condition (SOSC)

Let (x^*, λ^*, μ^*) satisfy the KKT conditions.

- ▶ The **Second-Order Sufficient Condition (SOSC)** holds if

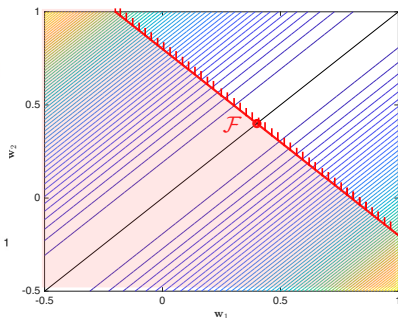
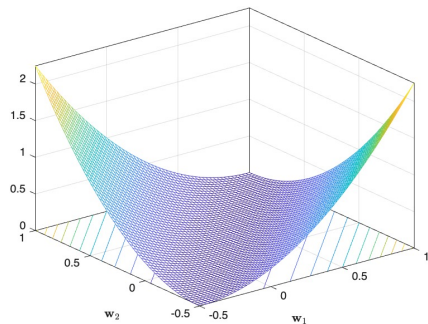
$$d^\top \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*, \mu^*) d > 0 \quad \forall d \in \mathcal{C}(x^*, \lambda^*, \mu^*) \setminus \{0\}$$

- ▶ Under a suitable constraint qualification, SOSC implies that x^* is a **strict local minimizer**.

Second order sufficient conditions

Set of feasible directions

$$\mathcal{F} := \{\mathbf{d} \mid \nabla h(\mathbf{x}^*)^\top \mathbf{d} = 0, \nabla g_{\mathbb{A}}(\mathbf{x}^*)^\top \mathbf{d} \leq 0\}$$

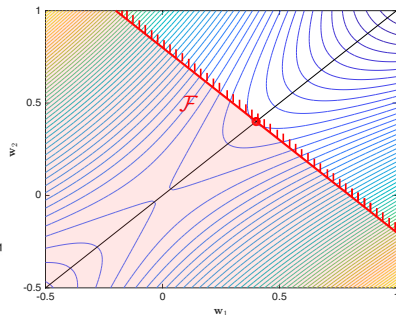
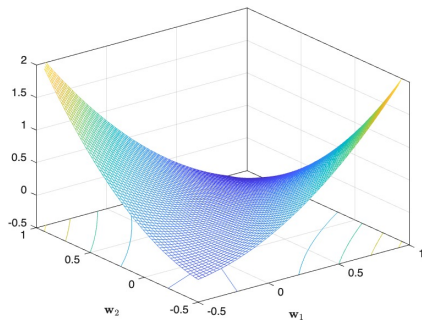


SOSC does not holds

Second order sufficient conditions

Set of feasible directions

$$\mathcal{F} := \{\mathbf{d} \mid \nabla h(\mathbf{x}^*)^\top \mathbf{d} = 0, \nabla g_{\mathbb{A}}(\mathbf{x}^*)^\top \mathbf{d} \leq 0\}$$



SOSC holds!

Duality

$$V^* = \min_x f(x) \quad \text{s.t. } g(x) \leq 0$$

The Lagrangian of this problem is

$$\mathcal{L}(x, \mu) = f(x) + \mu^T g(x)$$

Define the dual objective function

$$q(\mu) := \inf_x \mathcal{L}(x, \mu)$$

Dual problem

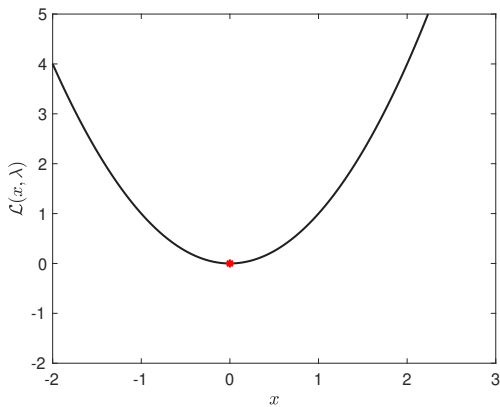
$$\max_{\mu} q(\mu) \quad \text{s.t. } \mu \geq 0$$

$$\Rightarrow V^* = \max_{\mu \geq 0} \min_x \mathcal{L}(x, \mu)$$

Duality

$$\mathcal{L}(x, \mu) = x^2 + \mu(1 - x)$$

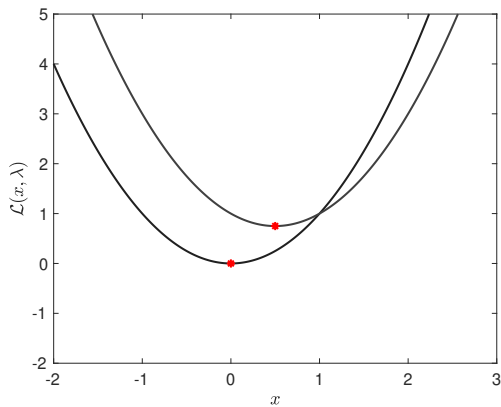
hold $\mu = 0$



Duality

$$\mathcal{L}(x, \mu) = x^2 + \mu(1 - x)$$

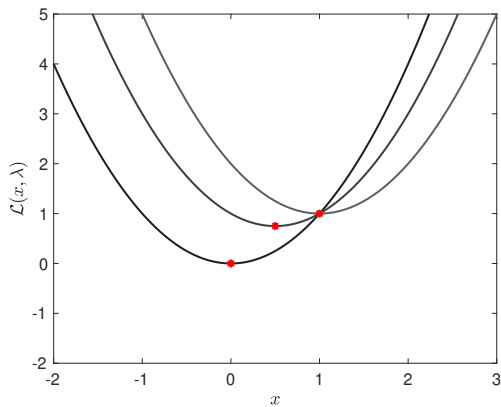
hold $\mu = 1$



Duality

$$\mathcal{L}(x, \mu) = x^2 + \mu(1 - x)$$

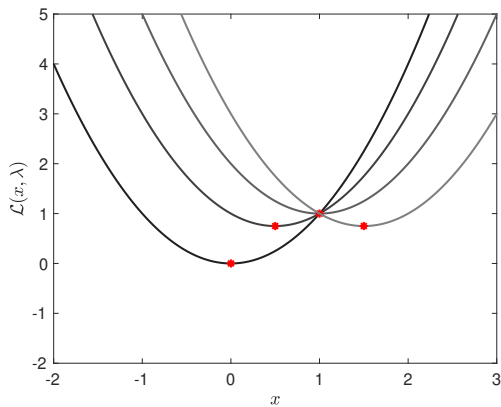
hold $\mu = 2$



Duality

$$\mathcal{L}(x, \mu) = x^2 + \mu(1 - x)$$

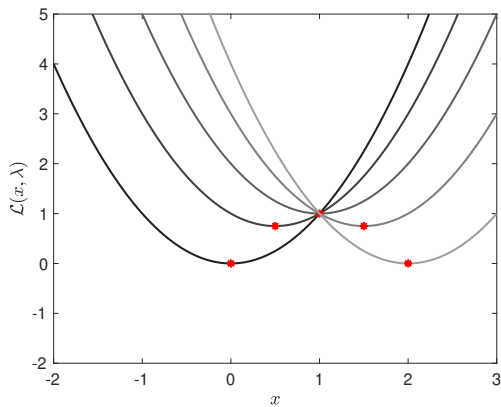
hold $\mu = 3$



Duality

$$\mathcal{L}(x, \mu) = x^2 + \mu(1 - x)$$

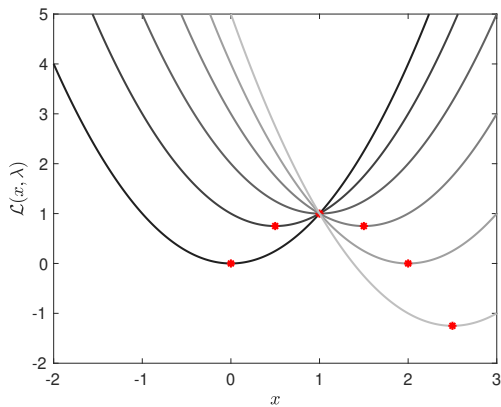
hold $\mu = 4$



Duality

$$\mathcal{L}(x, \mu) = x^2 + \mu(1 - x)$$

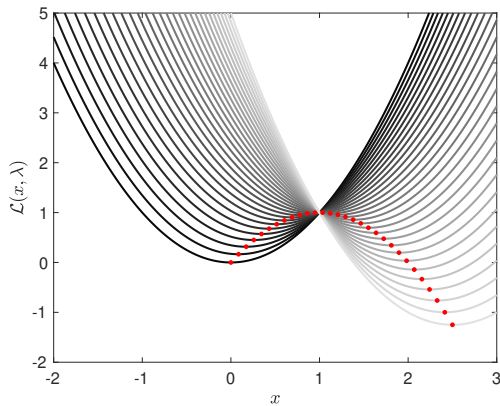
hold $\mu = 5$



Duality

$$\mathcal{L}(x, \mu) = x^2 + \mu(1 - x)$$

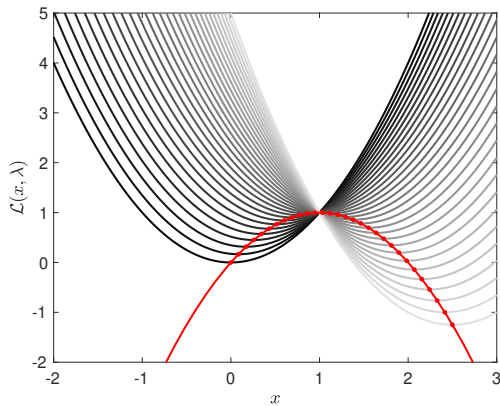
hold μ between $[0, 5]$



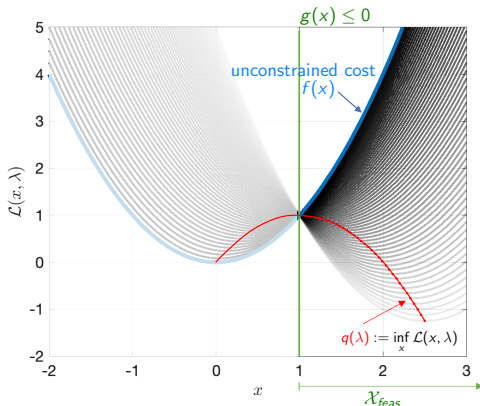
Duality

$$\mathcal{L}(x, \mu) = x^2 + \mu(1 - x)$$

$$q(\mu) = \inf_x \mathcal{L}(x, \mu)$$



Duality



$$V^* = \min_x f(x) \quad \text{s.t. } g(x) \leq 0$$

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

Define the dual function

$$q(\lambda) := \inf_x \mathcal{L}(x, \lambda)$$

$q(\lambda)$ is always concave and has the property

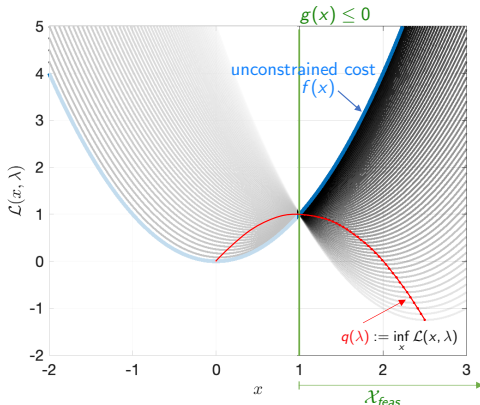
$$V^* \geq q(\lambda)$$

Proof: For any feasible $x \in \mathcal{X}_{feas}$ and $\lambda \geq 0$

$$f(x) \geq \mathcal{L}(x, \lambda)$$

Why? Because for any feasible x , $g(x) \leq 0$ (-ve) and $\lambda \geq 0$ (+ve)

Duality



$$V^* = \min_x f(x) \quad \text{s.t. } g(x) \leq 0$$

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

Define the dual function

$$q(\lambda) := \inf_x \mathcal{L}(x, \lambda)$$

$q(\lambda)$ is always concave and has the property

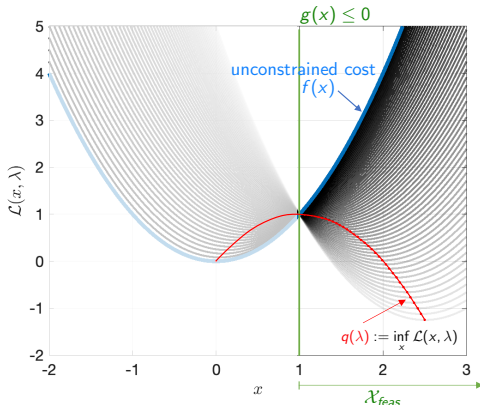
$$V^* \geq q(\lambda)$$

Proof: For any feasible $x \in \mathcal{X}_{feas}$ and $\lambda \geq 0$

$$\Rightarrow f(x) \geq \mathcal{L}(x, \lambda) \geq \inf_{x \in \mathcal{X}_{feas}} \mathcal{L}(x, \lambda) = q(\lambda)$$

At the optimal $x^* \in \mathcal{X}_{feas}$, $f(x^*) = V^*$, which proves our lower bound property

Duality



$$V^* = \min_x f(x) \quad \text{s.t. } g(x) \leq 0$$

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

Define the dual function

$$q(\lambda) := \inf_x \mathcal{L}(x, \lambda)$$

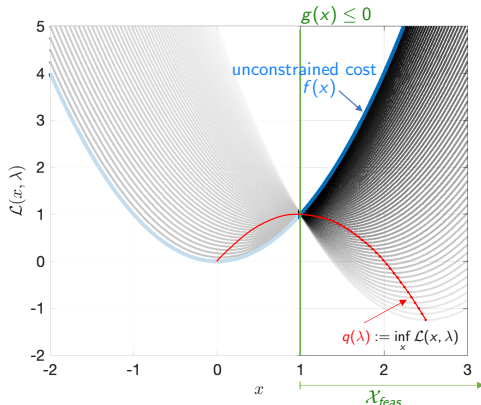
$q(\lambda)$ is always concave and has the property

$$V^* \geq q(\lambda)$$

Now that we have proven, that $q(\lambda)$ is always lower than the optimal cost V^* , in order to get the best lower bound (i.e. get as close to V^* as possible), we need to maximize the dual function.

$$\Rightarrow \max_{\lambda} q(\lambda)$$

Duality



$$V^* = \min_x f(x) \quad \text{s.t. } g(x) \leq 0$$

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

Define the dual function

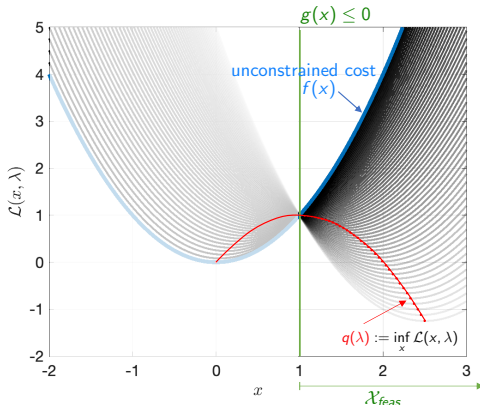
$$q(\lambda) := \inf_x \mathcal{L}(x, \lambda)$$

$q(\lambda)$ is always concave and has the property

$$V^* \geq q(\lambda)$$



Duality



$$V^* = \min_x f(x) \quad \text{s.t. } g(x) \leq 0$$

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

Define the dual function

$$q(\lambda) := \inf_x \mathcal{L}(x, \lambda)$$

$q(\lambda)$ is always concave and has the property

$$V^* \geq q(\lambda)$$

Why infimum in $q(\lambda) = \inf_x \mathcal{L}(x, \lambda)$?

For some very large λ in the feasible region, $\mathcal{L}(x, \lambda) \rightarrow -\infty$. Infimum of $\mathcal{L}(x, \lambda)$ denotes to a point which is smaller than every other value $\mathcal{L}(x, \lambda)$ can take for some very large $\lambda \rightarrow \infty$.

Duality

The Lagrangian dual function is defined as

$$q(\lambda, \mu) := \inf_{x \in \mathcal{F}} L(x, \lambda, \mu) = \inf_{x \in \mathcal{F}} \{f(x) + \lambda^\top h(x) + \mu^\top g(x)\}$$

and is **concave** even when the optimization problem is not.

It gives a **lower bound** on the optimal value p^* of the primal problem

$$q(\lambda, \mu) \leq V^* := f(x^*)$$

The Lagrange Dual Problem

What is the best lower bound we could get?

The Lagrange Dual Problem

What is the best lower bound we could get?

The one with the highest value.

The Lagrange dual problem is in fact defined as

$$\max_{\lambda, \mu} q(\lambda, \mu) =: q^* \quad (1)$$

$$\text{s.t. } \mu \geq 0 \quad (2)$$

What is a nice property of this problem?

The Lagrange Dual Problem

What is the best lower bound we could get?

The one with the highest value.

The Lagrange dual problem is in fact defined as

$$\max_{\lambda, \mu} q(\lambda, \mu) =: q^* \quad (1)$$

$$\text{s.t. } \mu \geq 0 \quad (2)$$

What is a nice property of this problem?

Convex feasible domain and concave objective function to maximize...It is a convex optimization problem!

We now know that

$$q^* \leq V^*,$$

i.e., the dual problem provides lower bound to the primal problem: We call it **weak duality**.

Strong Duality

In some cases, **strong duality** holds, i.e., $q^* = V^*$.

Why is it useful?

Strong Duality

In some cases, **strong duality** holds, i.e., $q^* = V^*$.

Why is it useful?

Can solve the dual problem (1) instead of the primal problem!

In general, strong duality does not hold. Yet, one special case where strong duality always holds is when

- ▶ the optimization problem is convex and
- ▶ **Slater's condition** holds