Transposition theorems and qualification-free optimality conditions

Hermann Schichl and Arnold Neumaier

Fakultät für Mathematik, Universität Wien
Nordbergstr. 15, A-1090 Wien, Austria
email: Hermann.Schichl@esi.ac.at, Arnold.Neumaier@univie.ac.at
WWW: http://www.mat.univie.ac.at/~neum/

February 20, 2006

Abstract. New theorems of the alternative for polynomial constraints (based on the Positivstellensatz from real algebraic geometry) and for linear constraints (generalizing the transposition theorems of Motzkin and Tucker) are proved.

Based on these, two Karush-John optimality conditions – holding without any constraint qualification – are proved for single- or multi-objective constrained optimization problems. The first condition applies to polynomial optimization problems only, and gives for the first time necessary and sufficient global optimality conditions for polynomial problems. The second condition applies to smooth local optimization problems and strengthens known local conditions. If some linear or concave constraints are present, the new version reduces the number of constraints for which a constraint qualification is needed to get the Kuhn-Tucker conditions.

Keywords: certificate of global optimality, first order optimality conditions, Fritz John conditions, Karush-John conditions, global optimality condition, global optimization, Kuhn-Tucker conditions, Mangasarian-Fromovitz constraint qualification, necessary and sufficient conditions, Positivstellensatz, second order optimality conditions, theorem of the alternative, transposition theorem

2000 MSC Classification: 90C30

1 Introduction

In this paper, we present a number of theorems that are useful for the global analysis of optimization problems, i.e., the assessment of their feasibility, and the construction and verification of a global solution. Several of the results are, however, also relevant for local optimization.

In constrained optimization, first and second-order optimality conditions play a central role, as they give necessary and/or sufficient conditions for a point to attain a local or global minimum of the problem considered, and thus define the goals that numerical methods should try to satisfy.

The various conditions currently available usually depend on qualitative conditions (concerning smoothness, linearity, convexity, etc.) that delineate the problem class, and on technical conditions, so-called constraint qualifications, that allow one to avoid certain difficulties in proofs, or certain known counterexamples.

The proof of the optimality conditions depend crucially on the availablility of certain theorems of the alternative, which state that among two alternative existence statements, exactly one can be satisfied. Thus a theorem of the alternative may serve to define certificates whose presence imply the solvability of one alternative and the unsolvability of the other alternative.

Recent advances in global optimization [27, 30] make it possible in many cases to find and verify the global optimality of a solution, or to verify that no feasible point exists. Certificates acquire in this case a special importance, particularly in the context of computer-assisted proofs.

However, in order to apply a necessary optimality condition to rule out candidate solutions, or a sufficient optimality condition to verify the existence of a solution, it is important that these conditions are valid under conditions that can be checked explicitly. Therefore the optimality conditions should not depend on any constraint qualification.

Optimality conditions characterizing the solutions of smooth nonlinear programming problems by first order necessary conditions are often called Fritz John conditions if they apply without constraint qualification, after Fritz John [15], who rediscovered unpublished earlier results of Karush [16]; see Kuhn [18, Section 6] for a history. Therefore, we shall refer to such conditions as **Karush-John optimality conditions**.

The importance of the Karush-John conditions stem from the fact that they apply with-

out any hypothesis on the optimization problem (apart from smoothness). For the known (second-order) sufficient conditions, a similar result was not known before, sufficiency requiring very strong nondegeneracy conditions. It is therefore remarkable that, *for polynomial optimization problems*, it is possible to formulate necessary and sufficient conditions for (global) optimality, valid without any restriction. These strong results are based on the so-called Positivstellensatz, a polynomial analogue of the transposition theorem for linear systems. The Positivstellensatz is a highly nontrivial tool from real algebraic geometry which has been applied recently also in an algorithmic way for the solution of global polynomial optimization problems. Some of the consequences of the Positivstellensatz are implemented in the packages GloptiPoly (Henrion & Lasserre [12, 13, 14]) and SOSTOOLS (Prajna et al. [32]).

Related results in this direction are in Lasserre [20]. He proved in Theorem 4.2 a sufficient condition for global optimality in polynomial optimization problems, which is a special case of our necessary and sufficient conditions. (The unconstrained minimization of the Motzkin polynomial shows that Lasserre's condition is not sufficient.) He shows that his certificates can be interpreted as polynomial multipliers in a fashion analogous to the Kuhn-Tucker optimality conditions. Instead of necessary conditions he obtains under some compactness assumption an infinite sequence of semidefinite relaxations whose optimal values converge to the global optimum.

In this article we derive in Section 2 polynomial transposition theorems and deduce from them a global Karush-John condition which is necessary and sufficient conditions for global optimality of *polynomial* programs.

Section 3 then proves a very general transposition theorem for *linear* constraints, establishing a theorem of the alternative from which the transposition theorems of Motzkin [24] and of Tucker [34] (as well as many weaker ones) can be obtained as corollaries. This level of generality is necessary to deduce in Section 4 a form of the constraint qualifications for the Kuhn-Tucker optimality conditions for general *smooth nonlinear* programming problems, which is stronger (i.e., makes more stringent assertions about the multipliers) than the known Karush-John conditions and also applies for multiple objectives.

Our Karush-John conditions imply derived Kuhn-Tucker conditions with linear independence constraint qualifications for fewer constraints than the conditions found in the literature. In particular, they imply the known result that for concavely (or linearly) constrained problems no constraint qualification is needed.

The new conditions will be incorporated in the COCONUT environment [9] for deterministic

global optimization; the local Karush-John conditions from Section 4 are already in place.

Notation. In the following, $\mathbb{R}$ is the field of real numbers, and $\mathbb{N}_0$ the set of nonnegative integers. To denote monomials and their degree, we use the multiindex notation

$$x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}, \quad |\alpha| = \alpha_1 + \cdots + \alpha_n$$

($x \in \mathbb{R}^n, \alpha \in \mathbb{N}_0^n$). Inequalities ($\leq$, $\geq$) and strict inequalities ($<$, $>$) between vectors and matrices are interpreted componentwise. However, disequality ($\neq$) is the negation of equality ($=$) and hence not interpreted componentwise. The infimum $\inf\{x, y\}$ of two vectors x, y of the same size is taken in the partial order $\leq$, and is equivalent to the componentwise minimum. In particular, the condition $\inf\{x, y\} = 0$ is just the complementarity condition $x \geq 0, y \geq 0, x_i y_i = 0$ for all i. By e we denote a column vector of arbitrary size all of whose entries have the value 1. $[A, B]$ denotes the $m \times (n + p)$-matrix formed by juxtaposition of the $m \times n$-matrix A and the $m \times p$-matrix B. Zero dimensional vectors and matrices (needed to avoid stating many special cases) are handled according to the conventions in de Boor [7]; in addition, any of the relations $=$, $<$, $\leq$ (but not $\neq$) between zero dimensional objects is considered to be valid.

2 Global optimality conditions for polynomials

It is well-known that first order (Kuhn-Tucker) optimality conditions for constrained (single-objective) optimization are sufficient for convex problems, but not in general. For nonconvex problems, they must be complemented by second order conditions, which come in two forms – as necessary conditions and as sufficient conditions, and they apply to local optimality only. Moreover, between necessary and sufficient conditions is a theoretical gap, in which various degenerate exceptional situations are possible. It is therefore remarkable that, for polynomial systems, it is possible to bridge this gap and formulate necessary and sufficient conditions for (global) optimality, valid without any restriction.

The following discussion is based on a polynomial analogue of the transposition theorem (Theorem 3.4), the so-called Positivstellensatz, a highly nontrivial result from real algebraic geometry. To present this result, we need some definitions.

$\mathbb{N}_0$ denotes the set of nonnegative integers. $\mathbb{R}[x_{1:n}] := \mathbb{R}[x_1, \ldots, x_n]$ denotes the algebra of polynomials in the indeterminates $x_1, \ldots, x_n$ with real coefficients. Let $R_i \in \mathbb{R}[x_{1:n}]$

$(i = 1 : k)$ be a finite family of polynomials, combined in the vector $R = (R_1, \ldots, R_k)^T$. The **ideal** generated by the R_i is the vector space

$$I\langle R\rangle = I\langle R_1, \ldots, R_k\rangle := \Big\{ \sum_{i=1}^{k} a_i R_i \;\Big|\; a_i \in \mathbb{R}[x_{1:n}] \Big\}.$$

The **multiplicative monoid** generated by the R_i is the semigroup

$$M\langle R\rangle = M\langle R_1, \ldots, R_k\rangle := \Big\{ \prod_{i=1}^{k} R_i^{e_i} \;\Big|\; e_i \in \mathbb{N}_0 \Big\}. \tag{1}$$

A **polynomial cone** C is a subset of $\mathbb{R}[x_{1:n}]$ containing all squares a^2 with $a \in \mathbb{R}[x_{1:n}]$, such that $r + s, rs \in C$ whenever $r, s \in C$. The smallest polynomial cone is the set SOS of polynomials which can be represented as sums of squares; we call such polynomials **SOS polynomials**. The **polynomial cone** generated by the R_i is the smallest polynomial cone containing $R_1, \ldots, R_k$; it is given by

$$C\langle R\rangle = C\langle R_1, \ldots, R_k\rangle = \Big\{ y_0 + Y^T R_S \;\Big|\; y_0 \in SOS,\ Y \in SOS^{2^k} \Big\}. \tag{2}$$

where R_S denotes the vector containing the 2^k polynomials in the **squarefree part**

$$S\langle R\rangle = S\langle R_1, \ldots, R_k\rangle := \Big\{ \prod_{i=1}^{k} R_i^{e_i} \;\Big|\; e_i \in \{0, 1\} \Big\}$$

of $M\langle R_1, \ldots, R_k\rangle$.

2.1 Theorem. (Polynomial transposition theorem I)

Let P, Q, and R be vectors of polynomials. Then exactly one of the following holds:

(i) $P(x) \geq 0$, $\quad Q(x) = 0$, $\quad R_i(x) \neq 0$ for $i = 1, \ldots, k$, $\quad$ for some $x \in \mathbb{R}^n$,

(ii) $f + g + h = 0$ for some $f \in C\langle P\rangle$, $g \in I\langle Q\rangle$, and $h \in M\langle R_1^2, \ldots, R_k^2\rangle$.

Proof. That conditions (i) and (ii) are mutually inconsistent can easily be seen. Indeed, if (i) holds then for any $f \in C\langle P\rangle$, $g \in I\langle Q\rangle$, and $h \in M\langle R_1^2, \ldots, R_k^2\rangle$, we have $f(x) \geq 0$, $g(x) = 0$, and $h(x) > 0$, whence $f(x) + g(x) + h(x) > 0$, contradicting (ii). That one of the two conditions can always be satisfied is the hard part. It follows from the statement

that the inconsistency of (i) implies the solvability of (ii), which is equivalent to the Weak Positivstellensatz stated and proved as Theorem 4.4.2 in BOCHNAK et al. [4]. □

For our application to optimality conditions, we need the following slightly different formulation.

2.2 Theorem. (Polynomial transposition theorem II)

Let P, Q, and R be vectors of polynomials. Then exactly one of the following holds:

(i) $P(x) \geq 0$, $Q(x) = 0$, and $R(x) > 0$ for some $x \in \mathbb{R}^n$,

(ii) $f + g + h = 0$ for some $f \in C\langle P, R\rangle$, $g \in I\langle Q\rangle$, and $h \in M\langle R\rangle$.

Proof. That conditions (i) and (ii) are mutually inconsistent can again easily be seen. Indeed, if (i) holds then for any $f \in C\langle P\rangle$, $g \in I\langle Q\rangle$, and $h \in M\langle R\rangle$, we have $f(x) \geq 0$, $g(x) = 0$, and $h(x) > 0$, whence $f(x) + g(x) + h(x) > 0$, contradicting (ii).

If, on the other hand, (i) is inconsistent, this implies that the system $\big(P(x), R(x)\big) \geq 0$, $Q(x) = 0$, and $R(x) \neq 0$ is inconsistent, and by Theorem 2.1, there exist $f \in C\langle P, R\rangle$, $g \in I\langle Q\rangle$, and $h \in M\langle R_1^2, \ldots, R_k^2\rangle$ with $f + g + h = 0$. Since $M\langle R_1^2, \ldots, R_k^2\rangle \subset M\langle R\rangle$, the result follows. □

Both versions have the following common generalization.

2.3 Theorem. (General polynomial transposition theorem)

Let P, Q, R, and $S_1, \ldots, S_k$ be vectors of polynomials. Then exactly one of the following holds:

(i) $P(x) \geq 0$, $Q(x) = 0$, $R(x) > 0$, and $S_i(x) \neq 0$ for $i = 1, \ldots, k$, for some $x \in \mathbb{R}^n$,

(ii) $f + g + h = 0$ for some $f \in C\langle P, R\rangle$, $g \in I\langle Q\rangle$, and $h \in M\langle R, S_1^T S_1, \ldots, S_k^T S_k\rangle$.

Proof. That conditions (i) and (ii) are mutually inconsistent can be proved as before. Given (i) holds, then for any $f \in C\langle P, R\rangle$, $g \in I\langle Q\rangle$, and $h \in M\langle R, S_1^T S_1, \ldots, S_k^T S_k\rangle$, we have $f(x) \geq 0$, $g(x) = 0$, and $h(x) > 0$, leading to $f(x) + g(x) + h(x) > 0$, contradicting (ii).

The fact that (i) is inconsistent implies that the system of constraints $R(x) \geq 0$, $Q(x) = 0$, and $\big(R(x), S_1(x)^T S_1(x), \ldots, S_k(x)^T S_k(x)\big) > 0$ is inconsistent, and by Theorem 2.2, there exist polynomials $f \in C\langle P, R\rangle$, $g \in I\langle Q\rangle$, and $h \in M\langle R, S_1^T S_1, \ldots, S_k^T S_k\rangle$ with $f+g+h = 0$. □

The equivalence of the three transposition theorems 2.1, 2.2, and 2.3 can be seen by taking $R(x) \equiv 1 \in \mathbb{R}$ in Theorem 2.3 and noting that $C\langle P, 1\rangle = C\langle P\rangle$ and $M\langle 1, S_1^T S_1, \ldots, S_k^T S_k\rangle = M\langle S_1^2, \ldots, S_k^2\rangle$ when all S_j are scalars.

The following result gives necessary and sufficient conditions for the global optimality of a feasible point of an optimization problem defined in terms of a polynomial objective function f and polynomial constraints. In most applications, f will be a real-valued function.

However, it is not difficult to state and prove analogous conditions for multiobjective optimization problems, by allowing f to be vector-valued. In this case, optimality is replaced by Pareto optimality, defined as follows. The point $\widehat{x}$ is called **weakly Pareto minimal** with respect to the continuous function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ for f on X if $x \in X$ and there exists a neighborhood of $\widehat{x}$ in X which does not contain a point y with $f(y) < f(\widehat{x})$.

2.4 Theorem. (Global Karush-John conditions)

Let $\hat{x}$ be a feasible point of the polynomial Pareto-optimization problem

$$\begin{array}{ll} \mathit{min} & f(x) \\ \mathit{s.t.} & C(x) \geq 0, \\ & F(x) = 0, \end{array} \tag{3}$$

where $f \in \mathbb{R}[x_{1:n}]^k$ and $C \in \mathbb{R}[x_{1:n}]^m$, $F \in \mathbb{R}[x_{1:n}]^r$ are vectors of polynomials in $x_{1:n}$. Write B for the vector obtained by concatenating C with the vector $G(x) = f(\hat{x}) - f(x)$, so that $B_k = C_k$ for $k \leq m$. Then the following are equivalent:

(i) The point $\hat{x}$ is a global weak Pareto-minimum of (3)*.*

(ii) There are a polynomial $y_0 \in SOS$, polynomial vectors $Y \in SOS^{2^{m+1}}$, $Z \in \mathbb{R}[x_{1:n}]^r$, and a multiindex $\alpha \in \mathbb{N}_0^k$ with $|\alpha| > 0$ such that

$$G(x)^\alpha + y_0(x) + Y(x)^T B_S(x) + Z(x)^T F(x) = 0 \tag{4}$$

identically in x.

Moreover, any solution of (4) *satisfies*

$$y_0(\hat{x}) = 0, \quad \inf\{Y(\hat{x}), B_S(\hat{x})\} = 0, \quad F(\hat{x}) = 0, \tag{5}$$

$$\delta_{|\alpha|\,1} f_i'(\hat{x})^T = B_S{}'(\hat{x})^T Y(\hat{x}) + F'(\hat{x})^T Z(\hat{x}), \tag{6}$$

where $\alpha_i = 1$.

Proof. $\hat{x}$ is a global weak Pareto-minimum of (3) iff the conditions

$$C(x) \geq 0, \quad F(x) = 0, \quad f(x) < f(\hat{x})$$

are inconsistent. Because $f(x) < f(\hat{x})$ iff $G > 0$, the polynomial transposition theorem, Theorem 2.2, applies and shows that this is equivalent to the existence of polynomials $q \in C\langle B\rangle$, $r \in I\langle F\rangle$ and $s \in M\langle G\rangle$ with $q + r + s = 0$. Expressing this more explicitly using (1) and (2) shows this to be equivalent to (ii) without the constraint (5), and α only restricted to be a nonnegative multiindex. The equivalence of (i) and (ii) follows if we show that $|\alpha| \neq 0$.

Since $\hat{x}$ is feasible, we have $B(\hat{x}) \geq 0$, $F(\hat{x}) = 0$, and by construction, $G(\hat{x}) = 0$. Moreover, as sum of squares, $y_0(\hat{x}) \geq 0$ and $Y(\hat{x}) \geq 0$. Inserting $x = \hat{x}$ into (4) gives, with the Kronecker δ,

$$\delta_{0\,|\alpha|} \leq \delta_{0\,|\alpha|} + y_0(\hat{x}) + Y(\hat{x})^T B_S(\hat{x}) = 0.$$

This indeed forces $|\alpha| > 0$.

We also get $y_0(\hat{x}) = 0$, and $Y(\hat{x})^T B_S(\hat{x}) = 0$. But the latter inner product is a sum of nonnegative terms, hence each product vanishes, giving the complementarity conditions (5). Differentiating the relation (4) and evaluating the result at $\hat{x}$ yields

$$\begin{aligned} 0 = \sum_{\substack{i=1 \\ \alpha_i > 0}}^{k} \alpha_i G^{\alpha - e_i}(\hat{x}) G_i'(\hat{x})^T + y_0'(\hat{x}) + Y'(\hat{x})^T B_S(\hat{x}) + B_S{}'(\hat{x})^T Y(\hat{x}) \\ + Z'(\hat{x})^T F(\hat{x}) + F'(\hat{x})^T Z(\hat{x}). \end{aligned} \tag{7}$$

We now note that $y_0'(\hat{x}) = 0$ because $y_0(\hat{x}) = 0$ and y_0 is SOS. Together with the facts that $|\alpha| > 0$, $G(\hat{x}) = 0$, and $F(\hat{x}) = 0$, we can simplify (7) and get

$$0 = \delta_{|\alpha|\,1} G_i'(\hat{x})^T + Y'(\hat{x})^T B_S(\hat{x}) + B_S{}'(\hat{x})^T Y(\hat{x}) + F'(\hat{x})^T Z(\hat{x}), \tag{8}$$

for that i with $\alpha_i = 1$. Finally, whenever $(B_S)_j(\hat{x}) \neq 0$, the complementarity conditions in (5) imply $Y_j(\hat{x}) = 0$ and then $Y_j'(\hat{x})^T = 0$ since Y_j is a SOS. Thus, $Y'(\hat{x})^T B_S(\hat{x}) = 0$, and (8) simplifies further to (6), upon noting that $G_i'(\hat{x}) = -f_i'(\hat{x})$. □

We may interpret the polynomials in (4) as a certificate that $\hat{x}$ is a global optimizer of (3). For applications in practice, one would first try to find $\hat{x}$ by local optimization or a heuristic global search, and then try to prove its globality by solving (4) with the side constraints (5). Note that the conditions are linear, except for the SOS conditions which give semidefinite constraints. Since the degree of the polynomials involved is not known a priori, one would solve a sequence of linear semidefinite feasibility problems on the finite-dimensional spaces of polynomials defined by limiting the total degree of the terms in (4) to $d = 1, 2, 3, ...$ Once a certificate is found one can stop.

Our theorem guarantees that this procedure will be finite if and only if $\hat{x}$ is indeed a global minimizer. In contrast, the method of Lasserre [20] yields an infinite sequence of semidefinite relaxations whose optimal values converge (under some compactness assumption) to the global optimum. There is no guarantee that the global optimum is found after finitely many steps. It would be interesting to combine the approaches to a constructive procedure for finding and verifying a global optimizer in finitely many steps.

For rigorous certification, one would have the additional problem of verifying the existence of an exact certificate close to the computed approximation.

We now relate the global Karush-John conditions to the traditional local conditions.

2.5 Corollary. (Kuhn-Tucker conditions) *If $\hat{x}$ is a global optimum of problem* (3) *with $k = 1$ and* (4) *holds with $\alpha = (1)$ then there are vectors $y \geq 0$ and z with*

$$\nabla f(\hat{x}) = C'(\hat{x})^T y + F'(\hat{x})^T z \tag{9}$$

and

$$\inf\{y, C(\hat{x})\} = 0. \tag{10}$$

Proof. We already know by Theorem 2.4 (6) that

$$\nabla f(\hat{x}) = {B_S}'(\hat{x})^T Y(\hat{x}) + F'(\hat{x})^T Z(\hat{x}).$$

we can write that in a slightly expanded way as follows

$$\nabla f(\hat{x}) = C_S'(\hat{x})^T Y^{(1)}(\hat{x}) + G_1'(\hat{x})^T C_S(\hat{x}) Y^{(2)}(\hat{x}) + F'(\hat{x})^T Z(\hat{x}).$$

Noting that $C_S(\hat{x})Y^{(2)}(\hat{x}) \geq 0$ and $G_1'(\hat{x})^T = -\nabla f(\hat{x})$ and expanding further we see that

$$\gamma \nabla f(\hat{x}) = \sum_{\beta \in \{0,1\}^m} \sum_{\substack{i=1 \\ \beta_i = 1}}^{m} C_i'(\hat{x})^T C^{\beta - e_i}(\hat{x}) Y_\beta^{(1)}(\hat{x}) + F'(\hat{x})^T Z(\hat{x}) = 0,$$

where $\gamma = 1 + C_S(\hat{x})Y^{(2)}(\hat{x}) > 0$. We reorder the sums and get

$$\gamma \nabla f(\hat{x}) = \sum_{i=1}^{m} C_i'(\hat{x})^T \sum_{\substack{\beta \in \{0,1\}^m \\ \beta_i = 1}} C^{\beta - e_i}(\hat{x}) Y_\beta^{(1)}(\hat{x}) + F'(\hat{x})^T Z(\hat{x}) = 0.$$

If we now set

$$y_i = \frac{1}{\gamma} \sum_{\substack{\beta \in \{0,1\}^m \\ \beta_i = 1}} C^{\beta - e_i}(\hat{x}) Y_\beta^{(1)}(\hat{x}), \quad z = \frac{1}{\gamma} Z(\hat{x}) \tag{11}$$

we get the required equality (9). For the complementarity conditions we calculate

$$C_i(\hat{x}) y_i = \frac{1}{\gamma} \sum_{\substack{\beta \in \{0,1\}^m \\ \beta_i = 1}} C^{\beta}(\hat{x}) Y_\beta^{(1)}(\hat{x}) = 0,$$

since by (5) all terms in the sum vanish. □

2.6 Example. We consider the simple optimization problem

$$\begin{aligned} &\min \ 2x - x^2 \\ &\text{s.t.} \ \ x \in [-1, 1]. \end{aligned}$$

Clearly, the objective function is concave, hence the global minimum $\hat{f} = -3$ is attained at the bound $\hat{x} = -1$. We can write

$$f(x) - f(\hat{x}) = 2x - x^2 + 3 = (1 + x)(3 - x) = 1(1 + x)(1 - x) + 2(1 + x).$$

Obviously, each term on the right is nonnegative, showing again that $\hat{x}$ is a global minimizer.

From this representation one reads off the certificate ($\alpha = 1, y_0 = 0, Y^T = (0, 0, 2, 0, 1, 0, 0, 0)$, $Z = ()$) satisfying (4), where the components of B_S are arranged in the order $1, 1 + x, 1 - x, x^2 - 2x - 3, 1 - x^2, \ldots$.

While this example is trivial, it shows the essentials. Higher dimensional examples only differ in the complexity of what has to be written. In many other situations, as e.g. in Example 2.7, the certificate will be very sparse and of low degree, thus simplifying the search for it.

By Theorem 2.4, for every global minimizer *all* terms in (4) have to vanish, i.e. for every global minimizer $\hat{x}$ we have that $G(\hat{x}) = 0$ and (5) and (6) are valid. Using this information frequently allows the identification of all global minimizers of an optimization problem when a certificate is available.

2.7 Example. We consider the optimization problem

$$\begin{array}{ll} \min & z - x \\ \text{s.t.} & x - y + z = 0 \\ & x^2 - 10x + y^2 \geq 0 \\ & x \in [0, 2], \quad y \geq 0. \end{array} \tag{12}$$

As Figure 1 shows, the point $(0, 0, 0)$ is a global optimizer with objective function value 0.

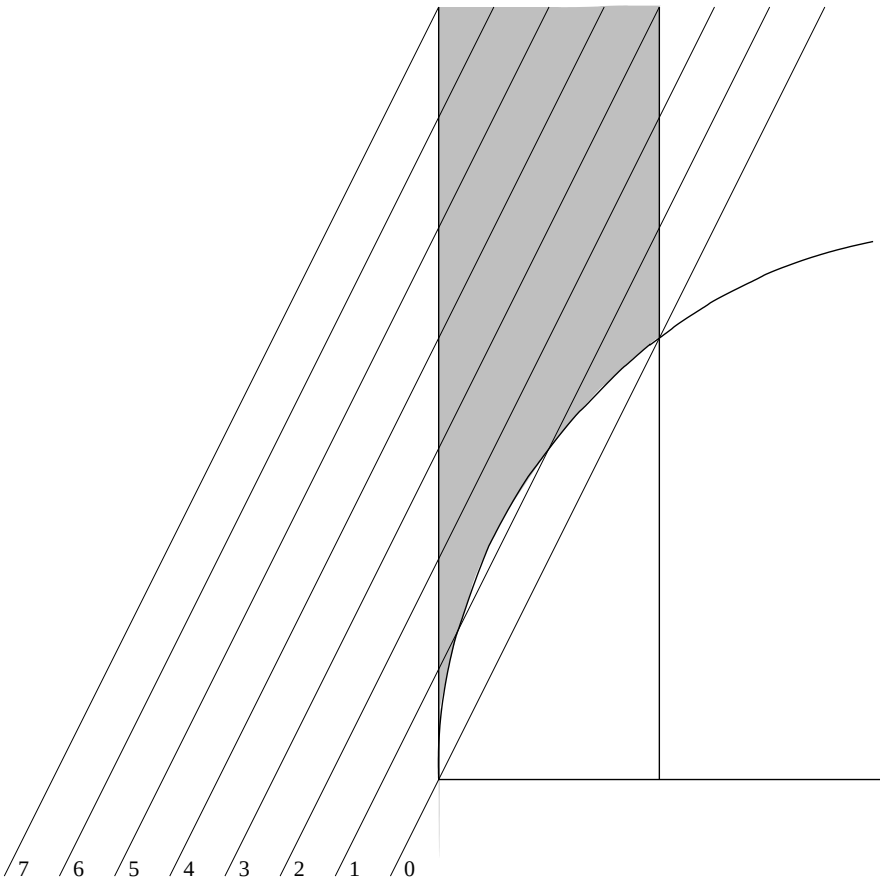

Figure 1: Feasible set (grey) and lines of constant value of f for Example 2.7, after eliminating z.

This can be validated formally by Theorem 2.4, using the following polynomial identity:

$$(x-z)+\tfrac{1}{2}y(2-x)+\tfrac{1}{4}y(x-z)+\tfrac{1}{4}x(2-x)+\tfrac{1}{4}(x^2-10x+y^2)+\tfrac{1}{4}(4+y)(x-y+z)=0.$$

By checking when in this identity all terms vanish, we see that for the point $(2,4,2)$ (and no third point), all terms of the certificate vanish as well. Hence (cf. also the figure), this point is another global minimizer of (12), and no other global minimizers exist.

2.8 Example. For the problem

$$\begin{array}{ll}\min & x+y+z\\ \text{s.t.} & y-y^2+z\geq 0\\ & 2x+y+y^2+z-2\geq 0\\ & x\geq 0\end{array}$$

we can find the polynomial identity

$$1-(x+y+z)+\tfrac{1}{2}\cdot(y-y^2+z)+\tfrac{1}{2}\cdot(2x+y+y^2+z-2)=0,$$

which implies that the global minimum value of the objective function is 1. By the complementarity conditions, we find that the two nonlinear inequality constraints must be active at every global minimum, i.e.

$$\begin{aligned} y-y^2+z&=0\\ 2x+y+y^2+z-2&=0,\end{aligned}$$

which implies that all

$$\hat{x}\in\left\{\begin{pmatrix}1-s^2\\ s\\ s(s-1)\end{pmatrix}\;\middle|\;s\in[-1,1]\right\}$$

are global optima.

Finally, the following example shows that in (4), the possibility $|\alpha|>1$ may indeed be necessary.

2.9 Example. For every nonnegative integer k the optimization problem

$$\begin{array}{ll}\min & x\\ \text{s.t.} & x^{2k+1}\geq 0\end{array}$$

admits the unique global optimizer $\hat{x} = 0$. The required polynomial identity of smallest degree is

$$(-x)^\alpha + 1 \cdot x^{2k+1} = 0, \quad \alpha = 2k+1.$$

In the GLOPTLAB package [28], an optimization package currently developed in MATLAB, methods are being implemented and tested, which work along the lines presented in this section. They use the SeDuMi [33] package for the semidefinite programming part and combine the present techniques with branch-and-bound, constraint propagation and linear relaxations. They have produced very promising results. At a later stage, the most successful techniques will be implemented as inference modules for the COCONUT environment [10].

3 Refined linear theorems of the alternative

In the linear case, there is a long tradition of theorems of the alternative, beginning with the Lemma of FARKAS [11], and culminating in the transposition theorems of MOTZKIN [24] and TUCKER [34]. These transposition theorems are concerned with the solvability of linear constraints of various forms (equations, inequalities, strict inequalities, disequalities); see, e.g., BROYDEN [6] for some history.

As we shall show, there is a single general transposition theorem, which contains the others as special cases. As for the latter, our starting point is the Lemma of Farkas:

3.1 Lemma. (Farkas) *Let $A \in \mathbb{R}^{m\times n}$ and $g \in \mathbb{R}^n$. Then exactly one of the following conditions can be satisfied.*

(i) $g^T p < 0$, $Ap \geq 0$ for some $p \in \mathbb{R}^n$.

(ii) $g = A^T q$, $q \geq 0$ for some $q \in \mathbb{R}^m$.

For the formulation of the transposition theorem and the constraint qualification we define $[\mathbf{1}, -u] =: E_u \in \mathbb{R}^{k\times(k+1)}$ with $0 < u \in \mathbb{R}^k$ and $\mathbf{1}$ being the identity matrix. We get the following result.

3.2 Lemma. *For $X \in \mathbb{R}^{n\times k}$ with $\mathrm{rk}X = n$, $0 < u \in \mathbb{R}^k$, $0 < v \in \mathbb{R}^\ell$, and $Y \in \mathbb{R}^{n\times\ell}$ there exists a matrix $0 \leq S \in \mathbb{R}^{(k+1)\times(\ell+1)}$ with*

$$XE_uS = YE_v.$$

Proof. Since $\mathrm{rk}X = n$ every $y \in \mathbb{R}^n$ can be written as a linear combination of the columns x_i of X:

$$y = \sum_{i=1}^{k} \lambda_i x_i.$$

Define

$$\mu = \max_{i=1}^{k} \left\{ -\frac{\lambda_i}{u_i}, 0 \right\}.$$

Then

$$y = \sum_{i=1}^{k} (\lambda_i + \mu u_i) x_i + \mu \left(- \sum_{i=1}^{k} u_i x_i \right) = X E_u s,$$

with $0 \geq s_i := \lambda_i + \mu u_i$ and $s_{k+1} := \mu$. Since all columns of $Y E_v \in \mathbb{R}^n$, the result follows. □

We prove the following general theorem of the alternative, and deduce from it the transposition theorems of Motzkin and Tucker.

3.3 Theorem. (General linear transposition theorem)

Consider matrices $A \in \mathbb{R}^{m_A \times n}$, $B \in \mathbb{R}^{m_B \times n}$, $C \in \mathbb{R}^{m_C \times n}$, and $D_j \in \mathbb{R}^{m_j \times n}$ with $m_j > 0$ for $j = 1, \ldots, N$. Then exactly one of the following holds.

(i) $Ax = 0, \quad Bx \geq 0, \quad Cx > 0, \quad$ and $\quad D_j x \neq 0, \;$ for $j = 1, \ldots, N, \quad$ for some $x \in \mathbb{R}^n$,

(ii) we have $m_C > 0$ and there exist $q \in \mathbb{R}^{m_A}$, $r \in \mathbb{R}^{m_B}$, and $s \in \mathbb{R}^{m_C}$ with

$$A^T q + B^T r + C^T s = 0, \quad r \geq 0, \quad s \geq 0, \; s \neq 0, \tag{13}$$

or for some $j \in \{1, \ldots, N\}$ there exist matrices $Q \in \mathbb{R}^{m_A \times (m_j + 1)}$ and $R \in \mathbb{R}^{m_B \times (m_j + 1)}$ with

$$A^T Q + B^T R = D_j^T E_u, \quad R \geq 0, \tag{14}$$

for some $u > 0$. Moreover, the same alternative holds if in (ii) a fixed vector $u > 0$ (such as the all-one vector $u = e$) is prescribed.

Proof. If (i) and (ii) hold simultaneously then multiplying (13) with x^T yields

$$(Bx)^T r + (Cx)^T s = 0.$$

Since $Bx \geq 0$ and $r \geq 0$ we have $(Cx)^T s \leq 0$, which is a contradiction to $Cx > 0$ and $s \geq 0$, $s \neq 0$. Multiplying, on the other hand, (14) by x^T we get

$$0 \leq (Bx)^T R = [(D_j x)^T, -(D_j x)^T u],$$

whence $D_j x \geq 0$ and $u^T D_j x < 0$ forces $D_j x = 0$, contradiction.

Now assume that (i) cannot be solved. Then, for all $j = 1, \ldots, N$ and all $v_j \in \mathbb{R}^{m_j}$, there is no $x \in \mathbb{R}^n$ with

$$Ax = 0, \quad Bx \geq 0, \quad Cx > 0, \quad \text{and} \quad v_j^T D_j x > 0. \tag{15}$$

Writing

$$g := \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad F := \begin{pmatrix} A & 0 \\ -A & 0 \\ B & 0 \\ C & e \\ v_1^T D_1 & 1 \\ \vdots & \vdots \\ v_N^T D_N & 1 \end{pmatrix}, \quad p := \begin{pmatrix} x \\ -\lambda \end{pmatrix},$$

we find that

$$g^T p < 0, \quad Fp \geq 0 \tag{16}$$

is unsolvable for $p \in \mathbb{R}^{n+1}$. By the Lemma of Farkas 3.1, we can find $q \in \mathbb{R}^{2m_A + m_B + m_C + N}$ with

$$F^T q = g, \quad q = \begin{pmatrix} \hat{a} \\ \bar{a} \\ b \\ c \\ \mu \end{pmatrix} \geq 0. \tag{17}$$

Writing $a := \hat{a} - \bar{a}$, we find the existence of vectors $a \in \mathbb{R}^{m_A}$, $b \in \mathbb{R}^{m_B}$, $c \in \mathbb{R}^{m_C}$, and $\mu \in \mathbb{R}^N$ (depending on the choice of the v_j) such that

$$A^T a + B^T b + C^T c + \sum_{j=1}^{N} \mu_j D_j^T v_j = 0, \quad e^T \begin{pmatrix} c \\ \mu \end{pmatrix} = 1, \quad b, c, \mu \geq 0. \tag{18}$$

For $M \leq N$, we consider the set S_M consisting of all $(v_1, \ldots, v_{M-1}) \in \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_{M-1}}$ for which (18) holds with $\mu_j = 0$ for $j \geq M$. Let $S_1 := \emptyset$. Let M be maximal with

$S_M \neq \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_{M-1}}$. If $M = 1$ we get $c \neq 0$, hence $m_C > 0$, and by setting $q := a$, $r := b$, and $s := c$ we find (13).

Hence we may assume that $M > 1$ and pick $(v_1, \ldots, v_{M-1}) \notin S_M$. Take an arbitrary $v_M \in \mathbb{R}^{m_M}$. We can find vectors a, a', $b \geq 0$, $b' \geq 0$, $c \geq 0$, $c' \geq 0$, $0 \leq \xi, \xi' \in \mathbb{R}^{M-1}$, and numbers $\lambda > 0$ and $\tilde{\lambda} > 0$ with

$$A^T a + B^T b + C^T c + \sum_{j=1}^{M-1} \xi_j D_j^T v_j + \lambda D_M^T v_M = 0, \qquad e^T \begin{pmatrix} c \\ \xi \end{pmatrix} + \lambda = 1, \tag{19}$$

$$A^T a' + B^T b' + C^T c' + \sum_{j=1}^{M-1} \xi'_j D_j^T v_j + \lambda' D_M^T (-v_M) = 0, \qquad e^T \begin{pmatrix} c' \\ \xi' \end{pmatrix} + \lambda' = 1. \tag{20}$$

Indeed, assume that we cannot find a, b, c, ξ, and λ with (19). We can get vectors a, b, c, μ satisfying (18). If there are only combinations with $\mu_{M+1:N} \neq 0$, then $(v_1, \ldots, v_M) \notin S_{M+1}$, contradicting the maximality of M. If there is a combination with $\mu_M = 0$, we find $(v_1, \ldots, v_{M-1}) \in S_M$, another contradiction. Thus $\mu_M \neq 0$, and we set $\xi := \mu_{1:M-1}$ and $\lambda := \mu_M$. The same argument gives (20). Combining (19) and (20) leads to

$$\begin{gathered} A^T \underbrace{\left(\tfrac{1}{\lambda} a + \tfrac{1}{\lambda'} a'\right)}_{:= q} + B^T \underbrace{\left(\tfrac{1}{\lambda} b + \tfrac{1}{\lambda'} b'\right)}_{:= r} + C^T \underbrace{\left(\tfrac{1}{\lambda} c + \tfrac{1}{\lambda'} c'\right)}_{:= s} + \sum_{j=1}^{M-1} \underbrace{\left(\tfrac{\xi_j}{\lambda} + \tfrac{\xi'_j}{\lambda'}\right)}_{:= \nu_j} D_j^T v_j = 0. \\ e^T \begin{pmatrix} s \\ \nu \end{pmatrix} = \frac{\lambda + \lambda'}{\lambda \lambda'} - 2 =: \sigma. \end{gathered} \tag{21}$$

If $s \neq 0$ or $\nu \neq 0$ the combination $(\mu_1, \ldots, \mu_{M-1}) := \sigma^{-1} \nu \geq 0$, $a := \sigma^{-1} q$, $b := \sigma^{-1} r \geq 0$, and $c := \sigma^{-1} s \geq 0$ proves that $(v_1, \ldots, v_{M-1}) \in S_M$, a contradiction. Thus $s = 0$, implying $c = c' = 0$, and $\nu = 0$, hence $\xi_1 = \xi'_1 = \cdots = \xi_{N-1} = \xi'_{N-1} = 0$, and $\lambda = 1$. Since v_M was arbitrary, we have shown that for all $v_M \in \mathbb{R}^{m_M}$ there exist vectors $a \in \mathbb{R}^{m_A}$ and $b \in \mathbb{R}^{m_B}$ with

$$A^T a + B^T b + D_M^T v_M = 0. \tag{22}$$

We set $j := M$ and choose for v_M in turn an arbitrary $u > 0$ and the vectors $w_k := -e_k$ $(k = 1, \ldots, m_j)$. This gives vectors q' and $r' \geq 0$ with

$$A^T q' + B^T r' = -D_j^T u.$$

and vectors q_k and $r_k \geq 0$ $(k = 1, \ldots, m_j)$ with

$$A^T q_j + B^T r_j = -D_j^T w_j.$$

Forming the matrices $Q := \left[q_1, \ldots, q_{m_j}, q'\right]$ and $R := \left[r_1, \ldots, r_{m_j}, r'\right]$ finally gives (14). □

The well known theorems of the alternative by MOTZKIN [24] and TUCKER [34] are consequences of this theorem.

3.4 Theorem. (Motzkin's linear transposition theorem)

Let $B \in \mathbb{R}^{m\times n}$, and let (I, J, K) be a partition of $\{1, \ldots, m\}$ with $K \neq \emptyset$. Then exactly one of the following holds:

(i) $(Bp)_I = 0$, $(Bp)_J \geq 0$, $(Bp)_K > 0$ *for some* $p \in \mathbb{R}^n$,

(ii) $B^T q = 0$, $q_{J\cup K} \geq 0$, $q_K \neq 0$ *for some* $q \in \mathbb{R}^m$.

Proof. We set $\tilde{A} := B_{I:}$, $\tilde{B} := B_{J:}$, $\tilde{C} := C_{K:}$, $N = 0$ and apply Theorem 3.3. □

3.5 Theorem. (Tucker's linear transposition theorem)

Let $B \in \mathbb{R}^{m\times n}$, and let (I, J, K) be a partition of $\{1, \ldots, m\}$ with $K \neq \emptyset$. Then exactly one of the following holds:

(i) $(Bp)_I = 0$, $(Bp)_{J\cup K} \geq 0$, $(Bp)_K \neq 0$ *for some* $p \in \mathbb{R}^n$,

(ii) $B^T q = 0$, $q_J \geq 0$, $q_K > 0$ *for some* $q \in \mathbb{R}^m$.

Proof. Set $\tilde{A} = -B^T$, define the matrix $\tilde{B}$ whose rows are indexed by $I \cup J \cup K$ and whose columns are indexed by J with $\tilde{B}_{J:} = \mathbf{1}$, $\tilde{B}_{I\cup K,:} = 0$, and introduce the matrix $\tilde{C}$ whose rows are indexed by $I \cup J \cup K$ with $\tilde{C}_{K:} = \mathbf{1}$, $\tilde{C}_{I\cup J,:} = 0$, and $N = 0$. Clearly, case (i) of Theorem 3.3 is equivalent to the solvability of the present (ii). On the other hand,, case (ii) of Theorem 3.3 is here equivalent to the existence of vectors q, $r \geq 0$, and $s \geq 0$, $s \neq 0$ with

$$\tilde{A}^T q + \tilde{B}^T r + \tilde{C}^T s = 0.$$

Plugging in the definitions of $\tilde{A}$, $\tilde{B}$, and $\tilde{C}$ this becomes

$$-B_{I:}q = 0, \quad -B_{J:}q + r = 0, \quad -B_{K:}q + s = 0,$$

which is clearly equivalent to (i). □

For the applications in the next section we need the following corollary of Theorem 3.3.

3.6 Corollary. *Let $B \in \mathbb{R}^{m\times n}$, and let (I, J, K) be a partition of $\{1, \dots, m\}$ of $i = |I|$, $j = |J|$, and $k = |K| > 0$ elements. Then exactly one of the following holds:*

(i) If $A = B^T_{K:}E_u$ for any $0 < u \in \mathbb{R}^k$, then $\operatorname{rk} A = k$ and for some matrix $P \in \mathbb{R}^{n\times(k+1)}$

$$(B(A+P))_{I:} = 0, \quad (B(A+P))_{J:} \geq 0, \quad \text{and} \quad (BP)_{K:} = 0.$$

(ii) $B^T q = 0$, $q_J \geq 0$, $q_K \neq 0$ for some $q \in \mathbb{R}^m$.

Proof. We set $\tilde{A} = -B^T$, define $\tilde{B} \in \mathbb{R}^{|I\cup J\cup K|\times|J|}$ with $\tilde{B}_{J:} = \mathbf{1}$, $\tilde{B}_{I\cup K,:} = 0$, construct $\tilde{D} \in \mathbb{R}^{|I\cup J\cup K|\times|K|}$ with $\tilde{D}_{K:} = \mathbf{1}$, $\tilde{D}_{I\cup J,:} = 0$, and set $N = 1$, $m_C = 0$, and $k = |K|$. Clearly, case (i) in Theorem 3.3 is equivalent to the present (ii).

On the other hand, (ii) in Theorem 3.3 is here equivalent to the existence of matrices Q and $R \geq 0$ with

$$B_{I:}Q = 0, \quad B_{J:}Q = R, \quad B_{K:}Q = E_u^T. \tag{23}$$

This, in turn, is equivalent to (i) by the following argument, for which we introduce the pseudo inverse $B^\dagger_{K:} = B^T_{K:}(B_{K:}B^T_{K:})^{-1}$ of $B_{K:}$.

Lets assume (i). By Lemma 3.2 we can find a matrix $S \geq 0$ with $(B_{K:}B^T_{K:})E_uS = E_u$, and we set $Q := B^\dagger_{K:}E_u + PS$. Then

$$\begin{gathered} B_{I:}Q = B_{I:}\big(B^T_{K:}(B_{K:}B^T_{K:})^{-1}E_u + PS\big) = B_{I:}(B^T_{K:}E_u + P)S = 0, \\ B_{J:}Q = B_{J:}\big(B^T_{K:}(B_{K:}B^T_{K:})^{-1}E_u + PS\big) = B_{J:}(B^T_{K:}E_u + P)S =: R \geq 0, \\ B_{K:}Q = B_{K:}B^\dagger_{K:}E_u + B_{K:}PS = E_u + 0. \end{gathered}$$

Now assume (23). The last equation implies $\operatorname{rk}B_{K:} = \operatorname{rk}A = k$, and so the pseudo inverse of $B_{K:}$ exists and Q is of the form $Q = B^\dagger_{K:}E_u + P'$ for some P' with $B_{K:}P' = 0$. By Lemma 3.2 we can find $S \geq 0$ with $(B_{K:}B^T_{K:})^{-1}E_uS = E_u$ and set $P := P'S$. Calculating

$$\begin{gathered} B_{I:}(A+P) = B_{I:}(B^T_{K:}E_u + P'S) = B_{I:}(B^\dagger_{K:}E_u + P')S = B_{I:}QS = 0, \\ B_{J:}(A+P) = B_{J:}(B^T_{K:}E_u + P'S) = B_{J:}(B^\dagger_{K:}E_u + P')S = B_{J:}QS = RS \geq 0, \\ B_{K:}P = B_{K:}P'S = 0, \end{gathered}$$

we prove (ii). □

4 A refinement of the Karush–John conditions

Karush-John conditions were originally derived – for single-objective optimization with inequality constraints only – by KARUSH [16], and rediscovered by JOHN [15]. They were subsequently generalized to mixed equality and inequality constraints, and to multiobjective optimization problems; there is a large literature on the subject, which can be accessed from the references below.

However, the Karush-John conditions in their most general form pose difficulties in applications, because the factor in front of the gradient term may be zero or very small. Therefore, most of the local solvers require a constraint qualification, like that of MANGASARIAN & FROMOVITZ [22] (MFCQ), to be able to reduce the Karush-John conditions to the much more convenient Kuhn-Tucker conditions [19]. Thorough discussions of such constraint qualifications can be found for single-objective optimization in BAZARAA et al. [3] and MANGASARIAN [21]. A more recent account is in BONNANS & SHAPIRO [5, Section 5.2]; there one can also find extensions to conic programming, semiinfinite programming, and infinite-dimensional problems (not considered in the present work). The Karush-John conditions have been investigated in the case of multiobjective optimization in MARUSCIAC [23], though the result implicitly contains a constraint qualification. Further reference can be found in PHAN & LE [31] and AGHEZZAF & HACHIMI [1], especially for connections to the constraint qualifications and, e.g., in CAMBINI [8] for second order conditions.

Deterministic global optimization algorithms cannot take this course, since it is not known beforehand whether the global optimum satisfies an assumed constraint qualification. Therefore, they have to use the Karush-John conditions in their general form (cf., e.g., KEARFOTT [17]). Unfortunately, the additional constraints needed involve all multipliers and are very inconvenient for the solution process.

In this section we prove a strong version of the Karush-John conditions for nonlinear programming and multiobjective optimization, and a corresponding relaxation of the Mangasarian-Fromovitz constraint qualification. Apart from the inverse function theorem, our main tools are the transposition theorems of the previous section. The treatment is along the lines of the special case of a single objective discussed in our unpublished paper [29].

We consider concave and nonconcave constraints separately, and introduce slack variables to transform all nonconcave constraints into equations. Thus we may write a general nonlinear

optimization problems without loss of generality in the form

$$\begin{aligned} &\min\ f(x) \\ &\text{s.t.}\ \ C(x) \geq 0, \quad F(x) = 0. \end{aligned} \tag{24}$$

In many applications, the objective function f will be a real-valued function. However, we allow f to be vector-valued; in this case, optimality is replaced by Pareto optimality.

The form (24), which separates the concave constraints (including bound constraints and general linear constraints) and the remaining nonlinear constraints, is most useful to prove our strong form of the Karush-John conditions. However, in computer implementations, a transformation to this form is not ideal, and the slack variables should be eliminated again from the optimality conditions.

4.1 Theorem. (General first order optimality conditions)

Let $f : U \to \mathbb{R}^k$, $C : U \to \mathbb{R}^m$, and $F : U \to \mathbb{R}^r$ be functions continuously differentiable on a neighborhood U of $\widehat{x} \in \mathbb{R}^n$. If C is convex on U and $\widehat{x}$ is a weakly Pareto minimal point of the nonlinear program (24)*, then there are vectors $\widehat{w} \geq 0 \in \mathbb{R}^k$, $\widehat{y} \in \mathbb{R}^m$, $\widehat{z} \in \mathbb{R}^r$ such that*

$$f'(\widehat{x})^T \widehat{w} = C'(\widehat{x})^T \widehat{y} + F'(\widehat{x})^T \widehat{z}, \tag{25}$$

$$\inf(\widehat{y}, C(\widehat{x})) = 0, \tag{26}$$

$$F(\widehat{x}) = 0, \tag{27}$$

and

$$\widehat{w}, \widehat{z} \text{ are not both zero.} \tag{28}$$

Proof. We begin by noting that a feasible point $\widehat{x}$ of (24) is also a feasible point for the optimization problem

$$\begin{aligned} &\min\ f(x) \\ &\text{s.t.}\ \ Ax \geq b \\ &\qquad F(x) = 0, \end{aligned} \tag{29}$$

where J is the set of all components j for which $C(\widehat{x})_j = 0$ and

$$A = C'(\widehat{x})_{J:}, \quad b = C'(\widehat{x})_{J:}\widehat{x}.$$

For the indices k corresponding to the set N of inactive constraints, we choose $y_N = 0$ to satisfy condition (26). Since C is convex, we have $C(x) \geq C(\widehat{x}) + C'(\widehat{x})(x - \widehat{x})$. Restricted to

the rows J we get $C(x)_J \geq C'(\widehat{x})_{J:}(x-\widehat{x})$. This fact implies that problem (24) is a relaxation of problem (29) on a neighborhood U of $\widehat{x}$. Note that since C is continuous we know that $C(x)_j > 0$ for $k \in N$ in a neighborhood of $\widehat{x}$ for all constraints with $C(\widehat{x})_j > 0$. Since, by assumption, $\widehat{x}$ is weakly Pareto minimal for a relaxation of (29) and a feasible point of (29), it is weakly Pareto minimal (29) as well. Together with the choice $y_N = 0$ the Karush-John conditions of problem (29) are again conditions (25)–(27). So we have successfully reduced the problem to the case where C is an affine function and all constraints are active at $\widehat{x}$.

Thus, in the following, we consider a weakly Pareto minimal point $\widehat{x}$ of the optimization problem (29) satisfying

$$A\widehat{x} = b. \tag{30}$$

If $\operatorname{rk} F'(\widehat{x}) < r$ then $z^T F'(\widehat{x}) = 0$ has a solution $z \neq 0$, and we can solve (25)–(28) with $\widehat{y} = 0$, $\widehat{w} = 0$. Hence we may assume that $\operatorname{rk} F'(\widehat{x}) = r$. This allows us to select a set R of r column indices such that $F'(\widehat{x})_{:R}$ is nonsingular. Let B be the $(0,1)$-matrix such that Bs is the vector obtained from $s \in \mathbb{R}^n$ by discarding the entries indexed by R. Then the function $\Phi : C \to \mathbb{R}^n$ defined by

$$\Phi(x) := \begin{pmatrix} F(x) \\ Bx - B\widehat{x} \end{pmatrix}$$

has at $x = \widehat{x}$ a nonsingular derivative

$$\Phi'(\widehat{x}) = \begin{pmatrix} F'(\widehat{x}) \\ B \end{pmatrix}.$$

Hence, by the inverse function theorem, Φ defines in a neighborhood of $0 = \Phi(\widehat{x})$ a unique continuously differentiable inverse function Φ^{-1} with $\Phi^{-1}(0) = \widehat{x}$. Using Φ we can define a curved search path with tangent vector $p \in \mathbb{R}^n$ tangent to the nonlinear constraints satisfying $F'(\widehat{x})p = 0$. Indeed, the function defined by

$$s_p(\alpha) := \Phi^{-1}\begin{pmatrix} 0 \\ \alpha Bp \end{pmatrix} - \widehat{x}$$

for sufficiently small $\alpha \geq 0$, is continuously differentiable, with

$$s_p(0) = \Phi^{-1}(0) - \widehat{x} = 0, \quad \begin{pmatrix} F(\widehat{x} + s_p(\alpha)) \\ Bs_p(\alpha) \end{pmatrix} = \Phi\left(\Phi^{-1}\begin{pmatrix} 0 \\ \alpha Bp \end{pmatrix}\right) = \begin{pmatrix} 0 \\ \alpha Bp \end{pmatrix},$$

hence

$$s_p(0) = 0, \quad F(\widehat{x} + s_p(\alpha)) = 0, \quad Bs_p(\alpha) = \alpha Bp. \tag{31}$$

Differentiation of (31) at $\alpha = 0$ yields

$$\begin{pmatrix} F'(\widehat{x}) \\ B \end{pmatrix} \dot{s}_p(0) = \begin{pmatrix} F'(\widehat{x})\dot{s}_p(0) \\ B\dot{s}_p(0) \end{pmatrix} = \begin{pmatrix} 0 \\ Bp \end{pmatrix} = \begin{pmatrix} F'(\widehat{x}) \\ B \end{pmatrix} p,$$

hence $\dot{s}_p(0) = p$, i.e., p is indeed a tangent vector to $\widehat{x} + s_p(\alpha)$ at $\alpha = 0$.

Since $\widehat{x}$ is weakly Pareto minimal, we know that there exists a neighborhood V of $\widehat{x}$ in the set of feasible points containing no y with $f(\widehat{x}) > f(y)$. Thus, for every $y \in V$ there exists an index j with $f_j(\widehat{x}) \leq f_j(y)$. Taking an arbitrary curved path γ in the feasible set with $\gamma(0) = \widehat{x}$ we conclude that there is an index j with $f_j'(\widehat{x})\dot{\gamma}(0) \geq 0$. Hence, there is no direction p along such a curved search path, for which $f'(\widehat{x})^T.p < 0$.

Now we consider a direction $p \in \mathbb{R}^n$ such that

$$Ap > 0, \tag{32}$$

$$F'(\widehat{x})p = 0. \tag{33}$$

(In contrast to the purely concave case, we need the strict inequality in (32) to take care of curvature terms.) Since $A\widehat{x} \geq b$ and (32) imply

$$A(\widehat{x} + s_p(\alpha)) = A(\widehat{x} + \alpha\dot{s}_p(0) + o(\alpha)) = A\widehat{x} + \alpha(Ap + o(1)) \geq b$$

for sufficiently small $\alpha \geq 0$, (31) implies feasibility of the points $\widehat{x} + s_p(\alpha)$ for small $\alpha \geq 0$.

Thus, s_p is a curved search path in the feasible set, and we conclude from the discussion above that there is no such p with $f'(\widehat{x})^T p > 0$. Thus, the equations (32), (33), and

$$f'(\widehat{x})^T p < 0, \tag{34}$$

are inconsistent.

Therefore, the transposition theorem 3.4 applies with

$$\begin{pmatrix} -f'(\hat{x}) \\ A \\ F'(\widehat{x}) \end{pmatrix}, \quad \begin{pmatrix} w \\ y \\ z \end{pmatrix} \quad \text{in place of } B, q,$$

and shows the solvability of

$$-f'(\widehat{x})^T w + A^T y + F'(\widehat{x})^T z = 0, \quad w \geq 0, \quad y \geq 0, \quad \begin{pmatrix} w \\ y \end{pmatrix} \neq 0.$$

If we put $\widehat{z} = z$, let $\widehat{y}$ be the vector with $\widehat{y}_J = y$ and zero entries elsewhere, and note that $\widehat{x}$ is feasible, we find (25)–(27).

Because of (28), it now suffices to discuss the case where $w = 0$ and $z = 0$, and therefore

$$A^T y = 0, \quad y \neq 0. \tag{35}$$

In this case, $b^T y = (A\widehat{x})^T y = \widehat{x}^T A^T y = 0$. Therefore any point $x \in U$ satisfies $(Ax - b)^T y = x^T A^T y - b^T y = 0$, and since $y \geq 0$, $Ax - b \geq 0$, we see that the set

$$K := \{i \mid (Ax)_i = b_i \text{ for all } x \in V \text{ with } Ax - b \geq 0\} \tag{36}$$

contains all indices i with $y_i \neq 0$ and hence is nonempty.

Since V is nonempty, the system $A_{K:}x = b_K$ is consistent, and hence equivalent to $A_{L:}x = b_L$, where L is a maximal subset of K such that the rows of A indexed by L are linearly independent. If M denotes the set of indices complementary to K, we can describe the feasible set equivalently by the constraints

$$A_{M:}x \geq b_M, \quad \begin{pmatrix} A_{L:}x - b_L \\ F(x) \end{pmatrix} = 0.$$

This modified description of the feasible set has no equality constraints implicit in the inequality $A_{M:}x \geq b_M$. For a solution $\widehat{x}$ of the equivalent optimization problem with these constraints, we find as before vectors $w \geq 0$, $\tilde{y}_M$ and $\begin{pmatrix} \tilde{y}_L \\ z \end{pmatrix}$ such that

$$f'(\widehat{x})^T w = A_{M:}^T \tilde{y}_M + \begin{pmatrix} A_{L:} \\ F'(\widehat{x}) \end{pmatrix}^T \begin{pmatrix} \tilde{y}_L \\ z \end{pmatrix}, \tag{37}$$

$$\inf(\tilde{y}_M, A_{M:}\widehat{x} - b_M) = 0, \tag{38}$$

$$F(x) = 0, \quad A_{K:}\widehat{x} - b_K = 0, \tag{39}$$

$$w, \ \begin{pmatrix} \tilde{y}_L \\ z \end{pmatrix} \text{ are not both zero.} \tag{40}$$

Clearly, this yields vectors $\widehat{w} = w$, $\widehat{y} = \tilde{y}$ and $\widehat{z} = z$ satisfying (25) and (26), but now $\tilde{y}_{K \setminus L} = 0$. The exceptional situation $w = 0$, $z = 0$ can no longer occur. Indeed, as before, all indices i with $\tilde{y}_i \neq 0$ lie in K; hence $\tilde{y}_M = 0$ and (37) gives $A_{L:}^T \tilde{y}_L = 0$. Since, by construction, the rows of $A_{L:}$ are linearly independent, this implies $\tilde{y}_L = 0$, contradicting (40). Hence, we have w and z are not both zero.

It remains to show that we can choose $y \geq 0$ with $A^T y = A^T_{M:}\tilde{y}_M + A^T_{L:}\tilde{y}_L$. From the definition (36) of K we know that the two relations $Ap \geq 0$ and $A_{K:}p \neq 0$ are inconsistent (set $x = \widehat{x}+p$). In particular, the relations $Ap \geq 0$ and $\tilde{y}^T_K A_{K:}p < 0$ are inconsistent. By the Lemma of Farkas 3.1 we conclude the existence of a vector $q \geq 0$ with $A^T q = A^T_{K:}\tilde{y}_K = A^T_{L:}\tilde{y}_L$. Setting $y_M = \tilde{y}_M + q_M$ and $y_K = q_K$ completes the proof. □

In contrast to our version of the Karush-John condition, the standard Karush-John condition asserts under our assumptions only that $\widehat{w}$, $\widehat{y}$, and $\widehat{z}$ are not simultaneously zero. Thus the present version gives more information in case that $\widehat{w} = 0$. Therefore, weaker constraint qualifications are needed to ensure that $\widehat{w} \neq 0$. In that case, the multipliers in equation (25) can be rescaled so that $\|\widehat{w}\| = 1$. However, from a numerical perspective, it may be better to keep the homogeneous formulations, since a tiny w in a well-scaled multiplier vector implies neardegeneracy and would give huge multipliers if normalized to $\|\widehat{w}\| = 1$.

Note that in view of (26), the condition (28) can be written (after rescaling) in the equivalent form

$$\widehat{w} \geq 0, \quad v^T\widehat{w} + u^T\widehat{y} + \widehat{z}^T D\widehat{z} = 1, \tag{41}$$

where $v \neq 0$ is an arbitrary nonnegative vector, u is an arbitrary nonnegative vector with $u_J > 0$, $u_N = 0$, and D is an arbitrary diagonal matrix with positive diagonal entries. This form is numerically stable in that all multipliers are bounded and near degeneracies – which would produce huge multipliers in the Kuhn-Tucker conditions – are revealed by a small norm of $\widehat{w}$. The lack of a constraint qualification (which generally cannot be established in finite precision arithmetic anyway) therefore simply appears as the limit $\widehat{w} = 0$.

The formulation (41) is particularly useful for the rigorous verification of the existence of a solution of our refined Karush-John conditions in the vicinity of an approximate solution; cf. Kearfott [17, Section 5.2.5] for the corresponding use of the standard Karush-John conditions. The advantage of our stronger formulation is that in case there are only few nonconcave constraints, condition (41) involves only a few variables and hence is a much stronger constraint if constraint propagation techniques [17, 35] are applied to the optimality conditions.

Let $B := F'(\widehat{x})^T E_u$ for some $u > 0$. We say that the **constraint qualification** (CQ) is satisfied if $\mathrm{rk}F'(\widehat{x}) = r$ and there exists a matrix $Q \in \mathbb{R}^{n\times(r+1)}$ with

$$\begin{aligned} C'(\widehat{x})_{J:}(B+Q) &\geq 0, \\ F'(\widehat{x})Q &= 0. \end{aligned} \tag{42}$$

4.2 Corollary. *If, under the assumptions of Theorem* 4.1*, the constraint qualification (CQ) is satisfied then the conclusion of Theorem* 4.1 *holds with* $\widehat{w} \neq 0$.

Proof. It is obvious that the conclusion of Theorem 4.1 holds with $\widehat{w} \neq 0$ if

$$C'(\widehat{x})^T_{J:} y_J + F'(\widehat{x})^T z = 0,\ y_J \geq 0 \quad \Rightarrow \quad z = 0. \tag{43}$$

If (43) is satisfied, we have $z \neq 0$, $y_J \geq 0$, and $C'(\widehat{x})^T_{J:} y_J + F'(\widehat{x})^T z = 0$ are inconsistent. By Corollary 3.6 this is equivalent to the constraint qualification. □

4.3 Theorem. (Kuhn-Tucker conditions)

Under the assumption of Theorem 4.1 *with* f *one-dimensional, if the constraint qualification (CQ) is satisfied, then there are vectors* $\widehat{y} \in \mathbb{R}^m, \widehat{z} \in \mathbb{R}^r$ *such that*

$$f'(\widehat{x})^T = C'(\widehat{x})^T \widehat{y} + F'(\widehat{x})^T \widehat{z}, \tag{44}$$

$$\inf(\widehat{y}, C(\widehat{x})) = 0, \tag{45}$$

$$F(\widehat{x}) = 0. \tag{46}$$

(44)–(46) are the **Kuhn-Tucker conditions** for the nonlinear program (24), cf. [19]. The traditional linear independence constraint qualification requires in place of the assumptions in Theorem 4.3 the stronger condition that the rows of $\begin{pmatrix} F'(\widehat{x}) \\ C'(\widehat{x})_{J:} \end{pmatrix}$ are independent. In contrast, our condition allows arbitrary dependence among the rows of $C'(\widehat{x})$.

Weaker than the constraint qualification (CQ) is the Mangasarian-Fromowitz constraint qualification (MFCQ), which asserts the existence of a vector q with $C'(\widehat{x})_{J:} q > 0$ and $F'(\widehat{x})q = 0$. It implies our constraint qualification (CQ), because $Q = q\lambda^T$ satisfies (42) for λ large enough. We now show that (MFCQ) is more restrictive than our new constraint qualification (CQ).

4.4 Example. We reconsider Example 2.7. As we have seen there, the point $\hat{x} = (0, 0, 0)^T$ is a local (even global) optimum of (12). Figure 1 reveals that there is a degeneracy since the two activities have a common tangent. Thus the constraint qualifications are nontrivial. Clearly, the mapping C defined by transforming the inequality constraints of to $C(x) \geq 0$ is convex, hence we can use Theorem 4.1. We have

$$f'(\hat{x}) = \begin{pmatrix} -1 & 0 & 1 \end{pmatrix}, \quad F'(\hat{x}) = \begin{pmatrix} 1 & -1 & 1 \end{pmatrix},$$

$$B = \begin{pmatrix} 1 & -1 \\ -1 & 1 \\ 1 & -1 \end{pmatrix}, \quad C'_{J:}(\hat{x}) = \begin{pmatrix} -10 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

We can use the formulas (11) to calculate the multipliers

$$\widehat{y} = \frac{1}{4} \begin{pmatrix} 1 \\ 4 \\ 2 \end{pmatrix}, \quad \widehat{z} = 1;$$

then (44) reduces to the identity

$$1 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} -10 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \\ 2 \end{pmatrix} + 1 \cdot \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

This shows that we can choose $\widehat{w} = 1$ in the Karush-John conditions. Indeed, our constraint qualification (CQ) is valid. For

$$Q = \frac{1}{19} \begin{pmatrix} -1 & 1 \\ 1 & -1 \\ 2 & -2 \end{pmatrix}$$

we have

$$C'_{J:}(\hat{x})(B + Q) = 0, \quad F'(\hat{x})Q = 0.$$

However, (MFCQ) is not satisfied since there is no vector q with $C'_{J:}(\hat{x})q > 0$.

5 Conclusions

We presented various theorems of the alternative, and, based on them, derived new optimality conditions that hold without any constraint qualification. These results strengthen known local conditions, but they are also suitable for use in a global optimization context, which was our main motivation for this work.

New and exciting is the fact that, for the first time, it is possible to give necessary and sufficient (global) optimality conditions for polynomial problems. In particular, it is possible to produce (under the idealized assumptions that all semidefinite programs can be solved

exactly) certificates for global optimality of a putative solution $\hat{x}$. However, these global results are probably not best possible.

The failure to find a certificate after all problems up to some maximum degree d have been solved makes it likely that $\hat{x}$ is not a global optimizer of (3). In this case, one would like to have a procedure that guarantees (for sufficiently large but a priori unknown d) to find a feasible point x with a better objective function value than the value at $\hat{x}$. Then a new local optimization could be started from x, resulting in a better candidate for a global optimizer. Work on this is in progress.

Also, at present we have no simple constraint qualification which would guarantee in the single-objective case that the exponent α in the global Karush-John condition of Theorem 2.4 takes the value 1, which is needed to construct from the certificate multipliers satisfying the Kuhn-Tucker conditions. We conjecture that the exponent $e = 1$ is possible iff the Kuhn-Tucker conditions can be satisfied at $\hat{x}$, in particular under the same (weakened Mangasarian-Fromovitz) constraint qualification as in our Theorem 4.3. This would strengthen the currently weak connectios between sections 2 and 3.

References

[1] B. Aghezzaf and M. Hachimi, On a gap between multiobjective optimization and scalar optimization. J. Opt. Theory and Appl., 109/2:431–435, 2001.

[2] K.J. Arrow, L. Hurwicz and H. Uzawa, Constraint qualifications in maximization problems, Naval Research Logistics Quarterly 8 (1961), 175–191.

[3] M.S. Bazaraa, H.D. Sherali and C.M. Shetty, Nonlinear Programming. Theory and Algorithms, 2nd edition, Wiley, New York 1993.

[4] J. Bochnak, M. Coste and M.-F. Roy, Real Algebraic Geometry. Volume **36** in Ergebnisse der Mathematik und ihrer Grenzgebiete, Folge 3, Springer, Berlin Heidelberg New York, 1998.

[5] J.F. Bonnans and A. Shapiro, Perturbation analysis of optimization problems, Springer, New York 2000.

[6] C.G. Broyden, A simple algebraic proof of Farkas's lemma and related theorems. Optimization Methods and Software, **8(3-4)**:185–199, 1998.

[7] C. de Boor, *An empty exercise*, ACM SIGNUM Newsletter, ACM Press New York, NY, USA, 1990.

[8] R. Cambini, Second order optimality conditions in multiobjective programming. Optimization, 44:139–160, 1998.

[9] COCONUT, COntinuous CONstraints – Updating the Technology. `http://www.mat.univie.ac.at/~neum/glopt/coconut/`

[10] The COCONUT Environment for global optimization, `http://www.mat.univie.ac.at/coconut-environment/`

[11] J. Farkas, Über die Theorie der einfachen Ungleichungen, J. reine angew. Math. 124 (1902), 1–24.

[12] D. Henrion and J.B. Lasserre, GloptiPoly: Global Optimization over Polynomials with Matlab and SeDuMi, ACM Trans. Math. Software 29 (2003), 165–194.

[13] D. Henrion and J.B. Lasserre, Solving Global Optimization Problems over Polynomials with GloptiPoly 2.1, pp. 43–58 in: Ch. Bliek et al. (eds.), Global Optimization and Constraint Satisfaction, Springer, Berlin 2003.

[14] D. Henrion and J.B. Lasserre, Detecting global optimality and extracting solutions in GloptiPoly, Manuscript (2003). `http://www.laas.fr/~henrion/papers/extract.pdf`

[15] F. John, Extremum problems with inequalities as subsidiary conditions. pp. 543–560 in: Fritz John, Collected Papers, Vol.2 (J. Moser, ed.), Birkhäuser, Boston (1985).

[16] W. Karush, Minima of Functions of Several Variables with Inequalities as Side Constraints. M.Sc. dissertation, Dept. of Mathematics, Univ. of Chicago, Chicago, Ill. (1939).

[17] R. B. Kearfott, Rigorous Global Search: Continuous Problems. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[18] H.W. Kuhn, Nonlinear programming: a historical note, pp. 82–96 in: History of Mathematical Programming (J.K. Lenstra et al., eds.) North Holland, Amsterdam 1991.

[19] H. W. Kuhn and A. W. Tucker, Nonlinear programming. In J. Neyman, editor, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, pages 481–492. Univ. of California Press, Berkeley, California, 1951.

[20] J.B. Lasserre, Global Optimization with Polynomials and the Problem of Moments, SIAM J. Optim. **11/3** (2001), 796–817.

[21] O.L. Mangasarian, Nonlinear Programming, McGraw-Hill New York 1969. Reprinted as: Classics in Applied Mathematics, SIAM, Philadelphia 1994.

[22] O. L. Mangasarian and S. Fromovitz, The Fritz John necessary optimality conditions in the presence of equality and inequality constraints, J. Math. Anal. Appl. 17 (1967), 37–47.

[23] I. Marusciac, On Fritz John type optimality criterion in multiobjective optimization. L'Analyse Numerique et la Theorie de l'Approximation, 1982.

[24] T.S. Motzkin, Beiträge zur Theorie der Linearen Ungleichungen, Inaugural Dissertation, Basel, Jerusalem, 1936.

[25] A. Neumaier, Interval Methods for Systems of Equations, Cambridge University Press, Cambridge, 1990.

[26] A. Neumaier, Introduction to Numerical Analysis. Cambridge Univ. Press, Cambridge, 2001.

[27] A. Neumaier, Complete Search in Continuous Global Optimization and Constraint Satisfaction, pp. 271–369 in: Acta Numerica 2004 (A. Iserles, ed.), Cambridge University Press 2004.

[28] A. Neumaier and F. Domes, GloptLab - a Global optimization laboratory, In preparation, 2006.

[29] A. Neumaier and H. Schichl, Sharpening the Karush-John optimality conditions, Manuscript (2003). `http://www.optimization-online.org/DB_HTML/2003/07/691.html`

[30] A. Neumaier, O. Shcherbina, W. Huyer and T. Vinko, A comparison of complete global optimization solvers, Mathematical Programming, to appear.

[31] Phan Quoc Khanh and Le Minh Luu, Multifunction optimization problems involving parameters: Necessary optimality conditions. Optimization, 26/3:577–595, 2002.

[32] S. Prajna, A. Papachristodoulou and A. Parrilo, SOSTOOLS: Sum of Squares Optimization Toolbox for MATLAB – User's Guide, Manuscript (2002), `http://www.optimization-online.org/DB_HTML/2002/05/483.html`

[33] SeDuMi, http://sedumi.mcmaster.ca

[34] A. W. Tucker. Dual Systems of Homogeneous Linear Relations in Linear Inequalities and Related Systems (H. W. Kuhn and A. W. Tucker, eds.), Annals of Mathematics Studies, vol. 38, Princeton University Press, Princeton, 1956.

[35] P. Van Hentenryck, L. Michel, and Y. Deville, Numerica – A Modeling Language for Global Optimization. MIT Press, Cambridge, MA, 1997.