# ON THE EXISTENCE OF LAGRANGE MULTIPLIERS IN NONLINEAR CONIC PROGRAMMING*

GABRIEL HAESER† AND DAIANA O. SANTOS‡

**Abstract.** The existence of Lagrange multipliers at a solution of a nonlinear optimization problem constitutes one of the cornerstones of modern optimization theory, with many important consequences for guiding algorithmic procedures towards a solution, defining stopping criteria, performing stability analysis, and several other aspects. However, the proof of this result is often intricate, relying on non-trivial tools such as Farkas' Lemma, duality theory, or the implicit function theorem. In this paper, we present a concise and accessible proof of the existence of Lagrange multipliers for nonlinear optimization problems with conic constraints, suitable for advanced undergraduate students or early graduate students. Using only elementary facts about sets and sequences together with Weierstrass' extreme value theorem, our approach employs a penalization technique combined with basic properties of the projection onto closed convex cones, which are presented in detail. The main result establishes that, under Robinson's constraint qualification, every local solution admits a Lagrange multiplier satisfying the Karush/Kuhn–Tucker (KKT) conditions.

**1. Introduction.** The KKT conditions are central to modern nonlinear optimization. They extend the classical method of Lagrange multipliers to problems with both equality and inequality constraints. Their development can be traced back to the independent contributions of Karush (1939) [18] and Kuhn and Tucker (1951) [20]. Later, Mangasarian (1969) [23] provided a clear exposition that helped consolidate the theory. These works have had a lasting influence on both the theoretical foundations and the practical applications of nonlinear optimization. We refer the reader to [19] for a historical account.

Formally, consider the *nonlinear optimization problem*

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} f(x), \\ &\text{s.t. } g_i(x) \le 0, \quad i = 1,\dots,p, \\ &\qquad h_j(x) = 0, \quad j = 1,\dots,q, \end{aligned} \tag{1.1}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$, $g\colon \mathbb{R}^n \to \mathbb{R}^p$ and $h\colon \mathbb{R}^n \to \mathbb{R}^q$ are continuously differentiable. We next recall the classical KKT conditions. They will serve as the starting point for our analysis.

THEOREM 1.1. *Let $\overline{x}$ be a local minimizer of* (1.1). *If $\overline{x}$ satisfies some regularity assumption, then there exist Lagrange multipliers*

$$\mu \in \mathbb{R}^p, \qquad \lambda \in \mathbb{R}^q,$$

*such that the following conditions are satisfied:*

(i) *Stationarity:*

$$\nabla f(\overline{x}) + \sum_{i=1}^{p} \mu_i \nabla g_i(\overline{x}) + \sum_{j=1}^{q} \lambda_j \nabla h_j(\overline{x}) = 0.$$

* August 17th, 2022. Revised manuscript on June 5th, 2026.

**Funding:** This work was funded by FAPESP grant 2023/08706-1 and CNPq grant 407147/2023-3.

†Institute of Mathematics, Statistics, and Computer Science, University of São Paulo, São Paulo-SP, Brazil. (ghaeser@ime.usp.br).

‡Paulista School of Politics, Economics and Business, Federal University of São Paulo, Osasco-SP, Brazil. (daiana.santos@unifesp.br).

(ii) *Primal feasibility:*

$$g_i(\overline{x}) \leq 0 \ \ (i = 1, \ldots, p), \qquad h_j(\overline{x}) = 0 \ \ (j = 1, \ldots, q).$$

(iii) *Dual feasibility:*

$$\mu_i \geq 0 \quad (i = 1, \ldots, p).$$

(iv) *Complementarity:*

$$\mu_i g_i(\overline{x}) = 0 \ \ (i = 1, \ldots, p).$$

Theorem 1.1 is often presented as a standard necessary optimality condition. However, its derivation contains several geometric ideas that become particularly important in more general optimization models, especially in conic programming.

From a pedagogical point of view, our goal is not only to prove existence of Lagrange multipliers, but also to isolate the few geometric and analytical ideas on which the argument relies. In particular, the result may be used in a first graduate course on nonlinear optimization as a bridge between the classical KKT theory and modern conic programming. For this reason, we emphasize projections, polar cones, and penalization, and include illustrative examples and several intermediate observations designed to clarify the role of the main tools involved in the proof.

Beyond their algebraic form, the KKT conditions also have an appealing geometric interpretation. Complementarity implies that $\mu_i = 0$ whenever the corresponding constraint is inactive, that is, whenever $g_i(\overline{x}) < 0$. Therefore, stationarity and dual feasibility show that $-\nabla f(\overline{x})$ belongs to the sum of the subspace generated by the equality constraint gradients and the cone generated by the gradients of active inequality constraints. By Farkas' lemma, this is equivalent to saying that there is no direction $d$ in the linearized feasible cone such that $\nabla f(\overline{x})^\top d < 0$, that is, no first-order feasible direction along which the objective function decreases. Under a suitable regularity condition, the linearized feasible cone faithfully represents the tangent geometry of the feasible set, and such decreasing directions cannot exist at a local minimizer. We refer the reader to the Introduction of [4] for this standard geometric viewpoint.

The approach developed in this paper is inspired by this geometric interpretation, but it is formulated in a way that is better suited to extensions beyond the classical nonlinear programming setting. This makes the proof sufficiently elementary in the nonlinear programming case, while also preparing the ground for conic optimization. As mentioned above, the validity of the KKT conditions as necessary optimality conditions depends on the satisfaction of an additional regularity assumption, known as a *constraint qualification* (CQ). From a geometric viewpoint, constraint qualifications ensure that the linearized model built from the constraint gradients correctly captures the local geometry of the feasible region. In the absence of a CQ, the feasible set may exhibit degeneracies that prevent the tangent feasible set from being described only in terms of the constraint gradients.

Among the many constraint qualifications proposed in the literature, we first recall the *Mangasarian–Fromovitz CQ* (MFCQ). This is one of the classical assumptions under which the KKT conditions hold for nonlinear programming problems. Later, when we move from standard nonlinear programming to problems with cone constraints, we will introduce Robinson's condition, which plays a role analogous to MFCQ in the conic setting.

Let $\overline{x}$ be a feasible point of (1.1). We denote by $\mathcal{A}(\overline{x}) := \{i \in \{1, \ldots, p\} : g_i(\overline{x}) = 0\}$ the set of indices of active inequality constraints at $\overline{x}$.

Definition 1.2. *The Mangasarian–Fromovitz CQ, or MFCQ, holds at $\overline{x}$ if the gradients $\nabla h_1(\overline{x}), \ldots, \nabla h_q(\overline{x})$ are linearly independent, and there exists a direction $d \in \mathbb{R}^n$ such that*

$$\nabla h_j(\overline{x})^\top d = 0 \quad (j = 1, \ldots, q), \qquad \textit{and} \qquad \nabla g_i(\overline{x})^\top d < 0 \quad (i \in \mathcal{A}(\overline{x})).$$

The direction $d$ in Definition 1.2 has a simple geometric meaning. At $\overline{x}$, it is tangent, to first order, to all equality constraints level-sets $\{x : h_i(x) = 0\}$. At the same time, it points strictly toward the interior of all active inequality constraints. Thus, MFCQ rules out the situation in which the active constraints form a degenerate boundary with no first-order direction entering the feasible region.

Under MFCQ, every local minimizer of (1.1) admits Lagrange multipliers satisfying the KKT conditions. The condition is more adequate than the *linear independence constraint qualification*, or LICQ, which requires the gradients of the equality constraints together with the gradients of all active inequality constraints to be linearly independent, which can be too strong, as the next example shows.

*Example* 1.3. Consider the feasible region in $\mathbb{R}^3$ given by the pyramid with square base with vertices at $(\pm 1, \pm 1, 0)$ and apex at $(0, 0, 1)$ as in Figure 1. This can be described by the following feasible set

$$g_1(x) := x_1 + x_3 - 1 \leq 0, \qquad g_2(x) := -x_1 + x_3 - 1 \leq 0,$$
$$g_3(x) := x_2 + x_3 - 1 \leq 0, \qquad g_4(x) := -x_2 + x_3 - 1 \leq 0, \qquad g_5(x) := -x_3 \leq 0.$$

Consider the apex $\overline{x} := (0, 0, 1)$. We have that $\mathcal{A}(\overline{x}) = \{1, 2, 3, 4\}$, therefore the four active gradients in $\mathbb{R}^3$ are necessarily linearly dependent, and LICQ fails. However, the direction $d := (0, 0, -\frac{1}{2})$ points inward the feasible set from $\overline{x}$ and MFCQ holds. An example where both conditions fail is the constraint set given by $x^2 \leq 0$, at the feasible point $\overline{x} := 0$, where the gradient vanishes.

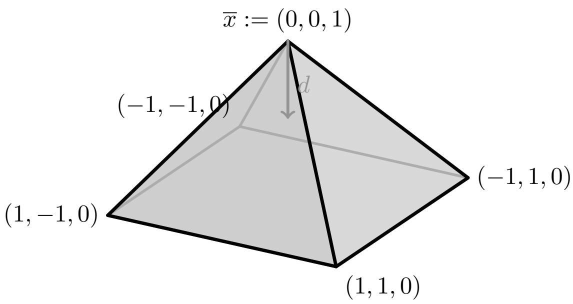


Fig. 1. *Pyramid with square base and apex at $\overline{x} := (0, 0, 1)$. Direction $d$ points inward the feasible set and MFCQ holds even though the gradients are not linearly independent.*

For standard nonlinear programming, MFCQ is one of the classical assumptions ensuring the existence of Lagrange multipliers at local minimizers. An elementary penalization-based proof of this fact can be found in [16]. The present paper may be viewed as a conic counterpart of that approach. In the conic setting, however, feasibility is no longer measured componentwise by scalar constraint violations; it is measured geometrically by the distance of the constraint function to the cone. This requires the use of polar cones, orthogonal projections, and Moreau's decomposition, which we introduce next. In the conic formulation considered in the next section, the corresponding regularity condition will be Robinson's condition.

**2. Optimization problems with cone constraints.** In this section, we consider an optimization problem with a cone constraint:

$$\begin{aligned} &\min_{x \in X} f(x), \\ &\text{s.t. } G(x) \in \mathcal{K}. \end{aligned} \tag{2.1}$$

Here $X$ and $Y$ are finite-dimensional real vector spaces, each endowed with an inner product and the associated norm. The mappings $f\colon X \to \mathbb{R}$ and $G\colon X \to Y$ are continuously differentiable, and $\mathcal{K} \subseteq Y$ is a nonempty closed convex cone.

We allow $X$ to be a general finite-dimensional vector space in order to include, for instance, problems with matrix variables. However, for a first reading, one may take $X = \mathbb{R}^n$ without losing the main ideas.

The formulation (2.1) should be understood as a standard and convenient language for writing different classes of constrained optimization problems. The mapping $G$ collects the constraints, while the set $\mathcal{K}$ describes the admissible values of these constraints. Thus, rather than treating equalities, inequalities, and matrix constraints separately, one can write a single geometric condition: $G(x) \in \mathcal{K}$. This point of view is useful because it allows several models to be treated within the same framework. For suitable choices of $G$ and $\mathcal{K}$, problem (2.1) includes standard nonlinear programming, second-order cone constrained problems, semidefinite programming, and other optimization problems with closed convex cone constraints. Optimization problems with cone constraints have a long history in nonlinear optimization, sensitivity analysis, and duality theory. General formulations in terms of set or cone constraints appear naturally in the classical works of Rockafellar [28, 29] and in the subsequent literature on optimization problems subject to cone constraints; see, for instance, Shapiro [31, 32] and Bonnans and Shapiro [8].

In what follows, we briefly recall some important examples of cones that fit this framework. These examples are included only to illustrate the range of models that can be written in the form (2.1).

The most classical example is the nonnegative orthant. When $Y := \mathbb{R}^p$, define

$$\mathbb{R}^p_+ := \{y \in \mathbb{R}^p \mid y_i \geq 0,\ i = 1, \ldots, p\}.$$

This cone allows us to recover standard nonlinear programming. Indeed, a system of inequality constraints $g_i(x) \leq 0$, for $i = 1, \ldots, p$, can be written as $G(x) := g(x) \in -\mathbb{R}^p_+$, where $g = (g_1, \ldots, g_p)$. A further important example is the second-order (Lorentz) cone

$$\mathcal{L}^p := \{(y_1, y_2, \ldots, y_p) \in \mathbb{R}^p \mid y_1 \geq \|(y_2, \ldots, y_p)\|\}.$$

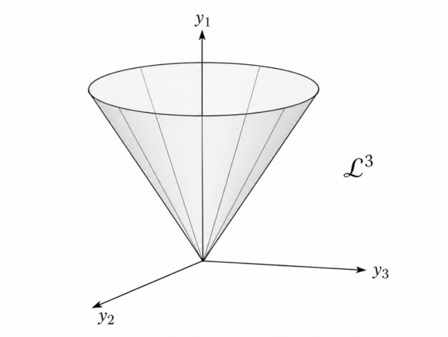


FIG. 2. *Geometric representation of the second-order cone* $\mathcal{L}^3 = \{(y_1, y_2, y_3) \ : \ y_1 \geq \sqrt{y_2^2 + y_3^2}\ \}$.

This cone gives rise to nonlinear second-order cone programming. See Figure 2. It appears naturally in problems involving Euclidean norm inequalities. For example, in facility location problems, one is interested in minimizing a sum of distances $\sum_{i=1}^{q} \|x - d_i\|$, where $x \in \mathbb{R}^n$ is the location of the facility sought, and $d_i, i = 1, \ldots, q$, are target points of interest. This can be formulated as minimizing $\sum_{i=1}^{q} t_i$ subject to second-order cone constraints $(t_i, x - d_i) \in \mathcal{L}^{n+1}, i = 1, \ldots, q$. See [22] for other interesting applications.

Another central case arises when $Y := \mathbb{S}^m$, the space of real $m \times m$ symmetric matrices. The positive semidefinite cone is defined by

$$\mathbb{S}^m_+ := \{A \in \mathbb{S}^m \mid u^\top A u \geq 0, \ \forall u \in \mathbb{R}^m\}.$$

The constraint $G(x) \in \mathbb{S}^m_+$ means that the symmetric matrix $G(x)$ must be positive semidefinite. This leads to nonlinear semidefinite programming. In contrast with scalar inequalities, semidefinite constraints impose infinitely many inequalities at once, since

$$G(x) \in \mathbb{S}^m_+ \quad \iff \quad u^\top G(x) u \geq 0, \quad \forall u \in \mathbb{R}^m.$$

Equivalently, relying on the spectral decomposition, the reader may verify that $G(x) \in \mathbb{S}^m_+$ if and only if all eigenvalues of $G(x)$ are nonnegative. This spectral characterization is often the most convenient way to visualize the cone $\mathbb{S}^m_+$, although treating the eigenvalue constraints explicitly can be difficult.

For $m = 2$, a positive semidefinite matrix $A$ has the form $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$, where, for all $u = (r, s) \in \mathbb{R}^2$, we have $0 \leq u^\top A u = ar^2 + 2brs + cs^2$. Taking $r = 0$, we get $c \geq 0$ while taking $s = 0$ we get $a \geq 0$. When $c = 0$, $b$ must also be zero since otherwise, for $r = 1$ we may choose $s$ to violate the above inequality. Completing the squares, we may write for $c > 0$:

$$0 \leq ar^2 + 2brs + cs^2 = c\left(s + \frac{b}{c} r\right)^2 + \left(a - \frac{b^2}{c}\right) r^2,$$

where the least value of the right-hand side expression corresponds to $s = -\frac{b}{c} r$. This implies that $a - \frac{b^2}{c} \geq 0$. A similar argument shows the reverse implication, that is, $A \in \mathbb{S}^2_+$ if, and only if

$$a \geq 0, \qquad c \geq 0, \qquad ac - b^2 \geq 0.$$

Thus, the cone $\mathbb{S}^2_+$ can be visualized in the three-dimensional space of coordinates $(a, b, c)$. The inequality $ac \geq b^2$ describes a curved cone, showing already in dimension two that semidefinite constraints are substantially different from componentwise scalar inequalities. See Figure 3.

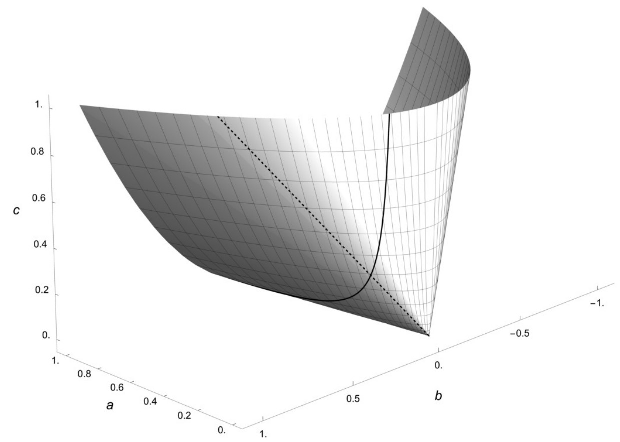


FIG. 3. *Boundary of the semidefinite cone. The set includes the half-lines* $(a, 0, 0), a \geq 0$ *and* $(0, 0, c), c \geq 0$. *One can see the top parabola* $a = b^2$ *at* $c = 1$, *the back parabola* $c = b^2$ *at* $a = 1$ *and the solid line depicts the hyperbola* $ac = b^2$ *at* $b = 0.2$. *The dashed line depicts the one-dimensional face of non-negative multiples of the rank-one matrix corresponding to* $(a, b, c) := (1, \sqrt{2}, 2)$.

For background on second-order cone programming and semidefinite programming, we refer the reader to Alizadeh and Goldfarb [1], Vandenberghe and Boyd [33, 34], and Shapiro [32].

Equality constraints can also be included in the conic form by taking Cartesian products with the zero cone. For instance, problem (1.1) can be written in the form (2.1) by setting

$$Y := \mathbb{R}^p \times \mathbb{R}^q, \qquad \mathcal{K} := -\mathbb{R}^p_+ \times \{0\}^q,$$

and $G(x) := (g(x), h(x))$. Then $G(x) \in \mathcal{K}$ is exactly equivalent to

$$g(x) \in -\mathbb{R}^p_+, \qquad h(x) = 0.$$

This example shows that the formulation (2.1) includes standard nonlinear programming as a particular case.

We now present a simple worked example that illustrates how a familiar matrix problem can be written using a semidefinite constraint.

*Example* 2.1. Let $A \in \mathbb{S}^m$. The largest eigenvalue of $A$ can be characterized by the Rayleigh quotient:

$$\lambda_{\max}(A) = \max_{u \neq 0} \frac{u^\top A u}{u^\top u}.$$

We claim that $\lambda_{\max}(A)$ is the optimal value of the semidefinite program

$$\min_{x \in \mathbb{R}} x \quad \text{subject to} \quad xI - A \in \mathbb{S}^m_+,$$

where $I$ denotes the $m \times m$ identity matrix. Indeed, by the definition of positive semidefiniteness,

$$xI - A \in \mathbb{S}^m_+ \quad \Longleftrightarrow \quad u^\top A u \le x\, u^\top u \quad \forall u \in \mathbb{R}^m.$$

Therefore $x$ is feasible if, and only if, $x \ge \lambda_{\max}(A)$. Minimizing $x$ corresponds to $x = \lambda_{\max}(A)$.

The previous example is pedagogically useful because it shows how a spectral condition can be represented by membership in the semidefinite cone. Similarly, the reader may verify that the smallest eigenvalue of $A$ can be computed through the semidefinite program

$$\max_{x \in \mathbb{R}} x \quad \text{subject to} \quad A - xI \in \mathbb{S}^m_+.$$

For other applications of semidefinite programming and more general cones we refer the reader to [12, 34].

**3. Projections.** The proof of the Lagrange multiplier theorem developed in this paper follows a classical projection-penalization strategy; see, for instance, [11, 13, 21, 35]. We measure infeasibility by the distance to the cone and then use this distance as a penalty term. This leads naturally to orthogonal projections. Our goal in this section is to collect, in a self-contained way, the elementary projection facts needed later in the penalization argument.

We recall that $Y$ denotes a finite-dimensional space and we fix on it an inner product $\langle \cdot, \cdot \rangle$ and the associated norm $\| \cdot \|$. All projections below are orthogonal projections with respect to this Euclidean structure. Let $\mathcal{C} \subseteq Y$ be a nonempty closed convex set. For $z \in Y$, consider the problem of finding a point of $\mathcal{C}$ closest to $z$:

$$\min_{c \in \mathcal{C}} \|z - c\|.$$

The next lemma shows the well-known fact that this problem has a unique solution, which is called the (orthogonal) projection of $z$ onto $\mathcal{C}$ and is denoted by $\Pi_{\mathcal{C}}(z)$. The lemma also shows a useful characterization of projections, which says that the projection $\Pi_{\mathcal{C}}(z)$ corresponds to a point $w$ where the direction from $w$ to $z$ forms an obtuse or right angle with every direction from $w$ to points $c \in \mathcal{C}$. See Figure 4.

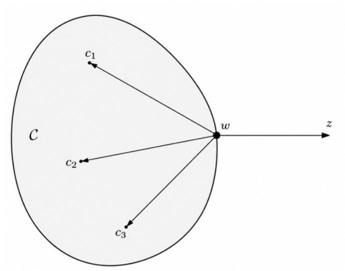


FIG. 4. *Characterization of the projection $w = \Pi_{\mathcal{C}}(z)$ as the point in which the directions $c - w$ for all $c \in \mathcal{C}$ form an angle $\ge 90°$ with $z - w$.*

Lemma 3.1. *[7, Proposition B.11] Let $\mathcal{C} \subseteq Y$ be a nonempty closed convex set and let $z \in Y$. Then the projection $\Pi_{\mathcal{C}}(z)$ is well-defined and unique. Moreover, a point $w \in \mathcal{C}$ satisfies $w = \Pi_{\mathcal{C}}(z)$ if and only if*

$$\langle z - w, c - w\rangle \leq 0 \qquad \forall c \in \mathcal{C}.$$

*In addition, the projection mapping is nonexpansive, that is,*

$$\|\Pi_{\mathcal{C}}(z_1) - \Pi_{\mathcal{C}}(z_2)\| \leq \|z_1 - z_2\| \qquad \forall z_1, z_2 \in Y.$$

*Proof.* We prove the three assertions separately.
*Existence and uniqueness.* Fix $\hat{c} \in \mathcal{C}$. Minimizing the continuous function $c \mapsto \|z-c\|$ over $\mathcal{C}$ is equivalent to minimizing it over the set

$$\mathcal{C} \cap \{c \in Y : \|z - c\| \leq \|z - \hat{c}\|\}.$$

This set is nonempty, closed, and bounded. Since $Y$ is finite dimensional, it is compact. Hence existence follows from Weierstrass' theorem.

To prove uniqueness, suppose that $c_1, c_2 \in \mathcal{C}$ are two distinct minimizers. By expanding $\frac{1}{4}\|(z - c_1) + (z - c_2)\|^2 + \frac{1}{4}\|(z - c_1) - (z - c_2)\|^2$, or, alternatively, using the parallelogram identity, we arrive at

$$\left\|z - \frac{c_1 + c_2}{2}\right\|^2 + \frac{1}{4}\|c_1 - c_2\|^2 = \frac{1}{2}\|z - c_1\|^2 + \frac{1}{2}\|z - c_2\|^2.$$

Since $c_1 \neq c_2$ and $\|z - c_1\|^2 = \|z - c_2\|^2$ by definition, it follows that

$$\left\|z - \frac{c_1 + c_2}{2}\right\|^2 < \|z - c_1\|^2.$$

By convexity, $(c_1 + c_2)/2 \in \mathcal{C}$, which contradicts the minimality of $c_1$. Thus the projection is unique.
*Characterization.* Let $w \in \mathcal{C}$ satisfy $\langle z - w, c - w\rangle \leq 0$, for all $c \in \mathcal{C}$. Then, for every $c \in \mathcal{C}$,

$$\begin{aligned}\|z - c\|^2 &= \|(z - w) - (c - w)\|^2 \\ &= \|z - w\|^2 + \|c - w\|^2 - 2\langle z - w, c - w\rangle \\ &\geq \|z - w\|^2.\end{aligned}$$

Therefore $w = \Pi_{\mathcal{C}}(z)$.

Conversely, assume that $w = \Pi_{\mathcal{C}}(z)$. Fix $c \in \mathcal{C}$. Since $\mathcal{C}$ is convex, the point $w + t(c - w)$ belongs to $\mathcal{C}$ for every $t \in [0, 1]$. Thus

$$\varphi(t) := \|z - (w + t(c - w))\|^2 = \|z - w\|^2 - 2t\langle z - w, c - w\rangle + t^2\|c - w\|^2$$

has a minimum at $t = 0$ over $[0, 1]$. Hence its right derivative at zero is nonnegative:

$$\varphi'(0) = -2\langle z - w, c - w\rangle \geq 0.$$

Thus $\langle z - w, c - w\rangle \leq 0$.
*Nonexpansiveness.* Set $p_i := \Pi_{\mathcal{C}}(z_i)$, for $i = 1, 2$. Using the characterization with $z = z_1$, $w = p_1$, and $c = p_2$, we obtain $\langle z_1 - p_1, p_2 - p_1\rangle \leq 0$. Using it again with

$z = z_2$, $w = p_2$, and $c = p_1$, we obtain $\langle z_2 - p_2, p_1 - p_2\rangle \le 0$. Adding these inequalities gives $\langle -z_1 + p_1 + z_2 - p_2, p_1 - p_2\rangle \le 0$, that is

$$\|p_1 - p_2\|^2 \le \langle z_1 - z_2, p_1 - p_2\rangle.$$

By the Cauchy–Schwarz inequality,

$$\|p_1 - p_2\|^2 \le \|z_1 - z_2\| \, \|p_1 - p_2\|.$$

If $p_1 = p_2$, the desired estimate is immediate. Otherwise, dividing by $\|p_1 - p_2\|$ yields

$$\|p_1 - p_2\| \le \|z_1 - z_2\|,$$

which completes the proof. ∎

The nonexpansiveness property implies, in particular, that projections onto closed convex sets are continuous. In the particular case where the nonempty closed convex set is a cone $\mathcal{K}$, that is, it contains all directions $tu$ whenever $u \in \mathcal{K}$ and $t \ge 0$, additional structure is gained. Specifically, every element $z \in Y$ can be written as the sum of its projections onto the cone $\mathcal{K}$ and onto the so-called polar cone of $\mathcal{K}$, namely

$$z = \Pi_{\mathcal{K}}(z) + \Pi_{\mathcal{K}^\circ}(z),$$

where the polar cone of $\mathcal{K}$ is defined as

$$\mathcal{K}^\circ := \{u \in Y : \langle u, d\rangle \le 0, \ \forall d \in \mathcal{K}\}.$$

The proof of this fact, known as Moreau's decomposition [25], will be given below in Lemma 3.4, item (i), where we also show that $\Pi_{\mathcal{K}}(z) \perp \Pi_{\mathcal{K}^\circ}(z)$.

The polar cone $\mathcal{K}^\circ$ consists of all vectors that make an angle of at least 90° with every vector in $\mathcal{K}$. Since it is an intersection of closed halfspaces containing the origin, $\mathcal{K}^\circ$ is itself a nonempty closed convex cone. The polar cone is sometimes called the negative dual cone; see, for instance, [28, 8]. See Figure 5. The fact that $\mathcal{K}^\circ := -\mathcal{K}$ when $\mathcal{K} := \mathcal{L}^3$ can be seen from the picture, and is left as an exercise. We give a proof of this fact for $\mathcal{K} := \mathbb{S}^m_+$ in the sequel.

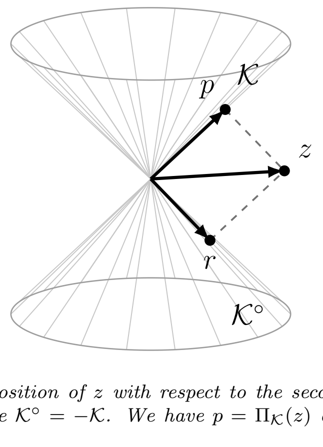


FIG. 5. *Moreau's decomposition of $z$ with respect to the second-order cone $\mathcal{K} := \mathcal{L}^3$ and the Euclidean inner product, where $\mathcal{K}^\circ = -\mathcal{K}$. We have $p = \Pi_{\mathcal{K}}(z)$ and $r = \Pi_{\mathcal{K}^\circ}(z)$ with $p \perp r$ and $z = p + r$.*

*Example* 3.2. Let $Y = \mathbb{S}^m$ endowed with the trace inner product

$$\langle A, B\rangle := \text{trace}(AB) = \sum_{i,j=1}^{m} A_{ij}B_{ij}.$$

Let $\mathcal{K} := \mathbb{S}^m_+$ and let us show that $(\mathbb{S}^m_+)^\circ = -\mathbb{S}^m_+$, the set of negative semidefinite matrices. For any $A \in \mathbb{S}^m_+$ and $U \in -\mathbb{S}^m_+$, we have $\langle U, A\rangle = \text{trace}(UA)$. Since the eigenvalues of $A$ are non-negative, we may define a matrix $A^{\frac{1}{2}} \in \mathbb{S}^m$ such that $A = A^{\frac{1}{2}}A^{\frac{1}{2}}$. Therefore, $\text{trace}(UA) = \text{trace}(UA^{\frac{1}{2}}A^{\frac{1}{2}}) = \text{trace}(A^{\frac{1}{2}}UA^{\frac{1}{2}})$. However, since $U$ is negative semidefinite, the same is true for $A^{\frac{1}{2}}UA^{\frac{1}{2}}$, therefore all its eigenvalues are non-positive. This implies that $\text{trace}(A^{\frac{1}{2}}UA^{\frac{1}{2}}) \leq 0$, hence $-\mathbb{S}^m_+ \subseteq (\mathbb{S}^m_+)^\circ$. For the converse inclusion, if $U \notin -\mathbb{S}^m_+$, then there exists $v \in \mathbb{R}^m$ such that $0 < v^\top Uv = \langle U, vv^\top\rangle$. Taking $A = vv^\top \in \mathbb{S}^m_+$ shows that $U \notin (\mathbb{S}^m_+)^\circ$, which completes the proof.

It may be useful to keep the following elementary example in mind.

*Example* 3.3. Let $\mathcal{K} := -\mathbb{R}^2_+ \times \{0\} \subseteq \mathbb{R}^3$. A vector $(x, y, z) \in \mathbb{R}^3$ belongs to $\mathcal{K}^\circ$ if, and only if, $xa + yb + zc \leq 0$ for all $(a, b, c) \in \mathcal{K}$, that is, $a \leq 0, b \leq 0, c = 0$. Therefore, we must have $x \geq 0, y \geq 0$ and $z \in \mathbb{R}$, which implies that $\mathcal{K}^\circ = \mathbb{R}^2_+ \times \mathbb{R}$. For $v = (x, y, z)$, the projections are given by

$$\Pi_{\mathcal{K}}(v) = \big(\min\{0, x\}, \min\{0, y\}, 0\big) \quad \text{and} \quad \Pi_{\mathcal{K}^\circ}(v) = \big(\max\{0, x\}, \max\{0, y\}, z\big).$$

The reader may verify directly that these two vectors are orthogonal and add up to $v$. Moreover,

$$\|\Pi_{\mathcal{K}^\circ}(v)\| = \|v - \Pi_{\mathcal{K}}(v)\| = \text{dist}(v, \mathcal{K}),$$

which implies that $v \mapsto \text{dist}(v, \mathcal{K})^2 = \max\{0, x\}^2 + \max\{0, y\}^2 + z^2$ is continuously differentiable with gradient vector equal to $2\Pi_{\mathcal{K}^\circ}(v)$. We will prove this fact for a general cone $\mathcal{K}$ in Lemma 3.4, item (ii) and (iii). See Figure 6.

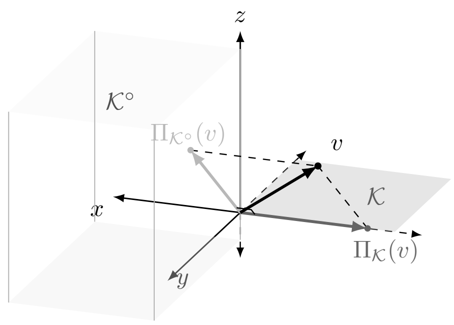


FIG. 6. *Given $\mathcal{K} := -\mathbb{R}^2_+ \times \{0\}$, $\mathcal{K}^\circ = \mathbb{R}^2_+ \times \mathbb{R}$. For $v := (-3, 2, 3)$, one has $\Pi_{\mathcal{K}}(v) = (-3, 0, 0)$ and $\Pi_{\mathcal{K}^\circ}(v) = (0, 2, 3)$. These vectors are orthogonal and add up to $v$. The distance from $v$ to $\mathcal{K}$ is equal to $\|\Pi_{\mathcal{K}^\circ}(v)\|$.*

The next lemma collects the projection identities that will be used in the penalized problem. The first item corresponds to Moreau's decomposition [25]. The differentiability statement is a consequence of results in [14]. We include the proof in order to keep the presentation self-contained.

LEMMA 3.4. *Let $\mathcal{K} \subseteq Y$ be a nonempty closed convex cone and let $z \in Y$. Then the following statements hold:*

(i) Moreau's decomposition: $z = \Pi_{\mathcal{K}}(z) + \Pi_{\mathcal{K}^\circ}(z)$ *and* $\langle \Pi_{\mathcal{K}}(z), \Pi_{\mathcal{K}^\circ}(z)\rangle = 0$.
(ii) Distance characterization: $\|\Pi_{\mathcal{K}^\circ}(z)\| = \operatorname{dist}(z, \mathcal{K}) := \inf_{u\in\mathcal{K}} \|z - u\|$.
(iii) Differentiability of the squared distance: *the function* $z \mapsto \operatorname{dist}(z, \mathcal{K})^2 = \|\Pi_{\mathcal{K}^\circ}(z)\|^2$ *is continuously differentiable and*

$$\nabla\|\Pi_{\mathcal{K}^\circ}(z)\|^2 = 2\Pi_{\mathcal{K}^\circ}(z).$$

*Proof.* We prove the three statements separately.

*Moreau's decomposition and orthogonality.* Set $p := \Pi_{\mathcal{K}}(z)$ and $r := z - p$. We first show that $r \in \mathcal{K}^\circ$ and that $p$ and $r$ are orthogonal. By Lemma 3.1,

$$\langle z - p, c - p\rangle \leq 0 \qquad \forall c \in \mathcal{K}.$$

Since $\mathcal{K}$ is a cone, we may take $c = tp$ for any $t \geq 0$. Hence $(t - 1)\langle r, p\rangle \leq 0$ for all $t \geq 0$. Taking $t = 0.5$ and $t = 1.5$, we obtain $\langle r, p\rangle = 0$. Now let $c \in \mathcal{K}$. Since $p + c \in \mathcal{K}$, the projection characterization gives

$$\langle r, c\rangle = \langle z - p, (p + c) - p\rangle \leq 0.$$

Therefore $r \in \mathcal{K}^\circ$.

It remains to show that $r$ is the projection of $z$ onto $\mathcal{K}^\circ$. Let $q \in \mathcal{K}^\circ$. Since $z - r = p$, we have

$$\langle z - r, q - r\rangle = \langle p, q - r\rangle = \langle p, q\rangle - \langle p, r\rangle.$$

Here $\langle p, q\rangle \leq 0$, because $p \in \mathcal{K}$ and $q \in \mathcal{K}^\circ$, while $\langle p, r\rangle = \langle r, p\rangle = 0$. Thus

$$\langle z - r, q - r\rangle \leq 0 \qquad \forall q \in \mathcal{K}^\circ.$$

By Lemma 3.1, this proves that $r = \Pi_{\mathcal{K}^\circ}(z)$. The result follows from $z = p + r$ with $\langle p, r\rangle = 0$.

*Distance characterization.* By the definition of the projection onto $\mathcal{K}$,

$$\operatorname{dist}(z, \mathcal{K}) = \|z - \Pi_{\mathcal{K}}(z)\|,$$

and the result follows from item (i).

*Differentiability of the squared distance.* Define

$$z \mapsto d_{\mathcal{K}}(z)^2 := \operatorname{dist}(z, \mathcal{K})^2 = \|z - \Pi_{\mathcal{K}}(z)\|^2.$$

Let $h \in Y$. We compare $d_{\mathcal{K}}(z+h)^2$ with $d_{\mathcal{K}}(z)^2$. Since $\Pi_{\mathcal{K}}(z+h) \in \mathcal{K}$, the minimality of $\Pi_{\mathcal{K}}(z)$ gives

$$\|z - \Pi_{\mathcal{K}}(z)\| \leq \|z - \Pi_{\mathcal{K}}(z + h)\|.$$

Similarly, since $\Pi_{\mathcal{K}}(z) \in \mathcal{K}$, the minimality of $\Pi_{\mathcal{K}}(z + h)$ gives

$$\|z + h - \Pi_{\mathcal{K}}(z + h)\| \leq \|z + h - \Pi_{\mathcal{K}}(z)\|.$$

After squaring and reversing the sign of the first inequality, by adding $d_{\mathcal{K}}(z + h)^2$ we get

$$\begin{aligned} d_{\mathcal{K}}(z + h)^2 - d_{\mathcal{K}}(z)^2 &\geq \|z + h - \Pi_{\mathcal{K}}(z + h)\|^2 - \|z - \Pi_{\mathcal{K}}(z + h)\|^2 \\ &= 2\langle z - \Pi_{\mathcal{K}}(z + h), h\rangle + \|h\|^2. \end{aligned}$$

Similarly, by squaring and subtracting $d_{\mathcal{K}}(z)^2$ to the second inequality, we get

$$\begin{aligned} d_{\mathcal{K}}(z+h)^2 - d_{\mathcal{K}}(z)^2 &\leq \|z+h-\Pi_{\mathcal{K}}(z)\|^2 - \|z-\Pi_{\mathcal{K}}(z)\|^2 \\ &= 2\langle z-\Pi_{\mathcal{K}}(z), h\rangle + \|h\|^2. \end{aligned}$$

Subtracting $2\langle z-\Pi_{\mathcal{K}}(z), h\rangle$ and combining both inequalities yields

$$\begin{aligned} &2\langle \Pi_{\mathcal{K}}(z) - \Pi_{\mathcal{K}}(z+h), h\rangle + \|h\|^2 \\ &\qquad \leq d_{\mathcal{K}}(z+h)^2 - d_{\mathcal{K}}(z)^2 - 2\langle z-\Pi_{\mathcal{K}}(z), h\rangle \leq \|h\|^2. \end{aligned}$$

By Cauchy-Schwarz and nonexpansiveness of the projection (Lemma 3.1),

$$\left|\langle \Pi_{\mathcal{K}}(z+h) - \Pi_{\mathcal{K}}(z), h\rangle\right| \leq \|\Pi_{\mathcal{K}}(z+h) - \Pi_{\mathcal{K}}(z)\| \, \|h\| \leq \|h\|^2.$$

Therefore, $-2\|h\|^2 \leq 2\langle \Pi_{\mathcal{K}}(z) - \Pi_{\mathcal{K}}(z+h), h\rangle$, which implies

$$d_{\mathcal{K}}(z+h)^2 = d_{\mathcal{K}}(z)^2 + 2\langle z-\Pi_{\mathcal{K}}(z), h\rangle + o(\|h\|),$$

where $o(\|h\|)/\|h\| \to 0$. Therefore, $d_{\mathcal{K}}(z)^2$ is differentiable at all points $z \in Y$ and the derivative acts on $h \in Y$ (by item (i)) as $\langle 2\Pi_{\mathcal{K}^\circ}(z), h\rangle$. Hence, the gradient of $d_{\mathcal{K}}^2(z)$ is given by $\nabla d_{\mathcal{K}}(z)^2 = 2\Pi_{\mathcal{K}^\circ}(z)$. The continuity of the gradient follows from the continuity of the projection onto $\mathcal{K}^\circ$. ∎

In the next section, we turn our attention to the notion of Lagrange multipliers, the central object of this work.

**4. Lagrange Multipliers.** We now introduce the KKT conditions associated with problem (2.1). Let us first recall how the classical inequality-constrained case fits this notation. Consider, for simplicity, the problem with only inequality constraints $g(x) \in -\mathbb{R}^p_+$, where $g\colon \mathbb{R}^n \to \mathbb{R}^p$ is given by $g = (g_1, \ldots, g_p)$. The derivative $Dg(\overline{x})$ is the $p \times n$ Jacobian matrix whose $i$-th row is $\nabla g_i(\overline{x})^\top$. Thus, for a multiplier vector $\mu = (\mu_1, \ldots, \mu_p)$, the stationarity condition in Theorem 1.1 can be written compactly as $\nabla f(\overline{x}) + Dg(\overline{x})^\top \mu = 0$. The dual feasibility condition $\mu \in \mathbb{R}^p_+$ can be written as

$$\mu \in (-\mathbb{R}^p_+)^\circ = \mathbb{R}^p_+,$$

and complementarity is equivalent to $\langle g(\overline{x}), \mu\rangle = 0$. This notation suggests the natural KKT system for the cone-constrained problem (2.1).

Definition 4.1. *A feasible point $\overline{x} \in X$ for problem* (2.1) *is called a* KKT point *if there exists a vector $\mu \in Y$, called a* Lagrange multiplier*, such that*

$$\nabla f(\overline{x}) + DG(\overline{x})^\top \mu = 0, \tag{4.1}$$

$$G(\overline{x}) \in \mathcal{K}, \tag{4.2}$$

$$\mu \in \mathcal{K}^\circ, \tag{4.3}$$

$$\langle G(\overline{x}), \mu\rangle = 0. \tag{4.4}$$

Here $DG(\overline{x})\colon X \to Y$ denotes the derivative of $G$ at $\overline{x}$. Its adjoint $DG(\overline{x})^\top : Y \to X$ is the unique linear operator satisfying

$$\langle DG(\overline{x})d, \mu\rangle = \langle d, DG(\overline{x})^\top \mu\rangle \qquad \forall d \in X,\ \forall \mu \in Y.$$

The gradient $\nabla f(\overline{x}) \in X$ is defined through the inner product on $X$ by

$$Df(\overline{x})d = \langle \nabla f(\overline{x}), d\rangle \qquad \forall d \in X.$$

Conditions (4.1)–(4.4) are the direct analogue of the classical KKT conditions. The polar cone $\mathcal{K}^\circ$ replaces the usual nonnegativity condition on the multipliers, and the identity $\langle G(\overline{x}), \mu\rangle = 0$ is the conic form of complementarity.

To illustrate the KKT conditions for semidefinite constraints, let $Y = \mathbb{S}^m$ be endowed with the trace inner product $\langle A, B\rangle := \operatorname{trace}(AB)$. Let $X = \mathbb{R}^n$ and let $G\colon \mathbb{R}^n \to \mathbb{S}^m$ be continuously differentiable. For each $x \in \mathbb{R}^n$, write

$$G(x) = \big(G_{ij}(x)\big)_{i,j=1}^m,$$

where each $G_{ij}\colon \mathbb{R}^n \to \mathbb{R}$ is a scalar component function and $G_{ij} = G_{ji}$. The derivative $DG(\overline{x})\colon \mathbb{R}^n \to \mathbb{S}^m$ is the linear map given by

$$DG(\overline{x})d = \sum_{\ell=1}^n \frac{\partial G(\overline{x})}{\partial x_\ell} d_\ell, \qquad d = (d_1, \dots, d_n) \in \mathbb{R}^n,$$

where

$$\frac{\partial G(\overline{x})}{\partial x_\ell} = \left(\frac{\partial G_{ij}(\overline{x})}{\partial x_\ell}\right)_{i,j=1}^m \in \mathbb{S}^m.$$

This is the matrix-valued analogue of multiplying a Jacobian matrix by a direction $d$, resulting in a component-wise derivative of each entry of $G$. The adjoint operator $DG(\overline{x})^\top\colon \mathbb{S}^m \to \mathbb{R}^n$ is defined by

$$\langle DG(\overline{x})d, \mu\rangle = \langle d, DG(\overline{x})^\top \mu\rangle \qquad \forall d \in \mathbb{R}^n,\ \forall \mu \in \mathbb{S}^m,$$

where at the left-hand side expression we use the trace inner-product and at the right-hand side expression we use the standard inner-product in $\mathbb{R}^n$. Thus, one obtains

$$\langle DG(\overline{x})d, \mu\rangle = \left\langle \sum_{\ell=1}^n \frac{\partial G(\overline{x})}{\partial x_\ell} d_\ell, \mu \right\rangle = \sum_{\ell=1}^n d_\ell \left\langle \frac{\partial G(\overline{x})}{\partial x_\ell}, \mu \right\rangle.$$

Therefore,

$$DG(\overline{x})^\top \mu = \left(\left\langle \frac{\partial G(\overline{x})}{\partial x_1}, \mu \right\rangle, \dots, \left\langle \frac{\partial G(\overline{x})}{\partial x_n}, \mu \right\rangle\right).$$

Equivalently,

$$DG(\overline{x})^\top \mu = \sum_{i,j=1}^m \mu_{ij} \nabla G_{ij}(\overline{x}),$$

where

$$\nabla G_{ij}(\overline{x}) = \left(\frac{\partial G_{ij}(\overline{x})}{\partial x_1}, \dots, \frac{\partial G_{ij}(\overline{x})}{\partial x_n}\right) \in \mathbb{R}^n$$

is the usual gradient of the scalar entry function $G_{ij}\colon \mathbb{R}^n \to \mathbb{R}$. We are now in a position to analyze the following illustrative example.

*Example* 4.2. Consider the semidefinite optimization problem

$$\begin{aligned} &\min_{x\in\mathbb{R}} 2x, \\ &\text{s.t. } G(x) := \begin{bmatrix} 0 & x \\ x & 1 \end{bmatrix} \in \mathcal{K} := \mathbb{S}^2_+. \end{aligned}$$

Let us first describe the feasible set. By the discussion concerning Figure 3, we must have $-x^2 \geq 0$. Thus the feasible set is the singleton $\{0\}$, and $\overline{x} = 0$ is the unique global minimizer. We now show that $\overline{x}$ does not satisfy the KKT conditions. Since

$$DG(\overline{x})d = \begin{bmatrix} 0 & d \\ d & 0 \end{bmatrix}, \qquad d \in \mathbb{R},$$

the adjoint $DG(\overline{x})^\top : \mathbb{S}^2 \to \mathbb{R}$ is given by

$$DG(\overline{x})^\top \mu = \left\langle \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \mu \right\rangle = 2\mu_{12}, \text{ where } \mu = \begin{bmatrix} \mu_{11} & \mu_{12} \\ \mu_{12} & \mu_{22} \end{bmatrix}.$$

Stationarity gives

$$\nabla f(\overline{x}) + DG(\overline{x})^\top \mu = 2 + 2\mu_{12} = 0,$$

hence $\mu_{12} = -1$. Complementarity gives

$$\langle G(\overline{x}), \mu \rangle = \left\langle \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \mu \right\rangle = \mu_{22} = 0.$$

Thus any Lagrange multiplier would necessarily have the form

$$\mu = \begin{bmatrix} \mu_{11} & -1 \\ -1 & 0 \end{bmatrix}.$$

Since $(\mathbb{S}^2_+)^\circ = -\mathbb{S}^2_+$, a Lagrange multiplier would have to satisfy $\mu \in -\mathbb{S}^2_+$. But this is not the case as this is always an indefinite matrix regardless of the value of $\mu_{11}$. In particular, $\mu \notin -\mathbb{S}^2_+$. Hence no matrix $\mu$ satisfies the KKT conditions at $\overline{x}$.

The previous example is noteworthy because it exhibits a solution to the problem that fails to satisfy the KKT conditions. This phenomenon is not unexpected, as we have seen in the previous section: in order for the KKT conditions to serve as genuine optimality conditions—meaning that every solution of problem (2.1) must satisfy them—an additional assumption on the constraints, known as a *constraint qualification*, is required. In the example, the linearized feasible set

$$G(\overline{x}) + DG(\overline{x})d = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + d \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \in \mathbb{S}^2_+$$

does not contain any interior point, that is, a positive definite matrix. The existence of such interior directions corresponds to the generalization of Definition 1.2 to semidefinite programming. In the more general conic setting, however, when $\mathcal{K}$ may have empty interior (for instance, if equality constraints are present), the generalization of Definition 1.2 takes a different form. In this work we shall adopt the following geometric formulation of Definition 1.2, known as *Robinson's condition* [26].

*Assumption* 4.3. At a point $\overline{x}$ such that $G(\overline{x}) \in \mathcal{K}$, one has

$$0 \in \operatorname{int}\big(DG(\overline{x})X + \mathcal{K} - G(\overline{x})\big),$$

where $DG(\overline{x})X$ denotes the image space of $DG(\overline{x})$, and $\operatorname{int}(\cdot)$ the topological interior of the set.

Notice that when $\mathcal{K} := \{0\}$, Assumption 4.3 reduces to the surjectivity of $DG(\overline{x})$, which is equivalent to the classical linear independence constraint qualification when $Y = \mathbb{R}^q$, whereas when $\mathcal{K}$ has non-empty interior, the condition reduces to $G(\overline{x}) + DG(\overline{x})d \in \operatorname{int}(\mathcal{K})$ similarly to Definition 1.2. The reader may prove as an exercise that Assumption 4.3 is equivalent to Definition 1.2 when $\mathcal{K} := -\mathbb{R}^p_+ \times \{0\}^q \subset \mathbb{R}^p \times \mathbb{R}^q$. In general, Robinson's condition expresses the idea that if the point $G(\overline{x}) \in \mathcal{K}$ is slightly perturbed, it is still possible to approximately restore feasibility in a neighborhood of $\overline{x}$. More precisely, for any sufficiently small perturbation $\varepsilon \in Y$, there exists some $d \in X$ such that

$$G(\overline{x}) + DG(\overline{x})d + \varepsilon \in \mathcal{K} \quad \text{where} \quad G(\overline{x} + d) \approx G(\overline{x}) + DG(\overline{x})d.$$

Equivalently, Assumption 4.3 can be interpreted as the *metric regularity* of the set-valued mapping $-G(\cdot) + \mathcal{K}$ at $(\overline{x}, 0)$, a classical concept in variational analysis [17].

*Example* 4.4. Consider the nonlinear cone-constrained problem

$$\min_{x \in \mathbb{R}} x \qquad \text{subject to} \qquad G(x) := (1 + x, x^2) \in \mathcal{L}^2.$$

The feasibility condition is $1 + x \geq |x^2|$, which is satisfied if, and only if $(1 - \sqrt{5})/2 \leq x \leq (1 + \sqrt{5})/2$. The unique global minimizer is $\overline{x} := (1 - \sqrt{5})/2$, which lies on the boundary of $\mathcal{L}^2$. Let us verify Robinson's condition. Since $DG(\overline{x})d = (d, 2\overline{x}d)$,

$$G(\overline{x}) + DG(\overline{x})d = (1 + \overline{x} + d, \overline{x}^2 + 2\overline{x}d).$$

Since $1 + \overline{x} = \overline{x}^2 > 0$, taking $d > 0$ sufficiently small with $2\overline{x}d < 0$ we have that $1 + \overline{x} + d > |\overline{x}^2 + 2\overline{x}d|$, that is, $G(\overline{x}) + DG(\overline{x})d \in \operatorname{int}(\mathcal{L}^2)$. We now compute a Lagrange multiplier. Since $(\mathcal{L}^2)^\circ = -\mathcal{L}^2$, a Lagrange multiplier $\mu = (\mu_1, \mu_2)$ must satisfy $\mu \in -\mathcal{L}^2$, that is, $\mu_1 \leq -|\mu_2|$. The complementarity condition gives

$$0 = \langle G(\overline{x}), \mu \rangle = (1 + \overline{x})\mu_1 + \overline{x}^2\mu_2 = \overline{x}^2(\mu_1 + \mu_2),$$

and hence $\mu_1 = -\mu_2$. The stationarity condition reduces to $1 + \mu_1 + 2\overline{x}\mu_2 = 0$. Using $\mu_1 = -\mu_2$ and $-1 + 2\overline{x} = -\sqrt{5}$, we obtain $\mu_1 = -\frac{1}{\sqrt{5}}$, $\mu_2 = \frac{1}{\sqrt{5}}$. Hence $\mu \in -\mathcal{L}^2$ and the KKT conditions hold at $\overline{x}$.

The next proposition shows an algebraic equivalent notion of Robinson's condition, which is very useful in order to check its validity.

PROPOSITION 4.5. *Robinson's condition holds at $\overline{x}$, with $G(\overline{x}) \in \mathcal{K}$, if and only if the following conic linear independence condition holds:*

$$DG(\overline{x})^\top \alpha = 0, \quad \langle G(\overline{x}), \alpha \rangle = 0, \quad \alpha \in \mathcal{K}^\circ \quad \Rightarrow \quad \alpha = 0. \tag{4.5}$$

*Proof.* Suppose first that Assumption 4.3 holds and that $\alpha$ satisfies the three relations in (4.5). Since 0 is an interior point of $DG(\overline{x})X + \mathcal{K} - G(\overline{x})$, for every sufficiently small $t > 0$ we have $t\alpha \in DG(\overline{x})X + \mathcal{K} - G(\overline{x})$. Hence there exist $d \in X$ and $w \in \mathcal{K}$ such that $t\alpha = DG(\overline{x})d + w - G(\overline{x})$. Taking the inner product with $\alpha$ gives

$$t\|\alpha\|^2 = \langle DG(\overline{x})d, \alpha \rangle + \langle w, \alpha \rangle - \langle G(\overline{x}), \alpha \rangle.$$

The first term on the right is zero because

$$\langle DG(\overline{x})d, \alpha \rangle = \langle d, DG(\overline{x})^\top \alpha \rangle$$

and $DG(\overline{x})^\top \alpha = 0$. The third term is zero by assumption, and the second term is nonpositive because $w \in \mathcal{K}$ and $\alpha \in \mathcal{K}^\circ$. Thus $t\|\alpha\|^2 \leq 0$, which implies $\alpha = 0$. Conversely, suppose that Robinson's condition fails. Since $DG(\overline{x})X + \mathcal{K} - G(\overline{x})$ is convex and contains the origin, but not in its interior, the convex separation theorem yields a nonzero vector $\alpha \in Y$ such that

$$\langle \alpha, DG(\overline{x})d + w - G(\overline{x})\rangle \leq 0, \qquad \forall d \in X,\ \forall w \in \mathcal{K}.$$

Taking $w = G(\overline{x})$ and both $d$ and $-d$ gives $\langle \alpha, DG(\overline{x})d\rangle = 0$, for all $d \in X$. Since $d$ is arbitrary and $\langle \alpha, DG(\overline{x})d\rangle = \langle d, DG(\overline{x})^\top \alpha\rangle$, we must have $DG(\overline{x})^\top \alpha = 0$. Since $G(\overline{x}) \in \mathcal{K}$ and $\mathcal{K}$ is a cone, choosing $w = 0$ and $w = 2G(\overline{x})$ gives, respectively,

$$\langle \alpha, G(\overline{x})\rangle \geq 0 \quad \text{and} \quad \langle \alpha, G(\overline{x})\rangle \leq 0.$$

Therefore $\langle \alpha, G(\overline{x})\rangle = 0$. Finally, taking $d = 0$ and using the fact that $\langle \alpha, G(\overline{x})\rangle = 0$ gives $\langle \alpha, w\rangle \leq 0$ for all $w \in \mathcal{K}$, that is, $\alpha \in K^\circ$. We have obtained a nonzero vector $\alpha$ satisfying all relations in (4.5), which proves the equivalence. ∎

*Remark* 4.6. Although the second part of the proof of Proposition 4.5 uses a separation theorem, this part is included only for completeness. The proof of the forthcoming theorem does not rely on it.

Our next goal is to establish that Robinson's condition constitutes a constraint qualification that guarantees the existence of Lagrange multipliers at a local solution of problem (2.1). To this end, we adapt the approach of [2], which relies on constructing a suitable penalization function. Consider the function $P(x) := \|\Pi_{\mathcal{K}^\circ}(G(x))\|^2$. From the second item of Lemma 3.4, $P(x)$ measures the squared distance of $G(x)$ to $\mathcal{K}$; in particular,

$$P(x) = 0 \quad \Longleftrightarrow \quad G(x) \in \mathcal{K}.$$

Thus $P$ is a smooth measure of infeasibility for the conic constraint. The derivative of $P(\cdot)$, obtained via the the chain rule and the third item of Lemma 3.4, plays a central role in defining a sequence of approximate Lagrange multipliers while the first item is responsible for complementarity.

The use of penalty and projection ideas in multiplier theory is classical. In particular, projection-penalty techniques appear in the work of Wierzbicki and Kurcyusz for problems with inequality constraints in Hilbert spaces, and exact penalty arguments have been widely used in optimization, see [35]. Our setting and purpose, however, are different. We consider a finite-dimensional nonlinear problem with an abstract closed convex cone constraint $G(x) \in \mathcal{K}$, and we give a self-contained derivation of the existence of Lagrange multipliers under Robinson's constraint qualification. In this setting, the approximate multipliers arise explicitly from the polar-cone projection $\mu^k = 2k\,\Pi_{\mathcal{K}^\circ}(G(x^k))$, and the proof uses only elementary projection facts, Moreau's decomposition, Weierstrass' theorem, and the algebraic form of Robinson's condition.

The proof of the forthcoming result, following [2, 5, 6], proceeds by constructing a sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ near a local minimizer $\overline{x}$ such that each $x^k$ is a local solution of an unconstrained minimization problem involving $f(x)$ plus an increasingly severe penalization term enforcing feasibility. Since the derivative of this penalized function vanishes at each $x^k$, one can naturally associate approximate multipliers, which under Assumption 4.3 form a bounded sequence. Limit points of this sequence are then shown to be genuine Lagrange multipliers.

THEOREM 4.7. *Let $\overline{x}$ be a local solution of problem* (2.1)*. Suppose that Robinson's condition holds at $\overline{x}$. Then there exists a Lagrange multiplier $\mu \in Y$ such that*

$$\nabla f(\overline{x}) + DG(\overline{x})^\top \mu = 0, \qquad G(\overline{x}) \in \mathcal{K}, \qquad \mu \in \mathcal{K}^\circ, \qquad \langle G(\overline{x}), \mu \rangle = 0.$$

*Proof.* Since $\overline{x}$ is a local solution, there exists $\delta > 0$ such that $f(\overline{x}) \leq f(x)$ for every feasible point $x$ satisfying $\|x - \overline{x}\| \leq \delta$. Recall that $P(x) := \|\Pi_{\mathcal{K}^\circ}(G(x))\|^2$. For each $k \in \mathbb{N}$, consider the auxiliary problem

$$\min_{x \in X} \varphi_k(x) := f(x) + \|x - \overline{x}\|^2 + kP(x) \quad \text{s.t.} \quad \|x - \overline{x}\| \leq \delta. \tag{4.6}$$

Since $\varphi_k$ is continuous and the feasible set in (4.6) is compact, Weierstrass' theorem ensures the existence of a global minimizer $x^k$ for every $k$.

**Claim 1.** The sequence $\{x^k\}$ converges to $\overline{x}$.

Indeed, since $x^k$ minimizes $\varphi_k$ over the ball $\|x - \overline{x}\| \leq \delta$, and since $P(\overline{x}) = 0$, we have

$$f(x^k) + \|x^k - \overline{x}\|^2 + kP(x^k) = \varphi_k(x^k) \leq \varphi_k(\overline{x}) = f(\overline{x}), \qquad \forall k. \tag{4.7}$$

In particular,

$$f(x^k) + \|x^k - \overline{x}\|^2 \leq f(\overline{x}), \qquad \forall k. \tag{4.8}$$

The sequence $\{x^k\}$ is bounded, because $x^k$ belongs to the closed ball $\|x - \overline{x}\| \leq \delta$ for all $k$. Let $x^*$ be an arbitrary accumulation point of $\{x^k\}$, which exists. That is, consider an infinite subsequence $k \in K_1$ such that $\lim_{k \in K_1} x^k = x^*$. By (4.7) and continuity of $f(\cdot) + \| \cdot + \overline{x}\|^2$ we have that $kP(x^k)$ is bounded from above on $K_1$. This proves that $P(x^*) = 0$. Thus, $\Pi_{\mathcal{K}^\circ}(G(x^*)) = 0$. By Lemma 3.4, this implies that $G(x^*) \in \mathcal{K}$. Hence $x^*$ is feasible. Moreover, $\|x^* - \overline{x}\| \leq \delta$, and so, by the local optimality of $\overline{x}$, $f(\overline{x}) \leq f(x^*)$. Taking limits in (4.8) along $K_1$ and using this bound gives

$$f(x^*) + \|x^* - \overline{x}\|^2 \leq f(x^*),$$

which implies $\|x^* - \overline{x}\|^2 = 0$. Therefore $x^* = \overline{x}$. Since every accumulation point of $\{x^k\}$ is equal to $\overline{x}$, we conclude that $x^k \to \overline{x}$.

**Claim 2.** For all sufficiently large $k$, the point $x^k$ is an unconstrained local minimizer of $\varphi_k$.

By Claim 1, $x^k \to \bar{x}$. Therefore, for all sufficiently large $k$, $\|x^k - \overline{x}\| < \delta$. Thus $x^k$ lies in the interior of the ball appearing in (4.6). Consequently, for all sufficiently large $k$, the usual first-order necessary condition for unconstrained local minimization gives $\nabla \varphi_k(x^k) = 0$. Using the chain rule, differentiability of $P$, and Lemma 3.4, we obtain

$$\nabla f(x^k) + 2(x^k - \overline{x}) + DG(x^k)^\top \left[2k\Pi_{\mathcal{K}^\circ}(G(x^k))\right] = 0. \tag{4.9}$$

Define

$$\mu^k := 2k\Pi_{\mathcal{K}^\circ}(G(x^k)). \tag{4.10}$$

Then $\mu^k \in \mathcal{K}^\circ$ and

$$\nabla f(x^k) + 2(x^k - \overline{x}) + DG(x^k)^T \mu^k = 0 \tag{4.11}$$

for all sufficiently large $k$. Moreover, by Lemma 3.4 item (i), we have

$$\langle \Pi_{\mathcal{K}}(G(x^k)), \mu^k \rangle = 0 \tag{4.12}$$

for all sufficiently large $k$.

**Claim 3.** The sequence $\{\mu^k\}$ is bounded.

Suppose, by contradiction, that $\{\mu^k\}$ is unbounded. Passing to a subsequence if necessary, we may assume that

$$\|\mu^k\| \to \infty \qquad \text{and} \qquad \frac{\mu^k}{\|\mu^k\|} \to \alpha$$

for some $\alpha \in Y$ with $\|\alpha\| = 1$. Since $\mathcal{K}^\circ$ is a closed cone and $\mu^k \in \mathcal{K}^\circ$, we have $\alpha \in \mathcal{K}^\circ$. Dividing (4.12) by $\|\mu^k\|$ and passing to the limit, we obtain $\langle G(\overline{x}), \alpha \rangle = 0$, because $x^k \to \overline{x}$, the projection $\Pi_{\mathcal{K}}$ is continuous, and $\Pi_{\mathcal{K}}(G(\overline{x})) = G(\overline{x})$. Similarly, dividing (4.11) by $\|\mu^k\|$ and passing to the limit gives $DG(\overline{x})^\top \alpha = 0$. Thus we have found $\alpha \neq 0$ such that

$$DG(\overline{x})^\top \alpha = 0, \qquad \langle G(\overline{x}), \alpha \rangle = 0, \qquad \alpha \in \mathcal{K}^\circ.$$

This contradicts the algebraic form of Robinson's condition given in (4.5). Therefore, $\{\mu^k\}$ is bounded.

**Claim 4.** Every accumulation point of $\{\mu^k\}$ is a Lagrange multiplier associated with $\bar{x}$.

By Claim 3, there exists a subsequence, still denoted by $\{\mu^k\}$, such that $\mu^k \to \mu$ for some $\mu \in Y$. Since $\mathcal{K}^\circ$ is closed and $\mu^k \in \mathcal{K}^\circ$, we have $\mu \in \mathcal{K}^\circ$. Passing to the limit in (4.11), and using $x^k \to \overline{x}$, we obtain

$$\nabla f(\overline{x}) + DG(\overline{x})^\top \mu = 0.$$

Furthermore, since $\overline{x}$ is feasible, $G(\overline{x}) \in \mathcal{K}$. Finally, passing to the limit in (4.12) gives $\langle G(\bar{x}), \mu \rangle = 0$. Therefore $\mu$ satisfies the KKT conditions. ∎

Under Assumption 4.3 for problem (2.1) a similar reasoning shows that the set of Lagrange multipliers is compact. Uniqueness is guaranteed if, instead of the conic linear independence condition (4.5), one imposes the standard linear independence requirement, namely, that the implication (4.5) holds for $\alpha \in Y$ instead of for $\alpha \in \mathcal{K}^\circ$. We leave these computations as an exercise to the reader. We conclude by noting that this line of reasoning has inspired the development of several new necessary optimality conditions and constraint qualifications in much broader settings; see, for instance, [9] for an extension to infinite-dimensional spaces and [3] for the case where $X$ is a Riemannian manifold.

**5. Concluding remarks.** The existence of Lagrange multipliers is one of the central results in constrained optimization. Although multiplier rules are often presented as standard consequences of advanced tools such as separation theorems, duality theory, or implicit-function arguments, their geometric content can be made visible through a much more elementary route.

In this paper, we presented a self-contained proof of the existence of Lagrange multipliers for finite-dimensional nonlinear optimization problems with a closed convex cone constraint. The proof is based on a simple penalization argument combined with elementary facts about projections onto closed convex cones. In particular, the squared distance to the cone is written as $\text{dist}(G(x), \mathcal{K})^2 = \|\Pi_{\mathcal{K}^\circ}(G(x))\|^2$, and the approximate multipliers arise naturally from the polar-cone projection $\mu^k = 2k\,\Pi_{\mathcal{K}^\circ}(G(x^k))$. Moreau's decomposition then provides the complementarity relation, while Robinson's constraint qualification ensures the boundedness needed to pass to the limit.

The penalty and projection ideas used here are classical, and related techniques appear in earlier works on multiplier rules, penalty methods, and duality. Our contribution is not to introduce a new multiplier theory, but to organize these ideas in a finite-dimensional conic setting in a way that is accessible to advanced undergraduate students and beginning graduate students. The resulting proof shows how the familiar KKT conditions for nonlinear programming extend naturally to cone-constrained problems, including second-order cone and semidefinite constraints.

From a pedagogical point of view, the argument highlights three useful lessons. First, infeasibility can be measured geometrically by the distance to the constraint cone. Second, multipliers can be interpreted as limiting objects generated by penalizing this infeasibility. Third, constraint qualifications such as Robinson's condition are precisely the assumptions that prevent these approximate multipliers from escaping to infinity, although more general assumptions may be used [30]. This viewpoint helps connect the classical theory of nonlinear programming with the broader framework of modern conic optimization.

We hope that the presentation can serve as a bridge between a first course in nonlinear optimization and more advanced topics in variational analysis, conic duality, and semidefinite programming. Possible continuations include comparing Robinson's condition with other constraint qualifications for cone-constrained problems, studying multiplier existence in infinite-dimensional settings, and exploring how similar projection-penalty ideas appear in numerical methods for conic optimization.

## REFERENCES


[1] Alizadeh, F., Goldfarb, D. (2003). Second-order cone programming. *Mathematical Programming*, 95, 3–51.

[2] Andreani, R., Gómez, W., Haeser, G., Mito, L. M., Ramos, A. (2022). On optimality conditions for nonlinear conic programming. *Mathematics of Operations Research*, 47(3), 2160–2185.

[3] Andreani, R., Couto, K. R., Ferreira, O. P., Haeser, G. (2024). Constraint qualifications and strong global convergence properties of an augmented Lagrangian method on Riemannian manifolds. *SIAM Journal on Optimization*, 34(2), 1799–1825.

[4] Andreani, R., Haeser, G., Schuverdt, M. L., Silva, P. J. S. (2012). Two new weak constraint qualifications and applications. *SIAM Journal on Optimization*, 22(3), 1109–1135.

[5] Andreani, R., Haeser, G., Viana, D. S. (2020). Optimality conditions and global convergence for nonlinear semidefinite programming. *Mathematical Programming*, 180(1), 203–235.

[6] Andreani, R., Haeser, G., Martínez, J. M. (2011). On sequential optimality conditions for smooth constrained optimization. *Optimization*, 60(5), 627–641.

[7] Bertsekas, D. P. (1999). *Nonlinear Programming.* Athena Scientific, Belmont, Massachusetts.

[8] Bonnans, J. F., Shapiro, A. (2000). *Perturbation Analysis of Optimization Problems.* Springer-Verlag, New York.

[9] Börgens, E., Kanzow, C., Mehlitz, P., Wachsmuth, G. (2020). New constraint qualifications for optimization problems in Banach spaces based on asymptotic KKT conditions. *SIAM Journal on Optimization*, 30(4), 2956–2982.

[10] Clarke, F. H. (1983). *Optimization and Nonsmooth Analysis.* John Wiley & Sons, New York.

[11] Courant, R. (1943). *Variational methods for the solution of problems of equilibrium and vibrations*, Bull. Amer. Math. Soc., 49, 1–23.

[12] Dür, M. (2010). Copositive programming – a survey. In M. Diehl, F. Glineur, E. Jarlebring, and W. Michiels (eds.), *Recent Advances in Optimization and its Applications in Engineering.* Springer, Berlin, Heidelberg.

[13] Fiacco, A. and McCormick, G. (1968). *Sequential Unconstrained Nonlinear Programming*, John Wiley & Sons, New York.

[14] Fitzpatrick, S., Phelps, R. R. (1982). Differentiability of the metric projection in Hilbert space. *Transactions of the American Mathematical Society*, 270(2), 483–501.

[15] Forsgren, A. (2000). Optimality conditions for nonlinear semidefinite programming. *Mathematical Programming*, 88, 105–128.

[16] Haeser, G., Santos, D. O. (2026). A 4-steps elementary proof of existence of Lagrange multipliers. *Computational and Applied Mathematics*, 45, 153.

[17] Ioffe, A. D. (2016). *Metric regularity—a survey. Part I. Theory*, Journal of the Australian Mathematical Society, 101, 188–243.

[18] Karush, W. (1939). Minima of functions of several variables with inequalities as side constraints. M.Sc. dissertation, Department of Mathematics, University of Chicago, Chicago, IL.

[19] Kjeldsen, T. H. (2000). A contextualized historical analysis of the Kuhn–Tucker theorem in nonlinear programming: The impact of World War II. *Historia Mathematica*, 27, 331–361.

[20] Kuhn, H. W., Tucker, A. W. (1951). Nonlinear programming. In J. Neyman (ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 481–492. University of California Press, Berkeley, CA.

[21] Levitin, E. S. and Poliak, B. T. (1967). *Convergence of minimizing sequences in conditional extremum problems*, Soviet Math. Dokl., 7, 764–767.

[22] Lobo, M. S., Vandenberghe, L., Boyd, S., Lebret, H. (1998). Applications of second-order cone programming. *Linear Algebra and its Applications*, 284, 193–228.

[23] Mangasarian, O. L. (1969). *Nonlinear Programming*. McGraw-Hill, New York.

[24] Mordukhovich, B. S., Outrata, J. V., Ramírez C., H. (2015). Second-order variational analysis in conic programming with applications to optimality and stability. *SIAM Journal on Optimization*, 25(1), 76–101.

[25] Moreau, J. J. (1962). Décomposition orthogonale d'un espace hilbertien selon deux cônes mutuellement polaires. *Comptes Rendus de l'Académie des Sciences de Paris*, 255, 238–240.

[26] Robinson, S. M. (1976). Stability theory for systems of inequalities, Part II: Differentiable nonlinear systems. *SIAM Journal on Numerical Analysis*, 13(4), 497–513.

[27] Robinson, S. M. (1976). First-order conditions for general nonlinear optimization. *SIAM Journal on Applied Mathematics*, 30(4), 597–607.

[28] Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton.

[29] Rockafellar, R. T. (1974). *Conjugate Duality and Optimization*. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 16, Society for Industrial and Applied Mathematics, Philadelphia.

[30] Santos, D. O. (2026). *Quasinormality and pseudonormality for nonlinear semidefinite programming*, Optimization Online.

[31] Shapiro, A. (1992). Sensitivity analysis of parametrized programs under cone constraints. *SIAM Journal on Control and Optimization*, 30(6), 1409–1422.

[32] Shapiro, A. (1997). On uniqueness of Lagrange multipliers in optimization problems subject to cone constraints. *SIAM Journal on Optimization*, 7(2), 508–518.

[33] Vandenberghe, L., Boyd, S. (1996). Semidefinite programming. *SIAM Review*, 38(1), 49–95.

[34] Vandenberghe, L., Boyd, S. (1999). Applications of semidefinite programming. *Applied Numerical Mathematics*, 29, 283–299.

[35] Wierzbicki, A. P., Kurcyusz, S. (1977). Projection on a cone, penalty functionals and duality theory for problems with inequality constraints in Hilbert space. *SIAM Journal on Control and Optimization*, 15(1), 25–56.