Lecture 7. Karush–Kuhn–Tucker conditions

Paul Delatte
delatte@usc.edu
University of Southern California

Last updated: 31 December, 2022

To prove this theorem, we use Farkas' lemma which was proved in Lecture 6 and which itself was proved using the Hahn-Banach extension theorem which was proved in Lecture 5.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. We consider two problems. Define (P) as

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && x \in S \end{aligned} \tag{P}$$

where $S \subset \mathbb{R}^n$ is a nonempty closed set (called feasible region).
Define (Q) as

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && g_i(x) \le 0, \quad i = 1, \dots, r \\ & && h_j(x) = 0, \quad j = 1, \dots, m \end{aligned} \tag{Q}$$

where the functions $g_i\colon \mathbb{R}^n \to \mathbb{R}$ and $h_j\colon \mathbb{R}^n \to \mathbb{R}$ are assumed continuously differentiable. (We take $\mathbb{R}^n$ as the domain of the functions for simplicity, but any open subset of $\mathbb{R}^n$ would work.) Problem (Q) is a version of problem (P) with $S = \{x \in \mathbb{R}^n : g_i(x) \le 0, i = 1, \dots, r,\ h_j(x) = 0, j = 1, \dots, m\}$.

Theorem 1 (Karush–Kuhn–Tucker (Necessary) Conditions). *Let x^* be a feasible point for (Q) and $I(x^*) = \{i \in \{1, \dots, r\} : g_i(x^*) = 0\}$. Suppose that the vectors $\{\nabla h_1(x^*), \dots, \nabla h_m(x^*)\} \cup \{\nabla g_i(x^*) : i \in I(x^*)\}$ are linearly independent. If x^* is a local minimizer for (Q), then there exists (λ, μ) such that*

$$\nabla f(x^*) + \sum_{i=1}^{r} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{m} \mu_j \nabla h_j(x^*) = 0,$$

and

$$\lambda \ge 0, \quad \textit{and} \quad \lambda_i g_i(x^*) = 0 \quad \textit{for all} \quad i = 1, \dots, r.$$

Proof. Under the assumptions of the theorem, the point x^* must satisfy the John Fritz conditions for $(\lambda_0, \lambda, \mu)$. If $\lambda_0 > 0$, then define $\lambda' = \lambda/\lambda_0$ and $\mu' = \mu/\lambda_0$. The KKT conditions then hold with (λ', μ') for multipliers. Suppose now that $\lambda_0 = 0$. Then $\sum_{i \in I(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{m} \mu_j \nabla h_j(x^*) = 0$, and so the corresponding gradients are linearly dependent, a contradiction. □

Remark. There exist several proofs for the KKT theorem; the two most standard proofs rely either: on some geometric arguments (using separation theorems for convex sets); or on a penalization approach (see Güler FO S.9.2.). (Other proofs exist; see, e.g., a proof through Ekeland's ε-variational principle as in Güler FO S.9.3.) Even though the penalization approach is more concise, we follow a geometric route because it provides the true foundation of the theory (especially when it comes to justifying constraint qualifications). Note that some care has to be taken when dealing with problems containing both inequality and equality constraints: in particular, one cannot simply derive the FJ and KKT conditions for inequality constraints and then include equality constraints by considering them as double inequalities (see p.121 in Bazaraa&Shetty FoO); in practice, the equality constraints have to be dealt directly (and this generally requires some form of the implicit function theorem or preliminary artillery that would lead to it – e.g., the penalization approach does not use the implicit function theorem but the penalization approach can be used to prove the implicit function theorem). As a consequence, it is standard to well separate problems with inequality constraints only and problems with both inequality and equality constraints, in particular when it comes to constraint qualification (see, e.g., "A Guided Tour in Constraint Qualifications for Nonlinear Programming under Differentiability Assumptions" by Giorgi). Finally, the assumptions of the theorem are not the weakest (in terms of differentiability for instance); we refer to Blot in "On the multiplier rules" (2016) for more details with respect to this specific set of assumptions.

Remark. (See C.5. in Bazaraa&Shetty FoO, C.9. in Güler FoO, S.5.3.-4. in Andreasson&Evgrafov&Patriksson ICO). The geometric proof proceeds in three steps: a geometric necessary condition for optimality is derived; the geometric condition is translated in an analytic condition by conic approximation (FT); the analytic solution is made effective (i.e., the multiplier for f in FT is nonzero) under the condition that the approximation is good enough (KKT). More precisely, the geometric proof is based on the basic idea that: "if the point $x^* \in S$ is a local minimum of f over S, it should not be possible to draw a curve starting at the point x^* inside S, such that f decreases along it arbitrarily close to x^*". This translates directly into the fact that the space of feasible directions at x^* and the space of strict descent direction at x^* are disjoint. By convex approximation through linearization, the sets can be separated and the existence of multipliers guaranteed. Under the conditions that the approximations are good enough (conditions known as constraint qualifications), the previous multipliers formulation can be made useful by guaranteeing that the optimality conditions depend on f (i.e., the multiplier μ_0 for f in FT is nonzero).

Remark. If one does not want to prove the Hahn–Banach theorem and Farkas' lemma, a very simple proof under a slightly stronger rank condition is given by Brezhneva et al. in "A simple and elementary proof of the Karush–Kuhn–Tucker theorem for inequality-constrained optimization" (2009). Moreover, if one is only interested in inequality constraints, the proof can be substantially reduced, for technicalities calling for the implicit function theorem are due to the inclusion of equality constraints. For our proof, we first prove a version of Farkas' lemma and then the implicit function theorem.

Reference. The proofs of KKT, FT, and Motzkin's lemma follow from Freund's lecture notes for 15.084J at MIT. Similar proofs for FT and KKT can be found in Güler FO (see T.9.4. and C.9.6.), as well as different proofs of Motzkin's lemma (see comment p.71-72 in S.3.3.). Other results are referenced after their statement.

Lemma 2 (Farkas' Lemma)**.** *Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. Then exactly one the following propositions is true, but not both.*

1. There exists $x \in \mathbb{R}^n$ such that $Ax = b$ and $x \geq 0_n$.

2. There exists $y \in \mathbb{R}^m$ such that $A^T y \geq 0_m$ and $b^T y < 0$.

Proof. See Lecture 6. □

Lemma 3 (**Motzkin's Transposition Lemma**)**.** *Given matrices A, B, and H, exactly one of the two systems has a solution:*

1. $Ax < 0,\ Bx < 0,\ Hx = 0$;

2. $A^T u + B^T v + H^T w = 0,\ (u, v) \geq 0,\ 1^T u = 1$.

Proof. We first prove that the two systems cannot be true simultaneously. Suppose (1.) is true for some x and (2.) holds, then $x^T(A^T u + B^T v + H^T w) = 0$, that is, $0 = u^T Ax + v^T Bx + w^T Hx = u^T Ax + v^T Bx < 0$, a contradiction.

Suppose now that (1.) does not have a solution. Then for some $\theta > 0$, the system $Ax + 1\theta \leq 0,\ Bx \leq 0,\ Hx \leq 0,\ -Hx \leq 0$ has no solution. This system rewrites as

$$\begin{bmatrix} A & 1 \\ B & 0 \\ H & 0 \\ -H & 0 \end{bmatrix} \begin{bmatrix} x \\ \theta \end{bmatrix} \leq 0\,, \quad \begin{bmatrix} 0 & \dots 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ \theta \end{bmatrix} > 0.$$

From Farkas' lemma, there exists $(u, v, w_1, w_2) \geq 0$ such that

$$\begin{bmatrix} A & 1 \\ B & 0 \\ H & 0 \\ -H & 0 \end{bmatrix}^T \begin{bmatrix} u \\ v \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

This rewrites as

$$A^T u + B^T v + H^T w_1 - H^T w_2 = 0, \quad 1^T u = 1.$$

Taking $w = w_1 - w_2$ concludes the proof. □

Lemma 4 (**Implicit Function Theorem**)**.** *Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be open sets. Let $f\colon X \times Y \to \mathbb{R}^m$ be a continuously differentiable function. If $(a, b) \in X \times Y$ is such that*

$$f(a, b) = 0 \quad \textit{and} \quad D_y f(a, b) := \left[\frac{\partial f_i}{\partial y_j}(a, b)\right]_{m\times m}$$

is invertible, then there exits open neighborhoods $U \subseteq X$ of a and $V \subseteq Y$ of b and a continuously differentiable function $g\colon U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^m$ such that a point $(x, y) \in U \times V$ satisfies $f(x, y) = 0$ if and only if $g(x) = y$. Moreover, the derivative of g at $x \in U$ is given by

$$Dg(x) = -D_y f(x, g(x))^{-1} D_x f(x, g(x)).$$

Proof. Assume without loss of generality that $(a, b) = (0, 0)$ (otherwise, consider $(x, y) \mapsto f(x + a, y + b) - f(a, b)$). Since Df is continuous and det is continuous, there exists open neighborhoods U of 0 and V of 0, such that $[\partial f_i(x, \zeta_{ij})/\partial y_j]_{m\times m}$ is invertible for all $(x, \zeta_{ij}) \in U \times V$ where $1 \leq i, j \leq m$. For each $x \in U$, there exists at most one $y \in V$ such that $f(x, y) = 0$. Suppose there exists $y, z \in V$, $z \neq y$, such that $f(x, y) = f(x, z) = 0$. The mean value theorem implies that there exists $w_i \in (y, z)$ such that

$$0 = f_i(x, z) - f_i(x, y) = [D_y f(x, w_i)]_i (z - y)$$

for all $i = 1, \dots, m$. Since the matrix with rows $[D_y f(x, w_i)]_i$ is invertible as shown above, we obtain that $y = z$, a contradiction.

Let $\bar{B}_r(0) \subseteq V$. Since $f(0, 0) = 0$, we have $f(0, y) \neq 0$ for any $y \in S_r(0) = \{y \in \mathbb{R}^m : \|y\| = r\}$. Since f is continuous, there exists $\alpha > 0$ such that $\|f(0, y)\| \geq \alpha$ for all $y \in S_r(0)$. Define $F(x, y) := \|f(x, y)\|^2 = \sum_{i=1}^m (f_i(x, y))^2$. Then $F(0, y) \geq \alpha > 0$ for $y \in S_r(0)$ and $F(0, 0) = 0$. Since F is continuous, there exists an open neighborhood $U' \subseteq U$ of 0 such that $F(x, y) \geq \alpha/2$, $F(x, 0) \leq \alpha/2$ for all $x \in U'$ and all $y \in S_r(0)$. Thus for a fixed $x \in U'$ the function $y \mapsto F(x, y)$ achieves its minimum on $\bar{B}_r(0)$ at a point $g(x)$ in the interior of $\bar{B}_r(0)$ and we have

$$0 = D_y f(x, g(x)) = 2 D_y f(x, g(x)) f(x, g(x)).$$

Since $D_y f(x, g(x))$ is nonsingular, we have $f(x, g(x)) = 0$ for all $x \in U'$.

Consider now $\Delta x \in \mathbb{R}^n$ such that $x + \Delta x \in U'$ and define $\Delta y = g(x + \Delta x) - g(x)$. Then by the mean value theorem we have

$$D_x f(x', y') \Delta x + D_y f(x', y') \Delta y = f(x + \Delta x, g(x + \Delta x)) - f(x, g(x)) = 0$$

for some (x', y') on the line segment between $(x, g(x))$ and $(x + \Delta x, g(x + \Delta x))$. From the triangle inequality and the consistency of the matrix norm, this and the invertibility of $D_y f(x', y')$ imply that if $\|\Delta x\| \to 0$, then $\|\Delta y\| \to 0$. This proves that g is a continuous function. To prove that g is C^1, we use Taylor's formula to get

$$\begin{aligned} 0 &= f(x + \Delta x, g(x + \Delta x)) - f(x, g(x)) \\ &= D_x f(x, g(x)) \Delta x + D_y f(x, g(x)) \Delta y + o((\Delta x, \Delta y)). \end{aligned}$$

Since g is continuous, we have $o((\Delta x, \Delta y)) = o(\Delta x)$, and so

$$\Delta y = -D_y f(x, g(x))^{-1} D_x f(x, g(x)) \Delta x + o(\Delta x)$$

This proves that g is differentiable at x with derivative

$$Dg(x) = -D_y f(x, g(x))^{-1} D_x f(x, g(x)).$$

The continuity of Df implies the continuity of Dg. □

Reference. We follow closely T.2.26. in Güler FO p.45 who expands the proof in S.9.-12. in Carathéodory CV&PDE (1935,1982) p.10-13. This is an elementary proof that simply uses Taylor's formula (which includes the mean value theorem for $k = 1$) whose proof can be found as T.1.1. in Güler FO p.2. For another elementary proof as well as references on other standard proofs (which generally prove first the inverse function theorem by resorting to the contracting mapping principle), see "The Implicit and Inverse Function Theorems: Easy Proofs" by de Oliveira (2013). For reminders on asymptotic comparisons, see Gourdon's Analyse S.2.2. p.86.

Remark. We collect a few definitions needed for the proof of KKT conditions proper. Consider f and S as in (P). Define $R_S(x)$ the set of all feasible directions for S at $x \in S$ by

$$R_S(x) = \Big\{d \in \mathbb{R}^n : \exists\, \delta > 0 \text{ s.t. } x + \lambda d \in S \text{ for all } \lambda \in [0, \delta]\Big\},$$

and $RD_f(x)$ the set of all (strict) descent directions of f at $x \in \mathbb{R}^n$ by

$$RD_f(x) = \Big\{d \in \mathbb{R}^n : \exists\, \delta > 0 \text{ s.t. } f(x + \lambda d) < f(x) \text{ for all } \lambda \in (0, \delta]\Big\}.$$

These sets which are cones, even if very natural, may be mis-behaved (e.g., for a circle in $\mathbb{R}^2$) and we need to consider slightly bigger cones containing them as developed by Bouligand[1]. Define then $T_S(x)$ the (Bouligand) tangent cone of the feasible set S at $x \in S$ by

$$T_S(x) = \Big\{d \in \mathbb{R}^n : \exists\, (x_n) \in S^\infty, \exists\, (\lambda_n) \in (0, \infty)^\infty \text{ s.t. } x_n \to x,\ \lambda_n(x_n - x) \to d\Big\},$$

and $TD_f(x)$ the set of all (strict) descent directions (in the sense of Bouligand) for f at $x \in \mathbb{R}^n$ by

$$\begin{aligned} TD_f(x) = \Big\{d \in \mathbb{R}^n : \exists\, (x_n) \in (\mathbb{R}^n)^\infty, \exists\, (\lambda_n) \in (0, \infty)^\infty \\ \text{s.t. } x_n \to x,\ \lambda_n(x_n - x) \to d, f(x_n) < f(x)\Big\}. \end{aligned}$$

It holds that: $R_S(x) \subseteq T_S(x)$; if S is convex and closed, then $R_S(x) = T_S(x)$ (see P.5.3. in Andreasson&Evgrafov&Patriksson ICO and P.2.55. in Bonnans&Shapiro PAOP); $RD_f(x) \subseteq TD_f(x)$ (see own proof).

Lemma 5 (Geometric (Necessary) Optimality Conditions). *If x^* is a local minimizer for (P), then*

$$T_S(x^*) \cap TD_f(x^*) = \emptyset.$$

Proof. If the intersection is not empty, then there exists a sequence of feasible points $(x_n) \in S^\infty$ such that $f(x_n) < f(x^*)$, which contradicts the fact that x^* is a local minimizer. □

Remark. The result above (see L.9.3. in Güler FO p.210) is not used directly for the proof but it is the initial result that motivates the geometric route. We will directly prove below a version for linear approximations of the sets above. In particular, it makes use of the sufficient condition for descent: $\nabla f^T d < 0$ (see P.4.16. in Andreasson&Evgrafov&Patriksson ICO). We nautrally define the linearized versions of the feasible directions and descent directions for problem (Q) and the associated cones

$$TD_f^0(x) = \{d \in \mathbb{R}^n : \nabla f(x)^T d < 0\},$$

and

$$T_S^0(x) := G^0(x) \cap H^0(x),$$

where

$$G^0(x) = \{d \in \mathbb{R}^n : \nabla g_i(x)^T d < 0,\ i \in I(x)\},$$

[1]"Comme pour le calcul de la dérivée d'une fonction, la définition des directions tangentes qui sont les éléments du cône tangent requiert un passage à la limite." (Wikipedia)

with $I(x) = \{i \in \{1, \dots, r : g_i(x) = 0\}$ the index set of active constraints at x, and

$$H^0(x) = \{d \in \mathbb{R}^n : \nabla h_j(x)^T d = 0,\ j = 1, \dots, m\}.$$

Lemma 6 (**Linearized Geometric (Necessary) Optimality Conditions**). *If x^* is a local minimizer for (Q) and the vectors $\{\nabla h_j(x^*)\}_{j=1}^m$ are linearly independent, then*

$$G^0(x^*) \cap H^0(x^*) \cap TD_f^0(x^*) = \emptyset.$$

Remark (S.9.1. in Güler FO p.211). If we only had inequality constraints, the result $G^0(x^*) \cap TD_f^0(x^*) = \emptyset$ could be proved much faster. Indeed, if $d \in \{d \in \mathbb{R}^n : \nabla F(x)^T d < 0\}$ for some continuously differentiable function $F\colon \mathbb{R}^n \to \mathbb{R}$, then $F(x + td) = F(x) + t[\nabla F(x)^T d + o(t)/t] < F(t)$ for all $t > 0$ small, since $\nabla F(x)^T d < 0$ and $\lim_{t\to 0} o(t)/t = 0$. Therefore, if $d \in G^0(x^*) \cap TD_f^0(x^*)$, then $f(x^* + td) < f(x^*)$ and $g(x^* + td) < g(x^*) = 0$ for $t > 0$ small, which contradicts that x^* is a local minimizer for (Q). The case with $H^0(x^*)$ is more technical; it can be handled by invoking the implicit function theorem.

Proof. The idea of the proof is similar. We show by contradiction that if $d \in G^0(x^*) \cap H^0(x^*) \cap TD_f^0(x^*)$, there is a point $x(\theta) \in S$ such that $\lim_{\theta\to 0} x(\theta) = x^*$ and $f(x(\theta)) < f(x^*)$ for $\theta > 0$ small, a contradiction with the fact that x^* is a local minimizer for (Q).

Let $A \in \mathbb{R}^{m\times n}$ be the matrix with rows $\nabla h_j(x^*)^T$. Since A has full row rank, its columns and the corresponding elements of x^* can be re-arranged so that $A = [B\ N]$ and $x^* = (y^*, z^*)$ where B is invertible. Then the implicit function theorem guarantees that there is an open set U containing z^* such that there is a function s such that $s(z^*) = y^*$ and $h(s(z), z) = 0$ for all $z \in U$. Suppose now that $d \in G^0(x^*) \cap H^0(x^*) \cap TD_f^0(x^*)$. Write $d = (q, p)$. Then $0 = Ad = Bq + Np$, so that $q = -B^{-1}Np$. Let $z(\theta) = z^* + \theta p$ and $y(\theta) = s(z(\theta)) = s(z^* + \theta p)$. Define $x(\theta) = (y(\theta), z(\theta))$.

We first show that $x(\theta) \in S$. For $\theta > 0$ sufficiently small, the implicit function theorem directly yields that $h(x(\theta)) = h(s(z(\theta)), z(\theta)) = 0$. Moreover, by differentiating with respect to θ, we get using the chain rule that for all $j = 1, \dots, m$,

$$0 = \sum_{k=1}^{m} \frac{\partial h_j(s(z(\theta)), z(\theta))}{\partial y_k} \frac{\partial s_k(z(\theta))}{\partial \theta} + \sum_{k=1}^{n-m} \frac{\partial h_j(s(z(\theta)), z(\theta))}{\partial z_k} \frac{\partial z_k(z(\theta))}{\partial \theta}.$$

Denote $r_k = \frac{\partial s_k(z(\theta))}{\partial \theta}$ and notice that $p_k = \frac{\partial z_k(z(\theta))}{\partial \theta}$. At $\theta = 0$, this system rewrites as $0 = Br + Np$, so that $r = -B^{-1}Np = q$. Then $\frac{\partial x_k(\theta)}{\partial \theta} = d_k$ for $k = 1, \dots, n$. Now, for all $i \in I(x^*)$,

$$\begin{aligned} g_i(x(\theta)) &= g_i(x^*) + \theta \frac{\partial g_i(x(\theta))}{\partial \theta}\bigg|_{\theta=0} + o(\theta) \\ &= 0 + \theta \sum_{k=1}^{n} \frac{\partial g_i(x(\theta))}{\partial x_k} frac\partial x_k(\theta)\partial\theta \bigg|_{\theta=0} + o(\theta) \\ &= \theta \nabla g_i(x^*)^T d + o(\theta). \end{aligned}$$

Hence, $g_i(x(\theta)) < 0$ for all $i = 1, \dots, r$ for $\theta > 0$ small enough. It follows that $x(\theta) \in S$ for $\theta > 0$ small enough. Moreover, $x(0) = x^*$ and so by continuity of s, $\lim_{\theta\to 0} x(\theta) = x^*$.

We finally prove that $x(\theta)$ improves on x^*. We have for the same reason that

$$f(x(\theta)) = f(x^*) + \theta \nabla f(x^*)^T d + o(\theta) < f(x^*)$$

for $\theta > 0$ small enough. This contradicts the local optimality of x^*. □

Reference. This is the difficult part in the derivation of the KKT conditions. We follow closely the proof of Freund for T.2.1. p.28 in his lecture notes for 15.084J at MIT. The same result and alternative proofs can be found in T.9.4. in Güler FO p.211 or T.5.2.1. in Bazaraa&Shetty p.122.

Lemma 7 (Fritz John (Necessary) Conditions). *If x^* is a local minimizer of (Q), then there exists $(\lambda_0, \lambda, \mu) \in \mathbb{R} \times \mathbb{R}^r \times \mathbb{R}^m$ such that*

$$\lambda_0 \nabla f(x^*) + \sum_{i=1}^{r} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{m} \mu_j \nabla h_j(x^*) = 0,$$

and

$$(\lambda_0, \lambda) \geq 0, \quad (\lambda_0, \lambda, \mu) \neq 0, \quad \textit{and} \quad \lambda_i g_i(x^*) = 0 \quad \textit{for all} \quad i = 1, \ldots, r.$$

Proof. If the vectors $\{\nabla h_j(x^*)\}_{j=1}^m$ are linearly dependent, then there exist $\mu \neq 0$ such that $\sum_{j=1}^m \mu_j \nabla h(x^*) = 0$. Set $(\lambda_0, \lambda) = 0$ gives the result.

Suppose now that $\{\nabla h_j(x^*)\}_{j=1}^m$ are linearly independent, then $G^0(x^*) \cap H^0(x^*) \cap TD_f^0(x^*) = \emptyset$. Without loss of generality, assume $I(x^*) = \{1, \ldots, p\}$. Define $A \in \mathbb{R}^{(p+1)\times n}$ the matrix with rows $\nabla f(x^*)^T$ and $\nabla g_i(x^*)^T$ for $i = 1, \ldots, p$ and $H \in \mathbb{R}^{m\times n}$ the matrix with rows $\nabla h_j(x^*)^T$ for $j = 1, \ldots, m$. The the empty intersection proposition implies that there is no $d \in \mathbb{R}^n$ such that $Ad < 0$ and $Hd = 0$. From Motzkin's lemma, there thus exists $(\lambda_0, \lambda_1, \ldots, \lambda_p)$ and $(\mu_1, \ldots, \mu_m)$ such that

$$\lambda_0 \nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{m} \mu_j \nabla h_j(x^*) = 0,$$

with $\sum_{i=0}^p \lambda_i = 1$ and $(\lambda_0, \lambda_1, \ldots, \lambda_p) \geq 0$. Define $(\lambda_{p+1}, \ldots, \lambda_r) = 0$. Then $(\lambda_0, \lambda) \geq 0$, $(\lambda_0, \lambda, \mu) \neq 0$, and by definition either $g_i(x^*)$ or $\lambda_i = 0$ for any $i = 1, \ldots, r$. Moreover,

$$\lambda_0 \nabla f(x^*) + \sum_{i=1}^{r} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{m} \mu_j \nabla h_j(x^*) = 0,$$

which concludes the proof. □

References

ANDRÉASSON, N., A. EVGRAFOV, AND M. PATRIKSSON (2005): *An introduction to continuous optimization: Foundations and fundamental algorithms*. Chalmers University of Technology Press.

BAZARAA, M. S., AND C. M. SHETTY (2012): *Foundations of optimization*, vol. 122. Springer.

BLOT, J. (2016): "On the multiplier rules," *Optimization*, 65(5), 947–955.

BONNANS, J. F., AND A. SHAPIRO (2013): *Perturbation analysis of optimization problems*. Springer.

BREZHNEVA, O. A., A. A. TRET'YAKOV, AND S. E. WRIGHT (2009): "A simple and elementary proof of the Karush–Kuhn–Tucker theorem for inequality-constrained optimization," *Optimization Letters*, 3, 7–10.

CARATHÉODORY, C. (1982): *Calculus of variations and partial differential equations of the first order*. Chelsea Publishing.

FREUND, R. M. (2016): "Optimality Conditions for Constrained Optimization Problems," *Lecture notes for 15.084J at MIT*.

GIORGI, G., ET AL. (2018): "A guided tour in constraint qualifications for nonlinear programming under differentiability assumptions," *Working Paper*.

GOURDON, X. (2020): *Analyse*. Ellipses.

GÜLER, O. (2010): *Foundations of Optimization*. Springer New York.