A Simple Canonical Form for Nonlinear Programming Problems and Its Use

Walter F. Mascarenhas

Received: 29 March 2018 / Accepted: 17 August 2018 / Published online: 22 August 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract
We argue that reducing nonlinear programming problems to a simple canonical form is an effective way to analyze them, specially when the gradients of the constraints are linearly dependent. To illustrate this fact, we solve an open problem about constraint qualifications using this canonical form.

Keywords Nonlinear programming · Second-order optimality conditions · Linear dependency

Mathematics Subject Classification 90C30 · 90C46

1 Introduction

Mathematical models are the foundation of optimization theory and applications, and in order to understand them we formulate theorems with reasonable hypothesis and relevant conclusions. When the optimization problem has constraints, and the gradients of these constraints exist, the linear independence of such gradients is the simplest hypothesis. Under this hypothesis, our intuition works well and we obtain clean optimality conditions, which are used in the construction of efficient and accurate algorithms. On the other hand, when the gradients of the constraints are linearly dependent, our understanding of the optimization problem is often unsatisfactory, and this lack of understanding can lead to incorrect theoretical results, as mentioned in [1].

The consequences of linear dependencies are relevant also in other areas of pure and applied mathematics, and several techniques have been developed for studying them. For instance, in bifurcation theory, “Normal Forms” [2] are relatively simple models illustrating how a system can change drastically due to a loss of linear independency of some vectors.

1 Universidade de São Paulo, São Paulo, Brazil
This idea of “Normal” or “Canonical” form is also helpful in optimization, and the Stable Manifold Theorem [3] is a good example in this regard. This theorem provides a “canonical form” for the set of points for which the iterates of a dynamical system converge to a saddle point, and in our research about nonlinear programming we have used it to analyze the convergence of significant algorithms, like the affine scaling method [4], Newton’s method [5], the BFGS method [6], and even to general families of methods [7,8]. In these articles, by studying low-dimensional problems with the right canonical form, we were able to provide satisfactory answers to relevant open problems.

In the present article, we propose a canonical form for nonlinear programming problems in which the gradients of the constraints are linearly dependent. Specifically, we look at problems in which at all points $y$ in a neighborhood of a point $x$ the Jacobian of the constraints has rank one more than the rank of the Jacobian at $x$. This is the simplest form of such linear dependency, and its relevance is illustrated by the article [9], by Andreani, Martínez and Schuverdt, which also states a conjecture regarding such nonlinear programming problems. We exemplify the effectiveness of our canonical form by using it to prove a conjecture stated by Andreani in [9].

In Sect. 2, we formulate precisely our mathematical models and Andreani’s conjecture. Section 3 emphasizes that usual optimality conditions in nonlinear programming are invariant under changes of variables. In particular, the Lagrange multipliers are the same in different coordinate systems, and the Mangasarian–Fromovitz constraint qualification is invariant under changes of these variables. Section 3 is obvious but relevant. In fact, we believe that changes in variables do not receive the attention they deserve in the mathematical programming literature. For instance, most textbooks do not mention the invariance above explicitly. Authors usually have these changes in variables on the back of their mind, and build good examples to illustrate their points based upon them. However, readers may not notice that, with the proper changes of variables, more general situations can be reduced to these good examples, and in many cases such examples are more enlightening than proofs.

Our canonical form is presented in Sect. 4, and in Sect. 5 we prove Andreani’s conjecture, using the previous sections and two linear algebraic lemmas proved in “Appendix.” Our results are summarized in the last section.

2 Mathematical Formulation

In this article, we consider the classical nonlinear programming problem

$$\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad h(x) = 0 \in \mathbb{R}^m, \\
& \quad g(x) \leq 0 \in \mathbb{R}^{r+p},
\end{align*}$$

(1)

where $f$, $g$ and $h$ are differentiable functions with domain $\mathbb{R}^n$. We are interested in situations in which there are points $x^*$ at which the derivatives $Dh(x^*)$ and $Dg(x^*)$ of the constraints are defective, in the sense that the rank of the combined matrix
\[ Dhg(x) := \left( \begin{array}{c} Dh(x) \\ Dg(x) \end{array} \right) \]  

(2)
is \( m + r \leq m + r + p \) at \( x = x^* \) (by “derivative” \( Dh(x) \) here we mean the linear transformation represented by the Jacobian matrix of the function \( h \)).

The analysis of problem (1) is usually based on a constraint qualification. For instance, the well-known Mangasarian–Fromovitz constraint qualification assumes that \( Dh(x^*) \) has rank \( m \) and requires the existence of a “strictly decreasing direction \( d \)” for the inequality constraints, which is also compatible with the equality constraints:

**Definition 2.1** (The Mangasarian–Fromovitz Constraint Qualification) We say that the functions \( h \) and \( g \) and the point \( x^* \) such that \( h(x^*) = 0 \) and \( g(x^*) = 0 \) satisfy the Mangasarian–Fromovitz constraint qualification for the problem (1) if \( \text{rank}(Dhg(x)) = m \) and there exists a vector \( d \in \mathbb{R}^n \) such that \( Dh(x^*)d = 0 \) and \( Dg(x^*)d < 0 \).

Andreati’s conjecture is stated using the Mangasarian–Fromovitz Constraint Qualification:

**Theorem 2.1** Suppose the functions \( f, h \) and \( g \) in problem (1) are of class \( C^2 \) in a neighborhood \( A \) of \( x^* \in \mathbb{R}^n \) and \( h(x^*) = 0 \) and \( g(x^*) = 0 \). If the Mangasarian–Fromovitz constraint qualification is satisfied and

\[ \text{rank}(Dhg(x)) \leq \text{rank}(Dhg(x^*)) + 1 \]

for \( x \in A \), then there exist \( \lambda \in \mathbb{R}^m \) and \( \mu \in \mathbb{R}^{r+p} \), with \( \mu \geq 0 \), such that

\[ S(d) := d^T \left( \nabla^2 f(x^*) + \sum_{j=1}^{m} \lambda_j \nabla^2 h_j(x^*) + \sum_{\ell=1}^{r+p} \mu_{\ell} \nabla^2 g_{\ell}(x^*) \right) d \geq 0 \]  

(3)

for all \( d \in \mathbb{R}^n \) with \( Dh(x^*)d = 0 \) and \( Dg(x^*)d = 0 \).

### 3 Changes of Variables

This section calls the reader’s attention to the fundamental fact that the usual optimality conditions in mathematical programming are invariant under changes of variables. In other words, in theory we can pick any coordinate system we please, as long as the coordinates are “consistent” and reflect correctly the problem we want to understand.

We emphasize that, for theoretical purposes, changes in coordinates do not need to be explicit: it suffices to know that they exist and have the nice properties required to deal with the problem at hand. Of course, things are different in practice, because we usually cannot afford, or know how, to compute such changes of coordinates.

The simplest changes of variables are the linear ones: we replace the coordinates \( x \) by \( Ay \), where \( A \) is a nonsingular square matrix. As a result, we can replace a function \( f \) by a more appropriate, or simpler, function \( \hat{f} \), as in
\[ \hat{f}(y) = f(Ay). \]  

The chain rule yields
\[ \nabla \hat{f}(y) = A^T \nabla f(Ay). \]  

and applying the formula above to the function \( \hat{f}_k(y) := \partial_k f(Ay) \), taking one more derivative, and doing the algebra we obtain that
\[ \nabla^2 \hat{f}(y) = A^T \nabla^2 f(Ay) A. \]  

Equations (5) and (6) are quite useful, but they are not enough to explore the full power of changes of variables. For that we need to replace \( x \) by a nonlinear function \( q(y) \) of a more convenient variable \( y \), as in the nonlinear version of Eq. (4):
\[ \hat{f}(y) := f(q(y)). \]  

In Eq. (7), the function \( q \) is a local diffeomorphism, that is a differentiable function defined in a neighborhood of the point with which we are concerned, and such that its inverse \( q^{-1} \) (in the sense that \( q^{-1}(q(y)) = y \)) exists and is also differentiable. In this nonlinear setting, we have the following version of Eq. (5)
\[ \nabla \hat{f}(y) = Dq(y)^T \nabla f(q(y)). \]  

Equation (8) is almost the same as (5): we only need to replace \( A \) by \( Dq(y) \) and \( Ay \) by \( q(y) \), and keep in mind that the matrix \( Dq(y) \) is square and nonsingular. Things are a bit more complicated for the Hessian. In this case, we need to introduce an extra sum due to the curvature in \( q \), and the resulting formula is:
\[ \nabla^2 \hat{f}(y) = Dq(y)^T \nabla^2 f(q(y)) Dq(y) + \sum_{k=1}^n \partial_k f(q(y)) \nabla^2 q_k(y). \]  

With Eqs. (8) and (9), we can find how the optimality conditions behave under nonlinear changes of coordinates \( x = q(y) \). To see why this is true, we consider the classical nonlinear programming problem (1). By making the change in variables \( x = q(y) \), we do not affect the satisfiability of the equalities and inequalities in (1), that is
\[ \hat{h}_j(y) := h_j(q(y)) \]  

will be equal to zero as long as \( h_j(x) \) is equal to zero.

Let us now analyze the first-order optimality conditions for problem (1) at a point \( x^* \). These conditions are written in terms of the Lagrange multiplies \( \lambda_j \in \mathbb{R} \) and \( \mu_\ell \geq 0 \):
\[ \nabla f(x^*) + \sum_{j=1}^m \lambda_j \nabla h_j(x^*) + \sum_{\ell=1}^{r+p} \mu_\ell \nabla g_\ell(x^*) = 0. \]  

\( \square \) Springer
Given a local diffeomorphism \( \mathbf{x} = q(\mathbf{y}) \), by defining
\[
\hat{f}(\mathbf{y}) := f(q(\mathbf{y})), \quad \hat{h}(\mathbf{y}) := h(q(\mathbf{y})), \quad \text{and} \quad \hat{g}(\mathbf{y}) := g(q(\mathbf{y})),
\]
(11)

Equation (8) yields
\[
\nabla \hat{f}(\mathbf{y}) = Dq(\mathbf{y})^T \nabla f(q(\mathbf{y}))
\]
(12)
and
\[
\nabla \hat{h}_j(\mathbf{y}) = Dq(\mathbf{y})^T \nabla h_j(q(\mathbf{y})) \quad \text{and} \quad \nabla \hat{g}_\ell(\mathbf{y}) = Dq(\mathbf{y})^T \nabla g_\ell(q(\mathbf{y})).
\]
(13)

Since our \( Dq(\mathbf{y}) \) is always nonsingular, for \( \mathbf{y}^* \) such that \( x^* = q(\mathbf{y}^*) \), the first-order condition (10) holds if and only if
\[
\nabla \hat{f}(\mathbf{y}^*) + m \sum_{j=1}^m \lambda_j \nabla \hat{h}_j(\mathbf{y}^*) + r + p \sum_{\ell=1}^{r+p} \mu_\ell \nabla \hat{g}_\ell(\mathbf{y}^*) = 0.
\]
(14)

In other words, the first-order conditions are invariant with respect to changes of coordinates, and we may study them by using Eq. (10), or Eq. (14), or both. In particular, the Lagrange multipliers do not change as we change coordinates as above.

For the second-order conditions, we consider directions \( \mathbf{d} \) such that
\[
Dh(x^*) \mathbf{d} = 0 \quad \text{and} \quad Dg(x^*) \mathbf{d} = 0,
\]
(15)
and Eq. (13) shows that Eq. (15) is equivalent to
\[
D\hat{h}(\mathbf{y}^*) \hat{\mathbf{d}} = 0 \quad \text{and} \quad D\hat{g}(\mathbf{y}^*) \hat{\mathbf{d}} = 0,
\]
(16)
for
\[
\hat{\mathbf{d}} := Dq(\mathbf{y}^*)^{-1} \mathbf{d}, \quad \text{or} \quad \mathbf{d} = Dq(\mathbf{y}^*) \hat{\mathbf{d}}.
\]

Therefore, the orthogonality condition (15) is invariant under changes of coordinates, that is, Eq. (15) for \( \mathbf{d} \) and Eq. (16) for \( \hat{\mathbf{d}} \) are equivalent.

Using the equations above, and the abbreviation \( S(\mathbf{d}) := S(\mathbf{d}, f, h, g, \lambda, \mu) \), we can write the second-order term
\[
S(\mathbf{d}) := \mathbf{d}^T \left( \nabla^2 f(x^*) + m \sum_{j=1}^m \lambda_j \nabla^2 h_j(x^*) + r + p \sum_{\ell=1}^{r+p} \mu_\ell \nabla^2 g_\ell(x^*) \right) \mathbf{d}
\]
(17)
as
\[
S(\mathbf{d}) = \hat{\mathbf{d}}^T \left( Dq(\mathbf{y}^*)^T \nabla^2 f(q(\mathbf{y}^*))Dq(\mathbf{y}^*) + m \sum_{j=1}^m \lambda_j Dq(\mathbf{y}^*)^T \nabla^2 h_j(q(\mathbf{y}^*))Dq(\mathbf{y}^*) \\
+ r + p \sum_{\ell=1}^{r+p} \mu_\ell Dq(\mathbf{y}^*)^T \nabla^2 g_\ell(q(\mathbf{y}^*))Dq(\mathbf{y}^*) \right) \hat{\mathbf{d}}.
\]
Equation (9) yields

\[ S(d, f, h, g, \lambda, \mu) = S(\hat{d}, \hat{f}, \hat{h}, \hat{g}, \lambda, \mu) - \Delta \]

for

\[
\Delta := \hat{d}^T \left( \sum_{k=1}^{n} \partial_k f(q(y^*)) \nabla^2 q_k(y^*) + \sum_{j=1}^{m} \lambda_j \sum_{k=1}^{n} \partial_k h_j(q(y^*)) \nabla^2 q_k(y^*) \right.
\]
\[
+ \sum_{\ell=1}^{r+p} \sum_{k=1}^{n} \partial_k g_\ell(q(y^*)) \nabla^2 q_k(y^*) \left) \hat{d} \right.
\]
\[
= \hat{d}^T \left( \sum_{k=1}^{n} \left( \partial_k f(q(y^*)) + \sum_{j=1}^{m} \lambda_j \partial_k h_j(q(y^*)) + \sum_{\ell=1}^{r+p} \mu_\ell \partial_k g_\ell(q(y^*)) \right) \nabla^2 q_k(y^*) \right) \hat{d}. \] (18)

When the first-order conditions (10) hold we have that

\[
\partial_k f(q(y^*)) + \sum_{j=1}^{m} \lambda_j \partial_k h_j(q(y^*)) + \sum_{\ell=1}^{r+p} \mu_\ell \partial_k g_\ell(q(y^*)) = 0
\]

and \( \Delta = 0 \). Therefore, when the first-order conditions hold, the second-order term \( S(d) \) in Eq. (17) is invariant with respect to changes of variables, and it can be evaluated using the expression

\[
\hat{S}(\hat{d}) := \hat{d}^T \left( \nabla^2 \hat{f}(y^*) + \sum_{j=1}^{m} \lambda_j \nabla^2 \hat{h}_j(y^*) + \sum_{\ell=1}^{r+p} \mu_\ell \nabla^2 \hat{g}_\ell(y^*) \right) \hat{d}. \] (18)

An analogous argument, starting from the conditions

\[
Dh(x^*) \ d = 0 \quad \text{and} \quad Dg(x^*) \ d < 0
\]

instead of Eq. (15), shows that the Mangasarian–Fromovitz constraint qualification is invariant under changes in coordinates, and similar arguments apply to many other constraint qualifications.

In summary, when trying to answer many theoretical questions regarding the first- and second-order optimality conditions, Lagrange multipliers and constraint qualifications for the nonlinear programming problem (1), we can analyze them in other coordinate systems, and reach correct conclusions by considering only simplified problems. As we show in the next sections, this observation has far-reaching consequences and can be used to give simpler proofs for some results in the nonlinear optimization literature.
4 The Canonical Form

In this section, we present a simple canonical form for the classical nonlinear programming problem (1), which can be used when the derivative of the equality constraints has full rank but the inequality constraints are degenerated. In this canonical form, the variables are \( y = (y_1, \ldots, y_m)^T \), \( z = (z_1, \ldots, z_r)^T \) and \( w = (w_1, \ldots, w_{n-m-r})^T \), where

\[
    r := \text{rank}(Dh(x^*)) - m.
\]

The equality constraints are given by

\[
    \hat{h}(y, z, w) = y = 0,
\]

and there are two groups of inequality constraints. The first one is given by

\[
    \hat{g}(y, z, w) = z \leq 0.
\]

The second group of inequalities is given by

\[
    c(y, z, w) \leq 0 \in \mathbb{R}^p,
\]

for a function \( c \) about which we know that

\[
    c(0, 0, 0) = 0 \quad \text{and} \quad D_wc(0, 0, 0) = 0, \tag{19}
\]

but which can be quite complicated. In particular, we have no control over \( D_y c(0, 0, 0) \) or \( D_z c(0, 0, 0) \). The domain of these functions is a product of open sets \( \mathcal{Y} \times \mathcal{Z} \times \mathcal{W} \) containing 0. In summary, we have the canonical nonlinear programming problem

\[
\begin{align*}
    \text{minimize} & \quad f(y, z, w) \\
    \text{subject to} & \quad y = 0 \in \mathbb{R}^m, \\
    & \quad z \leq 0 \in \mathbb{R}^r, \\
    & \quad c(y, z, w) \leq 0 \in \mathbb{R}^p, \\
    & \quad y \in \mathcal{Y} \subset \mathbb{R}^m, \quad z \in \mathcal{Z} \subset \mathbb{R}^r \quad \text{and} \quad w \in \mathcal{W} \subset \mathbb{R}^{n-m-r}. \tag{20}
\end{align*}
\]

In this problem, the derivative of the constraints has the simple form

\[
    Dhgc(y, z, w) = \begin{pmatrix}
        I_{m \times m} & 0 & 0 \\
        0 & I_{r \times r} & 0 \\
        D_y c(y, z, w) & D_z c(y, z, w) & D_w c(y, z, w)
    \end{pmatrix}, \tag{21}
\]

with \( D_w c(0, 0, 0) = 0 \), where \( I_{m \times m} \) is the \( m \times m \) identity matrix. The next theorem shows that nonlinear programming problems can be reduced to the canonical form (20) under mild assumptions, and the discussion in the previous section shows that the form of the first- and second-order conditions and the Lagrange multipliers do not

\[ \text{ Springer} \]
change in this reduction. In many relevant situations, we can then use Theorem 4.1 below and say rigorously

without loss of generality, we can assume that our nonlinear programming problem is of the form (20)

The purpose of the present article is to call the readers attention to this simple intuitive idea, which is formalized by Theorem 4.1.

**Theorem 4.1** (The Canonical Form) Suppose that the functions $h$ and $g$ in problem (1) are of class $C^s$, with $s \geq 1$, in a neighborhood $A$ of $x^* \in \mathbb{R}^n$, and $h(x^*) = 0$ and $g(x^*) = 0$. If $h$ then $\times (m + r)$ matrix

$$
\begin{pmatrix}
\nabla h_1(x^*) , \ldots , \nabla h_m(x^*) , \nabla g_1(x^*) , \ldots , \nabla g_r(x^*)
\end{pmatrix}
$$

has rank $m + r$, then there exist open sets $Y \subset \mathbb{R}^m$, $Z \subset \mathbb{R}^r$ and $W \in \mathbb{R}^{n - m - r}$, with $(0, 0, 0) \in Y \times Z \times W$, and a diffeomorphism

$q : Y \times Z \times W \to q(Y \times Z \times W) \subset A$

of class $C^s$ with $q(0, 0, 0) = x^*$ such that, for $1 \leq \ell \leq r$,

$$
\hat{h}(y, z, w) := h(q(y, z, w)) = y \text{ and } \hat{g}_\ell(y, z, w) := g_\ell(q(y, z, w)) = z_\ell
$$

and for $1 \leq \ell \leq p$ the functions

$$
c_\ell(y, z, w) := g_{r+\ell}(q(y, z, w))
$$

are such that $c(0, 0, 0) = 0$, $D_w c(0, 0, 0) = 0$ and $q(0, 0) = x^*$.

Theorem 4.1 is a direct consequence of the Projection Lemma below, which is a corollary of the inverse function theorem (see Thm 2-13 in p. 43 of [10].) Besides Lemma 4.1, we only need to note that the equality $D_h g(x^*)$ has rank $m + r$.

**Lemma 4.1** (The Projection Lemma) Let $A$ be a neighborhood of $x^* \in \mathbb{R}^{n+k}$ and let $f : A \to \mathbb{R}^n$ a function of class $C^s$. If $Df(x^*)$ has rank $n$, then there exists a neighborhood $Y$ of $0 \in \mathbb{R}^n$, a neighborhood $Z$ of $0 \in \mathbb{R}^k$ and a diffeomorphism

$q : Y \times Z \to q(Y \times Z) \subset A$

of class $C^s$ such that $f(q(y, z)) = y$ and $q(0, 0) = x^*$.
5 A Proof of Andreani’s Conjecture

In this section, we use the canonical form (20) to prove the conjecture by Andreani, Martínez and Schuverdt mentioned in the introduction. This is an interesting application of the canonical form because there were several failed attempts to find a proof of this conjecture by other means.

For instance, the authors of [1] attempted to obtain an appropriate coordinate system, by using a version of the singular value decomposition, and succeeded in proving new particular cases of Andreani’s conjecture with this approach. However, they did not prove the conjecture because their decomposition is not as effective as the canonical form: it has “high order terms” in places in which the canonical decomposition has exact zeros, and the technicalities required to handle these terms precluded them from obtaining a proof for which they had found all the other ingredients. This shows that good choices of variables go beyond controlling “high order terms”: we actually want to eliminate them, and Andreani’s conjecture is one of the fortunate cases in which this is possible.

Our proof of Andreani’s Conjecture uses the arguments presented in the previous sections and two linear algebraic lemmas. These lemmas are variations of results already presented in other references [1,11] and are quite technical. They are proved in “Appendix,” but it is best to read their proofs after understanding how they fit into the main proof, in order not to loose track of the arguments presented in the main proof.

Proof of Theorem 2.1 As a first step, we use Theorem 4.1 to reduce problem (1) to the canonical form (20). In order to do that we note that the Mangasarian–Fromovitz constraint qualification requires that $Dh(x^*)$ has rank $m$ and recall that we use $m + r$ to denote the rank of the matrix $Dhg(x^*)$ in Eq. (2). Therefore, by changing the order of the inequality constraints if necessary, we can assume that the matrix in Eq. (22) in the statement of Theorem 4.1 has rank $m + r$. Since in Sect. 3, we have shown that Eq. (3) and the Mangasarian–Fromovitz constraint qualification are invariant under changes of coordinates, we can then use Theorem 4.1 and assume without loss of generality that our nonlinear programming problem has the form (20).

Due to Eqs. (15) and (21), we can assume that the vector $d$ in Eq. (3) has the form

$$d = \begin{pmatrix} 0_m \\ 0_r \\ \tilde{d} \end{pmatrix} \quad \text{with } \tilde{d} \in \mathbb{R}^p,$$

and defining $\tilde{f} : \mathcal{W} \to \mathbb{R}$ and $\tilde{c} : \mathcal{W} \to \mathbb{R}$ by

$$\tilde{f}(w) := f(0_m, 0_r, w) \quad \text{and} \quad \tilde{c}(w) := c(0_m, 0_r, w),$$

we can rewrite Eq. (3) as

$$\tilde{d}^T \left( \nabla^2 \tilde{f}(0) + \sum_{\ell=1}^{p} \mu_{r+\ell} \nabla^2 \tilde{c}_{\ell}(0) \right) \tilde{d} \geq 0. \quad (23)$$
Equation (21) yields
\[
\text{rank}(Dhgc(y, z, w)) \leq m + r + 1 \iff \text{rank}(Dw_c(y, z, w)) \leq 1,
\]
and combining the observation that
\[
D\tilde{c}(w) = Dw_c(0, 0, w)
\]
with the next lemma we conclude that there exist \(\alpha_1, \ldots, \alpha_p \in \mathbb{R}\) and a symmetric matrix \(H \in \mathbb{R}^{p \times p}\) such that
\[
\nabla^2_{\tilde{c}_\ell}(0) = \alpha_\ell H \quad \text{for } \ell = 1, \ldots, p.
\]

**Lemma 5.1** (Hessians with rank at most one) Let \(A\) be a neighborhood of 0 \(\in \mathbb{R}^n\), and let \(c_1, \ldots, c_m\) be functions from \(A\) to \(\mathbb{R}\) of class \(C^2\). If \(Dc(0) = 0\) and \(\text{rank}(Dc(x)) \leq 1\) for all \(x \in A\), then there exist \(\alpha_1, \ldots, \alpha_m \in \mathbb{R}\) and a symmetric matrix \(H \in \mathbb{R}^{n \times n}\) such that \(\nabla^2_{\tilde{c}_\ell}(0) = \alpha_\ell H\) for \(\ell = 1, \ldots, m\).

It follows that Eq. (23) is equivalent to
\[
\tilde{S}(\tilde{d}, \gamma) := \tilde{d}^T \left( \nabla^2 f(0) + \gamma H \right) \tilde{d} \geq 0 \quad \text{for } \gamma = \sum_{\ell=1}^p \alpha_\ell \mu_{r+\ell}.
\] (24)

The Mangasarian–Fromovitz constraint qualification implies that for every \(d\) there exist \(\lambda\) and \(\mu > 0\) such that \(S(d) \geq 0\) in Eq. (3) (see [1]). Since this equation is equivalent to Eq. (24), for every \(\tilde{d}\) there exists \(\gamma\) such that \(\tilde{S}(\tilde{d}, \gamma) \geq 0\) in Eq. (24), and in order to complete the proof we use the following lemma:

**Lemma 5.2** (Semidefinite separation) Let \(I \subset \mathbb{R}\) be a compact interval. If the symmetric matrices \(A, B \in \mathbb{R}^{n \times n}\) are such that for all \(x \in \mathbb{R}^n\) there exists \(\gamma_x \in I\) such that \(x^T (A + \gamma_x B) x \geq 0\), then there exists \(\gamma^* \in I\) such that
\[
x^T (A + \gamma^* B) x \geq 0
\]
for all \(x \in \mathbb{R}^n\).

Lemma 5.2 implies that there exists
\[
\gamma^* = \sum_{\ell=1}^p \alpha_\ell \mu^*_{r+\ell}
\]
such that \(\tilde{S}(\tilde{d}, \gamma^*) \geq 0\) in Eq. (24) for all \(\tilde{d}\). The full set \(\lambda^*\) and \(\mu^*\) of Lagrange multipliers containing these \(\mu^*_{r+1}, \ldots, \mu^*_{r+p}\) is as required by Andreani’s Conjecture and we are done. \(\square\)
6 Conclusions

We presented a simple canonical form for nonlinear programming problems with the simplest kind of linear dependency of the gradients of the constraints, which was considered by Andreani et al. We then used this canonical form to prove Andreani’s conjecture.

Appendix

In this appendix, we prove the results involving Linear Algebra used in Sect. 5. The proof of Lemma 5.1 is based on the next two lemmas:

Lemma A.1 (Rank one columns) If the symmetric matrices $H_1, \ldots, H_m \in \mathbb{R}^{n \times n}$ are such that the $n \times m$ matrix

$$A_v := (H_j v, \ldots, H_m v)$$

has rank at most one for all $v \in \mathbb{R}^n$, then there exist $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ and a symmetric matrix $H \in \mathbb{R}^{n \times n} \setminus \{0\}$ such that $H_j = \alpha_j H$ for $j = 1 \ldots m$.

Lemma A.2 (Directional derivatives of rank one) Let $A$ be a neighborhood of $0 \in \mathbb{R}^n$ and let $h : A \to \mathbb{R}^{n \times m}$ be a function of class $C^1$, and for $1 \leq \ell \leq m$ let $h_\ell(x)$ be the $\ell$th column of $h(x)$. If $h(0) = 0$ and $\text{rank}(h(x)) \leq 1$ for all $x \in A$, then for every $v \in \mathbb{R}^n$ the $n \times m$ matrix

$$A_v := (Dh_1(x)v, \ldots, Dh_m(x)v)$$

has rank at most one.

The proof of Lemma 5.2 uses the following auxiliary results:

Theorem A.1 (Dines’ Theorem) If the symmetric matrices $A, B \in \mathbb{R}^{n \times n}$ are such that for all $x \in \mathbb{R}^n \setminus \{0\}$ either $x^T Ax \neq 0$ or $x^T Bx \neq 0$, then the set

$$\mathcal{R}(A, B) := \left\{ (x^T Ax, x^T Bx) : x \in \mathbb{R}^n \right\} \subset \mathbb{R}^2$$

(25)

is either $\mathbb{R}^2$ itself or a closed angular sector of angle less than $\pi$.

Lemma A.3 (Definite separation) Let $I \subset \mathbb{R}$ be a compact interval. If the symmetric matrices $A, B \in \mathbb{R}^{n \times n}$ are such that for all $x \in \mathbb{R}^n \setminus \{0\}$ there exists $\gamma_x \in I$ such that

$$x^T (A + \gamma_x B) x > 0,$$

for all $x \in \mathbb{R}^n \setminus \{0\}$.
Dines’ Theorem is proved in [12], and in the rest of this appendix we prove our lemmas in the order in which they were stated.

**Proof of Lemma 5.1** Applying Lemma A.2 to the function \( h(x) := Dc(x) \) we conclude that for every \( v \in \mathbb{R}^n \) the \( n \times m \) matrix

\[
A_v := \left( \nabla^2 c_1(0)v, \ldots, \nabla^2 c_m(0)v \right)
\]

has rank at most one, and Lemma A.1 yields the coefficients \( \alpha_j \) and the matrix \( H \). □

**Proof of Lemma 5.2** For every \( k \in \mathbb{N} \), Lemma A.3 yields \( \gamma_k \in I \) such that

\[
x^T \left( A + \gamma_k B + \frac{1}{k} I_{n \times n} \right) x > 0
\]

for all \( x \in \mathbb{R}^n \). Since the sequence \( \gamma_k \) is bounded, it has a subsequence which converges to some \( \gamma^* \in I \). This \( \gamma^* \) is as required by Lemma 5.2. □

**Proof of Lemma A.1** The Lemma holds when \( n = 1 \) or \( m = 1 \). Let us then assume that it holds when for \( n - 1 \geq 1 \) or \( m - 1 \geq 1 \) and show that it also holds for \( m \) and \( n \). If some \( H_j \) is zero then we can take \( \alpha_j = 0 \) and use induction for \( H_1, \ldots, H_{j-1}, H_{j+1}, \ldots, H_m \).

Therefore, we can assume that \( H_j \neq 0 \) for all \( j \). It follows that \( H_1 \) has an eigenvalue decomposition \( H_1 = QDQ^T \) with \( d_{11} \neq 0 \). By replacing \( H_j \) by \( Q^T H_j Q \) for all \( j \), we can assume that \( Q = I_{n \times n} \). Taking \( v = 1/d_{11} e_1 \), we obtain that \( H_1 v = e_1 \) and the hypothesis that the matrix \( A_v \) has rank one implies that \( H_j v = \alpha_j e_1 \) for all \( j \). Since the matrices \( H_j \) are symmetric, all of them have the form

\[
H_j = \begin{pmatrix}
\alpha_j & 0^T_{n-1} \\
0_{n-1} & H_j'
\end{pmatrix},
\]

for symmetric matrices \( H_j' \in \mathbb{R}^{(n-1) \times (n-1)} \setminus \{0\} \) which satisfy the hypothesis of Lemma A.1 with \( n = n - 1 \). Therefore, there exists a matrix \( H' \neq 0 \) and \( \alpha'_1, \ldots, \alpha'_m \) such that \( H_j' = \alpha'_j H' \) for \( j = 1, \ldots, m \). Since \( H_1 e_1 = d_{11} e_1 = \alpha_1 e_1 \neq 0 \), we have that \( \alpha_1 \neq 0 \). Let then \( v' \in \mathbb{R}^{n-1} \) be such that \( H' v' \neq 0 \), and write

\[
v = \begin{pmatrix} 1 \\ v' \end{pmatrix}.
\]

We have that

\[
H_1 v = \alpha_1 \begin{pmatrix} 1 \\ \alpha'_j H' v' \end{pmatrix} \quad \text{and} \quad H_j v = \begin{pmatrix} \alpha_j \\ \alpha'_j H' v' \end{pmatrix} \quad \text{for } j > 1,
\]
and the vectors $H_1 v$ and $H_j v$ are aligned by hypothesis. Moreover, either $\alpha_j \neq 0$ or $\alpha_j' \neq 0$, because $H_j \neq 0$. It follows that $\alpha_j' = \alpha_1' \alpha_j / \alpha_1$ for all $j$. As a result, $H_j = \alpha_1 H$, where

$$H := \begin{pmatrix} 1 & 0^T_{n-1} \\ 0_{n-1} & \alpha_1' \alpha_j H' \end{pmatrix}$$

and we are done.

**Proof of Lemma A.2** For every $v \in \mathbb{R}^n$ and $1 \leq \ell \leq m$, the facts that $h \in C^1$ and $h(0) = 0$ imply that

$$\lim_{\delta \to 0} \frac{1}{\delta} h(\delta v) = Dh(0) v = A_v.$$ 

Let $\rho > 0$ be such that if $\|B - A_v\| \leq \rho$ then $\text{rank}(B) \geq \text{rank}(A_v)$. Taking $\delta$ such that

$$\left\| \frac{1}{\delta} h(\delta v) - A_v \right\| < \rho$$

we obtain that

$$1 \geq \text{rank}(h(\delta v)) = \text{rank}\left( \frac{1}{\delta} h(\delta v) \right) \geq \text{rank}(A_v).$$

Therefore, $\text{rank}(A_v) \leq 1$, and we are done. \qed

**Proof of Lemma A.3** Let us write $I = [a, b]$. If $a = b$ then we could simply take $\gamma^* = a$ and we would be done. Therefore, we can assume that $a < b$. If $(u, v) \in \mathcal{R}(A, B)$ then $u + \gamma v \geq 0$ for some $\gamma \in [a, b]$. This implies that

$$u \geq -|\gamma| |v| \geq -(1 + |a| + |b|) |v|$$

Therefore, $(-2 (1 + |a| + |b|), 1) \notin \mathcal{R}(A, B)$ and by Dines’ Theorem $\mathcal{R}(A, B)$ is a closed pointed cone. For each $\gamma \in I$, the set

$$\mathcal{C}_\gamma := \{(u, v) \in \mathbb{R}^2 : u + \gamma v \leq 0\}$$

is a closed convex cone, and

$$\mathcal{C} := \bigcap_{\gamma \in I} \mathcal{C}_\gamma = \bigcap_{\gamma \in [a, b]} \mathcal{C}_\gamma$$

is also a closed cone. Moreover, $\mathcal{C}$ is pointed because $a < b$. The hypothesis tells us that for every $(u, v) \in \mathcal{R}(A, B) \setminus \{(0, 0)\}$ there exists $\gamma \in I$ such that $(u, v) \notin \mathcal{C}_\gamma$. \qed

Springer
and this implies that $C \cap \mathcal{R}(\mathbf{A}, \mathbf{B}) = \{(0, 0)\}$. Therefore, there exists $(c, d) \in \mathbb{R}^2$ such that
\[
cu + dv > 0 \quad \text{for } (u, v) \in \mathcal{R}(\mathbf{A}, \mathbf{B}) \setminus \{(0, 0)\}
\]
and
\[
cu + dv < 0 \quad \text{for } (u, v) \in C \setminus \{(0, 0)\}.
\]
Equation (26) shows that $(-1, 0) \in C_\gamma$ for all $\gamma$. Therefore, $(-1, 0) \in C$ and Eq. (28) implies that $c > 0$, and by dividing Eqs. (27) and (28) by $c$ if necessary, we can assume that $c = 1$. The point $(a, -1)$ belongs to $C_\gamma$ for all $\gamma \geq a$. Therefore, $(a, -1) \in C$ and Eq. (28) implies that $a - d < 0$, that is $d > a$. Similarly, The point $(-b, 1)$ belongs to $C_\gamma$ for all $\gamma \leq b$, and $-b + d < 0$, that is, $d < b$. In summary, we have that $d \in [a, b]$. Finally, we take $\gamma^* = d$, and Eq. (27) with $c = 1$ shows that this is a valid choice.

**References**

1. Behling, R., Haeser, G., Ramos, A., Viana, D.: On a conjecture in second-order optimality conditions (2017). arXiv:1706.07833v1 [math.OC]
2. Guckenheimer, J., Holmes, P.: Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields. Springer, Berlin (1983)
3. Perko, L.: Differential Equations and Dynamical Systems. Springer, New York (1991)
4. Mascarenhas, W.F.: The affine scaling algorithm fails for stepsize 0.999. SIAM J. Optim. 7, 34–46 (1997)
5. Mascarenhas, W.F.: Newton’s iterates can converge to non-stationary points. Math. Program. 112, 327–334 (2008)
6. Mascarenhas, W.F.: The BFGS algorithm with exact line searches fails for nonconvex functions. Math. Program. 99, 49–61 (2004)
7. Mascarenhas, W.F.: On the divergence of line search methods. Comput. Appl. Math. 26, 129–169 (2007)
8. Mascarenhas, W.F.: The divergence of the BFGS and Gauss Newton methods. Math. Program. 147, 253–276 (2014)
9. Andreani, R., Martínez, J.M., Schuverdt, M.L.: On second-order optimality conditions for nonlinear programming. Optimization 56, 529–542 (2007)
10. Spivak, M.: Calculus on Manifolds, A Modern Approach to Classical Theorems of Advanced Calculus. Addison-Wesley, Boston (1965)
11. Haeser, G.: An extension of Yuan’s lemma and its applications in optimization. J. Optim. Theory Appl. 174, 641–649 (2017)
12. Dines, L.L.: On the mapping of quadratic forms. Bull. Am. Math. Soc. 47, 494–498 (1941)