Introduction to Optimization

Vincent Leclère (CERMICS, ENPC)

Autumn 2026

Why should I bother to learn this stuff?

- Real designs are limited by many constraints: capacities, safety limits, budgets, balances, operating bounds.
- KKT conditions give a systematic way to understand which constraints determine an optimum.
- Optimization algorithms often compute more than a candidate solution: they also produce bounds and dual information.
- Duality is one of the basic mechanisms used by optimization solvers to assess solution quality.

Contents

First-order necessary optimality condition — recall

Consider a differentiable function f and

$$\underset{x \in X}{\text{Min}} \; f(x).$$

Recall the tangent cone at a feasible point x:

$$T_X(x) := \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \exists\, t_k \searrow 0, \;\exists\, d^k \to d, \\ x + t_k d^k \in X \quad \forall k \end{array} \right\}.$$

If $x^\sharp$ is a local minimum, then

$$\boxed{\nabla f(x^\sharp)^\top d \geq 0 \qquad \forall d \in T_X(x^\sharp).}$$

In the unconstrained case $X = \mathbb{R}^n$, this reduces to $\nabla f(x^\sharp) = 0$.

First-order necessary optimality condition — recall

Consider a differentiable function f and

$$\underset{x \in X}{\mathrm{Min}}\; f(x).$$

Recall the tangent cone at a feasible point x:

$$T_X(x) := \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \exists\, t_k \searrow 0,\ \exists\, d^k \to d, \\ x + t_k d^k \in X \quad \forall k \end{array} \right\}.$$

If $x^\sharp$ is a local minimum, then

$$\boxed{\nabla f(x^\sharp)^\top d \geq 0 \qquad \forall d \in T_X(x^\sharp).}$$

In the unconstrained case $X = \mathbb{R}^n$, this reduces to $\nabla f(x^\sharp) = 0$.

First-order necessary optimality condition — recall

Consider a differentiable function f and

$$\underset{x \in X}{\text{Min}} \, f(x).$$

Recall the tangent cone at a feasible point x:

$$T_X(x) := \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \exists\, t_k \searrow 0, \;\exists\, d^k \to d, \\ x + t_k d^k \in X \quad \forall k \end{array} \right\}.$$

If $x^\sharp$ is a local minimum, then

$$\boxed{\nabla f(x^\sharp)^\top d \geq 0 \qquad \forall d \in T_X(x^\sharp).}$$

In the unconstrained case $X = \mathbb{R}^n$, this reduces to $\nabla f(x^\sharp) = 0$.

Contents

Differentiable constrained problems

We now consider

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Min}} \quad & f(x) && (P) \\ \text{s.t.} \quad & g_i(x) = 0 && \forall i \in [n_E], \\ & h_j(x) \leq 0 && \forall j \in [n_I], \end{aligned}$$

where f, the g_i, and the h_j are differentiable.

The feasible set is

$$X := \{x \in \mathbb{R}^n \mid g_i(x) = 0 \ \forall i \in [n_E], \quad h_j(x) \leq 0 \ \forall j \in [n_I]\}.$$

The active inequality constraints at $x \in X$ are indexed by

$$J_0(x) := \{j \in [n_I] \mid h_j(x) = 0\}.$$

Differentiable constrained problems

We now consider

$$
\begin{aligned}
\underset{x\in\mathbb{R}^n}{\text{Min}} \quad & f(x) && (P)\\
\text{s.t.} \quad & g_i(x) = 0 && \forall i \in [n_E],\\
& h_j(x) \leq 0 && \forall j \in [n_I],
\end{aligned}
$$

where f, the g_i, and the h_j are differentiable.

The feasible set is

$$X := \{x \in \mathbb{R}^n \mid g_i(x) = 0 \ \forall i \in [n_E], \quad h_j(x) \leq 0 \ \forall j \in [n_I]\}.$$

The active inequality constraints at $x \in X$ are indexed by

$$J_0(x) := \{j \in [n_I] \mid h_j(x) = 0\}.$$

Differentiable constrained problems

We now consider

$$
\begin{aligned}
\underset{x\in\mathbb{R}^n}{\text{Min}} \quad & f(x) && (P)\\
\text{s.t.} \quad & g_i(x) = 0 && \forall i \in [n_E],\\
& h_j(x) \leq 0 && \forall j \in [n_I],
\end{aligned}
$$

where f, the g_i, and the h_j are differentiable.

The feasible set is

$$X := \{x \in \mathbb{R}^n \mid g_i(x) = 0 \ \forall i \in [n_E], \quad h_j(x) \leq 0 \ \forall j \in [n_I]\}.$$

The active inequality constraints at $x \in X$ are indexed by

$$J_0(x) := \{j \in [n_I] \mid h_j(x) = 0\}.$$

Linearized tangent cone

At $x \in X$, first-order feasibility suggests

$$\nabla g_i(x)^\top d = 0, \qquad \nabla h_j(x)^\top d \leq 0 \quad \text{for } j \in J_0(x).$$

We therefore define the linearized tangent cone

$$T_X^\ell(x) := \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} \nabla g_i(x)^\top d = 0 & \forall i \in [n_E], \\ \nabla h_j(x)^\top d \leq 0 & \forall j \in J_0(x) \end{array} \right\}.$$

One always has

$$T_X(x) \subseteq T_X^\ell(x).$$

Exercise. Prove it.

Linearized tangent cone

At $x \in X$, first-order feasibility suggests

$$\nabla g_i(x)^\top d = 0, \qquad \nabla h_j(x)^\top d \leq 0 \quad \text{for } j \in J_0(x).$$

We therefore define the linearized tangent cone

$$T_X^\ell(x) := \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} \nabla g_i(x)^\top d = 0 & \forall i \in [n_E], \\ \nabla h_j(x)^\top d \leq 0 & \forall j \in J_0(x) \end{array} \right\}.$$

One always has

$$T_X(x) \subseteq T_X^\ell(x).$$

Exercise. Prove it.

Contents

Constraint qualification

The linearized cone is explicit, while the true tangent cone may be difficult to compute.

We say that the constraints are qualified at x if

$$\boxed{T_X(x) = T_X^{\ell}(x).}$$

Under qualification, the first-order necessary condition becomes

$$\nabla f(x^\sharp)^\top d \geq 0 \qquad \forall d \in T_X^{\ell}(x^\sharp),$$

which is described entirely by the gradients of the active constraints.

Constraint qualification

The linearized cone is explicit, while the true tangent cone may be difficult to compute.

We say that the constraints are qualified at x if

$$\boxed{T_X(x) = T_X^\ell(x).}$$

Under qualification, the first-order necessary condition becomes

$$\nabla f(x^\sharp)^\top d \geq 0 \qquad \forall d \in T_X^\ell(x^\sharp),$$

which is described entirely by the gradients of the active constraints.

Sufficient qualification conditions

Standard sufficient conditions for qualification at a feasible point x include:

1. **Affine constraints:** all g_i and h_j are affine.
2. **Slater's condition:** the g_i are affine, the h_j are convex, and there exists x_S such that
$$g_i(x_S) = 0 \quad \forall i, \qquad h_j(x_S) < 0 \quad \forall j.$$
3. **LICQ:** the active constraint gradients
$$\{\nabla g_i(x)\}_{i \in [n_E]} \cup \{\nabla h_j(x)\}_{j \in J_0(x)}$$
are linearly independent.

Contents

Karush–Kuhn–Tucker conditions

Theorem (KKT — first-order necessary condition)

Assume that f, the g_i, and the h_j are differentiable. Let $x^\sharp$ be a local minimum and assume that the constraints are qualified at $x^\sharp$.
Then there exist dual variables $\lambda \in \mathbb{R}^{n_E}$ and $\mu \in \mathbb{R}_+^{n_I}$ such that

$$\begin{cases} \nabla f(x^\sharp) + \sum_{i=1}^{n_E} \lambda_i \nabla g_i(x^\sharp) + \sum_{j=1}^{n_I} \mu_j \nabla h_j(x^\sharp) = 0 & \textit{stationarity,} \\ g(x^\sharp) = 0, \quad h(x^\sharp) \leq 0 & \textit{primal feasibility,} \\ \mu \geq 0 & \textit{dual feasibility,} \\ \mu_j h_j(x^\sharp) = 0 \quad \forall j \in [n_I] & \textit{complementary slackness.} \end{cases}$$

For each inequality, complementary slackness implies

$$h_j(x^\sharp) = 0 \quad \text{or} \quad \mu_j = 0.$$

Karush–Kuhn–Tucker conditions

Theorem (KKT — first-order necessary condition)

Assume that f, the g_i, and the h_j are differentiable. Let $x^\sharp$ be a local minimum and assume that the constraints are qualified at $x^\sharp$.
Then there exist dual variables $\lambda \in \mathbb{R}^{n_E}$ and $\mu \in \mathbb{R}^{n_I}_+$ such that

$$\begin{cases} \nabla f(x^\sharp) + \sum_{i=1}^{n_E} \lambda_i \nabla g_i(x^\sharp) + \sum_{j=1}^{n_I} \mu_j \nabla h_j(x^\sharp) = 0 & \textit{stationarity,} \\ g(x^\sharp) = 0, \quad h(x^\sharp) \leq 0 & \textit{primal feasibility,} \\ \mu \geq 0 & \textit{dual feasibility,} \\ \mu_j h_j(x^\sharp) = 0 \quad \forall j \in [n_I] & \textit{complementary slackness.} \end{cases}$$

For each inequality, complementary slackness implies

$$h_j(x^\sharp) = 0 \quad \text{or} \quad \mu_j = 0.$$

Some remarks about KKT conditions

- KKT extends the unconstrained condition $\nabla f(x^\sharp) = 0$ to differentiable constrained problems.
- Under qualification, KKT conditions are necessary for local optimality; non-minimizing points may also satisfy them.
- Without qualification, a local minimum may fail to satisfy the KKT conditions.
- In convex optimization (f and the h_j convex, the g_i affine) KKT conditions become sufficient for global optimality (see next lecture).
- In general, the KKT system is nonlinear and complementary slackness creates multiple active-set cases. It is nevertheless a central tool for analysis and algorithm design.

Some remarks about KKT conditions

- KKT extends the unconstrained condition $\nabla f(x^\sharp) = 0$ to differentiable constrained problems.
- Under qualification, KKT conditions are necessary for local optimality; non-minimizing points may also satisfy them.
- Without qualification, a local minimum may fail to satisfy the KKT conditions.
- In convex optimization (f and the h_j convex, the g_i affine) KKT conditions become sufficient for global optimality (see next lecture).
- In general, the KKT system is nonlinear and complementary slackness creates multiple active-set cases. It is nevertheless a central tool for analysis and algorithm design.

Some remarks about KKT conditions

- KKT extends the unconstrained condition $\nabla f(x^\sharp) = 0$ to differentiable constrained problems.
- Under qualification, KKT conditions are necessary for local optimality; non-minimizing points may also satisfy them.
- Without qualification, a local minimum may fail to satisfy the KKT conditions.
- In convex optimization (f and the h_j convex, the g_i affine) KKT conditions become sufficient for global optimality (see next lecture).
- In general, the KKT system is nonlinear and complementary slackness creates multiple active-set cases. It is nevertheless a central tool for analysis and algorithm design.

Contents

Indicator functions for constraints

For a set $C \subseteq \mathbb{R}^n$, define

$$\mathbb{I}_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{otherwise.} \end{cases}$$

Two useful representations are

$$\mathbb{I}_{g(x)=0} = \sup_{\lambda \in \mathbb{R}^{n_E}} \lambda^\top g(x),$$

and

$$\mathbb{I}_{h(x)\leq 0} = \sup_{\mu \in \mathbb{R}_+^{n_I}} \mu^\top h(x).$$

Equality dual variables are free; inequality dual variables are nonnegative.

Indicator functions for constraints

For a set $C \subseteq \mathbb{R}^n$, define

$$\mathbb{I}_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{otherwise.} \end{cases}$$

Two useful representations are

$$\mathbb{I}_{g(x)=0} = \sup_{\lambda \in \mathbb{R}^{n_E}} \lambda^\top g(x),$$

and

$$\mathbb{I}_{h(x)\leq 0} = \sup_{\mu \in \mathbb{R}_+^{n_I}} \mu^\top h(x).$$

Equality dual variables are free; inequality dual variables are nonnegative.

Indicator functions for constraints

For a set $C \subseteq \mathbb{R}^n$, define

$$\mathbb{I}_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{otherwise.} \end{cases}$$

Two useful representations are

$$\mathbb{I}_{g(x)=0} = \sup_{\lambda \in \mathbb{R}^{n_E}} \lambda^\top g(x),$$

and

$$\mathbb{I}_{h(x)\leq 0} = \sup_{\mu \in \mathbb{R}_+^{n_I}} \mu^\top h(x).$$

Equality dual variables are free; inequality dual variables are nonnegative.

From a constrained problem to a min–sup problem

Consider

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Min}} \quad & f(x) && (P) \\ \text{s.t.} \quad & g(x) = 0, \\ & h(x) \leq 0. \end{aligned}$$

It is exactly equivalent to

$$\underset{x\in\mathbb{R}^n}{\text{Min}}\, f(x) + \mathbb{I}_{g(x)=0} + \mathbb{I}_{h(x)\leq 0}.$$

Define the Lagrangian

$$\mathcal{L}(x;\lambda,\mu) := f(x) + \lambda^\top g(x) + \mu^\top h(x).$$

Then

$$p^\sharp = \inf_{x\in\mathbb{R}^n} \sup_{\substack{\lambda\in\mathbb{R}^{n_E} \\ \mu\in\mathbb{R}^{n_I}_+}} \mathcal{L}(x;\lambda,\mu).$$

From a constrained problem to a min–sup problem ♡

Consider

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\mathrm{Min}} \quad & f(x) \\ \text{s.t.} \quad & g(x) = 0, \\ & h(x) \leq 0. \end{aligned} \tag{P}$$

It is exactly equivalent to

$$\underset{x\in\mathbb{R}^n}{\mathrm{Min}}\; f(x) + \mathbb{I}_{g(x)=0} + \mathbb{I}_{h(x)\leq 0}.$$

Define the Lagrangian

$$\mathcal{L}(x;\lambda,\mu) := f(x) + \lambda^\top g(x) + \mu^\top h(x).$$

Then

$$p^\sharp = \inf_{x\in\mathbb{R}^n} \sup_{\substack{\lambda\in\mathbb{R}^{n_E} \\ \mu\in\mathbb{R}_+^{n_I}}} \mathcal{L}(x;\lambda,\mu).$$

Contents

The min–max inequality

Let $\Phi : \mathcal{X} \times \mathcal{Y} \to \overline{\mathbb{R}}$ be any function. Then

$$\boxed{\sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} \Phi(x, y) \leq \inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \Phi(x, y).}$$

Exercise. Prove this result.

Thus, exchanging the order of *inf* and *sup* always provides a lower bound for a minimization problem.

The min–max inequality

Let $\Phi : \mathcal{X} \times \mathcal{Y} \to \overline{\mathbb{R}}$ be any function. Then

$$\boxed{\sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} \Phi(x, y) \leq \inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \Phi(x, y).}$$

Exercise. Prove this result.

Thus, exchanging the order of *inf* and *sup* always provides a lower bound for a minimization problem.

Lagrangian dual problem

♡

Define the dual function

$$d(\lambda, \mu) := \inf_{x \in \mathbb{R}^n} \mathcal{L}(x; \lambda, \mu).$$

The Lagrangian dual problem is

$$(D) \qquad d^\sharp := \sup_{\substack{\lambda \in \mathbb{R}^{n_E} \\ \mu \in \mathbb{R}_+^{n_I}}} d(\lambda, \mu).$$

The min–max inequality gives weak duality:

$$\boxed{d^\sharp \leq p^\sharp.}$$

No convexity or regularity assumption is required.

Lagrangian dual problem

♡

Define the dual function

$$d(\lambda, \mu) := \inf_{x \in \mathbb{R}^n} \mathcal{L}(x; \lambda, \mu).$$

The Lagrangian dual problem is

$$(D) \qquad d^\sharp := \sup_{\substack{\lambda \in \mathbb{R}^{n_E} \\ \mu \in \mathbb{R}^{n_I}_+}} d(\lambda, \mu).$$

The min–max inequality gives weak duality:

$$\boxed{d^\sharp \leq p^\sharp.}$$

No convexity or regularity assumption is required.

Lower bounds, upper bounds, and certificates

Any dual-feasible pair $\lambda \in \mathbb{R}^{n_E}$, $\mu \in \mathbb{R}_+^{n_I}$ gives a lower bound:

$$d(\lambda, \mu) \leq \underbrace{d^\sharp}_{=\sup_{\lambda,\mu} d(\lambda,\mu)} \leq p^\sharp.$$

Any primal-feasible point x gives an upper bound:

$$p^\sharp = \min_{x \in X} f(x) \leq f(x).$$

Hence

$$0 \leq f(x) - p^\sharp \leq f(x) - d(\lambda, \mu).$$

In particular,

$$f(x) - d(\lambda, \mu) \leq \varepsilon \quad \Longrightarrow \quad x \text{ is certified } \varepsilon\text{-optimal.}$$

Lower bounds, upper bounds, and certificates

Any dual-feasible pair $\lambda \in \mathbb{R}^{n_E}$, $\mu \in \mathbb{R}_+^{n_I}$ gives a lower bound:

$$d(\lambda, \mu) \leq \underbrace{d^\sharp}_{=\sup_{\lambda,\mu} d(\lambda,\mu)} \leq p^\sharp.$$

Any primal-feasible point x gives an upper bound:

$$p^\sharp = \min_{x \in X} f(x) \leq f(x).$$

Hence

$$0 \leq f(x) - p^\sharp \leq f(x) - d(\lambda, \mu).$$

In particular,

$$f(x) - d(\lambda, \mu) \leq \varepsilon \quad \Longrightarrow \quad x \text{ is certified } \varepsilon\text{-optimal.}$$

Lower bounds, upper bounds, and certificates

♡

Any dual-feasible pair $\lambda \in \mathbb{R}^{n_E}$, $\mu \in \mathbb{R}_+^{n_I}$ gives a lower bound:

$$d(\lambda, \mu) \leq \underbrace{d^\sharp}_{=\sup_{\lambda,\mu} d(\lambda,\mu)} \leq p^\sharp.$$

Any primal-feasible point x gives an upper bound:

$$p^\sharp = \min_{x \in X} f(x) \leq f(x).$$

Hence

$$0 \leq f(x) - p^\sharp \leq f(x) - d(\lambda, \mu).$$

In particular,

$$f(x) - d(\lambda, \mu) \leq \varepsilon \quad \Longrightarrow \quad x \text{ is certified } \varepsilon\text{-optimal.}$$

Contents

Strong duality

Recall that the primal and dual values are

$$p^\sharp := \inf_{x \in \mathbb{R}^n} \sup_{\substack{\lambda \in \mathbb{R}^{n_E} \\ \mu \in \mathbb{R}_+^{n_I}}} L(x, \lambda, \mu),$$

$$d^\sharp := \sup_{\substack{\lambda \in \mathbb{R}^{n_E} \\ \mu \in \mathbb{R}_+^{n_I}}} \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu).$$

Weak duality gives

$$d^\sharp \leq p^\sharp.$$

The duality gap is

$$p^\sharp - d^\sharp \geq 0.$$

We have strong duality when

$$\boxed{p^\sharp = d^\sharp.}$$

Equivalently: exchanging inf and sup does not change the value.

Strong duality

Recall that the primal and dual values are

$$p^\sharp := \inf_{x \in \mathbb{R}^n} \sup_{\substack{\lambda \in \mathbb{R}^{n_E} \\ \mu \in \mathbb{R}_+^{n_I}}} L(x, \lambda, \mu),$$

$$d^\sharp := \sup_{\substack{\lambda \in \mathbb{R}^{n_E} \\ \mu \in \mathbb{R}_+^{n_I}}} \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu).$$

Weak duality gives

$$d^\sharp \leq p^\sharp.$$

The duality gap is

$$p^\sharp - d^\sharp \geq 0.$$

We have strong duality when

$$\boxed{p^\sharp = d^\sharp.}$$

Equivalently: exchanging inf and sup does not change the value.

Strong duality

Recall that the primal and dual values are

$$p^\sharp := \inf_{x \in \mathbb{R}^n} \sup_{\substack{\lambda \in \mathbb{R}^{n_E} \\ \mu \in \mathbb{R}_+^{n_I}}} L(x, \lambda, \mu),$$

$$d^\sharp := \sup_{\substack{\lambda \in \mathbb{R}^{n_E} \\ \mu \in \mathbb{R}_+^{n_I}}} \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu).$$

Weak duality gives

$$d^\sharp \leq p^\sharp.$$

The duality gap is

$$p^\sharp - d^\sharp \geq 0.$$

We have strong duality when

$$\boxed{p^\sharp = d^\sharp.}$$

Equivalently: exchanging inf and sup does not change the value.

Strong duality

Recall that the primal and dual values are

$$p^\sharp := \inf_{x\in\mathbb{R}^n} \sup_{\substack{\lambda\in\mathbb{R}^{n_E}\\ \mu\in\mathbb{R}^{n_I}_+}} L(x,\lambda,\mu),$$

$$d^\sharp := \sup_{\substack{\lambda\in\mathbb{R}^{n_E}\\ \mu\in\mathbb{R}^{n_I}_+}} \inf_{x\in\mathbb{R}^n} L(x,\lambda,\mu).$$

Weak duality gives

$$d^\sharp \leq p^\sharp.$$

The duality gap is

$$p^\sharp - d^\sharp \geq 0.$$

We have strong duality when

$$\boxed{p^\sharp = d^\sharp.}$$

Equivalently: exchanging inf and sup does not change the value.

Saddle point

Recall

$$p^\sharp = \inf_{x} \sup_{\lambda,\mu \geq 0} L(x, \lambda, \mu), \qquad d^\sharp = \sup_{\lambda,\mu \geq 0} \inf_{x} L(x, \lambda, \mu).$$

A triplet $(x^\sharp, \lambda^\sharp, \mu^\sharp)$ is a saddle point of the Lagrangian if

$$L(x^\sharp, \lambda, \mu) \leq L(x^\sharp, \lambda^\sharp, \mu^\sharp) \leq L(x, \lambda^\sharp, \mu^\sharp)$$

for all x, λ, and $\mu \geq 0$.

Then

$$d^\sharp = L(x^\sharp, \lambda^\sharp, \mu^\sharp) = p^\sharp.$$

Thus $x^\sharp$ is primal optimal, $(\lambda^\sharp, \mu^\sharp)$ is dual optimal, and strong duality holds.

Figure: Geometric intuition: a saddle point is a minimum in one direction and a maximum in another.

Saddle point

Recall

$$p^\sharp = \inf_{x} \sup_{\lambda,\mu\geq 0} L(x,\lambda,\mu), \qquad d^\sharp = \sup_{\lambda,\mu\geq 0} \inf_{x} L(x,\lambda,\mu).$$

A triplet $(x^\sharp, \lambda^\sharp, \mu^\sharp)$ is a saddle point of the Lagrangian if

$$L(x^\sharp,\lambda,\mu) \leq L(x^\sharp,\lambda^\sharp,\mu^\sharp) \leq L(x,\lambda^\sharp,\mu^\sharp)$$

for all x, λ, and $\mu \geq 0$.

Then

$$d^\sharp = L(x^\sharp,\lambda^\sharp,\mu^\sharp) = p^\sharp.$$

Thus $x^\sharp$ is primal optimal, $(\lambda^\sharp,\mu^\sharp)$ is dual optimal, and strong duality holds.

Figure: Geometric intuition: a saddle point is a minimum in one direction and a maximum in another.

Saddle point

Recall

$$p^\sharp = \inf_{x} \sup_{\lambda,\mu \geq 0} L(x,\lambda,\mu), \qquad d^\sharp = \sup_{\lambda,\mu\geq 0} \inf_{x} L(x,\lambda,\mu).$$

A triplet $(x^\sharp, \lambda^\sharp, \mu^\sharp)$ is a saddle point of the Lagrangian if

$$L(x^\sharp,\lambda,\mu) \leq L(x^\sharp,\lambda^\sharp,\mu^\sharp) \leq L(x,\lambda^\sharp,\mu^\sharp)$$

for all x, λ, and $\mu \geq 0$.

Then

$$d^\sharp = L(x^\sharp,\lambda^\sharp,\mu^\sharp) = p^\sharp.$$

Thus $x^\sharp$ is primal optimal, $(\lambda^\sharp,\mu^\sharp)$ is dual optimal, and strong duality holds.

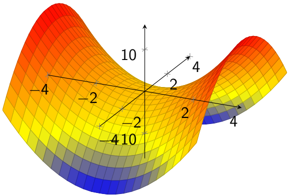


Figure: Geometric intuition: a saddle point is a minimum in one direction and a maximum in another.

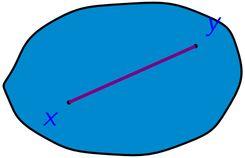
Contents

Convex sets

A set $C \subseteq \mathbb{R}^n$ is convex if

$$\forall x, y \in C, \quad \forall \theta \in [0,1], \qquad \theta x + (1-\theta) y \in C.$$

Equivalently, the segment joining x and y is contained in C.



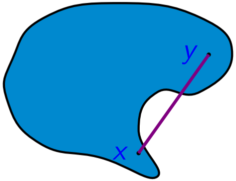
convex

Convex sets

A set $C \subseteq \mathbb{R}^n$ is convex if

$$\forall x, y \in C, \quad \forall \theta \in [0,1], \qquad \theta x + (1-\theta) y \in C.$$

Equivalently, the segment joining x and y is contained in C.



The convex hull of C is

$$\mathrm{conv}(C) := \left\{ \sum_{i=1}^{K} \theta_i x_i \;\middle|\; \begin{array}{l} x_i \in C,\ \theta_i \geq 0, \\ \sum_{i=1}^{K} \theta_i = 1 \end{array} \right\}.$$

It is the smallest convex set containing C.

Contents

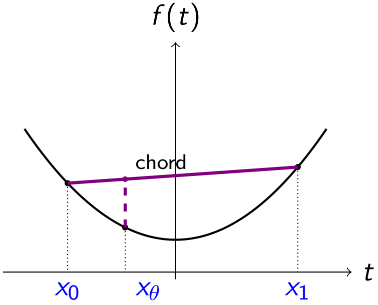
Convex functions

A function $f : C \to \mathbb{R}$ is convex if C is convex and

$$f\Big(\underbrace{\theta x_0 + (1-\theta)x_1}_{=:\, x_\theta}\Big) \leq \theta f(x_0) + (1-\theta) f(x_1)$$

for all $x_0, x_1 \in C$ and all $\theta \in [0,1]$.

Geometrically: the graph of f lies below its chords.



Convex functions

A function $f : C \to \mathbb{R}$ is convex if C is convex and

$$f\Big(\underbrace{\theta x_0 + (1-\theta)x_1}_{=: x_\theta}\Big) \leq \theta f(x_0) + (1-\theta) f(x_1)$$

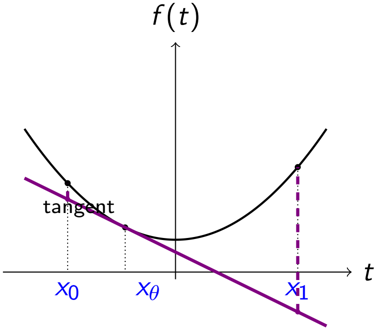
for all $x_0, x_1 \in C$ and all $\theta \in [0,1]$.

If f is differentiable, convexity is also equivalent to

$$f(x_0) \geq f(x_\theta) + \nabla f(x_\theta)^\top (x_0 - x_\theta)$$

and similarly for x_1.

So a differentiable convex function lies above its tangents.



What you should know

- The four KKT conditions: stationarity, primal feasibility, dual feasibility, complementary slackness.
- KKT conditions are necessary only under appropriate constraint qualification, and are not sufficient in general.
- The definition of the Lagrangian, dual function and dual problem.
- Weak duality and the meaning of primal and dual bounds.
- What strong duality and a saddle point mean.

What you should be able to do

- Write the KKT conditions of a differentiable constrained problem.
- Use KKT conditions to solve small problems by hand.
- Check the simple constraint qualification conditions used in the course.
- Derive the Lagrangian dual of a simple problem.