How Stringent is the Linear Independence Kink Qualification in Abs-Smooth Optimization?

Lukas Baumgärtner Franz Bethke
Ganna Shyshkanova Andrea Walther

December 12, 2024

Abs-smooth functions are given by compositions of smooth functions and the evaluation of the absolute value. The linear independence kink qualification (LIKQ) is a fundamental assumption in optimization problems governed by these abs-smooth functions, generalizing the well-known LICQ from smooth optimization. In particular, provided that LIKQ holds it is possible to derive optimality conditions for abs-smooth optimization problems that can be checked in polynomial time. Utilizing tools from differential topology, namely a version of the jet-transversality theorem, it is shown that assuming LIKQ for all feasible points of an abs-smooth optimization problem is a generic assumption.

Keywords: abs-normal form, genericity, jet-transversality, linear independence kink qualification, nonsmooth optimization, piecewise-smooth constraints

1 Introduction

Many real-world applications lead to tasks with nonsmooth structures challenging significantly the corresponding analysis. Up to now there are hardly any off-the-shelf solution algorithms or software packages to solve such problems, which is mainly due to the lack of computationally tractable optimality and stationarity conditions. For this reason, researchers concentrate on certain classes of nonsmooth problems like, e.g., semismooth functions to derive new analytical results or novel solution approaches.

One class of nonsmooth functions that gained some attention in the past years are functions that are defined by a suitable composition of smooth functions and the evaluation of the absolute value function, see, e.g., [GW16; HKS21; HS20; WG19]. The main motivation to consider this set of functions was an extended version of algorithmic differentiation [Gri13] to provide generalized derivatives in a convenient way for nonsmooth functions given as computer programs where the arguments of absolute value evaluations are evaluated one after the other. Conceptually, a function $\varphi\colon \mathbb{R}^n \to \mathbb{R}$ is called *abs-smooth* if for every input $x \in \mathbb{R}^n$ there is a unique vector $z \in \mathbb{R}^s$ of $s \in \mathbb{N}$ intermediate values such that the function value $\varphi(x)$ can be expressed by means of two smooth functions $f\colon \mathbb{R}^n \times \mathbb{R}^s \times \mathbb{R}^s \to \mathbb{R}$ and $c\colon \mathbb{R}^n \times \mathbb{R}^s \times \mathbb{R}^s \to \mathbb{R}^s$ by

$$\begin{aligned} \varphi(x) &= f(x, |z|, z), \\ z &= c(x, |z|, z). \end{aligned} \tag{1}$$

Here and throughout, the application of the absolute value function $|\cdot|$ to a vector $z \in \mathbb{R}^s$ is to be understood component wise, i.e., $|z| \in \mathbb{R}^s$ and $|z|_i = |z_i|$, $i \in \{1, \ldots, s\}$. The function c will be referred to as *switchting function* of φ and the vector z corresponding to the input x will be referred to as *switchting vector* for x. Each component of c captures a step a program that implements φ takes to compute an intermediate value whose absolute value is used in the overall computation. As example consider $\varphi(x) \coloneqq ||x_1| - |x_2||$ which can be represented by $c(x, y, z) \coloneqq (x_1, x_2, z_1 - z_2)$ and $f(x, y, z) \coloneqq y_3$.

An abs-smooth function could serve as target function in a nonsmooth optimization task as considered, e.g., in [GW16]. A collection of abs-smooth functions could represent a system of nonsmooth equations considered, e.g., in [Gri+15], or may be used to describe nonsmooth constraints as in [HS20]. In the literature, such functions are also called abs-normal. However, since the representation in (1) is by no means unique and φ is a composition of smooth functions and the absolute value, the term abs-smooth is used throughout this paper. Furthermore, the list of arguments of an abs-smooth function φ varies in the various contributions. It was shown in [Shy+25] that these different formulations are equivalent regarding the regularity condition introduced below.

As can be seen from the equations in (1), the nonsmoothness is located just in the evaluation of the absolute value. This is an important property of abs-smooth functions that can be exploited to derive and state theoretical properties of an abs-smooth function. A nonsmoothness occurs only if there exists an index i, $1 \leq i \leq s$, with $z_i = 0$, and probably z_i switches the sign in an appropriate neighborhood of this point. Motivated by the graphical representation in low dimensions, therefore the sets of points $x \in \mathbb{R}^n$, where at least one component of the corresponding vector z is equal to zero, are called kinks. In [GW16], the *linear independence kink qualification (LIKQ)* that is detailed below was introduced. Showing a close relation to the linear independence constraint qualification (LICQ) in smooth optimization, LIKQ allows to formulate optimality conditions that can be verified with polynomial complexity [GW16; HS20]. In [GW16; WG19], several examples were presented to illustrate the prerequisites required such that LIKQ holds. In [GW16, p. 8] the authors write: “In general, there is no reason why LIKQ should be violated and locally it can always be achieved by an arbitrary small perturbation …”, however, the argument given there is rather informal.

To formalize the concept of a property that “usually” holds for a problem class, said class is typically parametrized along the defining functions of a problem. A problem instance is then a point in the vector space of possible problems and the property is said to hold, if it holds at all feasible points.

For smooth optimization, the question whether LICQ is “usually” true was analyzed in different contributions. In [SR79], the authors consider a condition like LICQ as *generic*, if it holds for almost all problem instances. This concept has the drawback that the set, where the condition holds, might be closed such that a small perturbation may lead to a situation where the condition is no longer true. For this reason, and more or less in parallel, in [JT79] the authors assumed high regularity of the involved functions and used the Whitney topology to prove that the set of problems where LICQ holds everywhere is dense and open. This is a very strong notion of genericity as it ensures that an arbitrarily small perturbation to the defining functions of a problem instance leads to a situation in which the LICQ holds at all feasible points, and that this situation in turn is stable with respect to further perturbations.

Since the concept of abs-smooth functions was motivated originally by algorithmic differentiation where the smooth components are usually C^∞ functions, this paper follows the lines of Jongen and co-workers to prove that LIKQ is generic in the sense that the set of problem formulations where this property holds is dense and open under the assumption that the involved functions are very regular. Given a “random” abs-smooth problem, it is then not a strong restriction to assume that LIKQ holds in particular at a local minimizer. For mathematical programs with complementarity constraints (MPCCs), it was shown in [SS01] that MPCC-LICQ is generic in the sense of Jongen, i.e., the proof is based on Sard’s theorem. In [HKS21], the authors prove that abs-smooth nonlinear optimizations problems are equivalent to a certain class of MPCCs. However, this class is just a subset of the MPCCs considered in [SS01], and hence, one can not apply the general perturbations that are needed in the proofs used in [SS01] to show that MPCC-LICQ is generic for the class of MPCCs that are equivalent to abs-smooth nonlinear optimizations problems. Therefore, these two results can not be easily combined to show the genericity of LIKQ for the abs-smooth problem class.

Genericity of suitable constraint qualifications for other nonsmooth optimization problem classes have been shown and applied in the critical point theory in [JRS09; DSS12; DJS13; LS22]. Specifically, [DJS13] invokes the structured jet-transversality theorem of [Gün08] for Nash games; a tool that will be crucial for the arguments presented below.

The rest of the paper is structured as follows. Section 2 introduces the basic concepts and notations for abs-smooth optimization problems while Section 3 presents some definitions and results from differential topology that are required for the application of the structured jet-transversality theorem of [Gün08] in the context of abs-smooth problems. The established formalism is used in Section 4 to encode LIKQ as a transversality condition. Finally, in Section 5, the genericity of LIKQ is shown. Section 6 gives an outlook to further research questions.

To derive the theoretical results of this paper, the following notation is used. For a finite set M the cardinality of M is denoted by $|M|$. The set of positive integers up to $n \in \mathbb{N}$ is denoted by $[n] \coloneqq \{1, \dots, n\}$. In particular $[0] = \emptyset$. For a point x on a manifold A the tangent space of A at x is denoted by $T_x A$.

Let $i, n \in \mathbb{N}$, $i \leq n$ and let $X_1, \dots, X_n, Y$ be finite dimensional, real vector spaces, then the i-th partial derivative of a differentiable function $f\colon X_1 \times \cdots \times X_n \to Y$ is denoted by $\partial_i f$, the corresponding Jacobian with respect to the canonical basis by $\mathrm{D}_i f$, and the full Jacobian by $\mathrm{D}f$. Is f is smooth enough, $\ell \in \mathbb{N}$ and $a \in [n]^\ell$ a multi-index, then $\partial_a f := \partial_{a_\ell} \cdots \partial_{a_1} f$.

For $n \in \mathbb{N}$ the identity matrix in $\mathbb{R}^{n\times n}$ is denoted by I_n.

2 Prerequisites from abs-smooth optimization

The properties of an abs-smooth function φ to a large extent depend on the function c. The rather involved definition of c presented below already prepares the ground for the main result of this paper and takes the fact into account that the intermediate value z_i can only be influenced by the intermediate value z_j if $j < i$ which reflects the nature of a computer program.

Let from here on onward $k \in \mathbb{N} \cup \{\infty\}$ and $n, s \in \mathbb{N}$ be given and let $d := n + s + s$ and $d_i := n + 2(i-1)$ for $i \in [s]$.

Definition 1 (Switching functions)**.** Given an s-tuple of functions $(c_1, c_2, \ldots, c_s)$ with $c_i \in C^k(\mathbb{R}^{d_i})$ for $i \in [s]$, the function $c\colon \mathbb{R}^d \cong \mathbb{R}^n \times \mathbb{R}^s \times \mathbb{R}^s \to \mathbb{R}^s$ defined by

$$c(x, y, z) := (c_1(x), c_2(x, y_1, z_1), \ldots, c_s(x, (y_1, \ldots, y_{s-1}), (z_1, \ldots, z_{s-1}))). \tag{2}$$

is called a *switching function* of class C^k. The set of all switching functions is denoted by $C^k_{\mathrm{sw}}(\mathbb{R}^d; \mathbb{R}^s) \subsetneq C^k(\mathbb{R}^d; \mathbb{R}^s)$.

It is important to note that, for a given $c \in C^k_{\mathrm{sw}}(\mathbb{R}^d; \mathbb{R}^s)$ and $x \in \mathbb{R}^n$, by construction there is a unique solution $z \in \mathbb{R}^s$ to the so-called *switching equation*

$$z = c(x, |z|, z) \tag{3}$$

by simply consecutively evaluating

$$z_i = c_i(x, (|z_1|, \ldots, |z_{i-1}|), (z_1, \ldots, z_{i-1})),$$

for all $i \in [s]$, where for $i = 1$ this is to be understood as $z_1 = c_1(x)$. In particular, the Jacobians $\mathrm{D}_2 c(x, |z|, z) \in \mathbb{R}^{s\times s}$ and $\mathrm{D}_3 c(x, |z|, z) \in \mathbb{R}^{s\times s}$ are lower triangular matrices.

Let $q, p \in \mathbb{N}$. For a given switching function $c \in C^k_{\mathrm{sw}}(\mathbb{R}^d; \mathbb{R}^s)$ and functions $g \in C^k(\mathbb{R}^d; \mathbb{R}^p)$ and $h \in C^k(\mathbb{R}^d; \mathbb{R}^q)$ let $\mathcal{F}(c, g, h)$ be defined by

$$\mathcal{F}(c, g, h) := \{(x, z) \in \mathbb{R}^n \times \mathbb{R}^s \colon z = c(x, |z|, z), g(x, |z|, z) \geq 0, h(x, |z|, z) = 0\}. \tag{4}$$

Together with $f \in C^k(\mathbb{R}^d)$ an abs-smooth optimization problem reads

$$\text{minimize} \quad f(x, |z|, z) \quad \text{s.t.} \quad (x, z) \in \mathcal{F}(c, g, h). \tag{5}$$

The pair of functions $(f, c) \in C^k(\mathbb{R}^d) \times C^k_{\mathrm{sw}}(\mathbb{R}^d; \mathbb{R}^s)$ can be used to represent a function mapping $\mathbb{R}^n$ to $\mathbb{R}$. Given $x \in \mathbb{R}^n$ the value $(f, c)[x] \in \mathbb{R}$ is obtained by first computing the unique $z \in \mathbb{R}^s$ according to the switching equation (3) and then evaluating

$$(f, c)[x] := f(x, |z|, z).$$

The above notation should indicate the difference to the usual component wise way of interpreting the evaluation for a tuple of functions, i.e, $(f, c)(w) = (f(w), c(w)) \in \mathbb{R}^{s+1}$ for $w \in \mathbb{R}^d$. If for a given function $\varphi\colon \mathbb{R}^n \to \mathbb{R}$ it holds that $\varphi(x) = (f, c)[x]$ for all $x \in \mathbb{R}^n$, then φ is called *abs-smooth* and (f, c) is called an *evaluation procedure* of φ. The set of abs-smooth functions, i.e., the set of functions $\varphi\colon \mathbb{R}^n \to \mathbb{R}$ for which there is an evaluation procedure $(f, c) \in C^k(\mathbb{R}^d) \times C^k_{\text{sw}}(\mathbb{R}^d; \mathbb{R}^s)$, is denoted by $C^{k,s}_{\text{abs}}(\mathbb{R}^n)$. By extending the switching function c, it is possible to also represent functions $\varphi_g\colon \mathbb{R}^n \to \mathbb{R}^p$ and $\varphi_h\colon \mathbb{R}^n \to \mathbb{R}^q$ whose evaluation involves the absolute value function such that $\varphi_g(x) = (g, c)[x]$ and $\varphi_h(x) = (h, c)[x]$. For all possible evaluation procedures that represent φ, φ_g and φ_h, the optimization problem in (5) is equivalent to the problem

$$\text{minimize} \quad \varphi(x) \quad \text{s.t.} \quad x \in \{\tilde{x} \in \mathbb{R}^n : \varphi_g(\tilde{x}) \geq 0, \varphi_h(\tilde{x}) = 0\}. \tag{6}$$

The LIKQ condition is useful to characterize local optima to (5) or (6) in terms of certain KKT type conditions. In order to define LIKQ it is useful to introduce further notation for the Jacobian matrices of the switching equation in (3), the inequality constraints g and the equality constraints h with respect to the independent variable x. To that end, let for a given $z \in \mathbb{R}^s$ the matrix of signs of z be denoted by $\Sigma(z) \coloneqq \operatorname{diag}(\operatorname{sign}(z)) \in \{-1, 0, 1\}^{s \times s}$ and, given $x \in \mathbb{R}^n$, define

$$\begin{aligned}
J_z(x, z) &\coloneqq (I_s - \mathrm{D}_2 c(x, |z|, z)\Sigma(z) - \mathrm{D}_3 c(x, |z|, z))^{-1} \mathrm{D}_1 c(x, |z|, z),\\
J_g(x, z) &\coloneqq \mathrm{D}_1 g(x, |z|, z) + (\mathrm{D}_2 g(x, |z|, z)\Sigma(z) + \mathrm{D}_3 g(x, |z|, z)) J_z(x, z),\\
J_h(x, z) &\coloneqq \mathrm{D}_1 h(x, |z|, z) + (\mathrm{D}_2 h(x, |z|, z)\Sigma(z) + \mathrm{D}_3 h(x, |z|, z)) J_z(x, z).
\end{aligned} \tag{7}$$

The set of indices corresponding to active switches and the set of indices corresponding to active inequality constraints are denoted by

$$\alpha(z) \coloneqq \{i \in [s] \colon z_i = 0\} \qquad \text{and} \qquad \beta(x, z) \coloneqq \{i \in [p] \colon g(x, |z|, z)_i = 0\}.$$

The projection onto the active switching indices is defined by $P_\alpha \coloneqq (e_i^T)_{i \in \alpha(z)} \in \mathbb{R}^{|\alpha(z)| \times s}$, where $e_i \in \mathbb{R}^s$ is the i-th unit vector while the complementary projection to the inactive indices is denoted by $Q_\alpha \coloneqq (e_i^T)_{i \in [s] \setminus \alpha(z)} \in \mathbb{R}^{(s - |\alpha(z)|) \times s}$. By a slight abuse of notation, the projection onto the active/inactive inequality constraints will similarly be denoted by $P_\beta \coloneqq (e_i^T)_{i \in \beta(x,z)} \in \mathbb{R}^{|\beta(x,z)| \times p}$ and $Q_\beta \coloneqq (e_i^T)_{i \in [p] \setminus \beta(x,z)} \in \mathbb{R}^{(p - |\beta(x,z)|) \times p}$.

Definition 2 (LIKQ)**.** Given c, g and h as before, a point $(x, z) \in \mathcal{F}(c, g, h)$ is said to satisfy the *linear independence kink qualification (LIKQ)* at (x, z), if

$$\operatorname{rank} \begin{bmatrix} P_{\alpha(z)} J_z(x, z) \\ P_{\beta(x,z)} J_g(x, z) \\ J_h(x, z) \end{bmatrix} = |\alpha(z)| + |\beta(x, z)| + q.$$

The examples below consider the simple case with just one switch and without equality and inequality constraints.

Example 3 (LIKQ everywhere)**.** Consider the switching function $c(x, y, z) := -\sin(x)$ which has an active switch at $x \in \{\ell\pi : \ell \in \mathbb{Z}\}$. Since $\mathrm{D}_2 c(x, 0, 0) = \mathrm{D}_3 c(x, 0, 0) = 0$ and $\mathrm{D}_1 c(x, 0, 0) = -\cos(x) \in \{-1, 1\}$ for all those points, the matrix $P_{\alpha(z)} J_z(x, y, z)$ is either empty or ± 1, and hence, has full rank everywhere.

Example 4 (no LIKQ at 0)**.** For the second example consider $c(x, y, z) := \sin(x) - x$. At $x = 0$ the switching equation yields $z = |z| = 0$ and the stratum containing $j^0(0, 0, 0)$ is $A_{(0)}$. LIKQ does not hold at $(0, 0)$, since again $\mathrm{D}_2 c(0, 0, 0) = \mathrm{D}_3 c(0, 0, 0) = 0$, and hence, $J_z(x, z) = \mathrm{D}_1 c(0, 0, 0) = \cos(0) - 1 = 0$.

3 Prerequisites from differential topology

The arguments in the later sections are based on an extension of the jet-transversality theorem, see [JJT00, Thm. 7.4.5], for structured jets in a paper by H. Günzel [Gün08]. For the sake of a complete and mostly self-consistent presentation the important definitions of Günzels paper are restated here.

To not overload the notation that was established in the first part the notation here deviates in some points from the one that is used for example in [JJT00] or [Gün08]. This of course also prepares the application of the results in this section to the abs-smooth problem class. In the following the reader may think of $w \in \mathbb{R}^d$ as a tuple (x, y, z), ϕ as the combined function $(c, g, h) \colon \mathbb{R}^d \to \mathbb{R}^m$, i.e., $\phi(w) = \phi(x, y, z) = (c(x, y, z), g(x, y, z), h(x, y, z))$. In contrast, Φ can be thought of as the $(s + 2)$-tuple of functions defining the abs-smooth problem in (6), i.e, $\Phi = (c_1, \ldots, c_s, g, h)$ and for $i \in [s + 2]$ the vector $w_i \in \mathbb{R}^{d_i}$ as a possible input to the i-th component of Φ. The vector W represents a vector of independent inputs to the function Φ, i.e, $W = (w_1, \ldots, w_{s+2})$. Moreover, bold font is used to indicate variables that should be thought of as values of the corresponding function, e.g., $\boldsymbol{\phi} = \phi(w)$.

Definition 5 (stratifications, jet-transversality [JJT00, Def. 7.3.33, Thm. 7.3.4])**.** Let $d, m \in \mathbb{N}$ and $A \subseteq \mathbb{R}^{d+m}$, then a locally finite partition $\mathcal{A}$ of A into pairwise disjoint C^k manifolds is called a *stratification* of A of class C^k. The elements of $\mathcal{A}$ are referred to as *strata* and are indexed with some index set J. The dimension of $\mathcal{A}$ is defined by

$$\dim(\mathcal{A}) := \max_{j \in J} \dim(A_j).$$

For a given $(\boldsymbol{w}, \boldsymbol{\phi}) \in A$ the stratum containing $(\boldsymbol{w}, \boldsymbol{\phi})$ is denoted by $A_{(\boldsymbol{w},\boldsymbol{\phi})}$ and the tangent space at $(\boldsymbol{w}, \boldsymbol{\phi})$ on $A_{(\boldsymbol{w},\boldsymbol{\phi})}$ simply by $T_{(\boldsymbol{w},\boldsymbol{\phi})} A := T_{(\boldsymbol{w},\boldsymbol{\phi})} A_{(\boldsymbol{w},\boldsymbol{\phi})}$.

The stratification of A is called *weakly Whitney regular* if for any $(\boldsymbol{w}, \boldsymbol{\phi}) \in A$ and any sequence $((\boldsymbol{w}, \boldsymbol{\phi})_\ell)_{\ell \in \mathbb{N}}$ that converges to $(\boldsymbol{w}, \boldsymbol{\phi})$ and consists of elements of the same stratum $A_j \in \mathcal{A}$, for some $j \in J$, the inclusion

$$T_{(\boldsymbol{w},\boldsymbol{\phi})} A \subseteq \lim_{\ell \to \infty} T_{(\boldsymbol{w},\boldsymbol{\phi})_\ell} A_j$$

holds, whenever the last limit exists in the Grassmann manifold, cf. [Lee12, Example 1.36].

For a function $\phi \in C^k(\mathbb{R}^d;\mathbb{R}^m)$ the 0-jet-extension $j^0(\phi)\colon \mathbb{R}^d \to \mathbb{R}^{d+m}$ is given by $j^0(\phi)(w) := (w, \phi(w))$. The "0" in the term 0-jet-extension refers to the fact that no derivative information of ϕ is used in the jet-extension; a setting that suffices for the arguments in this paper.

The 0-jet-extension $j^0(\phi)$ is said to meet $\mathcal{A}$ *transversally*, denoted $j^0(\phi) \pitchfork \mathcal{A}$, if for all $w \in j^0(\phi)^{-1}(A)$

$$\operatorname{img}(\mathrm{D}j^0(\phi)(w)) + T_{j^0(\phi)(w)}A = \mathbb{R}^{d+m}.$$

Roughly speaking, the classical jet-transversality theorem states that for an open and dense set of functions $\phi \in C^k(\mathbb{R}^d;\mathbb{R}^m)$ the 0-jet-extension of ϕ meets any weakly Whitney regular stratified set transversally. Transversality theorems can be leveraged by encoding certain properties of the inputs w, e.g., feasibility with respect to ϕ, into a stratified set A and then reformulating the transversality condition into an algebraic condition, e.g., a constraint qualification. The transversality theorem then states on the one hand that the acquired condition will hold true for all inputs w, possibly after an arbitrarily small perturbation of ϕ (density), and that it is stable (openness).

The more flexible version for structured jets deals with a finite collection of functions and allows for the description of A to depend differently on different variables. For the remainder of this section let $\tilde{s} \in \mathbb{N}$ be the number of involved functions and let $d_1, \ldots, d_{\tilde{s}} \in \mathbb{N}$ and $m_1, \ldots, m_{\tilde{s}} \in \mathbb{N}$ denote the dimensions of their respective domain and image spaces. Furthermore, let $D \in \mathbb{N}$ be the total amount of independent inputs to the $\tilde{s}$ functions and $m \in \mathbb{N}$ the total number of their values, i.e,

$$D := d_1 + \ldots + d_{\tilde{s}} \quad \text{and} \quad m := m_1 + \ldots + m_{\tilde{s}}.$$

Definition 6 (structured 0-jet-extensions [Gün08, Definition 2.4])**.** For $i \in [\tilde{s}]$ let $\phi_i \in C^k(\mathbb{R}^{d_i};\mathbb{R}^{m_i})$ and $\Phi := (\phi_1, \ldots, \phi_{\tilde{s}})$, then the function

$$j^0(\Phi)\colon \mathbb{R}^{d_1} \times \ldots \times \mathbb{R}^{d_{\tilde{s}}} \to \mathbb{R}^{d_1} \times \ldots \times \mathbb{R}^{d_{\tilde{s}}} \times \mathbb{R}^{m_1} \times \ldots \times \mathbb{R}^{m_{\tilde{s}}}$$

given by

$$j^0(\Phi)(w_1, \ldots, w_{\tilde{s}}) := (w_1, \ldots, w_{\tilde{s}}, \phi_1(w_1), \ldots, \phi_{\tilde{s}}(w_{\tilde{s}}))$$

is called *structured 0-jet-extension* of Φ.

The tuple of functions Φ in the above definition can be understood as an overall function in $C^k(\mathbb{R}^D;\mathbb{R}^m)$ by defining its value for $W := (w_1, \ldots, w_{\tilde{s}}) \in \mathbb{R}^{d_1} \times \ldots \times \mathbb{R}^{d_{\tilde{s}}} \cong \mathbb{R}^D$ as

$$\Phi(W) := (\phi_1(w_1), \ldots, \phi_{\tilde{s}}(w_{\tilde{s}})). \tag{8}$$

Note that this evaluation of Φ allows to view the space $C^k(\mathbb{R}^{d_1};\mathbb{R}^{m_1}) \times \ldots \times C^k(\mathbb{R}^{d_{\tilde{s}}};\mathbb{R}^{m_{\tilde{s}}})$ as a subspace of $C^k(\mathbb{R}^D;\mathbb{R}^m)$. In fact, the values of the structured jet $j^0(\Phi)$ are identical to the values of the classical jet $W \mapsto (W, \Phi(W))$ when using this *structured* way of evaluating Φ. This justifies using the same notation for structured jets as for classical jets. Moreover, the notation of jet-transversality from Definition 5 extends naturally to structured jets. However, since the input w_i does not affect the output of ϕ_j whenever

$j \neq i$, the previously described subspace relation is strict. Specifically, the possible perturbations of Φ viewed as an element in $C^k(\mathbb{R}^{d_1};\mathbb{R}^{m_1})\times\ldots\times C^k(\mathbb{R}^{d_{\tilde{s}}};\mathbb{R}^{m_{\tilde{s}}})$ are different from the possible perturbations of Φ as an element of $C^k(\mathbb{R}^D;\mathbb{R}^m)$. The topology in which the perturbations have to be understood is clarified by the following definition.

Definition 7 ((strong) Whitney topology)**.** Let $\ell \in \{0,\ldots,k\}$, $i \in [\tilde{s}]$. The sets $U_{\phi_i,\varepsilon}$, which are indexed by continuous functions $\varepsilon\colon \mathbb{R}^{d_i} \to (0,\infty)$ and points $\phi_i \in C^k(\mathbb{R}^{d_i};\mathbb{R}^{m_i})$, and which are defined by

$$U_{\phi_i,\varepsilon} := \left\{ \psi \in C^k(\mathbb{R}^{d_i};\mathbb{R}^{m_i}) \ : \ \begin{array}{c} \|\partial_a \phi(w_i) - \partial_a \psi(w_i)\| < \varepsilon(w_i) \\ \text{for all } w_i \in \mathbb{R}^{d_i}, |a| \leq \ell \end{array} \right\},$$

form a basis of neighborhoods of the *(strong) Whitney C^ℓ-topology* of $C^k(\mathbb{R}^{d_i};\mathbb{R}^{m_i})$.

If in the definition one uses only positive constants ε instead of positive functions one obtains a basis of the coarser weak Whitney topology. However, since $\mathbb{R}^{d_i}$ is not compact, the convergence with respect to this simpler topology does not capture the behavior “at infinity” very well. In contrast, the strong topology is an extremely fine topology and in fact is not metrizable [Kri69]. Nevertheless, it forms a Baire space, which allows to meaningfully define *generic* subsets as sets which contain the intersection of a countable number of open and dense subsets. In particular, open and dense sets themselves are generic. For details see, e.g., [Hir76, p. 34–36, Theorem 4.4.] and [JJT00, p. 306].

For $\ell \leq k$ the product topology on $C^k(\mathbb{R}^{d_1};\mathbb{R}^{m_1}) \times \ldots \times C^k(\mathbb{R}^{d_{\tilde{s}}};\mathbb{R}^{m_{\tilde{s}}})$ induced by the Whitney C^ℓ-topologies on the respective spaces will be referred to as the Whitney C^ℓ-topology of $C^k(\mathbb{R}^{d_1};\mathbb{R}^{m_1}) \times \ldots \times C^k(\mathbb{R}^{d_{\tilde{s}}};\mathbb{R}^{m_{\tilde{s}}})$. This completes the prerequisites required to state the structured jet-transversality theorem.

Theorem 8 (structured jet-transversality [Gün08, Theorem 2.5])**.** *Let $\tilde{\mathcal{A}}$ be a weakly Whitney regular stratification of $\tilde{A} \subseteq \mathbb{R}^{d_1} \times \mathbb{R}^{m_1} \times \ldots \times \mathbb{R}^{d_{\tilde{s}}} \times \mathbb{R}^{m_{\tilde{s}}}$ with* $\dim(\tilde{\mathcal{A}}) < k+m$. *Then,*

$$\mathcal{G} := \{\Phi \in C^k(\mathbb{R}^{d_1};\mathbb{R}^{m_1}) \times \ldots \times C^k(\mathbb{R}^{d_{\tilde{s}}};\mathbb{R}^{m_{\tilde{s}}}) \ : \ j^0(\Phi) \pitchfork \tilde{\mathcal{A}}\}$$

is a dense subset of $C^k(\mathbb{R}^{d_1};\mathbb{R}^{m_1}) \times \ldots \times C^k(\mathbb{R}^{d_{\tilde{s}}};\mathbb{R}^{m_{\tilde{s}}})$ with respect to the Whitney C^k-topology. If $\tilde{A}$ is closed, then $\mathcal{G}$ is additionally open with respect to the Whitney C^1-topology, and hence, certainly open with respect to the Whitney C^k-topology.

As already stated, the concept of a structured jet and the structured jet-transversality theorem are useful to restrict the perturbations considered in the desired space of functions. However, from a technical point of view that comes with the downside of having independent variables for every function that is part of the tuple ϕ. Mathematically this is not a problem as relations between these variables can also be expressed in the stratified set A, see, e.g., [Gün08, Example 3.2]. The following lemma represents a tool to alleviate this problem by means of a left-invertible linear operator Π that relates between a few actual variables and the many formally required variables of the structured jet. The proof is rather elementary and similar, although not equivalent statements can be found as exercises in the book [JJT00]. Nevertheless, the proof is presented here also with the intent to provide a reference to this tool for future works.

Lemma 9. *Let $\Pi \in \mathbb{R}^{D\times d}$ be a left-invertible matrix and J an index set for a stratification $\mathcal{A} := \{A_j\}_{j\in J}$ of $A \subseteq \mathbb{R}^{d+m}$ of class C^k. Additionally, for $i \in [\tilde{s}]$ let $\phi_i \in C^k(\mathbb{R}^{d_i};\mathbb{R}^{m_i})$ and $\Phi := (\phi_1, \dots, \phi_{\tilde{s}})$. Then,*

a) $\tilde{\mathcal{A}} := \{\tilde{A}_j\}_{j\in J}$ is a stratification of $\tilde{A} \subseteq \mathbb{R}^{D+m}$ into C^k manifolds, where

$$\tilde{A}_j := \begin{bmatrix} \Pi & 0 \\ 0 & I_m \end{bmatrix} A_j \qquad \text{and} \qquad \tilde{A} = \begin{bmatrix} \Pi & 0 \\ 0 & I_m \end{bmatrix} A.$$

b) if $\mathcal{A}$ is weakly Whitney regular, so is $\tilde{\mathcal{A}}$.

c) $j^0(\Phi) \pitchfork \tilde{\mathcal{A}}$ if and only if $j^0(\phi) \pitchfork \mathcal{A}$, where $\phi := \Phi \circ \Pi$ using (8).

Proof. a) Clearly, $\tilde{\mathcal{A}}$ is a partition of $\tilde{A}$. For any $j \in J$ the mapping

$$\mathbb{R}^{d+m} \supset A_j \to \mathbb{R}^{D+m}, \qquad (\boldsymbol{w}, \boldsymbol{\phi}) \mapsto (\boldsymbol{W}, \boldsymbol{\Phi}) := \begin{bmatrix} \Pi & 0 \\ 0 & I_m \end{bmatrix} (\boldsymbol{w}, \boldsymbol{\phi})$$

is an injective smooth immersion. Thus, [Lee12, Proposition 5.18] ensures that $\tilde{A}_j$ is a C^k manifold. To show that $\tilde{\mathcal{A}}$ is locally finite, let some $(\boldsymbol{W}, \boldsymbol{\Phi}) \in \mathbb{R}^{D+m}$ be given. If $\boldsymbol{W} \notin \operatorname{img}\Pi$, then there is a whole neighborhood $U_{\boldsymbol{W}} \subset \mathbb{R}^D$ of $\boldsymbol{W}$ with $U_{\boldsymbol{W}} \cap \operatorname{img}\Pi = \emptyset$ due to $\operatorname{img}\Pi$ being a closed subspace of $\mathbb{R}^D$. Then, for any neighborhood $U_{\boldsymbol{\Phi}} \subset \mathbb{R}^m$ of $\boldsymbol{\Phi}$, clearly $(U_{\boldsymbol{W}} \times U_{\boldsymbol{\Phi}}) \cap \tilde{A}_j = \emptyset$ for any $j \in J$, since $\tilde{A}_j \subseteq \operatorname{img}\Pi \times \mathbb{R}^m$.

If otherwise $\boldsymbol{W} \in \operatorname{img}\Pi$, then there is $\boldsymbol{w} \in \mathbb{R}^d$ with $\Pi\boldsymbol{w} = \boldsymbol{W}$. Since $\mathcal{A}$ is locally finite there is a neighborhood $U_{(\boldsymbol{w},\boldsymbol{\phi})} \subset \mathbb{R}^{d+m}$ of $(\boldsymbol{w}, \boldsymbol{\Phi})$ such that $\{j \in J : U_{(\boldsymbol{w},\boldsymbol{\Phi})} \cap A_j \neq \emptyset\}$ is finite. Set

$$U_{(\boldsymbol{W},\boldsymbol{\Phi})} := \begin{bmatrix} \Pi & 0 \\ 0 & I_m \end{bmatrix} U_{(\boldsymbol{w},\boldsymbol{\Phi})},$$

then, for any $j \in J$ with $U_{(\boldsymbol{w},\boldsymbol{\Phi})} \cap A_j = \emptyset$, it holds

$$U_{(\boldsymbol{W},\boldsymbol{\Phi})} \cap \tilde{A}_j = \begin{bmatrix} \Pi & 0 \\ 0 & I_m \end{bmatrix} U_{(\boldsymbol{w},\boldsymbol{\Phi})} \cap \begin{bmatrix} \Pi & 0 \\ 0 & I_m \end{bmatrix} A_j = \begin{bmatrix} \Pi & 0 \\ 0 & I_m \end{bmatrix} (U_{(\boldsymbol{w},\boldsymbol{\Phi})} \cap A_j) = \emptyset,$$

where the penultimate equality holds due to Π being injective. Therefore, the set $\{j \in J : U_{(\boldsymbol{W},\boldsymbol{\Phi})} \cap \tilde{A}_j \neq \emptyset\}$ is finite.

b) Let $(\boldsymbol{w}, \boldsymbol{\phi}) \in A$, then

$$T_{(\Pi\boldsymbol{w},\boldsymbol{\phi})}\tilde{A} = \begin{bmatrix} \Pi & 0 \\ 0 & I_m \end{bmatrix} T_{(\boldsymbol{w},\boldsymbol{\phi})}A. \tag{9}$$

Let $(\boldsymbol{W}_\ell, \boldsymbol{\Phi}_\ell)_{\ell\in\mathbb{N}} \subset \tilde{A}_j$ for some $j \in J$ with $(\boldsymbol{W}_\ell, \boldsymbol{\Phi}_\ell) \to (\boldsymbol{W}, \boldsymbol{\Phi})$ as $\ell \to \infty$ and let $\tilde{T}$ be the limit of $T_{(\boldsymbol{W}_\ell,\boldsymbol{\Phi}_\ell)}\tilde{A}_j$ in the corresponding Grassmannian. Since, for every $\ell \in \mathbb{N}$, $(\boldsymbol{W}_\ell, \boldsymbol{\Phi}_\ell) \in \tilde{A}_j$, in particular $\boldsymbol{W}_\ell \in \operatorname{img}\Pi$, there is $\boldsymbol{w}_\ell \in \mathbb{R}^d$ with $\boldsymbol{W}_\ell = \Pi\boldsymbol{w}_\ell$ and $(\boldsymbol{w}_\ell, \boldsymbol{\Phi}_\ell) \in A_j$. It holds $(\boldsymbol{w}_\ell, \boldsymbol{\Phi}_\ell) \to (\Pi^\dagger\boldsymbol{W}, \boldsymbol{\Phi})$ as $\ell \to \infty$ and

$$T := \begin{bmatrix} \Pi^\dagger & 0 \\ 0 & I_m \end{bmatrix} \tilde{T} = \begin{bmatrix} \Pi^\dagger & 0 \\ 0 & I_m \end{bmatrix} \lim_{\ell\to\infty} T_{(\boldsymbol{W}_\ell,\boldsymbol{\Phi}_\ell)}\tilde{A} = \lim_{\ell\to\infty} T_{(\boldsymbol{w}_\ell,\boldsymbol{\Phi}_\ell)}\tilde{A}.$$

Since $\mathcal{A}$ is weakly Whitney regular, it follows $T_{(\Pi^\dagger \boldsymbol{W},\boldsymbol{\Phi})}A \subseteq T$. Thereby,

$$T_{(\boldsymbol{W},\boldsymbol{\Phi})}\tilde{A} = \begin{bmatrix}\Pi & 0\\ 0 & I_m\end{bmatrix} T_{(\Pi^\dagger \boldsymbol{W},\boldsymbol{\Phi})}A \subseteq \begin{bmatrix}\Pi & 0\\ 0 & I_m\end{bmatrix} T = \tilde{T}.$$

c) A vector W is an element of $j^0(\Phi)^{-1}(\tilde{A})$ if and only if $(W, \Phi(W)) \in \tilde{A}$. The latter is the case if and only if $W \in \operatorname{img}\Pi$ and $(\Pi^\dagger W, \Phi(W)) \in A$. Since for any $W \in \operatorname{img}\Pi$, it holds that

$$(\Pi^\dagger W, \Phi(W)) = (\Pi^\dagger W, \Phi(\Pi\,\Pi^\dagger W)) = j^0(\Phi \circ \Pi)(\Pi^\dagger W) = j^0(\phi)(\Pi^\dagger W),$$

the former conditions are all equivalent to $W \in \operatorname{img}\Pi$ and $j^0(\phi)(\Pi^\dagger W) \in A$. Finally, this is equivalent to $W \in \Pi j^0(\phi)^{-1}(A)$, which shows that

$$j^0(\Phi)^{-1}(\tilde{A}) = \Pi j^0(\phi)^{-1}(A). \tag{10}$$

Now fix $w \in j^0(\phi)^{-1}(A)$ and let $t := \dim(T_{j^0(\phi)(w)}A)$ and $B_d \in \mathbb{R}^{d\times t}$ and $B_m \in \mathbb{R}^{m\times t}$ be such that the columns of

$$\begin{bmatrix}B_d\\ B_m\end{bmatrix} \in \mathbb{R}^{(d+m)\times t}$$

form a basis of $T_{j^0(\phi)(w)}A$ and the condition $\mathbb{R}^{d+m} = \operatorname{img}(\mathrm{D}j^0(\phi)(w)) + T_{j^0(\phi)(w)}A$ is equivalent to

$$\operatorname{rank}\begin{bmatrix}I_d & B_d\\ \mathrm{D}\phi(w) & B_m\end{bmatrix} = \operatorname{rank}\begin{bmatrix}I_d & B_d\\ \mathrm{D}\Phi(\Pi w)\Pi & B_m\end{bmatrix} = d + m. \tag{11}$$

Let further $\Lambda \in \mathbb{R}^{D\times(D-d)}$ be left-invertible such that the columns of Λ span the $(D-d)$-dimensional orthogonal complement of $\operatorname{img}\Pi$, i.e., $\operatorname{img}\Lambda = (\operatorname{img}\Pi)^\perp$. Then, $\Pi^T\Lambda = 0$ and $[\Lambda\ \Pi] \in \mathbb{R}^{D\times D}$ has full rank. Using this split, (11) is equivalent to

$$\operatorname{rank}\begin{bmatrix}I_{D-d} & 0 & 0\\ 0 & I_d & B_d\\ \mathrm{D}\Phi(\Pi w)\Lambda & \mathrm{D}\Phi(\Pi w)\Pi & B_m\end{bmatrix} = (D-d) + d + m = D + m.$$

Moreover, since

$$\begin{bmatrix}I_{D-d} & 0 & 0\\ 0 & I_d & B_d\\ \mathrm{D}\Phi(\Pi w)\Lambda & \mathrm{D}\Phi(\Pi w)\Pi & B_m\end{bmatrix} = \begin{bmatrix}\Lambda^\dagger & 0\\ \Pi^\dagger & 0\\ 0 & I_m\end{bmatrix}\begin{bmatrix}I_d & \Pi B_d\\ \mathrm{D}\Phi(\Pi w) & B_m\end{bmatrix}\begin{bmatrix}\Lambda & \Pi & 0\\ 0 & 0 & I_m\end{bmatrix}$$

and the transformation matrices have full rank $D+m$ the above is equivalent to

$$\operatorname{rank}\begin{bmatrix}I_d & \Pi B_d\\ \mathrm{D}\Phi(\Pi w) & B_m\end{bmatrix} = D + m$$

which is equivalent to

$$\mathbb{R}^{D+m} = \operatorname{img}(\mathrm{D}j^0(\Phi)(\Pi w)) + \begin{bmatrix} \Pi & 0 \\ 0 & I_m \end{bmatrix} T_{j^0(\Phi\circ\Pi)(w)} A.$$

The equivalence of the transversality conditions then follows by

$$\begin{aligned}
& j^0(\Phi) \pitchfork \tilde{\mathcal{A}} \\
\overset{(10)}{\iff} & \forall\, W \in \Pi j^0(\phi)^{-1}(A)\colon \mathbb{R}^{D+m} = \operatorname{img}(\mathrm{D}j^0(\Phi)(W)) + T_{j^0(\Phi)(W)} \tilde{A} \\
\overset{(9)}{\iff} & \forall\, w \in j^0(\phi)^{-1}(A)\colon \mathbb{R}^{D+m} = \operatorname{img}(\mathrm{D}j^0(\Phi)(\Pi w)) + \begin{bmatrix} \Pi & 0 \\ 0 & I_m \end{bmatrix} T_{j^0(\Phi\circ\Pi)(w)} A \\
\iff & \forall\, w \in j^0(\phi)^{-1}(A)\colon \mathbb{R}^{d+m} = \operatorname{img}(\mathrm{D}j^0(\phi)(w)) + T_{j^0(\phi)(w)} A \\
\overset{\text{def.}}{\iff} & j^0(\phi) \pitchfork \mathcal{A}.
\end{aligned}$$

□

4 LIKQ as a transversality condition

For given functions $c \in C^k_{\mathrm{sw}}(\mathbb{R}^d; \mathbb{R}^s)$, $g \in C^k(\mathbb{R}^d; \mathbb{R}^p)$ and $h \in C^k(\mathbb{R}^d; \mathbb{R}^q)$, this section establishes a transversality condition for the remarkable property that all feasible points $(x, z) \in \mathcal{F}(c, g, h)$ satisfy the LIKQ condition. To simplify notation set $m := s+p+q$ and let $j^0(c, g, h)$ be the standard 0-jet for the combined function $(c, g, h)\colon \mathbb{R}^d \to \mathbb{R}^m$, i.e.,

$$j^0(c, g, h)(x, y, z) := (x, y, z, c(x, y, z), g(x, y, z), h(x, y, z)). \tag{12}$$

The definition of the set $A \subseteq \mathbb{R}^{d+m}$ as

$$A := \{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{h})\colon \boldsymbol{y} = |\boldsymbol{z}|, \boldsymbol{z} = \boldsymbol{c}, \boldsymbol{g} \geq 0, \boldsymbol{h} = 0\} \tag{13}$$

then ensures that $(x, z) \in \mathcal{F}(c, g, h)$ is equivalent to $(x, |z|, z) \in j^0(c, g, h)^{-1}(A)$; which is a fundamental prerequisite to the subsequent analysis. A stratification of A is given by $\mathcal{A} := \{A_{\sigma,\omega}\colon \sigma \in \{-1, 0, 1\}^s, \omega \in \{0, 1\}^p\}$, i.e.,

$$A = \bigcup_{\substack{\sigma \in \{-1,0,1\}^s \\ \omega \in \{0,1\}^p}} A_{\sigma,\omega} \qquad \text{with} \qquad A_{\sigma,\omega} := \mathbb{R}^n \times A_\sigma \times A_\omega \times \{0\}^q$$

and

$$\begin{aligned}
A_\sigma &:= \{(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{c}) \in \mathbb{R}^s \times \mathbb{R}^s \times \mathbb{R}^s\colon \boldsymbol{y} = |\boldsymbol{z}|, \boldsymbol{z} = \boldsymbol{c}, \operatorname{sign}(\boldsymbol{z}) = \sigma\}, \\
A_\omega &:= \{\boldsymbol{g} \in \mathbb{R}^p\colon\ \operatorname{sign}(\boldsymbol{g}) = \omega\}.
\end{aligned} \tag{14}$$

To analyze the transversality condition $j^0(c, g, h) \pitchfork A$ it is useful to characterize the tangent spaces of the strata as images. This is achieved by the following lemma.

Lemma 10 (Tangent spaces and Whitney regularity)**.** *Define for fixed* $\sigma \in \{-1,0,1\}^s$ *and* $\omega \in \{0,1\}^p$ *the matrices* $\Sigma := \operatorname{diag}(\sigma)$ *and* $\Omega := \operatorname{diag}(\omega)$*. Then, for* $(\boldsymbol{y},\boldsymbol{z},\boldsymbol{c}) \in A_\sigma$*, and* $\boldsymbol{g} \in A_\omega$*, the tangent spaces are given by*

$$T_{(\boldsymbol{y},\boldsymbol{z},\boldsymbol{c})}A_\sigma = \operatorname{img}\begin{bmatrix}|\Sigma|\\ \Sigma\\ \Sigma\end{bmatrix} \tag{15}$$

and $T_{\boldsymbol{g}}A_\omega = \operatorname{img}\Omega$*. Moreover, the stratification* $\mathcal{A}$ *is weakly Whitney regular.*

Proof. Let $\nu \in T_{(\boldsymbol{y},\boldsymbol{z},\boldsymbol{c})}A_\sigma$, then there is $\varepsilon > 0$ and a differentiable curve $\gamma\colon (-\varepsilon,\varepsilon) \to A_\sigma$ with $\gamma(0) = (\boldsymbol{y},\boldsymbol{z},\boldsymbol{c})$ and $\gamma'(0) = \nu$. The definition of A_σ implies

$$\gamma(t) =: \begin{bmatrix}\gamma_y(t)\\ \gamma_z(t)\\ \gamma_c(t)\end{bmatrix} = \begin{bmatrix}|\gamma_z(t)|\\ \gamma_z(t)\\ \gamma_z(t)\end{bmatrix} = \begin{bmatrix}I_s\\ \operatorname{diag}(\sigma)\\ \operatorname{diag}(\sigma)\end{bmatrix}|\gamma_z(t)| = \begin{bmatrix}|\Sigma|\\ \Sigma\\ \Sigma\end{bmatrix}\Sigma\gamma_z(t),$$

and hence,

$$\nu = \gamma'(0) = \begin{bmatrix}|\Sigma|\\ \Sigma\\ \Sigma\end{bmatrix}\Sigma\gamma_z'(0) \in \operatorname{img}\begin{bmatrix}|\Sigma|\\ \Sigma\\ \Sigma\end{bmatrix}.$$

For ν in the right-hand side of (15) there is $\eta \in \mathbb{R}^s$ such that

$$\gamma(t) := \begin{bmatrix}\gamma_y(t)\\ \gamma_z(t)\\ \gamma_c(t)\end{bmatrix} := \begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}\\ \boldsymbol{c}\end{bmatrix} + t\nu = \begin{bmatrix}|\boldsymbol{z}|\\ \boldsymbol{z}\\ \boldsymbol{z}\end{bmatrix} + t\begin{bmatrix}|\Sigma|\\ \Sigma\\ \Sigma\end{bmatrix}\eta.$$

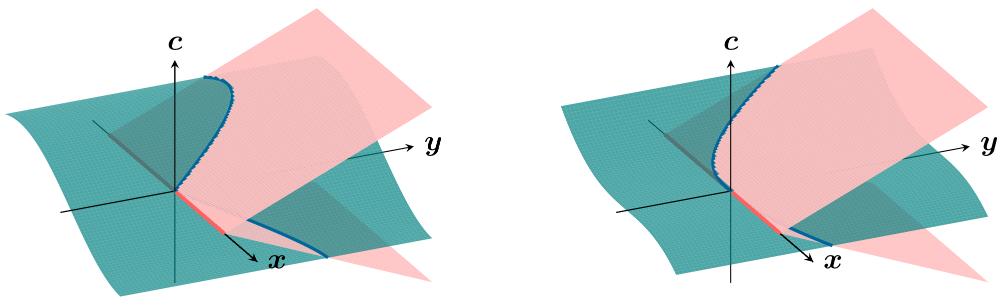
The definition of γ ensures $\gamma(0) = (\boldsymbol{y},\boldsymbol{z},\boldsymbol{c})$ and $\gamma'(0) = \nu$. Therefore, it remains to show that in the vicinity of the origin γ is a curve on A_σ. One has $\gamma_z = \gamma_c$. Since for $i \in [s]$ the term $(t\Sigma\eta)_i$ vanishes whenever $\sigma_i = 0$, there is a radius $\varepsilon > 0$ such that for $t < \varepsilon$

$$\operatorname{sign}(\gamma_z(t)) = \operatorname{sign}(\boldsymbol{z} + t\Sigma\eta) = \operatorname{sign}(\boldsymbol{z}) = \sigma.$$

In particular, $|\gamma_z(t)| = \Sigma\gamma_z(t) = \Sigma\boldsymbol{z} + t\Sigma^2\eta = |\boldsymbol{z}| + t|\Sigma|\eta = \gamma_y(t)$, whence, $\gamma(t) \in A_\sigma$. The assertion for the tangent space of A_ω follows analogously.

To show that $\mathcal{A}$ is weakly Whitney regular, let $\ell \in \mathbb{N}$ and $\boldsymbol{w}_\ell := (\boldsymbol{x}_\ell,\boldsymbol{y}_\ell,\boldsymbol{z}_\ell)$, $\boldsymbol{\phi}_\ell := (\boldsymbol{c}_\ell,\boldsymbol{g}_\ell,\boldsymbol{h}_\ell)$ such that $(\boldsymbol{w}_\ell,\boldsymbol{\phi}_\ell)$ is a sequence in $A_{\sigma,\omega}$ that converges to some limit point $(\mathring{\boldsymbol{w}},\mathring{\boldsymbol{\phi}}) := (\mathring{\boldsymbol{x}},\mathring{\boldsymbol{y}},\mathring{\boldsymbol{z}},\mathring{\boldsymbol{c}},\mathring{\boldsymbol{g}},\mathring{\boldsymbol{h}}) \in A$. Let $\mathring{\sigma} := \operatorname{sign}(\mathring{\boldsymbol{z}})$, $\mathring{\Sigma} := \operatorname{diag}(\mathring{\sigma})$, $\mathring{\omega} := \operatorname{sign}(\mathring{\boldsymbol{g}})$ and $\mathring{\Omega} := \operatorname{diag}(\mathring{\omega})$. Since $(\mathring{\boldsymbol{y}},\mathring{\boldsymbol{z}},\mathring{\boldsymbol{c}}) \in A_{\mathring{\sigma}}$ and $\mathring{\boldsymbol{g}} \in A_{\mathring{\omega}}$, it suffices to relate σ to $\mathring{\sigma}$, and ω to $\mathring{\omega}$. Now assume there is $i \in [s]$ with $\sigma_i = -\mathring{\sigma}_i \neq 0$, then there is a neighborhood of $\mathring{\boldsymbol{z}}$ in which the same relation holds true, which contradicts $\boldsymbol{z}_\ell \to \mathring{\boldsymbol{z}}$, as $\sigma = \operatorname{sign}(\boldsymbol{z}_\ell)$ for all $\ell \in \mathbb{N}$. Therefore, $\mathring{\sigma}_i = \sigma_i$ or $\mathring{\sigma}_i = 0$ for all $i \in [s]$. The same argument shows $\mathring{\omega}_i = \omega_i$ or $\mathring{\omega}_i = 0$. The previously established characterizations of the tangent spaces then yield

$$T_{(\boldsymbol{w},\boldsymbol{\phi})}A = \operatorname{img}\begin{bmatrix}I_n & 0 & 0 & 0\\ 0 & |\mathring{\Sigma}| & 0 & 0\\ 0 & \mathring{\Sigma} & 0 & 0\\ 0 & \mathring{\Sigma} & 0 & 0\\ 0 & 0 & \mathring{\Omega} & 0\\ 0 & 0 & 0 & 0\end{bmatrix} \subseteq \operatorname{img}\begin{bmatrix}I_n & 0 & 0 & 0\\ 0 & |\Sigma| & 0 & 0\\ 0 & \Sigma & 0 & 0\\ 0 & \Sigma & 0 & 0\\ 0 & 0 & \Omega & 0\\ 0 & 0 & 0 & 0\end{bmatrix} = \lim_{\ell\to\infty} T_{(\boldsymbol{w}_\ell,\boldsymbol{\phi}_\ell)}A_{\sigma,\omega},$$



(a) Example with LIKQ at all feasible points.

(b) Example without LIKQ at 0.

Figure 1: Two examples that show different situations with respect to the LIKQ condition. The pink area ( ) indicates the stratified set A and the teal area ( ) indicates the image of the jet $j^0(c)$. The intersection of the former two, the feasible set $\mathcal{F}(c)$, is depicted as the blue line ( ). Finally, the red line indicates ( ) the stratum $A_{(0)}$.

where the last limit is taken over a sequence of constant tangent spaces. □

Theorem 11 (LIKQ as a tranversality condition)**.** *For any tuple of functions* $(c, g, h) \in C^k_{\mathrm{sw}}(\mathbb{R}^d; \mathbb{R}^s) \times C^k(\mathbb{R}^d; \mathbb{R}^p) \times C^k(\mathbb{R}^d; \mathbb{R}^q)$ *one has*

$$\forall (x, z) \in \mathcal{F}(c, g, h) \colon \textit{LIKQ holds at } (x, z) \qquad \Longleftrightarrow \qquad j^0(c, g, h) \pitchfork \mathcal{A}.$$

Before the proof of Theorem 11 is presented here, the two examples of Section 2 are revisited to provide an intuition of the relation between transversality $j^0(c, g, h) \pitchfork A$ and the LIKQ condition of Definition 2.

Example 12 (LIKQ everywhere)**.** Figure 1 (a) shows the jet-space of Example 3 without the $\boldsymbol{z}$-dimension. The feasible points can be recognized in the jet-space as the intersection of the stratified set A with the image of the jet $j^0(c)$. At all points, in particular the feasible ones, the $\boldsymbol{y}$-axis is part of the linearization of the image of the jet $j^0(c)$. Moreover, the $\boldsymbol{x}$-axis is part of the tangent space at any (feasible) point in a stratum of A. Finally, for all feasible points $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{c})$ with $\boldsymbol{c} \neq 0$ the tangent spaces contain a third linear independent direction $(0, 1, \pm 1)$ which completes the combined dimensions to 3. If on the other hand $\boldsymbol{c} = 0$ for a feasible point then it is a point on the stratum $A_{(0)}$ and the tangent space is only one-dimensional. However, then the $\boldsymbol{c}$-axis is part of the linearization of the jet which again ensures transversality.

Example 13 (no LIKQ at 0)**.** Figure 1 (b) depicts the situation of Example 4. As in Example 12, the tangent space of $A_{(0)}$ is just the $\boldsymbol{x}$-axis. However, here the linearization of $j^0(x, y, z)$ at 0 spans the $\boldsymbol{x}$-$\boldsymbol{y}$-plane leaving the $\boldsymbol{c}$-axis as a linear independent dimension.

Proof. By definition the transversality $j^0(c,g,h) \pitchfork \mathcal{A}$, holds if and only if for all $(x,y,z) \in j^0(c,g,h)^{-1}(A)$

$$\operatorname{img}(\mathrm{D}j^0(c,g,h)(x,y,z)) + T_{j^0(c,g,h)(x,y,z)}A_{\sigma,\omega} = \mathbb{R}^{d+m}, \tag{16}$$

where $A_{\sigma,\omega}$ is the stratum containing $(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{c},\boldsymbol{g},\boldsymbol{h}) := j^0(c,g,h)(x,y,z)$. By the definitions of the sets A_σ and A_ω in Equation (14), one obtains $\sigma = \operatorname{sign}(\boldsymbol{z}) = \operatorname{sign}(z)$, $\omega = \operatorname{sign}(\boldsymbol{g})$ and $y = |z|$. Define $\Sigma := \operatorname{diag}(\sigma)$ and $\Omega := \operatorname{diag}(\omega)$, then by Lemma 10 the condition (16) is equivalent to

$$\operatorname{rank}\begin{bmatrix} I_n & 0 & 0 & I_d & 0 & 0 \\ 0 & I_s & 0 & 0 & |\Sigma| & 0 \\ 0 & 0 & I_s & 0 & \Sigma & 0 \\ \mathrm{D}_1c & \mathrm{D}_2c & \mathrm{D}_3c & 0 & \Sigma & 0 \\ \mathrm{D}_1g & \mathrm{D}_2g & \mathrm{D}_3g & 0 & 0 & \Omega \\ \mathrm{D}_1h & \mathrm{D}_2h & \mathrm{D}_3h & 0 & 0 & 0 \end{bmatrix} = d+m, \tag{17}$$

where the trivial last column in the characterization of the tangent spaces is left out, since it does not add to the overall rank, and the dependencies on the evaluation point $(x,|z|,z)$ is omitted in the partial derivatives of c, g and h for notational ease. Multiplication from the left of the above matrix with the full rank matrix

$$\begin{bmatrix} I_n & 0 & 0 & 0 & 0 & 0 \\ 0 & I_s & 0 & 0 & 0 & 0 \\ 0 & 0 & I_s & 0 & 0 & 0 \\ 0 & -S\mathrm{D}_2c & -S\mathrm{D}_3c & S & 0 & 0 \\ 0 & -(\mathrm{D}_2g\Sigma + \mathrm{D}_3g)S\mathrm{D}_2c - \mathrm{D}_2g & -(\mathrm{D}_2g\Sigma + \mathrm{D}_3g)S\mathrm{D}_3c - \mathrm{D}_3g & (\mathrm{D}_2g\Sigma + \mathrm{D}_3g)S & I_p & 0 \\ 0 & -(\mathrm{D}_2h\Sigma + \mathrm{D}_3h)S\mathrm{D}_2c - \mathrm{D}_2h & -(\mathrm{D}_2h\Sigma + \mathrm{D}_3h)S\mathrm{D}_3c - \mathrm{D}_3h & (\mathrm{D}_2h\Sigma + \mathrm{D}_3h)S & 0 & I_q \end{bmatrix},$$

where $S := (I_s - \mathrm{D}_2c\Sigma - \mathrm{D}_3c)^{-1} \in \mathbb{R}^{s\times s}$, yields

$$\operatorname{rank}\begin{bmatrix} I_n & 0 & 0 & I_n & 0 & 0 \\ 0 & I_s & 0 & 0 & |\Sigma| & 0 \\ 0 & 0 & I_s & 0 & \Sigma & 0 \\ J_z(x,z) & 0 & 0 & 0 & \Sigma & 0 \\ J_g(x,z) & 0 & 0 & 0 & 0 & \Omega \\ J_h(x,z) & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = d+m$$

as an equivalent characterization of transversality. Clearly, the first $d = n+s+s$ rows of this matrix are independent, and thus, the final reformulation of the transversality condition in (16) reads

$$\operatorname{rank}\begin{bmatrix} J_z(x,z) & \Sigma & 0 \\ J_g(x,z) & 0 & \Omega \\ J_h(x,z) & 0 & 0 \end{bmatrix} = m.$$

This is exactly the LIKQ condition in Definition 2 if the rows that are trivially linear independent from the rest, due to nonzero entries in Σ or Ω, are removed from the matrix. □

Remark 14 (No addtional constraints)**.** An analogous statement to Theorem 11 can be derived, when considering the problem class without additional inequality and equality constraints. In that case the LIKQ condition of Definition 2 reduces to a full rank condition of $P_\alpha \alpha(z) J_z(x, z)$. Changing of the definition of the feasible and stratified set accordingly, i.e., to $\mathcal{F}(c) \coloneqq \{(x, z) : \ z = c(x, |z|, z)\}$ and

$$A \coloneqq \bigcup_{\sigma \in \{-1,0,1\}^s} \{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{c})) : \ \boldsymbol{z} = \boldsymbol{c}, \boldsymbol{y} = |\boldsymbol{z}|, \operatorname{sign}(\boldsymbol{z}) = \sigma\}$$

one can use virtually the same arguments as in the proof of Theorem 11. The only difference being that the last two rows and the last column of the matrix in (17) are not present.

Remark 15 (No free switching variable)**.** Earlier papers on abs-smooth optimization, in particular [Gri13; GW16; WG19], used a different formalization to represent an abs-smooth function. Therein, the functions in the evaluation procedure do not depend explicitly on the switching variable z but only on its absolute value. That is, the representation of $\varphi \in C^{k,s}_{\text{abs}}(\mathbb{R}^n)$ reads

$$\begin{aligned} \varphi(x) &= f(x, |z|), \\ z &= c(x, |z|). \end{aligned}$$

The same can be done for abs-smooth inequality constraints and abs-smooth equality constraints prescribed by functions φ_g and φ_h respectively. Consequently, in such a formulation the matrices in (7) would not involve the corresponding partial derivatives with respect to a third argument. However, the definition of the feasible set, the jet and the strata in (4), (12) and (14) would need to remain unchanged to properly encode feasibility. The then seemingly unjustified input variable z of the jet can be explained by introducing artificial function $\psi(x, y, z)$ and considering the reduced structured jet that drops the actual value of ψ, c.f. [Gün08]. In the later context on genericity this additional freedom in the possible pertubations to the problem description needs to be justified, which will be done in Remark 17. In a proof of a theorem analogous to Theorem 11 basically nothing would change, except that all partial derivatives with respect to z are replaced by 0-matrices of the corresponding dimension.

5 Genericity of LIKQ

As already discussed after Definition 6 the perturbations that the standard jet-transversality theorem [JJT00, Theorem 7.4.5] for $j^0(c, g, h)$ and A as defined in (12) and (13) considers are too general. In particular, when perturbing c in $C^k(\mathbb{R}^d; \mathbb{R}^m)$ even slightly with respect to the strong Whitney topology, the result may not be a valid switching function anymore. In contrast, perturbing the functions $c_1, \ldots, c_s$ that define c via (2) individually will always result in a valid switching function. However, since the signatures of the c_i, $i \in [s]$ differ, this requires the use of the structured jet-transversality Theorem 8.

Theorem 16 (LIKQ is a generic assumption)**.** *Assume $k \geq n - q$. Then, the set $\mathcal{G} \subset C^k(\mathbb{R}^{d_1}) \times \ldots \times C^k(\mathbb{R}^{d_s}) \times C^k(\mathbb{R}^d; \mathbb{R}^p) \times C^k(\mathbb{R}^d; \mathbb{R}^q)$ defined by*

$$\mathcal{G} := \{(c_1, \ldots, c_s, g, h) \,:\, \forall (x, z) \in \mathcal{F}(c, g, h) \text{ LIKQ holds at } (x, z)\}$$

is an open and dense subset of $C^k(\mathbb{R}^{d_1}) \times \ldots \times C^k(\mathbb{R}^{d_s}) \times C^k(\mathbb{R}^d; \mathbb{R}^p) \times C^k(\mathbb{R}^d; \mathbb{R}^q)$ with respect to the strong C^k-Whitney topology.

Proof. Let $\tilde{s} := s + 2$. For $i \in [s]$ let $d_i := n + 2(i-1)$ and $d_{s+1} := d_{s+2} := d$. The matrices $\mathbb{R}^{d_i \times d}$ with

$$\Pi_i \begin{bmatrix} x & y & z \end{bmatrix}^T := \begin{bmatrix} x & y_1 & \ldots & y_{i-1} & z_1 & \ldots & z_{i-1} \end{bmatrix}^T, \quad i \in [s]$$

generate the required inputs for i-th line of the switching function c_i. By setting $\Pi_{s+1} := \Pi_{s+2} := I_d$ and $D := d_1 + \ldots + d_{\tilde{s}}$ the combined matrix

$$\Pi := \begin{bmatrix} \Pi_1 \\ \vdots \\ \Pi_{\tilde{s}} \end{bmatrix} \in \mathbb{R}^{D \times d}$$

when applied to $(x, y, z) \in \mathbb{R}^d$ provides valid inputs for the structured evaluation of $\Phi := (c_1, \ldots, c_s, g, h)$ in the sense that

$$\Phi(\Pi(x, y, z)) = (c_1(x), \ldots, c_s(x, y_1, \ldots, y_{s-1}, z_1, \ldots, z_{s-1}), g(x, y, z), h(x, y, z)).$$

Moreover, Π is left-invertible, as the last submatrix Π_{s+2} is an identity on the input, and it holds $\Phi \circ \Pi = (c, g, h) \in C^k(\mathbb{R}^d; \mathbb{R}^m)$. Since the stratification $\mathcal{A}$ of the A defined in (13) is weakly Whitney regular, the Assertion a) and b) of Lemma 9 show that $\tilde{\mathcal{A}}$ is a Whitney regular stratification of $\tilde{A}$, where

$$\tilde{\mathcal{A}} := \left\{ \begin{bmatrix} \Pi & 0 \\ 0 & I_m \end{bmatrix} A_{\sigma,\omega} \,:\, \sigma \in \{-1, 0, 1\}^s, \omega \in \{0, 1\}^p \right\} \quad \text{and} \quad \tilde{A} := \begin{bmatrix} \Pi & 0 \\ 0 & I_m \end{bmatrix} A.$$

Since the transformation that is applied to each stratum is left-invertible by construction, it is bijective onto its image, and hence, $\dim(\tilde{\mathcal{A}}) = \dim(\mathcal{A}) \leq n + s + p$. Now Theorem 8 shows that the set $\{(c_1, \ldots, c_s, g, h) \,:\, j^0(c_1, \ldots, c_s, g, h) \pitchfork \tilde{\mathcal{A}}\}$ is an open and dense subset of $C^k(\mathbb{R}^{d_1}) \times \ldots \times C^k(\mathbb{R}^{d_s}) \times C^k(\mathbb{R}^d; \mathbb{R}^p) \times C^k(\mathbb{R}^d; \mathbb{R}^q)$ with respect to the strong Whitney topology. Further, Lemma 9 c) ensures that

$$\{(c_1, \ldots, c_s, g, h) \,:\, j^0(c_1, \ldots, c_s, g, h) \pitchfork \tilde{\mathcal{A}}\} = \{(c_1, \ldots, c_s, g, h) \,:\, j^0((c, g, h)) \pitchfork \mathcal{A}\},$$

while Theorem 11 provides the required connection to the LIKQ condition. □

Remark 17 (genericity in other settings)**.** In view of Remark 14 and Remark 15 one might wonder if the result of Theorem 16 still holds if the problems are formulated without additional inequality and equality constraints, or when the functions c, g and h are formulated without an explicit dependence on the switching variable z.

In the first case of problem formulations without additional constraints g, h, the injectivity of the operator Π can not be concluded as in the previous proof. Similar to the argument in Remark 15, this can be fixed by adding an artificial function ψ that takes all inputs x, y and z and whose value is reduced in a structured jet that together with A encodes feasibility. The inputs to ψ are then again taken care of by an identity block in the last d rows of Π which ensures the injectivity just as before.

Thus, in both cases the final application of the structured-jet-transversality theorem in the proof of Theorem 16 ensures that the functions that have LIKQ at every feasible points form an open and dense set. Removing the artificial function ψ to obtain a clear result can simply be done by projecting onto the other components. This projection preserves the openness and the density.

6 Conclusion and Outlook

For abs-smooth optimization problems, the property LIKQ serves as a qualification of the nonsmoothness that allows to verify the optimality of a given point in polynomial time. The main result of the paper at hand states that requiring LIKQ at all feasible points of an abs-smooth problem is a generic assumption in that the set of problems for which this is true is dense and open in the strong Whitney topology.

Optimality conditions for large classes of nonsmooth optimization problems were derived in [GW16] and [HS20]. As a next step, future research could aim at developing a topologically meaningful stationarity definition, i.e., one that corresponds to a topological change of the sub-level sets, and prove that generically all stationary points in this sense satisfy an associated non-degenericity condition. Most likely this can be achieved using similar arguments as the once used in this paper. In particular one requires the 1-jet, i.e., a jet the incorporates the first derivatives of the involved functions and a stratified set that encodes the topological stationarity condition on images of that jet. Using again Lemma 9 this jet should be relatable to a structured jet and the application of Theorem 8 then should provide the genericity result. Similar results for other classes of nonsmooth optimization problems have been published in [JRS09; DSS12; DJS13; LS22].

Acknowledgements

The authors are very grateful Vladimir Shikhman for his thorough introduction to the theory of jet-transversality and numerous discussions on this subject.

The work on this paper was partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – The Berlin Mathematics Research Center MATH+ (EXC-2046/1, project ID: 390685689).

References

[DJS13] D. Dorsch, H. Jongen, and V. Shikhman. "On Intrinsic Complexity of Nash Equilibrium Problems and Bilevel Optimization". In: *Journal of Optimization Theory and Applications* 159 (Dec. 2013), pp. 606–634.

[DSS12] D. Dorsch, V. Shikhman, and O. Stein. "Mathematical programs with vanishing constraints: critical point theory". In: *J. of Global Optimization* 52.3 (Mar. 2012), pp. 591–605. URL: https://doi.org/10.1007/s10898-011-9805-z.

[Gri+15] A. Griewank et al. "Solving piecewise linear systems in abs-normal form". In: *Linear Algebra and its Applications* 471 (2015), pp. 500–530. URL: https://www.sciencedirect.com/science/article/pii/S0024379514008209.

[Gri13] A. Griewank. "On stable piecewise linearization and generalized algorithmic differentiation". In: *Optimization Methods and Software* 28.6 (Apr. 2013), pp. 1139–1178.

[Gün08] H. Günzel. "The structured jet transversality theorem". In: *Optimization* 57.1 (2008), pp. 159–164. URL: https://doi.org/10.1080/02331930701779039.

[GW16] A. Griewank and A. Walther. "First- and second-order optimality conditions for piecewise smooth objective functions". In: *Optimization Methods and Software* 31.5 (2016), pp. 904–930. URL: https://doi.org/10.1080/10556788.2016.1189549.

[Hir76] M. W. Hirsch. *Differential Topology*. 1st ed. Graduate Texts in Mathematics. Springer New York, NY, 1976, pp. x+222. URL: https://doi.org/10.1007/978-1-4684-9449-5.

[HKS21] L. C. Hegerhorst-Schultchen, C. Kirches, and M. C. Steinbach. "Relations between Abs-Normal NLPs and MPCCs. Part 1: Strong Constraint Qualifications". In: *Journal of Nonsmooth Analysis and Optimization* 2 (Feb. 2021). URL: https://jnsao.episciences.org/7196.

[HS20] L. C. Hegerhorst-Schultchen and M. C. Steinbach. "On first and second order optimality conditions for abs-Normal NLP". In: *Optimization* 69.12 (2020), pp. 2629–2656. URL: https://doi.org/10.1080/02331934.2019.1626386.

[JJT00] H. Jongen, P. Jonker, and F. Twilt. *Nonlinear Optimization in Finite Dimensions: Morse Theory, Chebyshev Approximation, Transversality, Flows, Parametric Aspects*. Nonconvex Optimization and Its Applications. Springer, 2000. URL: https://books.google.de/books?id=CY3n9w3upbUC.

[JRS09] H. T. Jongen, J.-J. Rückmann, and V. Shikhman. "MPCC: Critical Point Theory". In: *SIAM Journal on Optimization* 20.1 (2009), pp. 473–484. URL: https://doi.org/10.1137/080733693.

[JT79] H. Jongen and F. Twilt. "On Decomposition and Structural Stability in Non-Convex Optimization". In: *Numerische Methoden bei graphentheoretischen und kombinatorischen Problemen: Band 2*. Ed. by L. Collatz, G. Meinardus, and W. Wetterling. 1979, pp. 162–183.

[Kri69] N. Krikorian. “A note concerning the fine topology on function spaces”. en. In: *Compositio Mathematica* 21.4 (1969), pp. 343–348. URL: http://www.numdam.org/item/CM_1969__21_4_343_0/.

[Lee12] J. M. Lee. *Introduction to Smooth Manifolds.* 2nd ed. Graduate Texts in Mathematics. Springer New York, NY, 2012, pp. xvi + 708.

[LS22] S. Lämmel and V. Shikhman. “On nondegenerate M-stationary points for sparsity constrained nonlinear optimization”. In: *Journal of Global Optimization* 82.2 (2022), pp. 219–242.

[Shy+25] G. Shyshkanova et al. “On the nonsmooth regularity condition LIKQ for different abs-normal representations”. In: *XXX*. De Gruyter, 2025.

[SR79] J. E. Spingarn and R. T. Rockafellar. “The Generic Nature of Optimality Conditions in Nonlinear Programming”. In: *Mathematics of Operations Research* 4.4 (1979), pp. 425–430. (Visited on 11/12/2024).

[SS01] S. Scholtes and M. Stöhr. “How stringent is the linear independence assumption for mathematical programs with complementarity constraints?” In: *Math. Oper. Res.* 26.4 (2001), pp. 851–863. URL: https://doi.org/10.1287/moor.26.4.851.10007.

[WG19] A. Walther and A. Griewank. “Characterizing and Testing Subdifferential Regularity in Piecewise Smooth Optimization”. In: *SIAM Journal on Optimization* 29.2 (2019), pp. 1473–1501. URL: https://doi.org/10.1137/17M115520X.