Tutorial notes

2024-2025 Optimization theory

These notes are organized according to the bookAndréasson et al. [2020].

1 Optimality in unconstrained optimization

In Theorem[Fundamental Theorem of global optimality], we have established that locally optimal solutions are also global in the convex case. What are the necessary and sufficient conditions for a vector x^* to be a local optimum? This is an important question, because the algorithms that we will investigate for solving important classes of optimization problems are always devised based on those conditions that we would like to fulfill. This is a statement that seems to be true universally: efficient, locally or globally convergent iterative algorithms for an optimization problem are directly based on its necessary and/or sufficient local optimality conditions.

We begin by establishing these conditions for the case of unconstrained optimization, where the objective function is in C^1. Every proof is based on the Taylor expansion of the objective function up to order one or two. Our problem here is the following:

$$\text{minimize } f(x), \quad x \in \mathbb{R}^n, \tag{1}$$

where f is in C^1 on $\mathbb{R}^n$ (for short we say: in C^1 or $C^1(\mathbb{R}^n)$).

Theorem 1 (Necessary Optimality Conditions, C^1 Case)**.** *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is in C^1 on $\mathbb{R}^n$. Then,*

$$x^* \textit{ is a local minimum of } f \textit{ over } \mathbb{R}^n \implies \nabla f(x^*) = 0.$$

Note that:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix},$$

so the requirement is:

$$\frac{\partial f(x^*)}{\partial x_j} = 0, \quad j = 1, \dots, n. \tag{2}$$

We refer to this condition as x^* being a stationary point of f.

Special Case: For $n = 1$, Theorem 1 reduces to:

$$x^* \in \mathbb{R} \text{ is a local minimum} \implies f'(x^*) = 0.$$

Proof. (By contradiction.) Suppose that x^* is a local minimum, but $\nabla f(x^*) \neq 0$. Let $p := -\nabla f(x^*)$, and consider the Taylor expansion of f around $x = x^*$ in the direction of p:

$$f(x^* + \alpha p) = f(x^*) + \alpha \nabla f(x^*)^\top p + o(\alpha), \tag{3}$$

where $o : \mathbb{R} \to \mathbb{R}$ satisfies $\frac{o(s)}{s} \to 0$ as $s \to 0$.

Substituting $p = -\nabla f(x^*)$, we get:

$$f(x^* + \alpha p) = f(x^*) - \alpha \|\nabla f(x^*)\|^2 + o(\alpha). \tag{4}$$

For sufficiently small $\alpha > 0$, this implies:

$$f(x^* + \alpha p) < f(x^*),$$

which contradicts the assumption that x^* is a local minimum. Hence, $\nabla f(x^*) = 0$. This completes the proof. □

The opposite direction is false: take $f(x) = x^3$; then, $\bar{x} = 0$ is stationary, but it is neither a local minimum nor a local maximum.

The proof is instrumental in that it provides a sufficient condition for a vector p to define a descent direction, that is, a direction such that a small step along it yields a lower objective value. We first define this notion properly.

Definition 1 (Descent Direction)**.** *Let the function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be given. Let $x \in \mathbb{R}^n$ be a vector such that $f(x)$ is finite. Let $p \in \mathbb{R}^n$. We say that the vector $p \in \mathbb{R}^n$ is a* descent direction *with respect to f at x if there exists $\delta > 0$ such that:*

$$f(x + \alpha p) < f(x) \quad \text{for every } \alpha \in (0, \delta]. \tag{5}$$

Proposition 1 (Sufficient Condition for Descent)**.** *Suppose that $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is in C^1 around a point x for which $f(x) < +\infty$, and that $p \in \mathbb{R}^n$. Then:*

$$\nabla f(x)^\top p < 0 \implies p \text{ is a descent direction with respect to } f \text{ at } x. \tag{6}$$

Proof. Since f is in C^1 around x, we can construct a Taylor expansion of f, as follows:

$$f(x + \alpha p) = f(x) + \alpha \nabla f(x)^\top p + o(\alpha), \tag{7}$$

where $o : \mathbb{R} \to \mathbb{R}$ satisfies $\frac{o(s)}{s} \to 0$ as $s \to 0$.

Since $\nabla f(x)^\top p < 0$, we have:

$$f(x + \alpha p) < f(x), \tag{8}$$

for all sufficiently small $\alpha > 0$. This completes the proof. □

Notice that at a point $x \in \mathbb{R}^n$, there may be other descent directions $p \in \mathbb{R}^n$ besides those satisfying $\nabla f(x)^\top p < 0$.

If f is additionally convex, then the opposite implication in the above proposition is true, thus making the descent property equivalent to the property that the directional derivative is negative. Since this result can also be stated for non-differentiable functions f (in which case we replace the expression $\nabla f(x)^\top p$ with the classic expression for the directional derivative):

$$f'(x;p) := \lim_{\alpha \to 0^+} \frac{1}{\alpha}\big[f(x+\alpha p) - f(x)\big], \tag{9}$$

If f has stronger differentiability properties, then we can make additional statements about the nature of a local optimum.

Theorem 2 (Necessary Optimality Conditions, C^2 Case)**.** *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is in C^2 on $\mathbb{R}^n$. Then,*

$$x^* \text{ is a local minimum of } f \implies \begin{cases} \nabla f(x^*) = 0, \\ \nabla^2 f(x^*) \text{ is positive semidefinite.} \end{cases}$$

Note: *For $n = 1$, Theorem 2 reduces to:*

$$x^* \in \mathbb{R} \text{ is a local minimum of } f \implies f'(x^*) = 0 \text{ and } f''(x^*) \geq 0.$$

Proof. Consider the Taylor expansion of f up to order two around x^* in the direction of a vector $p \in \mathbb{R}^n$:

$$f(x^* + \alpha p) = f(x^*) + \alpha \nabla f(x^*)^\top p + \frac{\alpha^2}{2} p^\top \nabla^2 f(x^*) p + o(\alpha^2). \tag{10}$$

Suppose that x^* satisfies $\nabla f(x^*) = 0$, but there exists a vector $p \neq 0$ such that $p^\top \nabla^2 f(x^*) p < 0$, meaning that $\nabla^2 f(x^*)$ is not positive semidefinite. Then the above expansion yields:

$$f(x^* + \alpha p) < f(x^*) \quad \text{for all sufficiently small } \alpha > 0,$$

which implies that x^* cannot be a local minimum. This completes the proof. □

The next result shows that under some circumstances, we can establish local optimality of a stationary point.

Theorem 3 (Sufficient Optimality Conditions, C^2 Case)**.** *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is in C^2 on $\mathbb{R}^n$. Then:*

$$\nabla f(x^*) = 0 \quad \text{and} \quad \nabla^2 f(x^*) \text{ is positive definite} \implies x^* \text{ is a strict local minimum of } f \text{ over } \mathbb{R}^n.$$

Note: *For $n = 1$, Theorem 3 reduces to:*

$$f'(x^*) = 0 \quad \text{and} \quad f''(x^*) > 0 \implies x^* \in \mathbb{R} \text{ is a strict local minimum of } f \text{ over } \mathbb{R}.$$

Proof. Suppose that $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive definite. Take an arbitrary vector $p \in \mathbb{R}^n, p \neq 0$. Then, using the Taylor expansion:

$$f(x^* + \alpha p) = f(x^*) + \alpha \nabla f(x^*)^\top p + \frac{\alpha^2}{2} p^\top \nabla^2 f(x^*) p + o(\alpha^2). \tag{11}$$

Since $\nabla f(x^*)^\top p = 0$ and $p^\top \nabla^2 f(x^*) p > 0$ (due to positive definiteness of $\nabla^2 f(x^*)$), we have:

$$f(x^* + \alpha p) > f(x^*) \quad \text{for all small enough } \alpha > 0. \tag{12}$$

As p was arbitrary, this implies that x^* is a strict local minimum of f over $\mathbb{R}^n$. □

We naturally face the following question: When is a stationary point a global minimum? The answer is given in the next theorem.

Theorem 4 (Necessary and Sufficient Global Optimality Conditions). *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is convex and in C^1 on $\mathbb{R}^n$. Then:*

$$x^* \textit{ is a global minimum of } f \textit{ over } \mathbb{R}^n \iff \nabla f(x^*) = 0.$$

Proof. [$\Longrightarrow$] This has already been shown in Theorem 1, since a global minimum is also a local minimum.

[$\Longleftarrow$] The convexity of f yields that for every $y \in \mathbb{R}^n$:

$$f(y) \geq f(x^*) + \nabla f(x^*)^\top (y - x^*) = f(x^*), \tag{13}$$

where the equality stems from the property that $\nabla f(x^*) = 0$. This proves that x^* is a global minimum of f over $\mathbb{R}^n$. □

2 The Karush–Kuhn–Tucker conditions

Definition 2 (Cone of Feasible Directions). *Let $S \subseteq \mathbb{R}^n$ be a nonempty closed set. The cone of feasible directions for S at $x \in \mathbb{R}^n$, known also as the radial cone, is defined as:*

$$R_S(x) := \left\{ p \in \mathbb{R}^n \;\middle|\; \exists \tilde{\delta} > 0 \textit{ such that } x + \delta p \in S,\, 0 \leq \delta \leq \tilde{\delta} \right\}. \tag{14}$$

Thus, this is nothing else but the cone containing all feasible directions at x in the sense of Definition 4.20.

This cone is used in some optimization algorithms, but unfortunately, it is too small to develop optimality conditions that are general enough. Therefore, we consider less intuitive, but bigger and more well-behaving sets (cf. Proposition 5.3 and the examples that follow).

Definition 3 (Tangent Cone). *Let $S \subseteq \mathbb{R}^n$ be a nonempty closed set. The tangent cone for S at $x \in \mathbb{R}^n$ is defined as*

$$T_S(x) := \left\{ p \in \mathbb{R}^n \;\middle|\; \exists \{x_k\} \subset S, \{\lambda_k\} \subset (0, \infty) : \lim_{k \to \infty} x_k = x;\ \lim_{k \to \infty} \lambda_k (x_k - x) = p \right\}. \tag{15}$$

In order to compute approximations to the tangent cone $T_S(x)$, we consider cones associated with the active constraints at a given point:

$$\mathring{G}(x) := \left\{ p \in \mathbb{R}^n \,\middle|\, \nabla g_i(x)^T p < 0,\, i \in I(x) \right\}, \tag{16}$$

and

$$G(x) := \left\{ p \in \mathbb{R}^n \,\middle|\, \nabla g_i(x)^T p \leq 0,\, i \in I(x) \right\}. \tag{17}$$

The following proposition verifies that $\mathring{G}(x)$ is an inner approximation for $R_S(x)$ (and, therefore, for $T_S(x)$ as well, see Proposition 5.3), and $G(x)$ is an outer approximation for $T_S(x)$.

Definition 4 (Abadie's Constraint Qualification)**.** *We say that at the point $x \in S$ Abadie's constraint qualification holds if*

$$T_S(x) = G(x), \tag{18}$$

where $T_S(x)$ is defined by Definition 5.2 and $G(x)$ by (5.7).

Theorem 5 (Karush–Kuhn–Tucker optimality conditions)**.** *Assume that at a given point $x^* \in S$ Abadie's constraint qualification holds. If $x^* \in S$ is a local minimum of f over S, then there exists a vector $\mu \in \mathbb{R}^m$ such that*

$$\nabla f(x^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) = 0, \tag{5.9a}$$

$$\mu_i g_i(x^*) = 0, \quad i = 1, \ldots, m, \tag{5.9b}$$

$$\mu \geq 0. \tag{5.9c}$$

In other words,

$$x^* \text{ is a local minimum of } f \text{ over } S \quad \text{and} \quad \text{Abadie's CQ holds at } x^* \quad \Longrightarrow \quad \exists \mu \in \mathbb{R}^m : \text{(5.9) holds.}$$

The system will be referred to as the Karush–Kuhn–Tucker optimality conditions.

Proof. By Theorem 5.8, we have that

$$\mathring{F}(x^*) \cap T_S(x^*) = \emptyset,$$

which due to our assumptions implies that

$$\mathring{F}(x^*) \cap G(x^*) = \emptyset.$$

As in the proof of Theorem 5.15, construct a matrix A with columns $\nabla g_i(x^*)$, $i \in I(x^*)$. Then, the system

$$AT_p \leq 0 \quad \text{for} \quad |I(x^*)|$$

and

$$-\nabla f(x^*) T_p > 0$$

has no solutions. By Farkas' Lemma (cf. Theorem 3.30), the system

$$A\xi = -\nabla f(x^*), \quad \xi \geq 0 \quad \text{for} \quad |I(x^*)|$$

has a solution. Define the vector

$$\mu_{I(x^*)} = \xi, \quad \mu_i = 0, \quad \text{for} \quad i \notin I(x^*).$$

Then, the so-defined μ verifies the KKT conditions (5.9). $\square$

Now we consider both inequality and equality constraints, that is, we assume that the feasible set S is given by

$$S := \{x \in \mathbb{R}^n \mid g_i(x) \leq 0,\, i = 1, \dots, m;\, h_j(x) = 0,\, j = 1, \dots, \ell\}\,.$$

Definition 5. ***(Mangasarian–Fromovitz CQ)*** *We say that at the point $x \in S$, where S is given by* (5.10)*, the Mangasarian–Fromovitz CQ holds if the gradients $\nabla h_j(x)$ of the functions h_j, $j = 1, \dots, \ell$, defining the equality constraints, are linearly independent, and the intersection*

$$\overset{\circ}{G}(x) \cap H(x)$$

is nonempty.

3 Example

3.1 Abadie's constraint qualification

Example 1 (Cone of Feasible Directions and Tangent Cone)**.** *Let $S := \{x \in \mathbb{R}^2 \mid -x_1 \leq 0, (x_1 - 1)^2 + x_2^2 \leq 1\}$. Then, the cone of feasible directions at 0^2 is given by:*

$$R_S(0^2) = \{p \in \mathbb{R}^2 \mid p_1 > 0\},$$

and the tangent cone at 0^2 is:

$$T_S(0^2) = \{p \in \mathbb{R}^2 \mid p_1 \geq 0\}.$$

That is, $T_S(0^2) = \overline{R_S(0^2)}$. In this example, Abadie's constraint qualification holds, ensuring that the KKT system must be solvable.

Solution: *The feasible set S consists of points satisfying $x_1 \geq 0$ and within a unit disk centered at $(1, 0)$. The feasible directions are those where $p_1 > 0$, ensuring movement remains inside S. Since the tangent cone is the closure of the feasible direction cone, we conclude $T_S(0^2) = \overline{R_S(0^2)}$.*

The KKT system is:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

which admits solutions:

$$\mu = \begin{bmatrix} \mu_1 \\ \frac{1-\mu_1}{2} \end{bmatrix}, \quad 0 \leq \mu_1 \leq 1.$$

Therefore, there exist infinitely many Lagrange multipliers, all belonging to a bounded set.

Example 2 (Complementarity Constraint)**.** *Let $S := \{x \in \mathbb{R}^2 \mid -x_1 \leq 0, -x_2 \leq 0, x_1 x_2 \leq 0\}$. In this case, S forms a non-convex cone, and it holds that:*

$$R_S(0^2) = T_S(0^2) = S.$$

This is one of the rare cases where Abadie's constraint qualification is violated, yet the KKT system remains solvable.

Solution: *The constraint $x_1x_2 \leq 0$ restricts feasible points to the second and fourth quadrants, along with the coordinate axes. This results in S being non-convex. Since S is a cone, we conclude $R_S(0^2) = T_S(0^2) = S$.*

The KKT system:

$$\begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}-1 & 0 & 0\\ 0 & -1 & 0\end{bmatrix}\begin{bmatrix}\mu_1\\ \mu_2\\ \mu_3\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix},$$

gives the solutions:

$$\mu = \begin{bmatrix}1\\0\\ \mu_3\end{bmatrix}, \quad \mu_3 \geq 0.$$

This means that the set of Lagrange multipliers is unbounded.

Example 3 (Complementarity Constraint)**.** *Let $S := \{x \in \mathbb{R}^2 \mid -x_1^3 + x_2 \leq 0, \quad x_1^5 - x_2 \leq 0, \quad -x_2 \leq 0\}$.*

Then, the cone of feasible directions at 0^2 is empty:

$$R_S(0^2) = \emptyset,$$

and the tangent cone at 0^2 is:

$$T_S(0^2) = \{p \in \mathbb{R}^2 \mid p_1 \geq 0, p_2 = 0\}.$$

Since, in the Fritz John system, the multiplier μ_0 is necessarily zero, the KKT system admits no solutions:

$$\begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}0 & 0 & 0\\ 1 & -1 & -1\end{bmatrix}\begin{bmatrix}\mu_1\\ \mu_2\\ \mu_3\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix},$$

$$\mu \geq 0.$$

This system is inconsistent, confirming that the KKT conditions have no valid solutions.

Solution: *The constraints impose severe restrictions on the feasible region. The conditions $x_2 \geq x_1^3$ and $x_2 \leq x_1^5$ confine feasible points to an increasingly narrow region as $x_1 \to 0$. The additional constraint $x_2 \geq 0$ further limits movement in the negative x_2 direction. As a result, no nontrivial feasible directions exist, leading to an empty feasible direction cone:*

$$R_S(0^2) = \emptyset.$$

However, the tangent cone, which considers limiting directions of feasible sequences, is given by:

$$T_S(0^2) = \{p \in \mathbb{R}^2 \mid p_1 \geq 0, p_2 = 0\}.$$

This means movement is only possible along the positive x_1 axis.

The inconsistency in the KKT system stems from the first equation $1 + 0 = 0$, which is clearly unsolvable. This shows that the KKT conditions cannot be used to find a valid solution.

Furthermore, Abadie's constraint qualification is violated since $T_S(0^2) \neq L_S(0^2)$. The non-convexity of the feasible region and the empty feasible direction cone lead to a situation where linearizing the constraints fails to provide valid first-order necessary conditions.

Example 4 (Non-Empty Tangent Cone and Unbounded Multipliers)**.** *Let $S := \{x \in \mathbb{R}^2 \mid -x_2 \leq 0, \quad (x_1-1)^2 + x_2^2 = 1\}$. Then, the cone of feasible directions at 0^2 is empty:*

$$R_S(0^2) = \emptyset.$$

However, the tangent cone is:

$$T_S(0^2) = \{p \in \mathbb{R}^2 \mid p_1 = 0, p_2 \geq 0\}.$$

Solution: *The feasible set S is defined by a unit circle centered at $(1,0)$, with the additional constraint $x_2 \geq 0$, limiting it to the upper half-circle. Since the point $(0,0)$ is on the boundary of this circle, any small step from the origin immediately violates the constraints. Therefore, no feasible directions exist, making $R_S(0^2) = \emptyset$. However, considering sequences approaching the origin from within S, the only limit directions belong to the vertical axis, leading to $T_S(0^2) = \{p \in \mathbb{R}^2 \mid p_1 = 0, p_2 \geq 0\}$.*

The KKT system:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & -2 & 2 \\ -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

admits the solutions:

$$\mu = \begin{bmatrix} 0 \\ \mu_2 \\ \mu_2 - \frac{1}{2} \end{bmatrix}, \quad \mu_2 \geq \frac{1}{2}.$$

This means that the set of Lagrange multipliers is unbounded. The unboundedness arises due to splitting the original equality constraint into two inequalities. In Section 5.6, an alternative formulation that keeps the original equality constraint reduces the number of multipliers to just one.

3.2 KKT Conditions

Example 5. *Consider a symmetric square matrix $A \in \mathbb{R}^{n \times n}$, and the optimization problem to minimize*

$$-x^T A x, \quad \textit{subject to} \quad x^T x \leq 1.$$

The only constraint of this problem is convex; furthermore,

$$(0^n)^T 0^n = 0 < 1,$$

and thus Slater's CQ (Definition 5.38) is verified. Therefore, the KKT conditions are necessary for the local optimality in this problem. We will find all the possible KKT points, and then choose a globally optimal point among them.

The KKT system is as follows:

$$\nabla(-x^T A x) = -2Ax \quad (\textit{since } A \textit{ is symmetric}), \quad \nabla(x^T x) = 2x.$$

Thus, the system is:

$$x^T x \leq 1 \quad \textit{and} \quad -2Ax + 2\mu x = 0^n,$$

$$\mu \geq 0, \quad \mu(x^T x - 1) = 0.$$

From the first two equations, we immediately see that either $x = 0^n$, or the pair (μ, x) is, respectively, a nonnegative eigenvalue and a corresponding eigenvector of A (recall that $Ax = \mu x$ holds). In the former case, from the complementarity condition, we deduce that $\mu = 0$.

Thus, we can characterize the KKT points of the problem into the following groups:

1. *Let $\mu_1, \ldots, \mu_k$ be all the positive eigenvalues of A (if any), and define*

   $$X_i := \left\{ x \in \mathbb{R}^n \mid x^T x = 1;\ Ax = \mu_i x \right\}$$

   to be the set of corresponding eigenvectors of length 1, for $i = 1, \ldots, k$. Then, (x, μ_i) is a KKT point with the corresponding multiplier for every $x \in X_i$, for $i = 1, \ldots, k$. Moreover,

   $$-x^T Ax = -\mu_i x^T x = -\mu_i < 0, \quad \text{for every } x \in X_i,\ i = 1, \ldots, k.$$

2. *Define also*

   $$X_0 := \left\{ x \in \mathbb{R}^n \mid x^T x \leq 1;\ Ax = 0^n \right\}.$$

   Then, the pair $(x, 0)$ is a KKT point with the corresponding multiplier for every $x \in X_0$. We note that if the matrix A is nonsingular, then

   $$X_0 = \{0^n\}.$$

   In any case, $-x^T Ax = 0$ for every $x \in X_0$.

Therefore, if the matrix A has any positive eigenvalue, then the global minima of the problem we consider are the eigenvectors of length one, corresponding to the largest positive eigenvalue; otherwise, every vector $x \in X_0$ is globally optimal.

Example 6. *Similarly to the previous example, consider the following equality-constrained minimization problem associated with a symmetric matrix $A \in \mathbb{R}^{n \times n}$:*

$$\text{minimize} \quad -x^T Ax, \quad \text{subject to} \quad x^T x = 1.$$

The gradient of the only equality constraint equals $2x$, and since 0^n is infeasible, LICQ is satisfied at every feasible point (see Definition 5.41), and the KKT conditions are necessary for local optimality. In this case, the KKT system is extremely simple:

$$x^T x = 1 \quad \text{and} \quad -2Ax + 2\lambda x = 0^n.$$

Let $\lambda_1 < \lambda_2 < \cdots < \lambda_k$ denote all distinct eigenvalues of A, and define as before

$$X_i := \left\{ x \in \mathbb{R}^n \mid x^T x = 1;\ Ax = \lambda_i x \right\}$$

to be the set of corresponding eigenvectors of length 1, for $i = 1, \ldots, k$. Then, (x, λ_i) is a KKT point with the corresponding multiplier for every $x \in X_i$, for $i = 1, \ldots, k$.

Furthermore, since $-x^T Ax = -\lambda_i$ for every $x \in X_i$, for $i = 1, \ldots, k$, it holds that every $x \in X_k$, that is, every eigenvector corresponding to the largest eigenvalue, is globally optimal.

Considering the problem for $A^T A$ and using the spectral theorem, we deduce the well-known fact that

$$\|A\|_k = \max_{1 \leq i \leq k} \{|\lambda_i|\}.$$

Example 7. *Consider the problem of finding the projection of a given point y onto the polyhedron $\{x \in \mathbb{R}^n \mid Ax = b\}$, where $A \in \mathbb{R}^{k\times n}$ and $b \in \mathbb{R}^k$. Thus, we consider the following minimization problem with affine constraints (so that the KKT conditions are necessary for the local optimality, see Section 5.7.4):*

$$\text{minimize} \quad \frac{1}{2}\|x-y\|^2, \quad \text{subject to} \quad Ax = b.$$

The KKT system in this case is written as follows:

$$Ax = b, \quad (x-y) + A^T\lambda = 0^n,$$

for some $\lambda \in \mathbb{R}^k$. Pre-multiplying the last equation by A, and using the fact that $Ax = b$, we get:

$$AA^T\lambda = Ay - b.$$

Substituting an arbitrary solution of this equation into the KKT system, we calculate x via

$$x := y - A^T\lambda.$$

It can be shown that the vector $A^T\lambda$ is the same constant for every Lagrange multiplier λ, so using this formula, we obtain the globally optimal solution to our minimization problem.

Now assume that the columns of A^T are linearly independent, i.e., LICQ holds. Then, the matrix AA^T is nonsingular, and the multiplier λ is therefore unique:

$$\lambda = (AA^T)^{-1}(Ay - b).$$

Substituting this into the KKT system, we finally obtain

$$x = y - A^T(AA^T)^{-1}(Ay - b),$$

the well-known formula for calculating the projection.

References

Niclas Andréasson, Anton Evgrafov, and Michael Patriksson. *An introduction to continuous optimization: foundations and fundamental algorithms.* Courier Dover Publications, 2020.