Nonlinear Optimization: Advanced

Technische Universität München

January 3, 2025

Contents

1 Summary

we condiser the general nonlinear optimization problem:

$$\min_{x\in\mathbb{R}^n} f(x) \quad s.t. \quad g(x) \le 0, \quad h(x) = 0 \tag{1}$$

with continuously differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^p$

1.1 Optimality Conditions

Definition 1.1 (jacobian). *$g : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable, the jacobian of g is defined as*

$$J = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \frac{\partial g_m}{\partial x_2} & \cdots & \frac{\partial g_m}{\partial x_n} \end{bmatrix}$$

if $m = 1$, then the gradient of g, $\nabla g = J^\top$

Definition 1.2 (active index/non-active index). *the set*

$$X = \{x \in \mathbb{R}^n : g(x) \le 0, h(x) = 0\}$$

is called feasible set of (1). a point $x \in \mathbb{R}^n$ is called feasible if $x \in X$. for a feasible point $x \in X$ we define the index set of active inequailty constraints $\mathcal{A}(x)$ and accordingly the index set of inactive inequality constraints $\mathcal{I}(x)$:

1. $\mathcal{A} = \{i : 1 \le i \le m, g_i(x) = 0\}$

2. $\mathcal{I}(x) = \{1,..,m\} \setminus \mathcal{A}(x) = \{i, 1 \leq i \leq m, g_i(x) < 0\}$

Definition 1.3 (cone). *the set $K \subset \mathbb{R}^n$ is called cone, if*

$$\lambda x \in K, \quad \forall \lambda > 0, \quad x \in K$$

Definition 1.4 (tangent cone). *let $M \subset \mathbb{R}^n$ be a non-empty set. The tangent cone of M at a point $x \in M$ is given by the set*

$$T(M,x) = \{d \in \mathbb{R}^n : \exists \eta_k > 0, x^k \in M, \lim_{k\to\infty} x^k = x, \lim_{k\to\infty} \eta_k(x^k - x) = d\}$$

Remark. *if $M \subset \mathbb{R}^n$ is convex and $x \in M$, then it holds that*

$$T(M,x) = \overline{\{d \in \mathbb{R}^n : \exists \eta > 0, y \in M, d = \eta(y-x)\}}$$

Definition 1.5 (linearized tangential cone). *we call*

$$T_l(g,h,x) = \{d \in \mathbb{R}^n : \nabla g_i(x)^\top d \leq 0, i \in \mathcal{A}(x), \nabla h(x)^\top d = 0\}$$

the linearized tangent cone at $x \in X$ for the representation $X = \{x \in \mathbb{R}^n : g(x) \leq 0, h(x) = 0\}$ of X

Definition 1.6 (ACQ). *the condition*

$$T_l(g,h,x) = T(X,x)$$

is called Abadie Constraint Qualification for $x \in X$

Definition 1.7 (polar cone/dual cone). *let $K \subset \mathbb{R}^n$ be a non-empty cone. the polar cone of K is defined as*

$$K^\circ = \{v \in \mathbb{R}^n : v^\top d \leq 0, \forall d \in K\}$$

Definition 1.8 (GCQ). *the condition*

$$T_l(g,h,x)^\circ = T(X,x)^\circ$$

is called Guignard Constraint Qualification for $x \in X$

Definition 1.9 (CQ). *let $x \in X$. a condition that implies GCQ is called constraint qualification (CQ) at x*

Definition 1.10 (KKT-triple). *if a triple $(\overline{x}, \overline{\lambda}, \overline{\mu}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ fulfills the KKT conditions, we call $\overline{x}$ a KKT point of (1) and $(\overline{x}, \overline{\lambda}, \overline{\mu})$ a KKT triple of (1)*

Definition 1.11 ((strict) complementarity condition). *let $(\overline{x}, \overline{\lambda}, \overline{\mu})$ be a KKT triple for (1).*

1. *if*

$$\overline{\lambda}_i > 0, \forall i \in \mathcal{A}(\overline{x})$$

   holds, we say that the strict complementarity condition holds.

2. *if there exists $i \in \{1, ..., m\}$ with*

$$\overline{\lambda}_i = g_i(\overline{x}) = 0$$

   we say that the strict complementarity condition is violated.

Definition 1.12 (Lagrangian function). *the function $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$,*

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

is called Lagrangian function for problem (1)

Definition 1.13 (MFCQ). *the point $x \in X$ is said to satisfy the Mangasarian-Fromovitz Constraint Qualification (MFCQ), if*

1. *$\nabla h(x)$ is full column rank, (or h is affine)*
2. *there exists $d \in \mathbb{R}^n$ with*

$$\nabla g_i(x)^\top d < 0, i \in \mathcal{A}(x), \nabla h(x)^\top d = 0$$

if $m = 0$ (no inequality constraints) or $\mathcal{A}(x) - \emptyset$, we can ignore 2., if we have no equality constraints (p=0), 1. vanishes and in 2. $\nabla h(x)^\top d = 0$ is deleted.

Definition 1.14 (PLICQ). *we say $x \in X$ satisfies the Positive Linear Independence CQ (PLICQ) if*

1. *$\nabla h(x)$ is full column rank.*

2. *there exists no vectors $u \in \mathbb{R}^m$, $v \in \mathbb{R}^p$ with*

$$\nabla g(x)u + \nabla h(x)v = 0, \quad u_{\mathcal{A}(x)} \geq 0, \quad \mu_{\mathcal{A}(x)} \neq 0, \quad u_{\mathcal{I}(x)} = 0$$

if $m = 0$ (no inequality constraints) or $\mathcal{A}(x) = \emptyset$, 2. is omitted. for having no equality constraints (p=0), we can drop all terms that contain h and v in 1. and 2.

Definition 1.15 (LICQ, regular)**.** *a point $x \in X$ is called regular if all columns of the matrix*

$$(\nabla_{g_{\mathcal{A}(x)}}(x), \nabla h(x))$$

are linearly independent. we also say that the Linear Independence Constraint Qualification (LICQ) holds at $x \in X$.

Theorem 1.16. *[necessary optimal condition] let f be differentiable and let $\overline{x}$ be a local solution of (1). then it holds*

1. *$\overline{x} \in X$*
2. *$\nabla f(\overline{x})^\top d \geq 0$ for all $d \in T(X, \overline{x})$*

Lemma 1.17. *for all $x \in X$, it holds*

$$T(X, x) \subset T_l(g, h, x)$$

Theorem 1.18. *let f, g, h be differentiable and let $\overline{x}$ be a local solution of (1) and let the ACQ hold at $\overline{x}$. then we have*

1. *$\overline{x} \in X$*
2. *$\nabla f(\overline{x})^\top d \geq 0$ for all $d \in T_l(g, h, \overline{x})$*

Lemma 1.19. *if ACQ holds at $x \in X$, then GCQ holds at $x \in X$.*

Theorem 1.20. *let $\overline{x}$ be a local solution of (1) and let the GCQ hold at $\overline{x}$, then it holds:*

1. *$\overline{x} \in X$*
2. *$\nabla f(\overline{x})^\top d \geq 0$ for all $d \in T_l(g, h, \overline{x})$*

Lemma 1.21 (Farkas' lemma)**.** *let $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{p \times n}$, $c \in \mathbb{R}^n$. then the following conditions are equivalent:*

1. *for all $d \in \mathbb{R}^n$ with $Ad \leq 0$ and $Bd = 0$ it holds $c^\top d \leq 0$*
2. *there exist $u \in \mathbb{R}^m, u \geq 0$, and $v \in \mathbb{R}^p$ with $c = A^\top u + B^\top v$*

Theorem 1.22 (necessary first order optimality conditions, KKT conditions)**.** *let f, g and h be differentiable and let $\overline{x} \in \mathbb{R}^n$ be a local solution of (1) at which a constraint qualification holds. then it holds: **Karush-Kuhn-Tucker-conditions** there exist Lagrange multipliers $\overline{\lambda} \in \mathbb{R}^m$ and $\overline{\mu} \in \mathbb{R}^p$ s.t.*

1. *$\nabla f(\overline{x}) + \nabla g(\overline{x})\overline{\lambda} + \nabla h(\overline{x})\overline{\mu} = 0$ (multiplier rule)*
2. *$h(\overline{x}) = 0$*
3. *$\overline{\lambda} \geq 0, g(\overline{x}), \overline{\lambda}^\top g(\overline{x}) = 0$ (complementarity condition)*

Theorem 1.23. *let g be differentiable. then the condition*

$$g_i \text{ concave, } i \in \mathcal{A}(x), \quad h \text{ affine linear}$$

is a constraint qualification at $x \in X$

Theorem 1.24. *let g and h be continuously differentiable and let $x \in X$ be such that the MFCQ holds at x. then the ACQ holds at x, i.e., the MFCQ is a constraint qualification at x.*

Lemma 1.25 (lemma of alternative)**.** *let $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{p \times n}$. then the following statements are equivalent:*

1. *there are no vectors $u \in \mathbb{R}^m$, $v \in \mathbb{R}^p$ with $A^\top u + B^\top v = 0$, $u \geq 0$ and $u \neq 0$*
2. *there exists $d \in \mathbb{R}^n$ with $Ad < 0$ and $Bd = 0$.*

this lemma also holds for $p = 0$: just omit all terms with B or v

Theorem 1.26. *the point $x \in X$ satisfies the MFCQ if and only if PLICQ holds at x. in particular, PLICQ is a constraint qualification. this holds for the generalizations of MFCQ and PLICQ, too.*

Lemma 1.27. *the condition LICQ is a constraint qualification*

1.2 Convex Optimization

Definition 1.28 (convex optimization problem). *if the function f, g_i are convex and h is linear, the problem is called convex optimization problem.*

Definition 1.29 (Slater's condition). *we say that Slater's condition is satisfied if there exists a point $y \in \mathbb{R}^n$ such that $g(y) < 0$ and $h(y) = 0$*

Theorem 1.30. *let problem (1) be convex. then every local solution of (1) is a global solution. furthermore it holds:*

1. *if $\overline{x}$ is a local solution and a constraint qualification is satisfied at $\overline{x} \in X$, then the KKT conditions of theorem1.16 hold at $\overline{x}$*

2. *on the other hand, if the KKT conditions hold at $\overline{x}$, then $\overline{x}$ is a global solution of (1).*

Lemma 1.31. *consider problem (1) where the feasible set X is convex, i.e. the functions g_i for $i = 1, 2, ..., m$ are convex and h is affine. then slater's condition is a constraint qualification for every feasible point x of problem (1).*

1.2.1 Second-order optimality conditions

Definition 1.32 (critical cone). *for $x \in X$ and $\lambda \in [0, \infty)^m$ we define the critical cone:*

$$T_x(g, h, x, \lambda) = \left\{ d \in \mathbb{R}^n : \nabla g_i(x)^T d \begin{cases} = 0, & \text{if } i \in A(x) \text{ and } \lambda_i > 0 \\ \leq 0, & \text{if } i \in A(x) \text{ and } \lambda_i = 0 \end{cases}, \nabla h(x)^T d = 0 \right\}$$

Remark. *the cone $T_+(g, h, x, \lambda)$ is contained in the linearized tangent cone $T_l(g, h, x)$ and includes the tangent space*

$$T_a(g, h, x) = \{d \in \mathbb{R}^n : \nabla g_i(x)^\top d = 0, i \in \mathcal{A}(x), \nabla h(x)^\top d = 0\}$$

of active constraints:

$$T_a(g, h, x) \subset T_+(g, h, x, \lambda) \subset T_l(g, h, x)$$

furthermore, we have $T_a(g, h, x) = T_+(g, h, x, \lambda)$ if the strict complementarity holds.

Definition 1.33 (CQ2). *we say that $(x, \lambda, \mu) \in X \times [0, \infty)^m \times \mathbb{R}^p$ satisfies a second-order constraint qualification (CQ2) if for all $d \in T_+(g, h, x, \lambda)$ there exists an open interval $J \supset \{0\}$ and a twice continuously differentiable curve $\gamma : J \to \mathbb{R}^n$ such that*

$$\gamma(0) = x, \quad \gamma'(0) = d,$$

$$g_{A_0(x,d)}(\gamma(t)) = 0, \quad h(\gamma(t)) = 0 \quad \forall t \in J, t \geq 0$$

where $\mathcal{A}_0(x, d) := \{i \in \mathcal{A}(x); \nabla g_i(x)^\top d = 0\}$

Lemma 1.34. *if $(\overline{x}, \overline{\lambda}, \overline{\mu})$ is a KKT triple, then it holds*

$$T_+(g, h, \overline{x}, \overline{\lambda}) = \{d \in T_l(g, h, \overline{x}), \nabla f(\overline{x})^\top d = 0\}$$

Theorem 1.35 (second-order sufficient optimality condition). *let $\overline{x} \in \mathbb{R}^n$ satisfy the KKT conditions with multipliers $\overline{\lambda} \in \mathbb{R}^m$ and $\overline{\mu} \in \mathbb{R}^p$. suppose that*

$$d^\top \nabla_{xx} L(\overline{x}, \overline{\lambda}, \overline{\mu}) d > 0, \quad d \in T_+(g, h, \overline{x}, \overline{\lambda}) \setminus \{0\}$$

holds. then $\overline{x}$ is a strict local solution of (1).

Lemma 1.36. *let g and h be twice continuously differentiable functions and $(x, \lambda, \mu) \in X \times [0, \infty)^m \times \mathbb{R}^p$. if $x \in X$ is regular, it holds (CQ2).*

Theorem 1.37 (second-order necessary optimality conditions). *let f, g and h be twice continuously differentiable. suppose that $\overline{x}$ is a local solution of (1) where the GCQ holds. then there exist Lagrange multipliers $\overline{\lambda} \in \mathbb{R}^m$ and $\overline{\mu} \in \mathbb{R}^p$ with:*

1. $\nabla_x L(\overline{x}, \overline{\lambda}, \overline{\mu}) = 0$

2. $h(\overline{x}) = 0$

3. $\overline{\lambda} \geq 0$, $g(\overline{x}) \leq 0, \overline{\lambda}^\top g(\overline{x}) = 0$.
   if, in addition, (CQ2) is satisfied at $(\overline{x}, \overline{\lambda}, \overline{\mu})$, we have

4. $d^\top \nabla_{xx} L(\overline{x}, \overline{\lambda}, \overline{\mu}) d \geq 0$ *for all* $d \in T_+(g, h, \overline{x}, \overline{\lambda})$

1.3 Duality

1.4 Optimization Algorithms

1.4.1 Penalty methods

1. the quadratic penalty method
   the quadratic penalty method for problem (1) uses the quadratic penalty function:

$$P_\alpha(x) = f(x) + \frac{\alpha}{2}\sum_{i=1}^{m}\max{}^2\{0, g_i(x)\} + \frac{\alpha}{2}\sum_{i=1}^{p} h_i(x)^2$$

$$= f(x) + \frac{\alpha}{2}\|g(x)_+\|^2 + \frac{\alpha}{2}\|h(x)\|^2$$

where $(v)_+ \in \mathbb{R}^n$ denotes the vector with components $((v)_+)_i = \max\{0, v_i\}$ for a vector $v \in \mathbb{R}^n$. the scalar $\alpha > 0$ is the penalty parameter. note $(t)_+^2$ is C^1 with derivative $2(t)_+$. we have

$$\nabla P_\alpha(x) = \nabla f(x) + \alpha\sum_{i=1}^{m}(g_i(x))_+\nabla g_i(x) + \alpha\sum_{i=1}^{p} h_i(x)\nabla h_i(x)$$

and thus

$$P_\alpha(x) = f(x), \quad \nabla P_\alpha(x) = \nabla f(x) \quad \forall x \in X$$

Algorithm 1.38.

(a) *choose* $\alpha_0 > 0$
for $k = 0, 1, 2, \ldots$:

(b) *find a global minimizer* x^k *of the penalty problem*

$$\min_{x\in\mathbb{R}^n} P_{\alpha_k}(x)$$

for $k > 0$ *we usually use* x^{k-1} *as a starting point.*

(c) *STOP if* $x^k \in X$

(d) *Choose* $\alpha_{k+1} > \alpha_k$

Theorem 1.39. *let* f, g, h *be continuous and the feasible set* X *be nonempty. Suppose the sequence* $(\alpha_k) \subset (0, \infty)$ *is strictly monotonically increasing to infinity and let the algorithm generate a sequence* (x^k). *we write* $\pi(x) \triangleq \frac{1}{2}\left(\|(g(x))_+\|^2 + \|h(x)\|^2\right)$ *for the penalty term. then it holds:*

(a) *the sequence* $(P_{\alpha_k}(x^k))$ *is monotonically increasing*

(b) *the sequence* $(\|(g(x^k))_+\|^2 + \|h(x^k)\|^2)$ *is monotonically decreasing.*

(c) *the sequence* $(f(x^k))$ *is monotonically increasing.*

(d) *it holds* $\lim_{k\to\infty}\pi(x^k) = 0$, $\lim_{k\to\infty}(g(x^k))_+ = 0$,$\lim_{k\to\infty} h(x^k) = 0$

(e) *every accumulation point of the sequence* (x^k) *is a global solution of (1)*

Definition 1.40. *we define the sequences* (λ^k) *and* (μ^k) *according to the following definitions:*

$$\lambda_i^k \triangleq \alpha_k \max\{0, g_i(x^k)\}, \quad \mu_i^k \triangleq \alpha_k h_i(x^k) \tag{9}$$

then every generated point x^k *is a stationary point of* P_{α_k} *and hence*

$$0 = \nabla P_{\alpha_k}(x^k) = \nabla f(x^k) + \sum_{i=1}^{m}\alpha_k\max\{0, g_i(x^k)\}\nabla g_i(x^k) + \sum_{i=1}^{p}\alpha_k h_i(x^k)\nabla h_i(x^k)$$

$$= \nabla f(x^k) + \nabla g(x^k)\lambda^k + \nabla h(x^k)\mu^k$$

Theorem 1.41. *let* f, g, h *be continuously differentiable and assume that the feasible set* X *is non-empty. suppose the sequence* $(\alpha_k) \subset (0, \infty)$ *is strictly monotonically increasing to infinity. let Algorithm1.38 generate a sequence* (x^k) *(we assume that this sequence exists). then the following holds:*

(a) *if* $(x^k, \lambda^k, \mu^k)_K$ *is a subsequence of* (x^k, λ^k, μ^k) *converging to* $(\overline{x}, \overline{\lambda}, \overline{\mu})$, *then* $\overline{x}$ *is a global solution of (1) and* $(\overline{x}, \overline{\lambda}, \overline{\mu})$ *is a KKT triple of (1)*

(b) *suppose $\overline{x}$ is an accumulation point of (x^k) and $(x^k)_K$ is a subsequence converging to $\overline{x}$. Furthermore, let $\overline{x}$ be a regular point. then $(x^k, \lambda^k, \mu^k)_K$ converges to a KKT triple of (1) and $\overline{x}$ is a global solution of (1)*

in algorithm1.38, we are supposed to find a global minimizer of the unconstrained subproblem. in practice, this may be difficult if we have a non-convex problem. therefore, we introduce a more general algorithm that also allows for the inexact solution of the unconstrained subproblems.

Algorithm 1.42. *(a) choose $\alpha_0 > 0$ and a nonnegative sequence $(\epsilon_k)_{k\in\mathbb{N}}$ with $\epsilon_k \to 0$*
for $k = 0, 1, 2, \ldots$:

(b) *find an approximate local minimizer x^k of the penalty problem*

$$\min_{x\in\mathbb{R}^n} P_{\alpha_k}(x)$$

and terminate when

$$\|\nabla P_{\alpha_k}(x^k)\| \leq \epsilon_k$$

for $k > 0$ we usually use x^{k-1} as a starting point.

(c) *choose $\alpha_{k+1} > \alpha_k$*

Theorem 1.43. *Let f, g, and h be continuously differentiable and assume that the feasible set X is non-empty. Suppose the sequence $(\alpha_k) \subset (0, \infty)$ is strictly monotonically increasing to infinity. Furthermore, let $(\epsilon_k) \subset (0, \infty)$ be a sequence with $\epsilon_k \to 0$. Let Algorithm 1.42 generate a sequence (x^k) (we assume that this sequence exists). We define the sequences (λ^k) and (μ^k) according to (9). Then, the following holds:*

(a) *Suppose $\bar{x}$ is an accumulation point of (x^k). Then $\bar{x}$ is a stationary point of the penalty term*

$$\pi(x) = \frac{1}{2}\left(\|g(x)\|^2 + \|h(x)\|^2\right).$$

(b) *If $(x^k, \lambda^k, \mu^k)_K$ is a subsequence of (x^k, λ^k, μ^k) converging to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ where $\bar{x} \in X$ is a feasible point, then $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a KKT triple of (1).*

(c) *Suppose a feasible point $\bar{x} \in X$ is an accumulation point of (x^k) and $(x^k)_K$ is a subsequence converging to $\bar{x}$. Furthermore, let $\bar{x}$ be a regular point. Then $(x^k, \lambda^k, \mu^k)_K$ converges to a KKT triple of (1).*

2. Exact penalty functions

Definition 1.44 (exact penalty function)**.** *Let $\bar{x} \in \mathbb{R}^n$ be a local solution of (1). The penalty function $P : \mathbb{R} \to \mathbb{R}$ is called exact at the point $\bar{x}$ if $\bar{x}$ is a local minimum of P.*

Under appropriate assumptions, the following ℓ_1 penalty function is exact if $\alpha > 0$ is sufficiently large:

$$P^1_\alpha(x) = f(x) + \alpha\sum_{i=1}^{m}(g_i(x))_+ + \alpha\sum_{i=1}^{p}|h_i(x)| = f(x) + \alpha\left(\|(g(x))_+\|_1 + \|h(x)\|_1\right).$$

However, a drawback of the ℓ_1 penalty function is the non-differentiability of P^1_α that comes from the non-differentiability of the functions $(\cdot)_+$ and $|\cdot|$.

We now show that P^1_α is exact for convex optimization problems:

Theorem 1.45. *Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ be a KKT triple of the optimization problem (1) with convex continuously differentiable functions $f, g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$ and an affine function $h : \mathbb{R}^n \to \mathbb{R}^p$. Then $\bar{x}$ is a global solution of (1) and furthermore $\bar{x}$ is a global minimum of P_α on $\mathbb{R}^n$ for all*

$$\alpha \geq \max\{\bar{\lambda}_1, \ldots, \bar{\lambda}_m, |\bar{\mu}_1|, \ldots, |\bar{\mu}_p|\}.$$

1.4.2 SQP methods

We begin by considering the equality constrained problem

$$\min f(x) \quad \text{s.t.} \quad h(x) = 0, \tag{11}$$

and then extend the ideas to problems that also have inequality constraints.

1. Lagrange-Newton method for equality constraints
   Let $\bar{x}$ be a local solution of (11), where a CQ holds (h affine or Rank$\nabla h(\bar{x}) = p$). Then the KKT conditions are satisfied: There exists $\bar{\mu} \in \mathbb{R}^p$ with

$$\begin{aligned} \nabla_x L(\bar{x}, \bar{\mu}) &= 0, \\ h(\bar{x}) &= 0. \end{aligned}$$

   This system has $n+p$ variables $(\bar{x}, \bar{\mu})$ and $n+p$ equations. Thus, for the computation of $(\bar{x}, \bar{\mu})$, it appears promising to apply Newton's method to the system of equations

$$F(x, \mu) := \begin{pmatrix} \nabla_x L(x, \mu) \\ h(x) \end{pmatrix} = 0.$$

   Let f and h be twice continuously differentiable. Then F is continuously differentiable with

$$F'(x, \mu) = \begin{pmatrix} \nabla_{xx} L(x, \mu) & \nabla_{x\mu} L(x, \mu) \\ \nabla h(x)^T & 0 \end{pmatrix} = \begin{pmatrix} \nabla_{xx} L(x, \mu) & \nabla h(x) \\ \nabla h(x)^T & 0 \end{pmatrix}.$$

   If we denote the current iterate by (x^k, μ^k), then the Newton step d^k for (12) is given by

$$F'(x^k, \mu^k) d^k = -F(x^k, \mu^k).$$

   Thus, we have

$$\begin{pmatrix} \nabla_{xx} L(x^k, \mu^k) & \nabla h(x^k) \\ \nabla h(x^k)^T & 0 \end{pmatrix} \begin{pmatrix} d_x^k \\ d_\mu^k \end{pmatrix} = \begin{pmatrix} -\nabla_x L(x^k, \mu^k) \\ -h(x^k) \end{pmatrix},$$

   with $d^k = \begin{pmatrix} d_x^k \\ d_\mu^k \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R}^p$.

   Algorithm 1.46 (Lagrange-Newton method)**.**

   (a) Choose $x^0 \in \mathbb{R}^n$ and $\mu^0 \in \mathbb{R}^p$.
   For $k = 0, 1, 2, \ldots$:

   (b) If $h(x^k) = 0$ and $\nabla_x L(x^k, \mu^k) = 0$: STOP

   (c) Solve

$$\begin{pmatrix} \nabla_{xx} L(x^k, \mu^k) & \nabla h(x^k) \\ \nabla h(x^k)^T & 0 \end{pmatrix} \begin{pmatrix} d_x^k \\ d_\mu^k \end{pmatrix} = \begin{pmatrix} -\nabla_x L(x^k, \mu^k) \\ -h(x^k) \end{pmatrix}$$

   to obtain $d^k = \begin{pmatrix} d_x^k \\ d_\mu^k \end{pmatrix}$.

   (d) Set $x^{k+1} = x^k + d_x^k$, $\mu^{k+1} = \mu^k + d_\mu^k$.

   Lemma 1.47. *Let f and h be twice differentiable and $x \in \mathbb{R}^n$, $\mu \in \mathbb{R}^p$ be arbitrary. If we have*

$$\text{Rank } \nabla h(x) = p \quad \text{and} \quad s^T \nabla_{xx} L(x, \mu) s > 0 \quad \forall s \in \mathbb{R}^n \setminus \{0\} \quad \text{with } \nabla h(x)^T s = 0, \tag{14}$$

   then the matrix

$$F'(x, \mu) = \begin{pmatrix} \nabla_{xx} L(x, \mu) & \nabla h(x) \\ \nabla h(x)^T & 0 \end{pmatrix}$$

   is invertible.

   Theorem 1.48. *Let f and h be twice continuously differentiable and $(\bar{x}, \bar{\mu})$ be a KKT pair with*

$$\text{Rank } \nabla h(\bar{x}) = p \quad \text{(regularity)},$$

$$s^T \nabla_{xx} L(\bar{x}, \bar{\mu}) s > 0 \quad \forall s \in \mathbb{R}^n \setminus \{0\} \quad \text{with } \nabla h(\bar{x})^T s = 0 \quad \text{(2nd order suff. cond.)}.$$

   Then there exists a $\delta > 0$ such that for all $(x^0, \mu^0) \in B_\delta(\bar{x}, \bar{\mu})$, Algorithm 4.8 either terminates with $(x^k, \mu^k) = (\bar{x}, \bar{\mu})$ or generates a sequence (x^k, μ^k) that converges superlinearly to $(\bar{x}, \bar{\mu})$:

$$\|(x^{k+1} - \bar{x}, \mu^{k+1} - \bar{\mu})\| = o(\|(x^k - \bar{x}, \mu^k - \bar{\mu})\|) \quad (k \to \infty).$$

   The convergence rate is quadratic if $\nabla^2 f$ and $\nabla^2 h_i$ are Lipschitz continuous on $B_\delta(\bar{x})$.

2. The local SQP method

3. SQP methods for equality and inequality constraints

4. Globalized SQP methods

5. Problems and further aspects

1.4.3 Barrier methods

1.4.4 Augmented Lagrangian methods

1.4.5 Projected gradient

2 Exercises

2.1 Exercise 1

1. Repetition: Existence of Solutions: Suppose that $X \subseteq \mathbb{R}^n$, $n \in \mathbb{N}$, is a nonempty set and that $f : X \to \mathbb{R}$ is a continuous function. We consider the constrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t. } x \in X. \tag{P1}$$

Decide whether the optimization problem (P1) has a global solution under either of the following conditions. Provide a proof or a counterexample.

(a) The set X is closed and $\lim_{x \in X, \|x\| \to \infty} f(x) = \infty$.

solution. let $y \in X$ be arbitrary and consider the sublevel set $N_f(y) = \{x \in X | f(x) \leq f(y)\}$. we show that $N_f(y)$ is compact. first show closedness, then show boundedness by contradiction then we can replace the constraint $x \in X$ by $x \in N_f(y)$. f is continuous on the compact sublevel set $N_f(y)$ yields the existence of an optimal solution. (extreme value theorem) □

(b) The set X is open and $\lim_{x \in X, \|x\| \to \infty} f(x) = \infty$.

solution. counterexample: choose $f(x) = x$ and $X = (0, \infty)$, clearly $f(x)$ has no minimum as $x = 0$ cannot be attained. □

(c) The set X is closed and bounded, and $\lim_{x \in X, \|x\| \to \infty} f(x) = -\infty$.

solution. by Weierstrass' theorem (extreme value theorem), f is continuous and X is bounded and closed $\Rightarrow$ compact. f attains its minimum in X □

(d) The set X and the function f are convex.

solution. counterexample: $\min_{x \in \mathbb{R}} e^x$, both objective function and set X are convex but the problem has no minimum. □

(e) The set X is convex and closed, and the function f is strictly convex.

solution. using the counterexample in (d) □

(f) The set X is convex and closed, and the function f is strongly convex.

solution. under this condition, (P1) indeed has a global minimum. the strongly convexity says that $\exists \mu > 0$ s.t. for all $x, y \in X$ and all $\lambda \in [0, 1]$, we have

$$f(\lambda x + (1 - \lambda)y) + \mu\lambda(1 - \lambda)\|x - y\|^2 \leq \lambda f(x) + (1 - \lambda)f(y)$$

we aim to show that $\lim_{x \in X, \|x\| \to \infty} f(x) = \infty$ and we can apply the conclusion of (a).
the idea is that let $\lambda = \frac{1}{\|x-y\|}$ in the def of strong convexity and for arbitrary fixed $y \in X$, suppose $x \in X$ with $\|x - y\| > 1$:

$$f(y + \frac{x - y}{\|x - y\|}) + \mu(\|x - y\| - 1) \leq \frac{1}{\|x - y\|} f(x) + (1 - \frac{1}{\|x - y\|})f(y)$$

this implies that

$$-[f(y)] - \left(\mu + |f(y)| + \sup_{z \in X, \|z - y\| = 1} |f(z)|\right) \|x - y\| + \mu\|x - y\|^2 \leq f(x)$$

for all $x \in X$ satisfying $\|x - y\| > 1$. Note that the supremum in the above estimate is finite since the set $X \cap \{z \in \mathbb{R}^n : \|z - y\| = 1\}$ is compact and f is continuous; see (c). We deduce the existence of constants $C_1, C_2 > 0$ (which depend on y) such that

$$-C_1 - C_2\|x-y\| + \mu\|x-y\|^2 \le f(x)$$

holds for all $x \in X$ with $\|x-y\| > 1$. The above derivation implies that

$$\lim_{x\in X, \|x\|\to\infty} f(x) = \infty$$

and this, in turn, yields, by the triangle inequality,

$$\lim_{x\in X, \|x\|\to\infty} f(x) = \infty.$$

□

2. Tangent Cone of a Convex Set: Let $n \in \mathbb{N}, M \subseteq \mathbb{R}^n$ be a nonempty, convex set and let $x \in M$. Show that

$$T(M,x) = \overline{\{d \in \mathbb{R}^n : \exists \eta > 0, y \in M : d = \eta(y-x)\}}.$$

solution. define $T \triangleq \{d \in \mathbb{R}^n : \exists \eta > 0, y \in M : d = \eta(y-x)\}$. we show that $T(M,x) \subset \overline{T}$ and $\overline{T} \subset T(M,x)$. recall the definition of tangent cone: $T(M,x) \triangleq \{d \in \mathbb{R}^n : d = \lim_{k\to\infty} d^k, d^k = \eta_k(x^k - x), \eta_k > 0, x^k \in M\}$.
we first show that $T(M,x) \subset \overline{T}$: let $d \in T(M,x)$ be arbitrary. there exists $\eta_k > 0$ and $x^k \in M$ s.t. $d = \lim_{k\to\infty} d^k$, $d^k = \eta_k(x^k - x)$, by definition of T, we find that $d^k \in T$, and thus $d_k \to \infty d \in \overline{T}$
Next, we show that $T \subseteq T(M,x)$ using two different approaches.

First approach: Let $d \in T$. By definition, there exists $\eta > 0$ and $y \in M$ such that $d = \eta(y-x)$. We define $x^k := (1/\alpha_k)y + (1 - 1/\alpha_k)x$ with $\alpha_k := k$. Since M is convex, $1/\alpha_k \in (0,1]$, and $y, x \in M$, we have $x^k \in M$ for each $k \in \mathbb{N}$ and $x^k - x = (1/\alpha_k)(y-x) \to 0$ as $k \to \infty$. Defining $\eta_k := \alpha_k \eta$, we have $\eta_k(x^k - x) = \eta(y-x) = d \to d$ as $k \to \infty$. Hence $d \in T(M,x)$.

Since $T(M,x)$ is closed (see T.2.4 (c)), the inclusion $T \subseteq T(M,x)$ ensures that $\overline{T} \subseteq T(M,x)$.

Second approach: Let $d \in \overline{T}$. By definition, there exists a sequence $(d^k) \subseteq T$ with $d = \lim_{k\to\infty} d^k$. Furthermore, there exist $\eta_k > 0$ and $y^k \in M$ such that $d^k = \eta_k(y^k - x)$. We define

$$\alpha_k := k(\|y^k\| + 1), \quad \tilde{\eta}_k := \alpha_k \eta_k, \quad \text{and} \quad x^k := \frac{1}{\alpha_k} y^k + \left(1 - \frac{1}{\alpha_k}\right) x.$$

Then $\alpha_k > k$ and $\tilde{\eta}_k > 0$. Since $1/\alpha_k \in (0,1]$ and M is convex, we have $x^k \in M$ for all $k \in \mathbb{N}$. For each $k \in \mathbb{N}$, we obtain that

$$x^k - x = \frac{1}{k(\|y^k\|+1)} y^k - \frac{1}{k(\|y^k\|+1)} x \quad \text{and} \quad \tilde{\eta}_k(x^k - x) = \eta_k(y^k - x) = d^k.$$

Consequently, $x^k \to x$ and $\tilde{\eta}_k(x^k - x) \to d$ as $k \to \infty$. Hence $d \in T(M,x)$.

□

3. Optimality Conditions for Convex Problems: Consider the minimization problem

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{s.t. } x \in X, \tag{P2}$$

where $n \in \mathbb{N}, X \subseteq \mathbb{R}^n$ is a nonempty, convex set, and $f : \mathbb{R}^n \to \mathbb{R}$ is a function that is continuously differentiable on a neighborhood U of the set X.

(a) Show that if $\bar{x}$ is a global solution to $(P2)$, then it satisfies

$$\bar{x} \in X \quad \text{and} \quad \nabla f(\bar{x})^\top (w - \bar{x}) \ge 0 \quad \text{for all } w \in X. \tag{VI}$$

Remark: Inequalities of the form (VI) are called *variational inequalities.*

solution. let $\overline{x}$ be a solution to (P2). let $w \in X$ be arbitrary. we define the line segment $v(t) = \overline{x} + t(w-x)$ for $t \in [0,1]$. since X is a convex set, $v(t) \in X$ for all $t \in [0,1]$. we obtain

$$\nabla f(\overline{x})^\top (w-x) = \lim_{t\to 0^+} \frac{1}{t}(f(v(t)) - f(\overline{x})) \ge 0$$

□

(b) Assume that f is convex. Show that if $\bar{x}$ satisfies (VI), then it is a global solution to $(P2)$.

solution. f is convex, we obtain

$$f(w) \geq f(\overline{x}) + \nabla f(\overline{x})^\top (w - \overline{x}) \geq f(\overline{x}), \quad \forall w \in X$$

hence x is a global solution to (P2). □

In the following, we additionally assume that the set X is closed.

(c) The *Euclidean projection* $P_X(y) \in X$ of $y \in \mathbb{R}^n$ onto the set X is defined as the global solution to the following special case of problem $(P2)$:

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|x - y\|^2 \quad \text{s.t. } x \in X. \tag{P3}$$

Prove that $P_X : \mathbb{R}^n \to X$ is well-defined by showing that, for each $y \in \mathbb{R}^n$, $(P3)$ has a unique global solution.

solution. Let $y \in \mathbb{R}^n$ be arbitrary but fixed. The objective function $h_y : \mathbb{R}^n \to \mathbb{R}$ defined by

$$h_y(x) = \frac{1}{2}\|x - y\|_2^2$$

of $(P3)$ is continuous and coercive on $\mathbb{R}^n$, i.e., $\lim_{x\in X, \|x\|\to\infty} h_y(x) = \infty$. Since X is nonempty and closed, Exercise 1.1(a) yields the existence of a global solution to problem $(P3)$. The uniqueness follows from the fact that h_y is strongly convex. (The function h_y is twice continuously differentiable on $\mathbb{R}^n$ and its Hessian matrix is positive definite for $x \in X$, and X is convex.)

□

(d) Let $\tau > 0$ be arbitrary and let f be convex. Show that $\bar{x}$ is a global solution to $(P2)$ if and only if

$$\bar{x} = P_X(\bar{x} - \tau\nabla f(\bar{x})).$$

solution. Let $\tau > 0$ be fixed and let $x \in \mathbb{R}^n$ satisfy $d_\tau(x) = 0$, where $d_\tau(x) := x - \Pi_X(x - \tau\nabla f(x))$. Using part (c), we obtain that $\bar{x}$ is the unique global solution to the problem

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|x - (x - \tau\nabla f(x))\|_2^2 \quad \text{s.t.} \quad x \in X.$$

Thus, we can use part (a) and (b) (f is assumed to be convex) to characterize the optimality of x:

$$\begin{aligned}
d_\tau(x) = 0 \quad &\iff \quad (x - (x - \tau\nabla f(x)))^\top (w - x) \geq 0 \quad \text{for all } w \in X \\
&\iff \quad \nabla f(x)^\top (w - x) \geq 0 \quad \text{for all } w \in X \\
&\iff \quad x \text{ is a global solution to the problem } (P2).
\end{aligned}$$

Remark: the "$\Longleftarrow$"-direction in the last step follows from (a) and hence does not require convexity of f.

□

2.2 Exercises 2

1. We consider

$$\min_{x\in X} f(x), \tag{P}$$

where $X \subset \mathbb{R}^n$ is a nonempty set and $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable function.

Suppose that $\bar{x} \in \mathbb{R}^n$ satisfies the first-order sufficient optimality conditions:

(a) $\bar{x} \in X$,

(b) $\nabla f(\bar{x})^\top d > 0$ for all $d \in T(X, \bar{x}) \setminus \{0\}$.

Show that $\bar{x}$ is a strict local solution to problem (P), that is, there exists $\varepsilon > 0$ such that $f(\bar{x}) < f(x)$ for all $x \in X \setminus \{\bar{x}\}$ with $\|x - \bar{x}\| < \varepsilon$.

solution. Assume that $\bar{x}$ satisfies the first-order sufficient optimality condition and that $\bar{x}$ is not a strict local solution of problem (P). Then, there exists a sequence $(x^k) \subset X$ with $x^k \neq \bar{x}$ such that

$$f(x^k) \leq f(\bar{x}) \quad \text{and} \quad x^k \to \bar{x} \quad \text{as } k \to \infty.$$

Let us define $t_k := \|x^k - \bar{x}\|$ and $d^k := t_k^{-1}(x^k - \bar{x})$. Since (d^k) is bounded, there exists a subsequence $(d^{k'})$ converging to some $d \in \mathbb{R}^n$. Since $x^k \to \bar{x}$ as $k \to \infty$, $t_k > 0$ and $\|d_k\| = 1$ for all $k \in \mathbb{N}$, we have $d \in T(X, \bar{x}) \setminus \{0\}$.

A first-order Taylor's expansion yields

$$0 \geq f(x^k) - f(\bar{x}) = f(\bar{x} + t_k d^k) - f(\bar{x}) = t_k \nabla f(\bar{x})^\top d^k + o(t_k) \quad \text{as } k \to \infty.$$

Dividing by t_k and taking limits as $k \to \infty$, we obtain that $\nabla f(\bar{x})^\top d \leq 0$, a contradiction to the first-order sufficient optimality conditions.

□

2. For $g : \mathbb{R}^2 \to \mathbb{R}^2$, we define $X = \{x \in \mathbb{R}^2 : g(x) \leq 0\}$. Decide whether the ACQ and the GCQ condition hold at $\bar{x} = 0$ for the following choices of g:

   (a) $g(x) = (-x_1, x_2^2)^\top$,

   (b) $g(x) = (x_2 - x_1^3, -x_2)^\top$.

Hints: Visualize X, $\nabla g_1(\bar{x})$ and $\nabla g_2(\bar{x})$, and use the inclusion $T(X, \bar{x}) \subset T_l(g, \bar{x})$.

solution. (a) $X = \{x \in \mathbb{R}^2 : x_1 \geq 0, x_2^2 \leq 0\} = \mathbb{R}_{\geq 0} \times \{0\}$. We have

$$\mathcal{A}(\bar{x}) = \{1, 2\}, \quad \nabla g_1(\bar{x}) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad \nabla g_2(\bar{x}) = \begin{pmatrix} 0 \\ 2x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and

$$T_l(g, \bar{x}) = \big\{d \in \mathbb{R}^2 : (-1, 0)d \leq 0,\ (0, 0)d \leq 0\big\} = \{d \in \mathbb{R}^2 : d_1 \geq 0\}.$$

We claim $T(X, \bar{x}) = X$.

"$\subset$": Let $d \in T(X, \bar{x})$ and let us consider arbitrary sequences $(x^k) \subset X$ and $(\eta_k) \subset \mathbb{R}_{>0}$ with $d^k = \eta_k(x^k - x) \to d$ as $k \to \infty$. Owing to $\bar{x} = 0$, $x^k \in X$, and $\eta_k > 0$, we have

$$d_1^k \geq 0 \quad \text{and} \quad d_2^k = 0 \quad \Rightarrow \quad d^k \in X \quad \text{for all } k \in \mathbb{N}.$$

Since the set X is closed, we obtain $d \in X$. Hence $T(X, \bar{x}) \subset X$.

"$\supset$": Now, let $d \in X$ be arbitrary and let us define $\eta_k = k$, $x^k = (1/k)d$. Then, it follows

$$(\eta_k) \subset \mathbb{R}_{>0}, \quad (x^k) \subset X, \quad \lim_{k\to\infty} x^k = 0 = \bar{x}, \quad \lim_{k\to\infty} \eta_k(x^k - \bar{x}) = d.$$

This shows $X \subset T(X, \bar{x})$.

Since $T(X, \bar{x}) \neq T_l(g, \bar{x})$, the ACQ is not satisfied at $\bar{x}$.

We have

$$T_l(g, \bar{x})^\circ = \big\{d \in \mathbb{R}^2 : v^\top d \leq 0 \text{ for all } v \in \mathbb{R}_{\geq 0} \times \mathbb{R}\big\} = \{(t, 0)^\top : t \leq 0\},$$
$$T(X, \bar{x})^\circ = \big\{d \in \mathbb{R}^2 : (v, 0)d \leq 0 \text{ for all } v \in \mathbb{R}_{\geq 0}\big\} = \mathbb{R}_{\leq 0} \times \mathbb{R}.$$

Thus, the GCQ is also not satisfied at $\bar{x}$.

Remark. *This example illustrates that the linearized tangent cone $T_l(g, \bar{x})$ strongly depends on the structure of the function g. Here, the feasible set X can also be described and defined via linear constraints. In this case and in contrast to the present situation, a constraint qualification would then hold at $\bar{x}$.*

(b) We have

$$\mathcal{A}(\bar{x}) = \{1, 2\}, \quad \nabla g_1(\bar{x}) = \begin{pmatrix} -3x_1^2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \nabla g_2(\bar{x}) = \begin{pmatrix} 0 \\ -1 \end{pmatrix},$$

and

$$T_l(g, \bar{x}) = \{d \in \mathbb{R}^2 : (0, 1)d \leq 0,\ (0, -1)d \leq 0\} = \{(t, 0)^\top : t \in \mathbb{R}\}.$$

Let $d \in T(X,\bar{x})$ be arbitrary. Since $T(X,\bar{x}) \subset T_l(g,\bar{x})$, there exists $t \in \mathbb{R}$ such that $d = (t,0)^\top$.

Next, let us consider $(x^k) \subset X$, $x^k \to \bar{x}$ and $\eta_k > 0$ with $\eta_k(x^k - \bar{x}) \to d$. Then, it holds

$$d^k = \eta_k(x^k - \bar{x}) = \eta_k x^k \geq 0,$$

where we used the nonnegativity of $x^k \in X$. Thus, it follows $d = \lim_{k\to\infty} d^k \geq 0$ and we can infer $T(X,\bar{x}) \subset \{(t,0)^\top : t \in \mathbb{R}_{\geq 0}\}$. Consequently, the ACQ is violated at $\bar{x}$. Finally, let us show $T(X,\bar{x}) = \{(t,0)^\top : t \in \mathbb{R}_{\geq 0}\}$. Let $t \geq 0$ be arbitrary. We define the sequences $x^k = (t/k, 0)^\top$ and $\eta_k = k$. Then

$$x^k \in X, \quad \eta_k > 0, \quad \lim_{k\to\infty} x^k = \bar{x}, \quad \text{and} \quad \lim_{k\to\infty} \eta_k(x^k - \bar{x}) = (t,0)^\top.$$

This establishes $(t,0)^\top \in T(X,\bar{x})$ and $T(X,\bar{x}) = \{(t,0)^\top : t \in \mathbb{R}_{\geq 0}\}$.

We have

$$\begin{aligned} T_l(g,\bar{x})^\circ &= \{d \in \mathbb{R}^2 : (t,0)d \leq 0 \text{ for all } t \in \mathbb{R}\} = \{d \in \mathbb{R}^2 : d_1 = 0, d_2 \in \mathbb{R}\}, \\ T(X,\bar{x})^\circ &= \{d \in \mathbb{R}^2 : (t,0)d \leq 0 \text{ for all } t \geq 0\} = \{d \in \mathbb{R}^2 : d_1 \leq 0, d_2 \in \mathbb{R}\}. \end{aligned}$$

Since $T_l(g,\bar{x})^\circ \neq T(X,\bar{x})^\circ$, the GCQ is not satisfied at $\bar{x}$.

□

3. We consider $X = \{x \in \mathbb{R}^2 : g(x) \leq 0\}$, where $g : \mathbb{R}^2 \to \mathbb{R}^3$ is defined by $g(x) = (-x_1, -x_2, x_1 x_2)^\top$.

   Show that the GCQ holds at 0, but the ACQ is violated at 0.

   Hints: Visualize X, $\nabla g_1(\bar{x})$ and $\nabla g_2(\bar{x})$, and use the inclusion $T(X,\bar{x}) \subset T_l(g,\bar{x})$, where $\bar{x} = 0$.

solution. We define $\bar{x} = 0$. We have $X = \{x \in \mathbb{R}^2_{\geq 0} : x_1 x_2 = 0\}$.

We have

$$\mathcal{A}(\bar{x}) = \{1,2,3\}, \quad \nabla g_1(\bar{x}) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad \nabla g_2(\bar{x}) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad \nabla g_3(\bar{x}) = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

and

$$T_l(g,\bar{x}) = \{d \in \mathbb{R}^2 : -d_1 \leq 0, \ -d_2 \leq 0\} = \mathbb{R}^2_{\geq 0}.$$

We show that $T(X,\bar{x}) = X$.

First, we show that $X \subset T(X,\bar{x})$. Let $d \in X$ be arbitrary. There exists $t \geq 0$ such that $d = (t,0)^\top$ or $d = (0,t)^\top$. Defining $x^k = (1/k)d$ and $\eta_k = k$, we have $\eta_k x^k = d$ for all $k \in \mathbb{N}$, $\lim_{k\to\infty} x^k = 0 = \bar{x}$, and $\lim_{k\to\infty} \eta_k(x^k - \bar{x}) = d$. Thus, $d \in T(X,\bar{x})$ and $X \subset T(X,\bar{x})$.

Next, we show that $T(X,\bar{x}) \subset X$ using two different approaches.

First approach: Let $d \in T(X,\bar{x})$. By definition, there exist $\eta_k > 0$ and $x^k \in X$ such that $\eta_k x^k = \eta_k(x^k - \bar{x}) \to d$ as $k \to \infty$. Since $\eta_k > 0$ and $x^k \in X$, we have $\eta_k x^k \in X$ for all $k \in \mathbb{N}$. Since X is closed, we have $d \in X$.

Second approach: Let us consider $d \in T_l(g,\bar{x}) \setminus X$. We have $d_1, d_2 > 0$. For $\varepsilon = \min\{d_1, d_2\}$ and every $x \in X$, $\eta > 0$, we obtain

$$\begin{aligned} \|d - \eta(x - \bar{x})\|_2 = \|d - \eta x\|_2 &= \begin{cases} \|(d_1, d_2 - \eta t)^\top\|_2, & \text{if } x = (0,t)^\top \\ \|(d_1 - \eta t, d_2)^\top\|_2, & \text{if } x = (t,0)^\top \end{cases} \\ &\geq \min\{|d_1|, |d_2|\} \geq \varepsilon > 0. \end{aligned}$$

Consequently, $d \notin T(X,\bar{x})$. Combined with $T(X,\bar{x}) \subset T_l(g,\bar{x})$, this shows that $X = T(X,\bar{x}) \subset T_l(g,\bar{x})$ and hence, the ACQ is violated at $\bar{x}$.

The following calculations show that the GCQ is satisfied at $\bar{x}$:

$$T_l(g,\bar{x})^\circ = \{d \in \mathbb{R}^2 : v^\top d \leq 0 \text{ for all } v \in \mathbb{R}^2_{\geq 0}\} = \mathbb{R}^2_{\leq 0},$$

$$\begin{aligned} T(X,\bar{x})^\circ &= \{d \in \mathbb{R}^2 : v^\top d \leq 0 \text{ for all } v \in X\} \\ &= \{d \in \mathbb{R}^2 : (0,t)^\top d \leq 0, \ (t,0)d \leq 0 \text{ for all } t \geq 0\} = \mathbb{R}^2_{\leq 0}. \end{aligned}$$

□

4. Let $n \in \mathbb{N}$, $M \subset \mathbb{R}^n$ be nonempty, and let $x \in M$. Prove the following statements.

   (a) It holds $0 \in T(M,x)$ and the tangent cone $T(M,x)$ is indeed a cone.

(b) If x is an interior point of M, then $T(M,x) = \mathbb{R}^n$.

(c) The tangent cone $T(M,x)$ is closed.

solution.

(a) We show that $0 \in T(M,x)$. We define $x^k = x$ and $\eta_k = 1$ for all $k \in \mathbb{N}$. Then $\eta_k(x^k - x) = 0$ for all $k \in \mathbb{N}$. Hence $0 \in T(M,x)$.

We show that $T(M,x)$ is a cone. Let $d \in T(M,x)$ and let $\alpha > 0$ be arbitrary but fixed. Since $d \in T(M,x)$, there exist $(\eta_k) \subset (0,\infty)$ and $(x^k) \subset M$ such that $x^k \to x$ and $\eta_k(x^k - x) \to d$ as $k \to \infty$. Hence $\eta_k(\alpha x^k - x) \to \alpha d$ as $k \to \infty$ with $\eta_k := \alpha\eta_k > 0$. Putting together the pieces, we conclude that $\alpha d \in T(M,x)$ for all $\alpha > 0$.

(b) We show that $\mathbb{R}^n \subset T(M,x)$. Since $x \in M$ is an interior point of M, there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subset M$. Fix $d \in \mathbb{R}^n \setminus \{0\}$. We define $\eta_k = (2k/\|d\|)\|d\|$ and $x^k = x + (1/\eta_k)d$ for $k \in \mathbb{N}$. We have $\|x^k - x\| = (1/\eta_k)\|d\| \leq \varepsilon/2 < \varepsilon$. Hence $x^k \subset M$. Moreover $x^k \to x$ as $k \to \infty$ and $\eta_k(x^k - x) = (\eta_k/\eta_k)d = d$ for all $k \in \mathbb{N}$. Hence $d \in T(M,x)$.

Combined with $0 \in T(M,x)$ (see part (a)) and $T(M,x) \subset \mathbb{R}^n$, we find that $T(M,x) = \mathbb{R}^n$.

(c) Let $(d^k) \subset T(M,x)$, $d \in \mathbb{R}^n$ fulfill $d^k \to d$ as $k \to \infty$. We demonstrate that $d \in T(M,x)$. By definition of the tangent cone, for each $k \in \mathbb{N}$, there exist $(x^{k,\ell}) \subset M$ and $(\eta_{k,\ell}) \subset (0,\infty)$ such that $x^{k,\ell} \to x$ and $\eta_k(x^{k,\ell} - x) \to d^k$ as $\ell \to \infty$. Hence, for each $k \in \mathbb{N}$, there exists $\ell(k) \in \mathbb{N}$ such that

$$\|\eta_{k,\ell}(x_{k,\ell(k)} - x) - d^k\|_2 \leq 1/k \quad \text{and} \quad \|x^{k,\ell(k)} - x\|_2 \leq 1/k.$$

Define $\eta_k := \eta_{k,\ell(k)}$ and $\overline{x_k} := x_{k,\ell(k)}$. Then $(\eta_k) \subset (0,\infty)$, $(\bar{x}_k) \subset M$, $\bar{x}_k \to x$ as $k \to \infty$ and

$$\|\bar{\eta}_k(\bar{x}_k - x) - d\|_2 \leq \|\bar{\eta}_k(\bar{x}_k - x) - d^k\|_2 + \|d - d^k\|_2 \leq 1/k + \|d - d^k\|_2.$$

Since $d_k \to d$ as $k \to \infty$, we find that $\eta_k(x_k - x) \to d$ as $k \to \infty$. Thus $d \in T(M,x)$.

□

5. Let $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ be differentiable. We define $X = \{x \in \mathbb{R}^n : g(x) \leq 0, h(x) = 0\}$. Let $\bar{x} \in X$, and define

$$X_l(\bar{x}) = \{x \in \mathbb{R}^n : g^l(x) \leq 0, h^l(x) = 0\},$$

where $g^l : \mathbb{R}^n \to \mathbb{R}^m$ and $h^l : \mathbb{R}^n \to \mathbb{R}^p$ are given by

$$g^l(x) = g(\bar{x}) + \nabla g(\bar{x})^\top(x - \bar{x}) \quad \text{and} \quad h^l(x) = h(\bar{x}) + \nabla h(\bar{x})^\top(x - \bar{x}).$$

(a) Prove that $T_l(g,h,\bar{x}) = T_l(g^l,h^l,\bar{x})$.

(b) Show that $T_l(g^l,h^l,\bar{x}) = T(X_l(\bar{x}),\bar{x})$.

(c) Deduce that $T_l(g,h,\bar{x}) = T(X_l(\bar{x}),\bar{x})$.

solution.

(a) By definition, g^l and h^l are the first-order Taylor polynomials corresponding to g and h about $\bar{x}$, respectively. Hence $\nabla g^l(\bar{x}) = \nabla g(\bar{x})$ and $\nabla h^l(\bar{x}) = \nabla h(\bar{x})$.

Let $\mathcal{A}^l(\bar{x})$ be the active set corresponding to g^l at $\bar{x}$. Since $g^l(\bar{x}) = g(\bar{x})$, we have

$$\mathcal{A}^l(\bar{x}) = \{i \in \{1,\ldots,m\} : g_i^l(\bar{x}) = 0\} = \mathcal{A}(\bar{x}).$$

Putting together the pieces, we find that

$$T_l(g^l,h^l,\bar{x}) = \{d \in \mathbb{R}^n : \nabla g_i^l(\bar{x})^\top d \leq 0,\ i \in \mathcal{A}(\bar{x}),\ \nabla h^l(\bar{x})^\top d = 0\}$$
$$= \{d \in \mathbb{R}^n : \nabla g_i(\bar{x})^\top d \leq 0,\ i \in \mathcal{A}(\bar{x}),\ \nabla h(\bar{x})^\top d = 0\} = T_l(g,h,\bar{x}).$$

(b) We present two different approaches to establish the assertion.

First approach: Since g^l and h^l are affine-linear, Theorem 2.19 from the lecture notes yields the assertion.

Second approach: We deduce $T(X_l(\bar{x}),\bar{x}) \subset T_l(g^l,h^l,x)$ from Lemma 2.6 in the lecture notes. Next, we show the opposite inclusion. Let $d \in T_l(g^l,h^l,x)$. We define $x^k = x + (1/\eta_k)d$ and $\eta_k = k$. We have $x^k \to x$ and $\eta_k(x^k - x) = d \to d$ as $k \to \infty$. Using $d \in T_l(g^l,h^l,x)$ and part (a), we have $d \in T_l(g,h,x)$.

We show that $x^k \in X_l(x)$ for all sufficiently large $k \in \mathbb{N}$. Using Taylor's expansion and part (a), we obtain $\mathcal{A}^l(x) = \mathcal{A}(x)$ and

$$h^l(x^k) = h^l(x) + \nabla h^l(\tilde{x})^\top (x^k - x) = h(\bar{x}) + (1/\eta_k)\nabla h(\tilde{x})^\top d,$$
$$g^l(x^k) = g^l(\tilde{x}) + \nabla g^l(x)^\top (x^k - \tilde{x}) = g(\bar{x}) + (1/\eta_k)\nabla g(\bar{x})^\top d.$$

Since $h(\bar{x}) = 0$ and $d \in T_l(g, h, \bar{x})$, we have $h^l(x^k) = 0$ for all $k \in \mathbb{N}$. If $i \in \mathcal{A}(\bar{x})$, then $g_i(\bar{x}) = 0$, and since $d \in T_l(g^l, h^l, \bar{x})$ and $\nabla g_i(x)^\top d \leq 0$, we ensure that $g_i^l(x^k) \leq 0$ for all $k \in \mathbb{N}$. If $i \notin \mathcal{A}(x)$, then $g_i(x) < 0$. Combined with $\nabla g_i(\bar{x})^\top d \in \mathbb{R}$ and (1), we find that $g_i^l(x^k) \leq 0$ for all sufficiently large $k \in \mathbb{N}$. Since the cardinality of $\mathcal{I}(x)$ is finite, we deduce that $x^k \in X_l(\bar{x})$ for all sufficiently large $k \in \mathbb{N}$. Putting together the pieces, we conclude that $d \in T(X_l(\bar{x}), \bar{x})$.

(c) Parts (a) and (b) imply the assertion.

□

2.3 Exercises 3

1. (KKT Conditions for an Example Problem) We define $\bar{x} := (2, 0)^\top$ and consider

$$\begin{aligned} \min_{x \in \mathbb{R}^2} \ & (4 - x_1)^2 + x_2^4 \\ \text{s.t.} \ & (x_1 - 1)^2 + (x_2 - 1)^2 - 2 \leq 0, \\ & (x_1 - 1)^2 + (x_2 + 1)^2 - 2 \leq 0, \\ & x_1 - 2 \leq 0. \end{aligned} \tag{P2}$$

(a) Show that the MFCQ holds at $\bar{x}$.

solution. recall def of MCCQ, there is no equality constraint, so we only need to find $d \in \mathbb{R}^2$ s.t. $\nabla g_i(\bar{x})^\top d < 0$ for $i = 1, 2, 3$:

$$\nabla g_1(\bar{x}) = \begin{pmatrix} 2(\bar{x}_1 - 1) \\ 2(\bar{x}_2 - 1) \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix}, \quad \nabla g_2(\bar{x}) = \begin{pmatrix} 2(\bar{x}_1 - 1) \\ 2(\bar{x}_2 + 1) \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad \nabla g_3(\bar{x}) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

choosing $d = (-1, 0)^\top$ we have $\nabla g_i(\bar{x})^\top d < 0$, so MFCQ holds at $\bar{x}$ □

(b) Derive the KKT conditions for (P_2).

solution.

$$\nabla f(\bar{x}) = \begin{pmatrix} -2(4 - \bar{x}_1) \\ 4\bar{x}_2^3 \end{pmatrix} = \begin{pmatrix} -4 \\ 0 \end{pmatrix}.$$

recall def of the KKT conditions, we need to check:

i. $\nabla f(\bar{x}) + \nabla g_i(\bar{x})^\top \bar{\lambda}_i = 0$:

$$\begin{pmatrix} -4 \\ 0 \end{pmatrix} + \bar{\lambda}_1 \begin{pmatrix} 2 \\ -2 \end{pmatrix} + \bar{\lambda}_2 \begin{pmatrix} 2 \\ 2 \end{pmatrix} + \bar{\lambda}_3 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{1}$$

ii. $g(\bar{x}) \leq 0$

iii. $g_i(\bar{x})^\top \bar{\lambda}_i = 0$ and $\bar{\lambda}_i \geq 0$

□

(c) Define $\bar{\lambda} := (1, 1, 0)^\top$. Show that $(\bar{x}, \bar{\lambda})$ is a KKT tuple of (P_2) and that the strict complementarity condition is violated.

solution. substitute $\bar{\lambda} = (1, 1, 0)^\top$ into

$$\begin{pmatrix} -4 \\ 0 \end{pmatrix} + \bar{\lambda}_1 \begin{pmatrix} 2 \\ -2 \end{pmatrix} + \bar{\lambda}_2 \begin{pmatrix} 2 \\ 2 \end{pmatrix} + \bar{\lambda}_3 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{1}$$

the first condition of KKT conditions is satisfied. then we check the third condition: note that $g_1(\bar{x}) = 0$, $g_2(\bar{x}) = 0$, and $g_3(\bar{x}) = 0$, so complementarity holds for all λ, but $\bar{\lambda}_3 = 0$, so strict complementarity condition is violated. □

(d) Compute the set of all $\bar{\lambda} \in \mathbb{R}^3$ such that $(\bar{x}, \bar{\lambda})$ is a KKT tuple of (P_2).

solution. we require:

$$-4 + 2\overline{\lambda}_1 + 2\overline{\lambda}_2 + \overline{\lambda}_3 = 0 \quad -2\overline{\lambda}_1 + 2\overline{\lambda}_2 = 0, \quad \overline{\lambda}_i \geq 0$$

so we have $\overline{\lambda} = (a, a, 4 - 4a)^\top$ for $a \in [0, 1]$ □

2. (Property of the Tangent Cone) Let $M \subset \mathbb{R}^n$ be a nonempty, convex set, and let $x \in M$. Show that $T(M, x)$ is convex.
Hints: Use the representation for $T(M, x)$ established in Tutorial Exercise T1.2 and the fact that the closure of a convex set is convex.

solution. recall the alternative definition of tangent cone:

$$T(M, x) = \overline{\{d \in \mathbb{R}^n | \exists \eta > 0, y \in M : d = \eta(y - x)\}}$$

since the closure of a convex set is convex, it suffices to show that the set

$$T = \{d \in \mathbb{R}^n | \exists \eta > 0, y \in M, d = \eta(y - x)\}$$

is convex: let $d_1, d_2 \in T$ and $\lambda \in (0, 1)$. by definition there exist $\eta_1, \eta_2 > 0$ and $y_1, y_2 \in M$ s.t. $d_1 = \eta_1(y_1 - x)$ and $d_2 = \eta_2(y_2 - x)$. we define $\overline{\lambda} \triangleq \lambda\eta_1 + (1 - \lambda)\eta_2$. since $\eta_1, \eta_2 > 0$, we obtain that $\overline{\lambda} > 0$, we have

$$\lambda d_1 + (1 - \lambda)d_2 = \lambda\eta_1 y_1 + (1 - \lambda)\eta_2 y_2 - (\lambda\eta_1 + (1 - \lambda)\eta_2)x = \overline{\lambda}\left(\frac{\lambda\eta_1}{\overline{\lambda}} y_1 \frac{(1 - \lambda)\eta_2}{\overline{\lambda}} y_2 - x\right)$$

since $\overline{y} \triangleq \frac{\lambda\eta_1}{\overline{\lambda}} + \frac{(1-\lambda)\eta_2}{\overline{\lambda}} y_2$ is a convex combination of $y_1, y_2 \in M$ and M is convex, we find that $\overline{y} \in M$. we conclude that $\lambda d_1 + (1 - \lambda)d_2 \in T$ □

3. (KKT Conditions and MFCQ)
   (a) Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ be a KKT triple of (P_1). Show that $\bar{x}$ satisfies the first-order necessary optimality conditions stated in Theorem 2.12, that is, conditions a) and b) in Theorem 2.12.

   solution. recall in theorem 2.12, we require $\overline{x}$ is feasible and $\nabla f(\overline{x})^\top d \geq 0, \forall d \in T_l(g, h, \overline{x})$. the first condition is true from KKT second condition.
   we show that $\nabla f(\overline{x})^\top d \geq 0$: for all $d \in T_l(g, h, \overline{x})$. by KKT first condition,

$$\nabla f(\overline{x})^\top d = (-\lambda^\top \nabla g(\overline{x}) - \mu^\top \nabla h(\overline{x}))^\top d$$

   if $i \in \mathcal{I}(\overline{x})$, then $g_i(\overline{x})\overline{\lambda}_i = 0$ yields $\overline{\lambda}_i = 0$. if $i \in \mathcal{A}(\overline{x})$, then $d \in T_l(g, h, \overline{x})$ ensures that $\nabla g_i(\overline{x})^\top d \leq 0$. moreover $\nabla h(\overline{x})^\top d = 0$ and $\overline{\lambda} \geq 0$. putting together the pieces, we find that $\nabla f(\overline{x})^\top d \geq 0$. hence the condition is verified. □

   (b) Suppose that the generalized MFCQ holds at a feasible point of (P_1). Deduce from the proof of Theorem 2.21 (MFCQ is a CQ) that there exists $y \in \mathbb{R}^n$ with $g(y) < 0$ and $h(y) = 0$.
   Hint: use the sequence x^k being constructed in the respective proof.

solution. generalized MFCQ is to replace '$\nabla h(x)$ is full-column rank' by '$\nabla h(x)$ is full-column rank or $h(x)$ is affine'. let x be feasible of (P_1), and suppose that the MFCQ holds at x. we consider $x^k = x + w^k + \phi(w^k)$, ϕ is chosen such that $h(x^k) = 0$ for all $k \in \mathbb{N}$ and $w^k \in B_\rho(0) \cap Ker(h'(x))$. furthermore, the proof shows that there exists $K > 0$ with $g_i(x^k) < 0$ for all $i \in \mathcal{A}(x)$, and $i \in \mathcal{I}(x)$. thus, we can choose $y = x^k$ for some sufficiently large $k \in \mathbb{N}$.

in the proof of theorem 2.21, we construct continuously differentiable function $\phi : B_\rho(0) \to \mathbb{R}^n, \rho > 0$ with:

$$h(x^k) = h(x + w + \phi(w)) = 0 \quad \text{for all } w \in B_\rho(0) \cap \text{Ker}(h'(x)),$$

$$\phi(0) = 0,$$

$$\phi'(0)z = 0 \quad \text{for all } z \in \text{Ker}(h'(x)).$$

In particular, we have for all $w \in B_\rho(0) \cap \text{Ker}(h'(x))$:

$$\phi(w) = \phi(0) + \phi'(0)w + o(\|w\|) = o(\|w\|) \quad (\|w\| \to 0).$$

let $w \to 0$, we have $h(x) \leftarrow h(x^k) = 0$ the proof of existence of ϕ is omitted here.

the next step is to prove that $g_i(x) \leq 0$. the idea is to consider $s^k = w^k + \phi(w^k)$: Furthermore, let d be the direction of the MFCQ. W.l.o.g. we can assume $\|d\| \leq \rho/2$, too.

- For all $i \in \mathcal{A}(x)$ we have
$$\lim_{\|v\|\to 0} \frac{|g_i(x+v) - g_i(x) - \nabla g_i(x)^T v|}{\|v\|} = 0.$$
Additionally, it holds
$$\lim_{w \in \mathrm{Ker}(h'(x)), \|w\|\to 0} \frac{\|\phi(w)\|}{\|w\|} = 0.$$
Thus, there exist $l > 0$ and a sequence $(\alpha_k)_{k\geq l} \subset (0,1)$ with $\alpha_k \to 0$ and
$$\alpha_k^2 \geq r_i(v) := \frac{|g_i(x+v) - g_i(x) - \nabla g_i(x)^T v|}{\|v\|} \leq 1/k, \quad \forall k \geq l, \forall i \in \mathcal{A}(x),$$
and simultaneously
$$\alpha_k^2 \geq \frac{\|\phi(w)\|}{\|w\|} \quad \forall w \in \mathrm{Ker}(h'(x)),\ \|w\| \leq 1/k,\ \forall k \geq l.$$
For $k \geq l$ consider $w^k = (s + \alpha_k d)/k$. We conclude
$$\|w^k\| \leq \frac{\|s\|}{k} + \frac{\alpha_k \|d\|}{k} \leq \frac{\rho}{2k} + \frac{\rho}{2k} = \frac{\rho}{k} \leq \frac{1}{k}.$$
Furthermore, we have $w^k \in \mathrm{Ker}(h'(x))$ since $s \in \mathrm{Ker}(h'(x))$ and $d \in \mathrm{Ker}(h'(x))$.
Thus for all $i \in \mathcal{A}(x)$ and all $k \geq l$ we arrive at
$$g_i(x + s^k) \leq g_i(x) + \nabla g_i(x)^T s^k + r_i(s^k)\|s^k\|$$
by Taylor expansion and recall $i \in \mathcal{A}(x) \to g_i(x) = 0$, $s^k = w^k + \phi(s^k)$, $w^k = \frac{s+\alpha_k d}{k}$:
$$= \frac{1}{k}\nabla g_i(x)^T s + \frac{\alpha_k}{k}\nabla g_i(x)^T d + \nabla g_i(x)^T \phi(w^k) + r_i(s^k)\|s^k\|$$
$$\leq \frac{\alpha_k}{k}\nabla g_i(x)^T d + \alpha_k^2 \|\nabla g_i(x)\| \|w^k\| + \alpha_k^2 \|s^k\|^2$$
$$\leq \frac{\alpha_k}{k}\nabla g_i(x)^T d + (1 + \|\nabla g_i(x)\|)\frac{\alpha_k^2}{k}.$$
Because of $\nabla g_i(x)^T d < 0$, and $\frac{\alpha_k^2}{k} \to 0$ for k large enough, it follows $g_i(x + s^k) \leq 0$ for all $k \geq l'$ and $l' \geq l$ large enough.
- For all $i \in \mathcal{I}(x)$, there exists $l'' \geq l'$ such that
$$g_i(x + s^k) < 0 \quad \forall k \geq l'', \forall i \in \mathcal{I}(x),$$
since $g_i(x + s^k) \to g_i(x) < 0$ for $k \to \infty$.

□

4. (Slater's Condition and MFCQ) Let $g : \mathbb{R}^n \to \mathbb{R}^m$ be convex and continuously differentiable, and let $h : \mathbb{R}^n \to \mathbb{R}^p$ be affine-linear. Define
$$X = \{x \in \mathbb{R}^n : g(x) \leq 0,\ h(x) = 0\}.$$
Slater's condition is satisfied if there exists $y \in \mathbb{R}^n$ such that $g(y) < 0$ and $h(y) = 0$.

(a) Show that Slater's condition is a CQ for every $x \in X$.
Hint: Show that Slater's condition implies the generalized MFCQ for each $x \in X$.

solution. Fix $x \in X$. We show that the MFCQ is satisfied at x. Let us define $d = y - x$, where $y \in \mathbb{R}^n$ satisfies $g(y) < 0$ and $h(y) = 0$. Using the convexity of g, we obtain for all $i \in \mathcal{A}(x)$,
$$\nabla g_i(x)^T d = \nabla g_i(x)^T (y - x) \leq g_i(y) - g_i(x) = g_i(y) < 0.$$

Moreover, we have $\nabla h(x)^T d = h(y) - h(x) = 0$ since h is affine-linear. We conclude that the MFCQ is satisfied at x.

$\square$

(b) Let the generalized MFCQ be satisfied at $x \in X$. Show that Slater's condition holds.

solution. Exercise T.3.3 b) implies the assertion. $\square$

5. (MFCQ and the Set of Lagrange Multipliers) Suppose that $\bar{x}$ is a KKT point of (P_1) and define the set of Lagrange multipliers by

$$\mathcal{M}(\bar{x}) = \{(\lambda, \mu)^\top \in \mathbb{R}^m \times \mathbb{R}^p : (\bar{x}, \lambda, \mu) \text{ is a KKT triple for } (P_1)\}.$$

(a) Show that $\mathcal{M}(\bar{x})$ is closed.

solution. Let $(\lambda^k, \mu^k)^T \subset \mathcal{M}(\bar{x})$ fulfill $(\lambda^k, \mu^k)^T \to (\lambda, \mu)^T$ as $k \to \infty$. Since $\lambda^k \geq 0, g(\bar{x})^T \lambda^k = 0$ for all $k \in \mathbb{N}$, we have $\lambda \geq 0, g(\bar{x})^T \lambda = 0$. Moreover, $\nabla f(\bar{x}) + \nabla g(\bar{x}) \lambda^k + \nabla h(\bar{x}) \mu^k = 0$ for all $k \in \mathbb{N}$ yields $\nabla f(\bar{x}) + \nabla g(\bar{x}) \lambda + \nabla h(\bar{x}) \mu = 0$.
Putting together the pieces, we find that $(\lambda, \mu)^T \in \mathcal{M}(\bar{x})$. $\square$

(b) Let $(\lambda^k, \mu^k)^\top \subset \mathcal{M}(\bar{x})$ fulfill $0 < \|(\lambda^k, \mu^k)^\top\|_2 \to \infty$ as $k \to \infty$. Deduce the existence of a tuple $(\hat{\lambda}, \hat{\mu})$ with $\|(\hat{\lambda}, \hat{\mu})^\top\|_2 = 1$ and

$$\hat{\lambda} \geq 0, \quad g(\bar{x})^\top \hat{\lambda} = 0, \quad \nabla g(\bar{x})^\top \hat{\lambda} + \nabla h(\bar{x}) \hat{\mu} = 0.$$

solution. We define $(\hat{\lambda}^k, \hat{\mu}^k) = (\lambda^k, \mu^k)/\|(\lambda^k, \mu^k)\|_2$. This sequence is bounded and, hence, it, as a convergent subsequence $(\hat{\lambda}^k, \hat{\mu}^k)_K$ converging to some $(\hat{\lambda}, \hat{\mu})$ with $\|(\hat{\lambda}, \hat{\mu})\|_2 = 1$.
Since $\lambda^k \geq 0$ and $g(\bar{x})\lambda^k = 0$ for all $k \in \mathbb{N}$, we have $\hat{\lambda} \geq 0$ and $g(\bar{x})\hat{\lambda} = 0$.
Using $\nabla f(\bar{x}) + \nabla g(\bar{x}) \lambda^k + \nabla h(\bar{x}) \mu^k = 0$ for all $k \in \mathbb{N}$, we obtain

$$0 = \lim_{K \to \infty} \frac{\nabla f(\bar{x})}{\|(\lambda^k, \mu^k)\|_2} = \lim_{K \to \infty} \frac{\nabla g(\bar{x})\lambda^k + \nabla h(\bar{x})\mu^k}{\|(\lambda^k, \mu^k)\|_2} = \nabla g(\bar{x})\hat{\lambda} + \nabla h(\bar{x})\hat{\mu}.$$

$\square$

(c) Suppose that the MFCQ holds at $\bar{x}$. Prove that $\mathcal{M}(\bar{x})$ is bounded.

solution. Suppose $\mathcal{M}(\bar{x})$ is unbounded. In this case, part (b) implies the existence of $(\hat{\lambda}, \hat{\mu})$ with $\|(\hat{\lambda}, \hat{\mu})\|_2 = 1$ and

$$\hat{\lambda} \geq 0, \quad g(\bar{x})\hat{\lambda} = 0, \quad \nabla g(\bar{x})\hat{\lambda} + \nabla h(\bar{x})\hat{\mu} = 0. \tag{2}$$

The MFCQ implies the PLICQ. Since the PLICQ holds at $\bar{x}$ and $\nabla h(\bar{x})$ has full column rank, there exist no vectors $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^p$ (see Definition 2.22 b) in the lecture notes) with

$$\nabla g(\bar{x})u + \nabla h(\bar{x})v = 0, \quad u_{\mathcal{A}(\bar{x})} \geq 0, \quad u_{\mathcal{A}(\bar{x})} \neq 0, \quad u_{\mathcal{I}(\bar{x})} = 0.$$

Combined with (2), we have $\hat{\lambda}_{\mathcal{A}(\bar{x})} = 0$. Hence $\hat{\lambda} = 0$. Since $\nabla h(\bar{x})$ has full column rank, the last equation in (2) implies $\hat{\mu} = 0$. These derivations contradict $\|(\hat{\lambda}, \hat{\mu})\|_2 = 1$. Hence $\mathcal{M}(\bar{x})$ is bounded. $\square$

2.4 Exercises 4

Throughout the exercise sheet, let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ be continuously differentiable. We consider

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t. } g(x) \leq 0, \quad h(x) = 0. \quad (P_1)$$

1. Uniqueness of KKT points and of Lagrange multipliers

(a) Let f be strictly convex, g be convex and h be affine-linear. Suppose that $\bar{x}$ and $\hat{x}$ are KKT points of (P_1). Show that $\bar{x} = \hat{x}$.

Proof. this is a convex optimization problem: by theorem 2.27 in the lecture note, we know that KKT points of (P_1) are global solutions. consider $y = \frac{1}{2}\hat{x} + \frac{1}{2}\bar{x}$ we aim to show that if $\bar{x} \neq \hat{x}$, then

$$f(y) < f(\bar{x})$$

which contradicts to the fact that $f(\bar{x})$ is global minimal. $\square$

(b) Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ and $(\bar{x}, \hat{\lambda}, \hat{\mu})$ be KKT triples of (P_1). Suppose that the LICQ holds at $\bar{x}$. Show that $(\bar{\lambda}, \bar{\mu}) = (\hat{\lambda}, \hat{\mu})$.

Proof. LICQ holds at $\overline{x}$ indicates that $(g_{\mathcal{A}(\overline{x})}(\overline{x}), h(\overline{x}))$ is linearly independent. $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a KKT triple indicates that

$$\nabla g(\overline{x})\overline{\lambda} + \nabla h(\overline{x})\overline{\mu} = 0$$

$(\nabla f(\overline{x}) = 0)$ and we also have

$$\nabla g(\overline{x})\hat{\lambda} + \nabla h(\overline{x})\hat{\mu} = 0$$

suppose $(\bar{\lambda}, \bar{\mu}) \neq (\hat{\lambda}, \hat{\mu})$ we have

$$\nabla g(\overline{x})(\overline{\lambda} - \hat{\lambda}) + \nabla h(\overline{x})(\overline{\mu} - \hat{\mu}) = 0$$

for inactive inequality constraints $\mathcal{I}(\overline{x})$, the related $\lambda = 0$. we only consider the active inequality constraints:

$$\nabla g_{\mathcal{A}(\overline{x})}(\overline{x})(\overline{u} - \hat{u}) + \nabla h(\overline{x})(\overline{v} - \hat{v}) = 0$$

the system $(g_{\mathcal{A}(\overline{x})}(\overline{x}), h(\overline{x}))$ is clearly not linear independent. □

2. (Second-Order Sufficient Condition and Quadratic Growth)
Let f, g and h be twice continuously differentiable and let $\bar{x}$ be a KKT point of (P_1) with multipliers $\bar{\lambda} \in \mathbb{R}^m$ and $\bar{\mu} \in \mathbb{R}^p$. We denote by X the feasible set of (P_1). Suppose that the second-order sufficient conditions stated in Theorem 2.33 hold true:
Let $\bar{x} \in \mathbb{R}^n$ satisfy the KKT conditions 2.15 a)–c) with multipliers $\bar{\lambda} \in \mathbb{R}^m$ and $\bar{\mu} \in \mathbb{R}^p$. Suppose that

$$d^\top \nabla_{xx} L(\bar{x}, \bar{\lambda}, \bar{\mu}) d > 0 \quad \forall d \in T_+(g, h, \bar{x}, \bar{\lambda}) \setminus \{0\}$$

holds. Then $\bar{x}$ is a strict local solution of (1).

Show that there exist $\alpha > 0$ and $\delta > 0$ such that

$$f(x) - f(\bar{x}) \geq \alpha \|x - \bar{x}\|^2 \quad \text{for all } x \in X \text{ with } \|x - \bar{x}\| < \delta.$$

Hint: Adapt the proof of Theorem 2.33.

Proof. prove by contradiction, we assume that $\exists \alpha_k > 0$ and $\exists \{x_k\} \subset X \setminus \{\overline{x}\}$ s.t. $f(x_k) - f(\overline{x}) < \alpha_k \|x_k - \overline{x}\|^2$. let $d_k = x_k - x$ and $y_k = \frac{d_k}{\|d_k\|}$. y_k is bounded and we assume that $y_k \to y \in \mathbb{R}^n$ (taking a subsequence if necessary). then we have $\|y\| = 1$ and

$$\frac{f(x_k) - f(\overline{x})}{\|d_k\|} = \frac{\nabla f(\overline{x})^T (x_k - x) + o(d_k)}{\|d_k\|} = \nabla f(\overline{x})^T y_k + \frac{o(d_k)}{\|d_k\|} \to \nabla f(\overline{x})^T y$$

similarly we have

$$\frac{g(x_k) - g(\overline{x})}{\|d_k\|} \to \nabla g(\overline{x})^T y, \quad \frac{h(x_k) - h(\overline{x})}{\|d_k\|} \to \nabla h(\overline{x})^T y$$

we have

$$\nabla g_i(\overline{x})^T y \leq 0, \quad \nabla h(\overline{x})^T y = 0, \quad i \in \mathcal{A}(\overline{x})$$

we show that $y \in T_+(g, h, \overline{x}, \overline{\lambda}) \setminus \{0\}$: From the KKT conditions we obtain

$$0 = \nabla_x L(\bar{x}, \bar{\lambda}, \bar{\mu})^\top y = \underbrace{\nabla f(\bar{x})^\top y}_{\leq 0} + \underbrace{\sum_{i \in \mathcal{A}(\bar{x})} \bar{\lambda}_i \nabla g_i(\bar{x})^\top y}_{\leq 0} + \underbrace{\sum_{i=1}^{p} \bar{\mu}_i \nabla h_i(\bar{x})^\top y}_{=0}.$$

Therefore, for all $i \in \mathcal{A}(\bar{x})$ with $\bar{\lambda}_i > 0$ we have $\nabla g_i(\bar{x})^\top y = 0$ (otherwise the right-hand side would be negative). This shows $y \in T_+(g, h, \bar{x}, \bar{\lambda})$.

from the assumption at beginning, we have

$$\alpha_k \|d_k\| > \frac{f(x) - f(\overline{x})}{\|d_k\|} = \nabla f(\overline{x})^T y_k + \frac{o(d_k)}{\|d_k\|} \to \nabla f(\overline{x})^T y, \quad k \to \infty$$

combining with $\alpha_k > 0$ and $d_k \to 0$ we have

$$\nabla f(\overline{x})^T y \leq 0$$

consider Lagrangian at x_k: recall $\overline{\lambda} \geq 0$, for $x_k \in X$ we have $g \leq 0$ and $h = 0$ and $f(x_k) - f(\overline{x}) < \alpha \|x_k - \overline{x}\|^2$:

$$L(x_k, \bar{\lambda}, \bar{\mu}) = f(x_k) + \bar{\lambda}^\top g(x_k) + \bar{\mu}^\top h(x_k) \leq f(x_k) < f(\bar{x}) + \alpha_k t_k^2 = L(\bar{x}, \bar{\lambda}, \bar{\mu}) + \alpha_k \|d_k\|^2.$$

using the fact that $\nabla_x L(\overline{x}, \overline{\lambda}, \overline{\mu}) = 0$, we have

$$\alpha_k > \frac{L(x_k, \overline{\lambda}, \overline{\mu}) - L(\overline{x}, \overline{\lambda}, \overline{\mu})}{\|d_k\|^2} = \frac{d_k^T \nabla^2 L(x_k, \overline{\lambda}, \overline{\mu}) d_k}{\|d_k\|^2} + \frac{o(\|d_k\|^2)}{\|d_k\|^2}$$

taking $k \to \infty$ we have $\alpha > y^T \nabla^2 L(\overline{x}, \overline{\lambda}, \overline{\mu}) y$, since $\alpha > 0$ we have $y^T \nabla^2 L(\overline{x}, \overline{\lambda}, \overline{\mu}) y \leq 0$, this contradicts to the second-order sufficient conditions in Theorem 2.33, which is assumed to be true in the problem. □

3. Second-Order Sufficient Optimality Conditions We define $\bar{x} := (1, \frac{1}{2}, \frac{1}{2})^\top$, $\bar{\lambda} := (1, 0)^\top$, and $\bar{\mu} := 2$. We consider

$$\begin{aligned} \min_{x \in \mathbb{R}^3} \ & f(x) = x_1 + x_2 - 2x_3 \\ \text{s.t.} \ & g_1(x) := \frac{1}{2}x_1^2 - x_2 \leq 0, \\ & g_2(x) := e^{x_1 - 1} - x_1 \leq 0, \\ & h(x) := x_3^2 - x_1 + \frac{3}{4} = 0. \quad (P_2) \end{aligned}$$

   (a) Show that $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a KKT triple of (P_2).
   (b) Prove that $T_+(g, h, \bar{x}, \bar{\lambda}) = \{(t, t, t) : t \in \mathbb{R}\}$.
   (c) Show that $\bar{x}$ is a local solution to (P_2).

reminder: how to compute $\nabla_x f$ for $f \in \mathbb{R}^{3\times1}$ and $x \in \mathbb{R}^{3\times1}$

4. LICQ and Slack Variables Let $p = 0$ and let $\bar{x}$ be a feasible for (P_1). Suppose that the LICQ holds at $\bar{x}$. We consider

$$\min_{(x,s)\in\mathbb{R}^n\times\mathbb{R}^m} f(x) \quad \text{s.t. } g(x) + s = 0, \quad -s \leq 0. \quad (P_3)$$

We define $\bar{s} := -g(\bar{x})$.
Show that $(\bar{x}, \bar{s})$ is feasible for (P_3) and that the LICQ holds at $(\bar{x}, \bar{s})$ for (P_3).

Proof. We have $g(\bar{x}) \leq 0$ and hence $\bar{s} \geq 0$. Moreover $g(\bar{x}) + \bar{s} = g(\bar{x}) - g(\bar{x}) = 0$. Thus, $(\bar{x}, \bar{s})$ is feasible for (P_3).
Since $\bar{s} = -g(\bar{x})$, the active set corresponding to (P_3) at $(\bar{x}, \bar{s})$ equals that of (P_1) at $\bar{x}$.
Let us define $\hat{h} : \mathbb{R}^{n+m} \to \mathbb{R}^m$ by $\hat{h}(x, s) := g(x) + s$ and $\hat{g} : \mathbb{R}^{n+m} \to \mathbb{R}^m$ by $\hat{g}(x, s) := -s$. We have

$$\nabla \hat{h}(\bar{x}, \bar{s}) = \begin{pmatrix} \nabla g(\bar{x}) \\ I \end{pmatrix} \quad \text{and} \quad \nabla \hat{g}(x, s) = \begin{pmatrix} 0 \\ -I \end{pmatrix},$$

where $I \in \mathbb{R}^{m\times m}$ is the $m \times m$-identity matrix and $0 \in \mathbb{R}^{n\times m}$ is the $n \times m$-zero matrix.
We show that $(\nabla \hat{g}_A(\bar{x}, \bar{s}), \nabla \hat{h}(\bar{x}, \bar{s}))$ has full column rank.
First approach: Let $(v, w)^\top \in \mathbb{R}^n \times \mathbb{R}^m$ and let

$$\begin{pmatrix} \nabla \hat{g}_A(\bar{x}, \bar{s}) & \nabla \hat{h}(\bar{x}, \bar{s}) \end{pmatrix} \begin{pmatrix} v_A(\bar{x}) \\ w \end{pmatrix} = 0.$$

This system of equations can also be expressed as

$$\sum_{i=1}^{m} w_i \begin{pmatrix} \nabla g_i(\bar{x}) \\ e_i \end{pmatrix} - \sum_{j\in A(\bar{x})} v_j \begin{pmatrix} 0 \\ e_j \end{pmatrix} = 0, \tag{2}$$

where e_i is the i-th canonical unit vector in $\mathbb{R}^m$ and 0 is the zero vector in $\mathbb{R}^n$. We show that $(v_A(\bar{x}), w)^\top = 0$. By definition, the sets $I(\bar{x})$ and $A(\bar{x})$ are disjoint. If $i \in I(\bar{x})$, then (2) ensures that $w_i = 0$. Combined with (2), we find

$$\sum_{i\in A(\bar{x})} w_i \nabla g_i(\bar{x}) = 0.$$

Since $\nabla g_A(\bar{x})$ has full column rank, we have $w_A(\bar{x}) = 0$. Putting together the pieces, we find that $w = 0$. Using $w = 0$, we deduce $v_A = 0$ from (2). Hence $(\nabla \hat{g}_A(\bar{x}, \bar{s}), \nabla \hat{h}(\bar{x}, \bar{s}))$ has full column rank.

□

5. KKT Conditions for the Celis–Dennis–Tapia Problem We consider

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & x^\top Hx + 2b^\top x \\ \text{s.t.} \quad & \|x\|_2^2 - \Delta^2 \le 0, \\ & \|A^\top x + c\|_2^2 - \xi^2 \le 0. \quad (P_4) \end{aligned}$$

where $\Delta > 0$, $\xi \ge 0$, $H \in \mathbb{R}^{n\times n}$ is symmetric, $A \in \mathbb{R}^{n\times m}$, $b \in \mathbb{R}^n$, and $c \in \mathbb{R}^m$.
Suppose that $\xi > \min_{\|u\|_2\le\Delta} \|A^\top u + c\|_2$.

(a) Verify that a CQ holds at each feasible point of (P_4).

(b) Derive the KKT conditions for (P_4).

Proof. (a) We verify Slater's condition. We observe that the inequality constraints in (P_4) are convex and continuously differentiable. Let y^* be an optimal solution of $\min_{\|x\|_2\le\Delta} \|A^\top x + c\|_2$. (Since the feasible set of this optimization problem is nonempty and compact, and the objective function is continuous, it has an optimal solution.)

If $\|y^*\|_2 < \Delta$, then $\xi > \|A^\top y^* + c\|_2$ ensures Slater's condition.

If $\|y^*\|_2 = \Delta$, we consider $x^k = (1 - 1/(2k))y^*$. For each $k \in \mathbb{N}$, we have $\|x^k\|_2 < \Delta$. Moreover, the continuity of $x \mapsto \|A^\top x + c\|_2$, and $\|A^\top y^* + c\|_2 < \xi$ ensure $\|A^\top x^k + c\|_2 < \xi$ for all sufficiently large $k \in \mathbb{N}$. Hence, Slater's condition holds.

Hence, a CQ is fulfilled for each feasible point of (P_4).

(b) Let $\bar{x} \in \mathbb{R}^n$. The KKT conditions for $\bar{x}$ are given by: there exists $\lambda \in \mathbb{R}^2$ such that

$$2H\bar{x} + 2b + 2\lambda_1\bar{x} + 2\lambda_2 A(A^\top\bar{x} + c) = 0,$$

$$\|\bar{x}\|_2^2 \le \Delta^2, \quad \|A^\top\bar{x} + c\|_2^2 \le \xi^2, \quad \lambda \ge 0,$$

$$\lambda_1(\|\bar{x}\|_2^2 - \Delta^2) = 0, \quad \lambda_2(\|A^\top\bar{x} + c\|_2^2 - \xi^2) = 0.$$

□

2.5 E-test 1

We consider the general constrained optimization problem:

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \quad g(x) \le 0, \quad h(x) = 0 \tag{P}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^p$ are twice continuously differentiable functions. The feasible set is defined as:

$$X := \{x \in \mathbb{R}^n : g(x) \le 0, h(x) = 0\}.$$

1. If f, g, h are convex, then the problem (P) is a convex optimization problem.

solution. This is not true since a convex optimization problem requires both a convex objective function and a convex feasible set. However, if h is convex, the feasible set is not necessarily convex. For example, consider $n = 2$, $h(x) = x_1^2 + x_2^2 - 1$, and $g(x) = 0$. Thus, for (P) to be a convex optimization problem:

- f and g_i must be convex,
- h must be linear.

□

2. Quadratic programs are convex optimization problems.

solution. This is only true if the constant matrix $\nabla^2 f(x)$ is **positive semidefinite**. For example, if $f(x) = -x^2$ and $g \equiv h \equiv 0$, then the problem is quadratic but **nonconvex**.

□

3. Let $b, c \in \mathbb{R}^n$, and $C, B \in \mathbb{R}^{n \times n}$, and:

$$f(x) := c^\top x + \frac{1}{2} x^\top C x$$
$$g(x) := b^\top x + \frac{1}{2} x^\top B x$$

and $h \equiv 0$.

Is this a quadratic optimization problem?

solution. By definition of quadratic programs, g and h must be **linear**. Here, $g(x)$ includes a quadratic term, so the problem is not a valid quadratic optimization problem.

□

4. Suppose that $T(X, \bar{x}) = \mathbb{R}^n$. Then, the first-order necessary condition $\nabla f(\bar{x})^\top d \geq 0$ for all $d \in T(X, \bar{x})$ is equivalent to $\nabla f(\bar{x}) = 0$.

solution. Given $T(X, \bar{x}) = \mathbb{R}^n$, the tangent cone is the entire space. Therefore:

$$\nabla f(\bar{x})^\top d \geq 0 \quad \forall d \in \mathbb{R}^n$$

implies:

$$\nabla f(\bar{x})^\top (-\nabla f(\bar{x})) \geq 0 \implies \|\nabla f(\bar{x})\|^2 \leq 0 \implies \nabla f(\bar{x}) = 0.$$

□

5. Let $\bar{x} \in X$ be a local minimizer of f over the feasible set X, and assume that there exists $\epsilon > 0$ such that $B_\epsilon(\bar{x}) \subset X$. Then $\nabla f(\bar{x}) = 0$.

solution. Since $\bar{x}$ lies in the interior of X, we have $T(X, \bar{x}) = \mathbb{R}^n$. By the first-order optimality condition, we obtain:

$$\nabla f(\bar{x}) = 0.$$

□

6. Let $x \in X$. Then $T(X, x) = \mathbb{R}^n$ implies that $x \notin \partial X$, where ∂X is the boundary of X.

solution. A counterexample is provided where $g(x_1, x_2) = x_2^2(x_1^2 - x_2)(x_1^2 + x_2)$ and no equality constraints. The feasible set is:

$$X = \{(x_1, x_2)^\top \in \mathbb{R}^2 : g(x_1, x_2) \leq 0\}.$$

At $x = 0$, the tangent cone $T(X, 0) = \mathbb{R}^2$, but 0 lies on the boundary ∂X.

□

7. Suppose that $h \equiv 0$. Then, the linearized tangent cone $\mathcal{T}_l(g, h, x)$ is the set of all directions $d \in \mathbb{R}^n$ such that the gradients of the active constraints and d encompass an angle of at most 90 degrees.

solution. By definition, the linearized tangent cone is:

$$\mathcal{T}_l(g, h, x) = \{d \in \mathbb{R}^n : \nabla g_i(x)^\top d \leq 0, i \in \mathcal{A}(x)\}.$$

This means the angle between the gradients of the active constraints and d must be **at least** 90 degrees, not at most.

□

8. Let $x \in X$. Then, the linearized tangent cone $T_l(g, h, x)$ is a nonempty, closed, and convex cone.

solution. By definition, the linearized tangent cone is:

$$T_l(g, h, x) = \{d \in \mathbb{R}^n : \nabla g_i(x)^\top d \leq 0, i \in \mathcal{A}(x), \nabla h_j(x)^\top d = 0\}.$$

The cone is **nonempty** because $0 \in T_l(g, h, x)$, and it can be shown through standard analysis that the cone is both closed and convex.

□

9. Let $\bar{x}$ be a local solution to (P) and f be continuously differentiable. The cone of descent directions at $\bar{x}$ is defined as:

$$V(\bar{x}) = \{d \in \mathbb{R}^n : \nabla f(\bar{x})^\top d < 0\}.$$

Then it holds that:

$$V(\bar{x}) \cap T(X, \bar{x}) = \emptyset.$$

solution. The first-order necessary optimality conditions ensure that $\bar{x} \in X$ and:

$$\nabla f(\bar{x})^\top d \geq 0 \quad \forall d \in T(X, \bar{x}).$$

Thus, no direction $d \in V(\bar{x})$ can also belong to $T(X, \bar{x})$. Therefore:

$$V(\bar{x}) \cap T(X, \bar{x}) = \emptyset.$$

□

10. Let $x \in X$. Then the tangent cone $T(X, x)$ is a nonempty, closed, and convex cone.

solution. While the tangent cone $T(X, x)$ is always **nonempty** and **closed**, it does not necessarily have to be **convex**.

□

11. Polar cones are nonempty, closed, and convex.

solution. The polar cone of a set K is defined as:

$$K^\circ = \{v \in \mathbb{R}^n : v^\top d \leq 0 \, \forall d \in K\}.$$

By construction, the polar cone K° is **nonempty** (contains 0), **closed**, and **convex**.

□

12. Suppose that $x \in X$ satisfies $\mathcal{A}(x) = \emptyset$. Then, a CQ (Constraint Qualification) holds at x.

solution. Consider $g(x) = -1$ and $h(x) = x^2$ at $x = 0$. Then:

- $\mathcal{A}(x) = \emptyset$,
- $T(X, 0) = \{0\}$,
- $T_l(g, h, 0) = \mathbb{R}$.

Since $T(X, 0)$ is the polar cone of T_l, the generalized constraint qualification (GCQ) cannot be satisfied.

□

13. Let us consider the constraints $x_2 \leq \sin^2(x_1)$ and $x_2 \geq 0$ in $\mathbb{R}^2$. Then, the ACQ (Abadie Constraint Qualification) is satisfied at $\bar{x} = 0$.

solution. The tangent cone $T(X, \bar{x})$ is shown to be equal to the linearized tangent cone $T_l(g, \bar{x})$, thereby satisfying the ACQ condition.

□

14. Let $K \subseteq \mathbb{R}^n$ be a closed cone. Then, the angle between arbitrary elements of K° (polar cone of K) and K must be greater than or equal to 90 degrees.

solution. The polar cone is defined as:

$$K^\circ = \{v \in \mathbb{R}^n : v^\top d \leq 0 \, \forall d \in K\}.$$

Thus, for every $v \in K^\circ$ and $d \in K$, we have $v^\top d \leq 0$, meaning the angle between v and d is greater than or equal to 90 degrees.

□

15. It is possible that the GCQ (Generalized Constraint Qualification) is satisfied at a certain point $x \in X$ while the ACQ fails to hold at x.

solution. It is possible that the GCQ (Generalized Constraint Qualification) is satisfied at a certain point $x \in X$ while the ACQ fails to hold at x.

$\square$

16. Suppose that $\bar{x}$ satisfies the KKT conditions but no CQ (Constraint Qualification) holds at $\bar{x}$. Then, $\bar{x}$ is no local solution.

solution. The KKT conditions can still hold even without a CQ. An example is given where the GCQ is not satisfied, but $\bar{x}$ is still a valid KKT point and a global solution to the problem.

$\square$

17. The KKT conditions at the point $(\bar{x}, \bar{\lambda}, \bar{\mu})$ are equivalent to the system:

- $\nabla_x L(\bar{x}, \bar{\lambda}, \bar{\mu}) = 0,$
- $\nabla_\mu L(\bar{x}, \bar{\lambda}, \bar{\mu}) = 0,$
- $\nabla_\lambda L(\bar{x}, \bar{\lambda}, \bar{\mu}) \leq 0,\ \bar{\lambda} \geq 0,\ \bar{\lambda}^\top g(\bar{x}) = 0.$

solution. We calculate the derivatives of the Lagrangian function $L(x, \lambda, \mu) = f(x) + \lambda^\top g(x) + \mu^\top h(x)$:

$$\begin{aligned}
\nabla_x L(x, \lambda, \mu) &= \nabla f(x) + \nabla g(x)\lambda + \nabla h(x)\mu, \\
\nabla_\lambda L(x, \lambda, \mu) &= g(x), \\
\nabla_\mu L(x, \lambda, \mu) &= h(x).
\end{aligned}$$

These conditions match the given system, proving the statement is true.

$\square$

18. $\bar{x} = (0, 0)$ is a KKT point of:

$$\min_{x \in \mathbb{R}^2} x_1^2 + (x_2 + 1)^2 \quad \text{s.t.} \quad x \geq 0$$

where strict complementarity holds.

solution. Both constraints are active at $\bar{x}$. The KKT conditions are satisfied with:

$$0 - \bar{\lambda}_1 = 0, \quad 2(0+1) - \bar{\lambda}_2 = 0, \quad \bar{\lambda} \geq 0, \quad \bar{x}_1\bar{\lambda}_1 = 0, \quad \bar{x}_2\bar{\lambda}_2 = 0.$$

From these conditions, $\bar{\lambda}_1 = 0$, meaning strict complementarity does not hold.

$\square$

19. The optimization problem:

$$\min_{x \in \mathbb{R}} x \quad \text{s.t.} \quad x^2 = 0$$

has KKT points.

solution. The only feasible point is $\bar{x} = 0$. If $\bar{x}, \bar{\lambda}$ formed a KKT tuple, the KKT conditions would require:

$$\nabla f(\bar{x}) + \bar{\lambda}\nabla h(\bar{x}) = 0.$$

Substituting values leads to a contradiction, thus no KKT points exist.

$\square$

20. Let $\bar{x}$ be a KKT point of (P). Then $\nabla f(\bar{x})^\top d \geq 0$ for all $d \in T(X, \bar{x})$.

solution. According to the conditions from Homework Exercise T3.4, $\nabla f(\bar{x})^\top d \geq 0$ holds for all $d \in T_l(g, h, \bar{x})$. Since $T(X, \bar{x}) \subseteq T_l(g, h, \bar{x})$, this implies the assertion.

$\square$

21. Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ be a KKT point. Then, the multipliers $\bar{\lambda}$ and $\bar{\mu}$ are unique.

solution. The uniqueness of multipliers $\bar{\lambda}$ and $\bar{\mu}$ is not generally guaranteed. Counterexamples can be found, such as in Exercise H3.2.

□

22. Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ be a KKT point that satisfies the strict complementarity condition. Then, it holds $\bar{\lambda}_j > 0$ for all $j \in \{1, \ldots, m\}$.

solution. Strict complementarity ensures $\bar{\lambda}_j > 0$ only for indices $j \in \mathcal{A}(\bar{x})$ (active constraints). For other indices where $g_j(\bar{x}) < 0$, the complementary slackness condition ensures $\bar{\lambda}_j = 0$.

□

2.6 Exericese 5

1. (CQ2 and Second-Order Necessary Conditions) We consider

$$\min_{x \in \mathbb{R}^2} f(x) := \frac{1}{2}(x_1^2 + x_2^2) \quad \text{s.t.} \quad g(x) := x_1^4 - x_2^2 \leq 0. \tag{P$_1$}$$

   (a) Show that $\bar{x} = 0$ is the unique, global solution to (P$_1$).
   (b) Show that $(\bar{x}, \lambda)$ is a KKT tuple of (P$_1$) for each $\lambda \geq 0$.
   (c) Show that the cones $T_a(g, \bar{x})$, $T_+(g, \bar{x}, \lambda)$, and $T_l(g, \bar{x})$ coincide for all $\lambda \geq 0$.
   (d) Prove that the CQ2 is violated at $(\bar{x}, \lambda)$ for all $\lambda \geq 0$.
   Hint: Show that for $d = (0, 1)^\top$ there does not exist a twice continuously differentiable curve that satisfies the properties required by the CQ2.
   (e) Show that $\nabla_{xx} L(\bar{x}, \bar{\lambda})$ is not positive semidefinite on $T_+(g, \bar{x}, \bar{\lambda})$ for $\bar{\lambda} = 1$. Here, L is the Lagrangian function of (P$_1$). Why does this not contradict the second-order necessary optimality conditions?
   Remark: A CQ holds at $\bar{x}$.

2. (Exactness and Differentiability of Penalty Functions) For a function $f : \mathbb{R}^n \to \mathbb{R}$ and a nonempty set $X \subset \mathbb{R}^n$, we consider

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X, \tag{P$_2$}$$

   Let $\alpha > 0$ and let $\pi : \mathbb{R}^n \to \mathbb{R}$ be a penalty term for X. We define the penalty function $P_\alpha^\pi : \mathbb{R}^n \to \mathbb{R}$ by $P_\alpha^\pi(x) = f(x) + \alpha\pi(x)$.
   Suppose that $\bar{x}$ is a local solution to (P$_2$) and that P_α^π is exact at $\bar{x}$ for some $\bar{\alpha} > 0$.

   (a) Show that P_α^π is exact at $\bar{x}$ for all $\alpha \geq \bar{\alpha}$.
   (b) Suppose that f and π are differentiable at $\bar{x}$. Show that $\nabla f(\bar{x}) = 0$.

3. (Augmented Lagrangian Function) Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ be twice continuously differentiable. We consider

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad h(x) = 0. \tag{P$_3$}$$

   Let L be the Lagrangian function corresponding to (P$_3$). For fixed $\mu \in \mathbb{R}^p$, we also consider

$$\min_{x \in \mathbb{R}^n} L(x, \mu) \quad \text{s.t.} \quad h(x) = 0. \tag{P$_4$}$$

   (a) Show that $\bar{x}$ is a local solution to (P$_3$) if and only if $\bar{x}$ is a local solution to (P$_4$).
   (b) Show that the quadratic penalty function $P_\alpha^\mu : \mathbb{R}^n \to \mathbb{R}$ corresponding to (P$_4$) is given by

$$P_\alpha^\mu(x) = f(x) + \mu^\top h(x) + \frac{\alpha}{2}\|h(x)\|_2^2.$$

   (c) Let $(\bar{x}, \bar{\mu})$ be a KKT tuple of (P$_3$). Suppose that the second-order sufficient conditions are satisfied at $(\bar{x}, \bar{\mu})$. Show that there exists $\bar{\alpha} > 0$ such that $P_\alpha^{\bar{\mu}}$ is exact for (P$_3$) for all $\alpha \geq \bar{\alpha}$.

   Hint: You can use without proof Debreu's lemma: Let $A \in \mathbb{R}^{k \times n}$ and let $H \in \mathbb{R}^{n \times n}$ be symmetric such that $d^\top H d > 0$ for all $d \in \mathbb{R}^n \setminus \{0\}$ with $Ad = 0$. Then there exists $\bar{\rho} \geq 0$ such that $H + \bar{\rho} A^\top A$ is positive definite for all $\rho \geq \bar{\rho}$.

4. (A Modified Penalty Method) We consider the optimization problem

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \quad h(x)=0, \tag{P$_5$}$$

where $f:\mathbb{R}^n\to\mathbb{R}$ and $h:\mathbb{R}^n\to\mathbb{R}^p$ are continuously differentiable, and let $P_\alpha:\mathbb{R}^n\to\mathbb{R}$ be the quadratic penalty function associated with (P_5).

Algorithm:

0. Choose $\alpha_0>0$.

(a) For $k=0,1,\dots$:

   i. Compute x^k such that $\nabla P_{\alpha_k}(x^k)=0$.

   ii. STOP if $x^k\in X$.

   iii. Choose $\alpha_{k+1}>\alpha_k$.

Let (x^k) be generated by the above algorithm with $\alpha_k\to\infty$ as $k\to\infty$. Suppose that $\lim_{k\to\infty} x^k=\bar{x}\in\ell^n$ and that the LICQ is satisfied at $\bar{x}$.

Define $(\mu^k)\subset\mathbb{R}^p$ by $\mu^k:=\alpha_k h(x^k)$.

(a) Show that $\mu^k=-(\nabla h(x^k)^\top\nabla h(x^k))^{-1}\nabla h(x^k)^\top\nabla f(x^k)$ for all sufficiently large $k\in\mathbb{N}$.

(b) Prove that $\bar{\mu}:=\lim_{k\to\infty}\mu^k$ exists and compute $\bar{\mu}$.

(c) Show that $\bar{x}$ is feasible for (P_5).

(d) Show that $(\bar{x},\bar{\mu})$ is a KKT pair of (P_5).