Chapter 6. Optimality Conditions & Duality

- Introduction
- Local first-order optimality conditions
- Duality

Why optimality conditions?

- Check whether a solution is an optimal solution or a KKT point of an optimization problem.
  - The satisfaction of the optimality conditions can be used as stopping criteria in optimization algorithms.
- Extremely useful in proving the convergence or convergence rate of an optimization algorithm.

Another way of checking the optimality of a solution is by the dual gap: $f(\mathbf{x}_k) - g(\boldsymbol{\lambda}_k, \boldsymbol{\nu}_k)$, where $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is the objective function of the dual problem.

Introduction

- Necessary and sufficient optimality conditions for unconstrained problems

We consider a real-valued function $f : D \to \mathbb{R}$ with domain $D \subset \mathbb{R}^n$ and define for a point $\mathbf{x}_0 \in D$:

1. f has a *local minimum* in $\mathbf{x}_0 \iff \exists U \in \mathcal{U}_{\mathbf{x}_0},\ \forall \mathbf{x} \in U \cap D,\ f(\mathbf{x}) \geq f(\mathbf{x}_0)$.
2. f has a *strict local minimum* in $\mathbf{x}_0 \iff \exists U \in \mathcal{U}_{\mathbf{x}_0},\ \forall \mathbf{x} \in U \cap D \setminus \{\mathbf{x}_0\},\ f(\mathbf{x}) > f(\mathbf{x}_0)$.
3. f has a *global minimum* in $\mathbf{x}_0 \iff \forall \mathbf{x} \in D,\ f(\mathbf{x}) \geq f(\mathbf{x}_0)$.
4. f has a *strict global minimum* in $\mathbf{x}_0 \iff \forall \mathbf{x} \in D \setminus \{\mathbf{x}_0\},\ f(\mathbf{x}) > f(\mathbf{x}_0)$.

Here, $\mathcal{U}_{\mathbf{x}_0}$ denotes the neighborhood system of $\mathbf{x}_0$.

We often say "$\mathbf{x}_0$ is a *local minimizer* of f" or "$\mathbf{x}_0$ is a *local minimum point of f*" instead of "f has a *local minimum* in $\mathbf{x}_0$" and so on. The *minimizer* is a point $\mathbf{x}_0 \in D$, the *minimum* is the corresponding value $f(\mathbf{x}_0)$.

Introduction

- Necessary optimality conditions for unconstrained problems

Suppose that the function f has a local minimum in $\mathbf{x}_0 \in D^\circ$, that is, in an interior point of D. Then:

a) If f is differentiable in $\mathbf{x}_0$, then $\nabla f(\mathbf{x}_0) = \mathbf{0}$ holds.

b) If f is twice continuously differentiable in a neighborhood of $\mathbf{x}_0$, then the Hessian $H_f(\mathbf{x}_0) = \nabla^2 f(\mathbf{x}_0) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0)\right)$ is positive semidefinite.

We will use the notation $f'(\mathbf{x}_0)$ (to denote the derivative of f at $\mathbf{x}_0$; as we know, this is a linear map from $\mathbb{R}^n$ to $\mathbb{R}$, read as a *row vector*) as well as the corresponding transposed vector $\nabla f(\mathbf{x}_0)$ (gradient, *column vector*).

Points $\mathbf{x} \in D^\circ$ with $\nabla f(\mathbf{x}) = \mathbf{0}$ are called *stationary points*. At a stationary point there can be a local minimum, a local maximum or a *saddle point*.

Introduction

- Sufficient optimality conditions for unconstrained problems

Suppose that the function f is twice continuously differentiable in a neighborhood of $\mathbf{x}_0 \in D$; also suppose that the necessary optimality condition $\nabla f(\mathbf{x}_0) = \mathbf{0}$ holds and that the Hessian $\nabla^2 f(\mathbf{x}_0)$ is positive definite. Then f has a strict local minimum in $\mathbf{x}_0$.

Introduction

- Necessary optimality conditions for equality constrained problems

Now let f be a real-valued function with domain $D \subset \mathbb{R}^n$ which we want to minimize subject to the *equality constraints*

$$h_j(\mathbf{x}) = 0, \quad j = 1, ..., p,$$

for $p < n$; here, let $h_1, ..., h_p$ also be defined on D. We are looking for local minimizers of f, that is, points $\mathbf{x}_0 \in D$ which belong to the *feasible region*

$$\mathcal{F} := \{\mathbf{x} \in D | h_j(\mathbf{x}) = 0, j = 1, ..., p\}$$

and to which a neighborhood U exists with $f(\mathbf{x}) \geq f(\mathbf{x}_0)$ for all $\mathbf{x} \in U \cap \mathcal{F}$.

Introduction

- Necessary optimality conditions for equality constrained problems

Theorem 1 (Lagrange Multiplier Rule). *Let $D \subset \mathbb{R}^n$ be open and $f, h_1, ..., h_p$ continuously differentiable in D. Suppose that f has a local minimum in $\mathbf{x}_0 \in \mathcal{F}$ subject to the constraints*

$$h_j(\mathbf{x}) = 0, \quad j = 1, ..., p.$$

Let also the Jacobian $\left(\frac{\partial h_j}{\partial \mathbf{x}_k}(\mathbf{x}_0)\right)_{p,n}$ have rank p. Then there exist real numbers $\mu_1, ..., \mu_p$ – the so-called Lagrange multipliers – such that

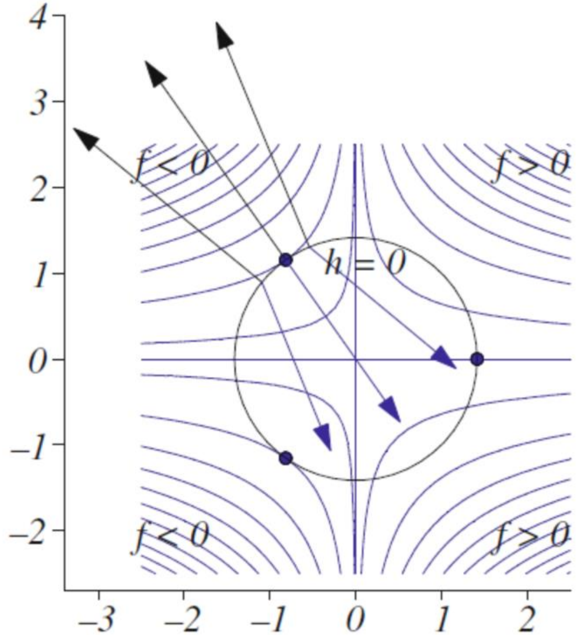
$$\nabla f(\mathbf{x}_0) + \sum_{j=1}^{p} \mu_j \nabla h_j(\mathbf{x}_0) = \mathbf{0}.$$

Introduction

- Necessary optimality conditions for equality constrained problems

Example: With $f(\mathbf{x}) := x_1 x_2^2$ and $h(\mathbf{x}) := h_1(\mathbf{x}) := x_1^2 + x_2^2 - 2$ for $\mathbf{x} = (x_1, x_2)^\top \in D := \mathbb{R}^2$ we consider the problem:

$$\min_{\mathbf{x}} f(\mathbf{x}), \quad s.t. \quad h(\mathbf{x}) = 0.$$



Introduction

- Minimization problems with inequality constraints

$$(P) \quad \begin{cases} \min\limits_{\mathbf{x}} \ f(\mathbf{x}), \\ s.t. \ g_i(\mathbf{x}) \leq 0, \ \text{for } i \in \mathcal{I} := \{1, ..., m\}, \\ \qquad h_j(\mathbf{x}) = 0, \ \text{for } j \in \mathcal{E} := \{1, ..., p\}. \end{cases} \tag{1}$$

With $m, p \in \mathbb{N}_0$ (hence, $\mathcal{E} = \emptyset$ or $\mathcal{I} = \emptyset$ are allowed), the functions $f, g_1, ..., g_m$, $h_1, ..., h_p$ are supposed to be continuously differentiable on an open subset D in $\mathbb{R}^n$ and $p \leq n$. The set

$$\mathcal{F} := \{\mathbf{x} \in D | g_i(\mathbf{x}) \leq 0 \text{ for } i \in \mathcal{I}, h_j(\mathbf{x}) = 0 \text{ for } j \in \mathcal{E}\}$$

is called the *feasible region* or *set of feasible points of* (P).

The *optimal value* $v(P)$ to problem (P) is defined as

$$v(P) := \inf\{f(\mathbf{x}) | \mathbf{x} \in \mathcal{F}\}.$$

Local First-Order Optimality Conditions

- Concepts

For $\mathbf{x}_0 \in \mathcal{F}$, the index set

$$\mathcal{A}(\mathbf{x}_0) := \{i \in \mathcal{I} | g_i(\mathbf{x}_0) = 0\}$$

describes the *inequality restrictions which are active at* $\mathbf{x}_0$.

Definition 1. *Let* $\mathbf{d} \in \mathbb{R}^n$ *and* $\mathbf{x}_0 \in \mathcal{F}$. *Then* $\mathbf{d}$ *is called the* feasible direction *of* $\mathcal{F}$ *at* $\mathbf{x}_0$ $:\Leftrightarrow \exists \delta > 0,\ \forall \tau \in [0, \delta],\ \mathbf{x}_0 + \tau \mathbf{d} \in \mathcal{F}$.

A 'small' movement from $\mathbf{x}_0$ along such a direction gives feasible points. The set of all feasible directions of $\mathcal{F}$ at $\mathbf{x}_0$ is a *cone*, denoted by

$$\mathcal{C}_{fd}(\mathbf{x}_0).$$

Local First-Order Optimality Conditions

- Concepts

Let $\mathbf{d}$ be a feasible direction of $\mathcal{F}$ at $\mathbf{x}_0$. If we choose a δ according to the definition, then we have

$$\underbrace{g_i(\mathbf{x}_0 + \tau\mathbf{d})}_{\leq 0} = \underbrace{g_i(\mathbf{x}_0)}_{=0} + \tau g_i'(\mathbf{x}_0)\mathbf{d} + o(\tau)$$

for $i \in \mathcal{A}(\mathbf{x}_0)$ and $0 < \tau \leq \delta$. Dividing by τ and passing to the limit as $\tau \to 0$ gives $g_i'(\mathbf{x}_0)\mathbf{d} \leq 0$. In the same way we get $h_j'(\mathbf{x}_0)\mathbf{d} = 0$ for all $j \in \mathcal{E}$.

Local First-Order Optimality Conditions

- Concepts

Definition 2. *For any* $\mathbf{x}_0 \in \mathcal{F}$,

$$\mathcal{C}_l(P, \mathbf{x}_0) := \left\{ \mathbf{d} \in \mathbb{R}^n \middle| \forall i \in \mathcal{A}(\mathbf{x}_0),\ g_i'(\mathbf{x}_0)\mathbf{d} \leq 0,\ \forall j \in \mathcal{E},\ h_j'(\mathbf{x}_0)\mathbf{d} = 0 \right\}$$

is called the linearizing cone *of* (P) *at* $\mathbf{x}_0$.

Hence, $\mathcal{C}_l(\mathbf{x}_0) := \mathcal{C}_l(P, \mathbf{x}_0)$ *contains at least all feasible directions of* $\mathcal{F}$ *at* $\mathbf{x}_0$:

$$\mathcal{C}_{fd}(\mathbf{x}_0) \subset \mathcal{C}_l(\mathbf{x}_0).$$

The linearizing cone is not only dependent on the *set* of feasible points $\mathcal{F}$ but also on the *representation* of $\mathcal{F}$. We therefore write more precisely $\mathcal{C}_l(P, \mathbf{x}_0)$.

Local First-Order Optimality Conditions

- Concepts

Definition 3. *For any* $\mathbf{x}_0 \in D$

$$\mathcal{C}_{dd}(\mathbf{x}_0) := \{\mathbf{d} \in \mathbb{R}^n | f'(\mathbf{x}_0)\mathbf{d} < 0\}$$

is called the cone of descent directions of f at $\mathbf{x}_0$.

Note that $\mathbf{0}$ is not in $\mathcal{C}_{dd}(\mathbf{x}_0)$; also, for all $\mathbf{d} \in \mathcal{C}_{dd}(\mathbf{x}_0)$

$$f(\mathbf{x}_0 + \tau\mathbf{d}) = f(\mathbf{x}_0) + \tau \underbrace{f'(\mathbf{x}_0)\mathbf{d}}_{<0} + o(\tau)$$

holds and therefore, $f(\mathbf{x}_0 + \tau\mathbf{d}) < f(\mathbf{x}_0)$ for sufficiently small $\tau > 0$.

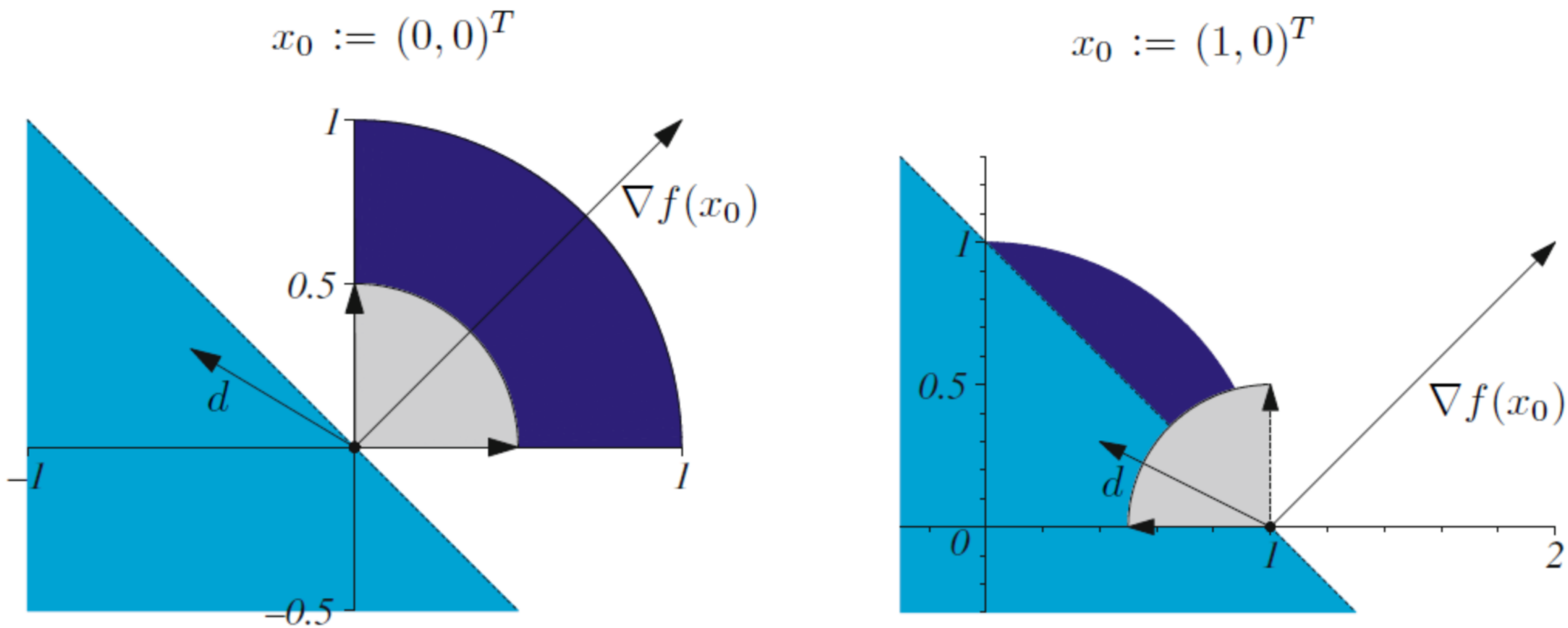
Thus, $\mathbf{d} \in \mathcal{C}_{dd}(\mathbf{x}_0)$ guarantees that the objective function f can be reduced along this direction. Hence, for a local minimizer $\mathbf{x}_0$ of (P) it necessarily holds that $\mathcal{C}_{dd}(\mathbf{x}_0) \cap \mathcal{C}_{fd}(\mathbf{x}_0) = \emptyset$.

Local First-Order Optimality Conditions

- Examples

$$\mathcal{F} := \left\{\mathbf{x} = (x_1, x_2)^\top \in \mathbb{R}^2 | x_1^2 + x_2^2 - 1 \leq 0, -x_1 \leq 0, -x_2 \leq 0\right\},$$

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



Local First-Order Optimality Conditions

- Karush-Kuhn-Tucker (KKT) conditions

Proposition 1. *For* $\mathbf{x}_0 \in \mathcal{F}$ *it holds that* $\mathcal{C}_l(\mathbf{x}_0) \cap \mathcal{C}_{dd}(\mathbf{x}_0) = \emptyset$ *if and only if there exist* $\boxed{\boldsymbol{\lambda}} \in \mathbb{R}_+^m$ *and* $\boxed{\boldsymbol{\mu}} \in \mathbb{R}^p$ *such that*

Lagrange multipliers

$$\nabla f(\mathbf{x}_0) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{x}_0) + \sum_{j=1}^{p} \mu_j \nabla h_j(\mathbf{x}_0) = \mathbf{0} \tag{1}$$

and

$$\boxed{\lambda_i g_i(\mathbf{x}_0) = 0 \text{ for all } i \in \mathcal{I}.} \tag{2}$$

complementary slackness condition or complementarity condition

$$\mathbf{x}_0 \in \mathcal{F} \tag{3}$$

$$\boldsymbol{\lambda} \in \mathbb{R}_+^m \tag{4}$$

KKT point: $(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ satisfying (1)-(4).

Local First-Order Optimality Conditions

- Karush-Kuhn-Tucker (KKT) conditions

$$\nabla f(\mathbf{x}_0) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{x}_0) + \sum_{j=1}^{p} \mu_j \nabla h_j(\mathbf{x}_0) = \mathbf{0}$$

$$\Updownarrow$$

$$\nabla f(\mathbf{x}_0) + \sum_{i \in \mathcal{A}(\mathbf{x}_0)} \lambda_i \nabla g_i(\mathbf{x}_0) + \sum_{j=1}^{p} \mu_j \nabla h_j(\mathbf{x}_0) = \mathbf{0}$$

Proof. By definition of $\mathcal{C}_l(\mathbf{x}_0)$ and $\mathcal{C}_{dd}(\mathbf{x}_0)$ it holds that:

$$\mathbf{d} \in \mathcal{C}_l(\mathbf{x}_0) \cap \mathcal{C}_{dd}(\mathbf{x}_0) \Leftrightarrow \begin{cases} f'(\mathbf{x}_0)\mathbf{d} < 0 \\ \forall i \in \mathcal{A}(\mathbf{x}_0),\ g_i'(\mathbf{x}_0)\mathbf{d} \le 0 \\ \forall j \in \mathcal{E},\ h_j'(\mathbf{x}_0)\mathbf{d} = 0. \end{cases}$$

$$\Leftrightarrow \begin{cases} f'(\mathbf{x}_0)\mathbf{d} < 0 \\ \forall i \in \mathcal{A}(\mathbf{x}_0),\ -g_i'(\mathbf{x}_0)\mathbf{d} \ge 0 \\ \forall j \in \mathcal{E},\ -h_j'(\mathbf{x}_0)\mathbf{d} \ge 0;\ \forall j \in \mathcal{E},\ h_j'(\mathbf{x}_0)\mathbf{d} \ge 0. \end{cases}$$

Strong alternatives:

1. $\exists \mathbf{x} \in \mathbb{R}_+^n,\ \mathbf{A}\mathbf{x} = \mathbf{b}$,
2. $\exists \mathbf{y} \in \mathbb{R}^m,\ \mathbf{A}^T\mathbf{y} \ge \mathbf{0}\ \&\ \mathbf{b}^T\mathbf{y} < 0$.

Local First-Order Optimality Conditions

- Karush-Kuhn-Tucker (KKT) conditions

With that the Theorem of the Alternative directly provides the following equivalence:

$\mathcal{C}_l(\mathbf{x}_0) \cap \mathcal{C}_{dd}(\mathbf{x}_0) = \emptyset$ iff there exist $\lambda_i \geq 0$ for $i \in \mathcal{A}(\mathbf{x}_0)$ and $\mu'_j \geq 0, \mu''_j \geq 0$ for $j \in \mathcal{E}$ such that

$$\nabla f(\mathbf{x}_0) = \sum_{i \in \mathcal{A}(\mathbf{x}_0)} \lambda_i(-\nabla g_i(\mathbf{x}_0)) + \sum_{j=1}^{p} \mu'_j(-\nabla h_j(\mathbf{x}_0)) + \sum_{j=1}^{p} \mu''_j \nabla h_j(\mathbf{x}_0).$$

If we now set $\lambda_i := 0$ for $i \in \mathcal{I} \setminus \mathcal{A}(\mathbf{x}_0)$ and $\mu_j := \mu'_j - \mu''_j$ for $j \in \mathcal{E}$, the above is equivalent to: There exist $\lambda_i \geq 0$ for $i \in \mathcal{I}$ and $\mu_j \in \mathbb{R}$ for $j \in \mathcal{E}$ with

$$\nabla f(\mathbf{x}_0) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{x}_0) + \sum_{j=1}^{p} \mu_j \nabla h_j(\mathbf{x}_0) = \mathbf{0}$$

and

$$\lambda_i g_i(\mathbf{x}_0) = 0 \text{ for all } i \in \mathcal{I}.$$

Local First-Order Optimality Conditions

- Karush-Kuhn-Tucker (KKT) conditions

The *Lagrange function* or *Lagrangian* of (P):

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) := f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j h_j(\mathbf{x}) = f(\mathbf{x}) + \boldsymbol{\lambda}^\top g(\mathbf{x}) + \boldsymbol{\mu}^\top h(\mathbf{x})$$

with $\mathbf{x} \in D, \boldsymbol{\lambda} \in \mathbb{R}_+^m$ and $\boldsymbol{\mu} \in \mathbb{R}^p$.

Local First-Order Optimality Conditions

- Karush-Kuhn-Tucker (KKT) conditions

Question: What is the relationship between local minimum and $\mathcal{C}_l(\mathbf{x}_0) \cap \mathcal{C}_{dd}(\mathbf{x}_0) = \emptyset$?

Lemma 1.
For $\mathbf{x}_0 \in \mathcal{F}$ *it holds that:* $\mathcal{C}_l(\mathbf{x}_0) \cap \mathcal{C}_{dd}(\mathbf{x}_0) = \emptyset \Leftrightarrow \nabla f(\mathbf{x}_0) \in \mathcal{C}_l(\mathbf{x}_0)^*$.

Proof.

$$\begin{aligned} \mathcal{C}_l(\mathbf{x}_0) \cap \mathcal{C}_{dd}(\mathbf{x}_0) = \emptyset &\Leftrightarrow \forall \mathbf{d} \in \mathcal{C}_l(\mathbf{x}_0),\ \langle \nabla f(\mathbf{x}_0), \mathbf{d} \rangle = f'(\mathbf{x}_0)\mathbf{d} \geq 0 \\ &\Leftrightarrow \nabla f(\mathbf{x}_0) \in \mathcal{C}_l(\mathbf{x}_0)^*. \end{aligned}$$

□

The cone $\mathcal{C}_{fd}(\mathbf{x}_0)$ of all feasible directions is too small to ensure general optimality conditions. Difficulties may occur when the boundary of $\mathcal{F}$ is curved. Therefore, we have to consider a set which is less intuitive but bigger and with more suitable properties, called *tangent cone*.

Local First-Order Optimality Conditions

- Karush-Kuhn-Tucker (KKT) conditions

Definition 4. *A sequence* $(\mathbf{x}_k)$ converges in direction $\mathbf{d}$ *to* $\mathbf{x}_0$

$$:\Leftrightarrow \mathbf{x}_k = \mathbf{x}_0 + \alpha_k(\mathbf{d} + \mathbf{r}_k) \text{ with } \alpha_k \downarrow 0 \text{ and } \mathbf{r}_k \to \mathbf{0}.$$

We will use the following notation: $\mathbf{x}_k \overset{\mathbf{d}}{\to} \mathbf{x}_0$. It simply means: There exists a sequence of positive numbers (α_k) such that $\alpha_k \downarrow 0$ and

$$\frac{1}{\alpha_k}(\mathbf{x}_k - \mathbf{x}_0) \longrightarrow \mathbf{d} \text{ for } k \longrightarrow \infty.$$

Definition 5. *Let M be a nonempty subset of $\mathbb{R}^n$ and $\mathbf{x}_0 \in M$. Then*

$$\mathcal{C}_t(M, \mathbf{x}_0) := \left\{ \mathbf{d} \in \mathbb{R}^n | \exists \{\mathbf{x}_k\} \in M^{\mathbb{N}},\ \mathbf{x}_k \overset{\mathbf{d}}{\to} \mathbf{x}_0 \right\}$$

is called the tangent cone *of M at $\mathbf{x}_0$. The vectors of $\mathcal{C}_t(M, \mathbf{x}_0)$ are called* tangents *or* tangent directions *of M at $\mathbf{x}_0$.*

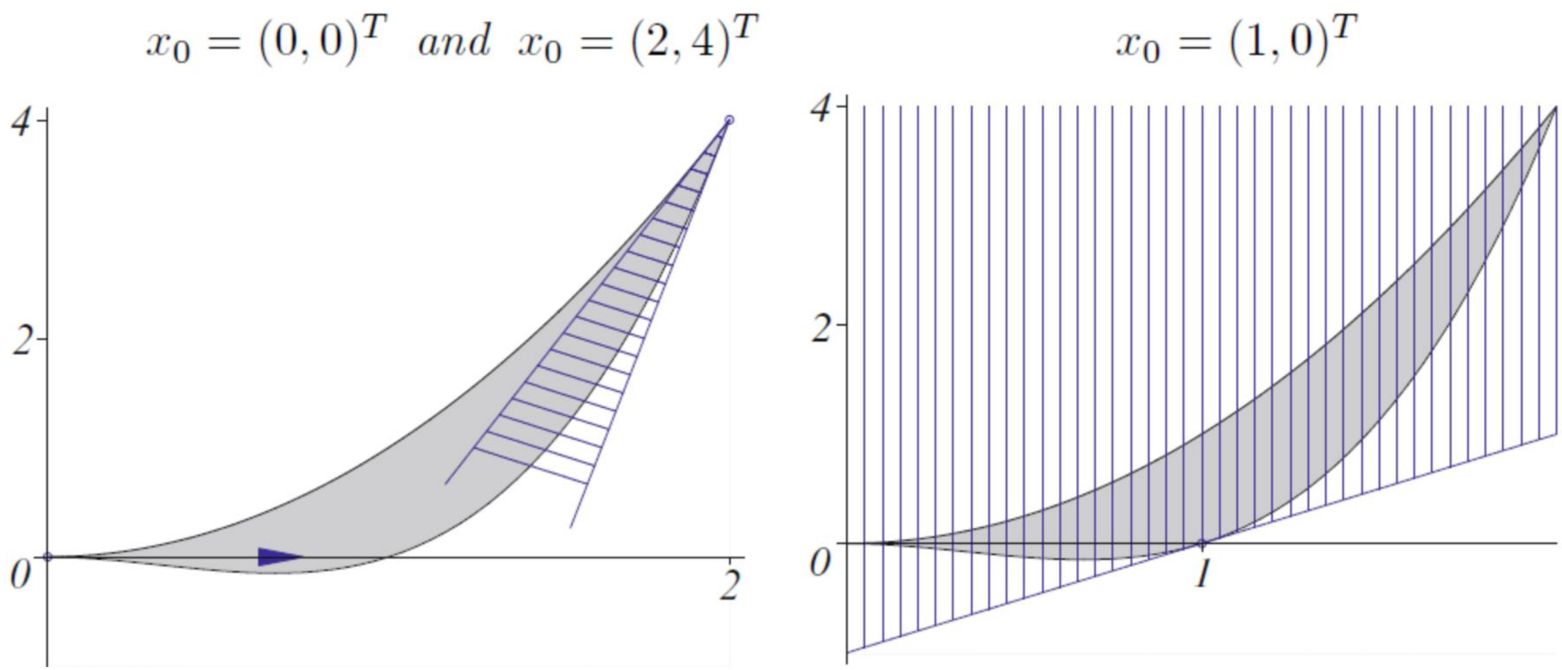
Local First-Order Optimality Conditions

- Karush-Kuhn-Tucker (KKT) conditions

Examples. a) The tangent cones of

$$\mathcal{F} := \left\{ \mathbf{x} = (x_1, x_2)^\top \in \mathbb{R}^2 | x_1 \geq 0, x_1^2 \geq x_2 \geq x_1^2(x_1 - 1) \right\}$$

and the points $\mathbf{x}_0 \in \left\{ (0,0)^\top, (2,4)^\top, (1,0)^\top \right\}$.



Local First-Order Optimality Conditions

- Karush-Kuhn-Tucker (KKT) conditions

b) $\mathcal{F} := \{\mathbf{x} \in \mathbb{R}^n | \, \|\mathbf{x}\|_2 = 1\} : \mathcal{C}_t(\mathbf{x}_0) = \{\mathbf{d} \in \mathbb{R}^n | \, \langle \mathbf{d}, \mathbf{x}_0 \rangle = 0\}$.

c) $\mathcal{F} := \{\mathbf{x} \in \mathbb{R}^n | \, \|\mathbf{x}\|_2 \leq 1\}$: Then $\mathcal{C}_t(\mathbf{x}_0) = \mathbb{R}^n$ if $\|\mathbf{x}_0\|_2 < 1$ holds, and $\mathcal{C}_t(\mathbf{x}_0) = \{\mathbf{d} \in \mathbb{R}^n | \, \langle \mathbf{d}, \mathbf{x}_0 \rangle \leq 0\}$ if $\|\mathbf{x}_0\|_2 = 1$.

Lemma 1. **1)** $\mathcal{C}_t(\mathbf{x}_0)$ *is a closed cone,* $\mathbf{0} \in \mathcal{C}_t(\mathbf{x}_0)$.

2) $\overline{\mathcal{C}_{fd}(\mathbf{x}_0)} \subset \mathcal{C}_t(\mathbf{x}_0) \subset \mathcal{C}_l(\mathbf{x}_0)$.

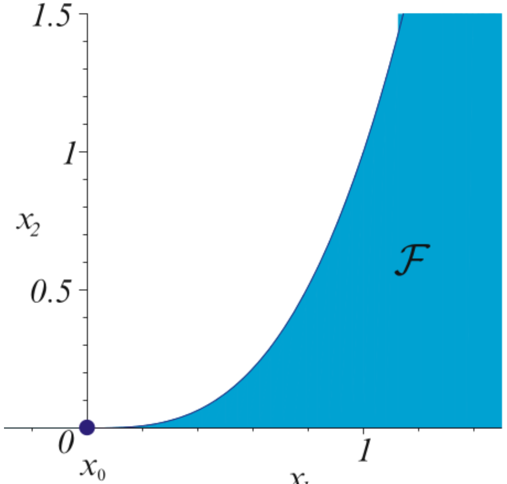
Local First-Order Optimality Conditions

- Karush-Kuhn-Tucker (KKT) conditions

Question: whether $\mathcal{C}_t(\mathbf{x}_0) = \mathcal{C}_l(\mathbf{x}_0)$ always holds? No!

Example 1. a) *Consider* $\mathcal{F} := \{\mathbf{x} \in \mathbb{R}^2 | -x_1^3 + x_2 \leq 0, -x_2 \leq 0\}$ *and* $\mathbf{x}_0 := (0,0)^\top$. *In this case* $\mathcal{A}(\mathbf{x}_0) = \{1,2\}$. *This gives* $\mathcal{C}_l(\mathbf{x}_0) = \{\mathbf{d} \in \mathbb{R}^2 | d_2 = 0\}$ *and* $\mathcal{C}_t(\mathbf{x}_0) = \{\mathbf{d} \in \mathbb{R}^2 | d_1 \geq 0,\ d_2 = 0\}$.

b) *Now let* $\mathcal{F} := \{\mathbf{x} \in \mathbb{R}^2 | -x_1^3 + x_2 \leq 0, -x_1 \leq 0, -x_2 \leq 0\}$ *and* $\mathbf{x}_0 := (0,0)^\top$. *Then* $\mathcal{A}(\mathbf{x}_0) = \{1,2,3\}$ *and therefore* $\mathcal{C}_l(\mathbf{x}_0) = \{\mathbf{d} \in \mathbb{R}^2 | d_1 \geq 0, d_2 = 0\} = \mathcal{C}_t(\mathbf{x}_0)$.



the linearizing cone is dependent on the representation of the set of feasible points!

Local First-Order Optimality Conditions

- Karush-Kuhn-Tucker (KKT) conditions

$$\boxed{\nabla f(\mathbf{x}_0) \in \mathcal{C}_l(\mathbf{x}_0)^* \Longleftrightarrow \mathcal{C}_{dd}(\mathbf{x}_0) \cap \mathcal{C}_l(\mathbf{x}_0) = \emptyset}$$

Lemma 2. *For a local minimizer* $\mathbf{x}_0$ *of* (P) *it holds that* $\nabla f(\mathbf{x}_0) \in \mathcal{C}_t(\mathbf{x}_0)^*$, *hence* $\mathcal{C}_{dd}(\mathbf{x}_0) \cap \mathcal{C}_t(\mathbf{x}_0) = \emptyset$.

Geometric meaning: for a local minimizer $\mathbf{x}_0$ of (P) the angle between the gradient and any tangent direction, especially any feasible direction, does not exceed $90°$.

Proof. Let $\mathbf{d} \in \mathcal{C}_t(\mathbf{x}_0)$. Then there exists a sequence $\{\mathbf{x}_k\} \in \mathcal{F}^{\mathbb{N}}$ such that $\mathbf{x}_k = \mathbf{x}_0 + \alpha_k(\mathbf{d} + \mathbf{r}_k)$, $\alpha_k \downarrow 0$ and $\mathbf{r}_k \to \mathbf{0}$.

$$0 \leq f(\mathbf{x}_k) - f(\mathbf{x}_0) = \alpha_k f'(\mathbf{x}_0)(\mathbf{d} + \mathbf{r}_k) + o(\alpha_k)$$

gives the result $f'(\mathbf{x}_0)\mathbf{d} \geq 0$. $\square$

Local First-Order Optimality Conditions

- Karush-Kuhn-Tucker (KKT) conditions

Theorem 2 (Karush-Kuhn-Tucker)**.** *Suppose that* $\mathbf{x}_0$ *is a local minimizer of* (P), *and the constraint qualification* $\mathcal{C}_l(\mathbf{x}_0)^* = \mathcal{C}_t(\mathbf{x}_0)^*$ *is fulfilled. Then there exist vectors* $\boldsymbol{\lambda} \in \mathbb{R}_+^m$ *and* $\boldsymbol{\mu} \in \mathbb{R}^p$ *such that*

$$\nabla f(\mathbf{x}_0) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{x}_0) + \sum_{j=1}^{p} \mu_j \nabla h_j(\mathbf{x}_0) = \mathbf{0} \textit{ and}$$

$$\lambda_i g_i(\mathbf{x}_0) = 0 \textit{ for } i = 1, ..., m.$$

Proof. If $\mathbf{x}_0$ is a local minimizer of (P), it follows from Lemma 1 with the help of the presupposed constraint qualification that

$$\nabla f(\mathbf{x}_0) \in \mathcal{C}_t(\mathbf{x}_0)^* = \mathcal{C}_l(\mathbf{x}_0)^*;$$

Lemma 2 yields $\mathcal{C}_l(\mathbf{x}_0) \cap \mathcal{C}_{dd}(\mathbf{x}_0) = \emptyset$ and the latter together with Proposition 1 gives the result. $\square$

Local First-Order Optimality Conditions

- Karush-Kuhn-Tucker (KKT) conditions

For $\mathbf{x} = (x_1, x_2)^\top \in \mathbb{R}^2$ whether the feasible points $\mathbf{x}_0 := (-1, 0)^\top$ and $\tilde{\mathbf{x}}_0 := (0, 1)^\top$ are local minimizers of consider the problem

$$\begin{aligned} \min_{\mathbf{x}} \ & f(\mathbf{x}) := x_1 + x_2, \\ s.t. \ & -x_1^3 + x_2 \leq 1, \\ & x_1 \leq 1, -x_2 \leq 0. \end{aligned}$$

Local First-Order Optimality Conditions

- Constraint Qualifications

The condition $\mathcal{C}_l(\mathbf{x}_0)^* = \mathcal{C}_t(\mathbf{x}_0)^*$ is very abstract, extremely general, but not easily verifiable. Therefore, for practical problems, we will try to find regularity assumptions called *constraint qualifications* (CQ) which are more specific, easily verifiable, but also somewhat restrictive.

Assuming that we only have inequality constraints.
(GCQ) Guignard Constraint Qualification: $\mathcal{C}_l(\mathbf{x}_0)^* = \mathcal{C}_t(\mathbf{x}_0)^*$.
(ACQ) Abadie Constraint Qualification: $\mathcal{C}_l(\mathbf{x}_0) = \mathcal{C}_t(\mathbf{x}_0)$.
(SCQ) Slater Constraint Qualification: The functions g_i are convex for all $i \in \mathcal{I}$ and

$$\exists \tilde{\mathbf{x}} \in \mathcal{F},\ g_i(\tilde{\mathbf{x}}) < 0 \text{ for } i \in \boxed{\mathcal{I}_1}.$$

$\mathcal{I}_1$ is the index set of nonlinear constraints.

$$\text{(SCQ)} \Longrightarrow \text{(ACQ)} \Longrightarrow \text{(GCQ)}.$$