The effect of smooth parametrizations on nonconvex optimization landscapes

Eitan Levin∗ Joe Kileel† Nicolas Boumal‡

September 4, 2023

Abstract

We develop new tools to study landscapes in nonconvex optimization. Given one optimization problem, we pair it with another by smoothly parametrizing the domain. This is either for practical purposes (e.g., to use smooth optimization algorithms with good guarantees) or for theoretical purposes (e.g., to reveal that the landscape satisfies a strict saddle property). In both cases, the central question is: how do the landscapes of the two problems relate? More precisely: how do desirable points such as local minima and critical points in one problem relate to those in the other problem? A key finding in this paper is that these relations are often determined by the parametrization itself, and are almost entirely independent of the cost function. Accordingly, we introduce a general framework to study parametrizations by their effect on landscapes. The framework enables us to obtain new guarantees for an array of problems, some of which were previously treated on a case-by-case basis in the literature. Applications include: optimizing low-rank matrices and tensors through factorizations; solving semidefinite programs via the Burer–Monteiro approach; training neural networks by optimizing their weights and biases; and quotienting out symmetries.

Keywords: Overparametrization, symmetries in landscapes, benign nonconvexity, strict saddles, low rank optimization, Hadamard parametrization of simplex, optimization on manifolds.

Contents

1 Introduction 2

2 Lifts and their properties 4
  2.1 The sphere-to-simplex Hadamard lift 6
  2.2 Smooth semidefinite programs via Burer–Monteiro 7
  2.3 Low-rank matrices 7

∗Dept. of Computing and Mathematical Sciences, California Institute of Technology (eitanl@caltech.edu)
†Dept. of Mathematics and Oden Institute for Computational Engineering and Sciences, University of Texas at Austin (jkileel@math.utexas.edu)
‡Institute of Mathematics, École polytechnique fédérale de Lausanne (EPFL) (nicolas.boumal@epfl.ch)
We consider pairs of optimization problems (P) and (Q) as defined below, where $\mathcal{E}$ is a linear space, $\mathcal{M}$ is a smooth manifold, and $\varphi: \mathcal{M} \to \mathcal{E}$ is a smooth (over)parametrization of the search space $\mathcal{X} = \varphi(\mathcal{M})$ of (P). Their optimal values are equal:

$$\min_{x \in \mathcal{X}} f(x) \quad (P)$$
$$\min_{y \in \mathcal{M}} g(y) \quad (Q)$$

\[ \mathcal{M} \xrightarrow{\varphi} \mathcal{X} \subseteq \mathcal{E} \xrightarrow{f} \mathbb{R} \]

The two-headed arrow $\rightarrow$ in the diagram denotes a surjection.
We usually assume \( f : \mathcal{E} \to \mathbb{R} \) is smooth \((C^\infty)\), hence so is \( g = f \circ \varphi : \mathcal{M} \to \mathbb{R} \) by composition.

Such pairs of problems \((P)\) and \((Q)\) arise in two scenarios (concrete examples follow):

(a) Our task is to minimize \( f \) on \( \mathcal{X} \) as in \((P)\), but we lack good algorithms to do so, e.g., because \( \mathcal{X} \) lacks regularity. In this case, we choose a smooth parametrization \( \varphi \) of \( \mathcal{X} \) and run algorithms on the smooth problem \((Q)\) instead.

(b) Our task is to minimize \( g \) on \( \mathcal{M} \) as in \((Q)\), but its landscape is complex (e.g., due to symmetries). In this case, we factor \( g \) through a smooth map \( \varphi \) in the hope of revealing a problem \((P)\) whose landscape is simpler and can be leveraged to analyze that of \((Q)\).

In both cases, we run an optimization algorithm on the smooth problem \((Q)\). This algorithm may find desirable points \( y \) on \( \mathcal{M} \) for \((Q)\) (global or local minima, stationary points). For example, certain trust-region algorithms are guaranteed to accumulate at second-order stationary points—see [18] and an extension to manifolds [37, §3]—and many first- and even zeroth-order methods converge to second-order stationary points from almost all initializations [4, 34, 55]. However, in general such points need not map to desirable points \( \varphi(y) \) on \( \mathcal{X} \) for \((P)\). Indeed, nonlinear parametrizations may severely distort landscapes, and notably may introduce spurious critical points. Algorithms running on \((Q)\) are liable to terminate at an approximately stationary point near such a spurious point, and return a point whose image through \( \varphi \) is nowhere near any stationary point for \((P)\).

In this paper, we characterize the properties that the parametrization \( \varphi \) needs to satisfy for desirable points of \((Q)\) to map to desirable points of \((P)\), that is, we develop a general framework to relate the landscapes of pairs of problems of the above form. Importantly, we observe that these properties are often entirely independent of the cost function \( f \) in \((P)\), since many parametrizations map desirable points for \((Q)\) to those for \((P)\) for any cost function. Our framework enables us to unify and strengthen the analysis of a wealth of parametrizations arising in applications, hitherto studied case-by-case and often only for specific costs.

Parametrizations are ubiquitous. They arise in semidefinite programming [9, 15], low-rank optimization [23, 36, 41], computer vision [19], inverse kinematics and trajectory planning [49, Chaps. 1,4], algebraic geometry [24, Chap. 17], training neural networks [43, 38], and risk minimization [51, 6, 7]. The following are two concrete examples that illustrate the above two scenarios.

For an example of scenario \((a)\), consider minimizing a cost \( f \) over the set \( \mathcal{X} = \mathbb{R}^{m \times n}_{\leq r} \) of all \( m \times n \) matrices of rank at most \( r \). Unfortunately, standard algorithms running on \((P)\) may converge to a non-stationary point because of the nonsmooth geometry of \( \mathcal{X} \) [37, 42]. Instead of trying to solve \((P)\) directly, it is common to parametrize \( \mathcal{X} \) by the linear space \( \mathcal{M} = \mathbb{R}^{m \times r} \times \mathbb{R}^{n \times r} \) using the rank factorization \( \varphi(L,R) = LR^\top \), and to solve \((Q)\) instead. The resulting problem \((Q)\) requires minimizing a smooth cost function over a linear space; there are several algorithms that converge to a second-order stationary point for such problems. Furthermore, any second-order stationary point for \((Q)\) maps under \( \varphi \) to a stationary point for \((P)\) by [23, Thm. 1]. Thus, parametrization of \( \mathcal{X} \) by \( \varphi \) gives us an algorithm converging to a stationary point for \((P)\) by running standard algorithms on \((Q)\), even though similarly reasonable algorithms may fail to produce a stationary point when applied directly to \((P)\).

For an illustration of scenario \((b)\), consider finding the smallest eigenvalue of a \( d \times d \) symmetric matrix \( A \), which can be written in the form \((Q)\) with \( \mathcal{M} \) the unit sphere in \( \mathbb{R}^d \) and \( g(y) = y^\top Ay \). This problem is not convex, hence it could have bad local minima. Here is one way to reason that it does not (as is well known). If \( \lambda \in \mathbb{R}^d \) denotes the vector of eigenvalues of \( A \) and \( U \in O(d) \) is an orthogonal matrix of eigenvectors satisfying \( A = U \text{diag}(\lambda) U^\top \) (both of which are unknown), define \( \varphi(y) = \text{diag}(U^\top yy^\top U) \in \mathbb{R}^d \) and \( f(x) = \lambda^\top x \). It is easy to check that \( g = f \circ \varphi \), and that
\( \mathcal{X} = \varphi(\mathcal{M}) \) is the standard simplex in \( \mathbb{R}^d \). The resulting problem (P) is convex in this case, hence each of its stationary points is a global minimum. A corollary of the theory we develop in this paper is that any second-order stationary point for (Q) with \( \varphi \) as above maps to a stationary point for (P), for any cost function \( f \)—see Example 4.13. Thus, we recover the well-known fact that any second-order stationary point for the eigenvalue problem (Q) is globally optimal. Even though the problem (P) cannot be solved directly in this case because \( f \) and \( \varphi \) are unknown, their mere existence can be used to show that the nonconvexity of (Q) is “benign”. From this perspective, problem (P) reveals hidden convexity in problem (Q). This hidden convexity is present more generally in lifts arising from Kostant’s convexity theorem, extending this example to optimization of certain linear functions over certain Lie group orbits [33].

We state our main definitions and results relating the landscapes of (P) and (Q) in general, and instantiate these results on a number of specific lifts arising in the literature, in Section 2. Table 1 collects the notations and definitions for several sets used throughout the paper. Other notation and basic definitions are introduced in the text as needed and are collected for the reader’s convenience in Appendix E.

| Bounded-rank matrices | \( \mathbb{R}_{\leq r}^{m \times n} = \{ X \in \mathbb{R}^{m \times n} : \text{rank}(X) \leq r \} \) |
|-----------------------|------------------------------------------------------------------|
| Fixed-rank matrices   | \( \mathbb{R}_{= r}^{m \times n} = \{ X \in \mathbb{R}^{m \times n} : \text{rank}(X) = r \} \) |
| Symmetric matrices    | \( \mathbb{S}^r = \{ X \in \mathbb{R}^{m \times m} : X = X^\top \} \) |
| Nonnegative orthant   | \( \mathbb{R}_{\geq 0}^n = \{ x \in \mathbb{R}^n : x_i \geq 0 \text{ for all } i \} \) |
| Positive orthant      | \( \mathbb{R}_{> 0}^n = \{ x \in \mathbb{R}^n : x_i > 0 \text{ for all } i \} \) |
| Positive-semidefinite (PSD) matrices | \( \mathbb{S}_{\geq 0}^m = \{ X \in \mathbb{S}^m : X \succeq 0 \} \) |
| Positive-definite (PD) matrices | \( \mathbb{S}_{> 0}^m = \{ X \in \mathbb{S}^m : X \succ 0 \} \) |
| Standard simplex      | \( \Delta^{n-1} = \{ x \in \mathbb{R}_{\geq 0}^n : \sum_{i=1}^n x_i = 1 \} \) |
| Stiefel matrices      | \( \text{St}(m, r) = \{ X \in \mathbb{R}^{m \times r} : X^\top X = I_r \} \) |
| Unit sphere           | \( \mathbb{S}^{m-1} = \text{St}(m, 1) \) |
| Orthogonal matrices   | \( O(m) = \text{St}(m, m) \) |
| Special orthogonal matrices | \( \text{SO}(m) = \{ X \in O(m) : \det(X) = 1 \} \) |
| Invertible matrices   | \( \text{GL}(m) = \{ X \in \mathbb{R}^{m \times m} : \det(X) \neq 0 \} \) |
| Grassmannian          | \( \text{Gr}(m, r) = \{ S \subseteq \mathbb{R}^m : S \text{ is a linear subspace, } \dim S = r \} \) |

Table 1: Sets used frequently in the paper. Other sets that appear include low rank tensors and functions representable by fixed neural network architectures.

## 2 Lifts and their properties

We call the parametrization in (Q) a (smooth) lift of \( \mathcal{X} \):

**Definition 2.1.** A smooth lift of \( \mathcal{X} \subseteq \mathcal{E} \) is a smooth manifold \( \mathcal{M} \) together with a smooth map \( \varphi : \mathcal{M} \to \mathcal{E} \) such that \( \varphi(\mathcal{M}) = \mathcal{X} \).
As the two scenarios in Section 1 illustrate, understanding when lifts map desirable points for (Q) to desirable points for (P) yields guarantees for algorithms running on (Q). Here desirable points might be minimizers (global or local) and stationary points (of first, second, or higher order). The relation between these two sets of desirable points has been studied for various specific lifts and cost functions. In this paper, we study this relation in general and answer the following question:

Which lifts have the property that desirable points of (P) map to desirable points of (Q), for all cost functions f?

Surprisingly, we find that many lifts arising in practice satisfy such properties, yielding guarantees for algorithms running on (Q) that are independent of the particular cost function involved, and only depend on the geometry of the lift. We further show that whenever a lift does not preserve desirable points for all cost functions, then it fails to do so already for quite simple costs. In this case our results identify obstructions to proofs of guarantees for algorithms, which must then exploit the structure of the particular cost function at hand.

To begin answering the above question, we note that global minima of (Q) always map under \( \varphi \) to global minima of (P), for all cost functions \( f \). This holds simply because \( \varphi(M) = \mathcal{X} \), see Proposition 3.5. Global minima are hard to find in general, so we study other types of desirable points such as local minima and stationary points. In contrast to global minima, these types of desirable points are not guaranteed to map\(^2\) to each other under smooth lifts. In fact, it is possible for a local minimum of (Q) to map under \( \varphi \) to a non-stationary point for (P), see Example 3.7. Thus, we define the following properties of smooth lifts that, when satisfied, yield a connection between desirable points for (Q) and those for (P).

**Definition 2.2** (Desirable properties of lifts). Suppose \( \varphi: \mathcal{M} \to \mathcal{X} \) is a smooth lift.

(a) The lift \( \varphi \) satisfies the \textit{local} \( \Rightarrow \) \textit{local} property at \( y \in \mathcal{M} \) if, for all continuous \( f: \mathcal{X} \to \mathbb{R} \), if \( y \) is a local minimum for (Q) then \( x = \varphi(y) \) is a local minimum for (P). We say \( \varphi \) satisfies the \textit{local} \( \Rightarrow \) \textit{local} property if it does so at all \( y \in \mathcal{M} \).

(b) The lift \( \varphi \) satisfies the \textit{k} \( \Rightarrow \textit{1} \) property at \( y \) for \( k = 1, 2 \) if for all \( k \)-times differentiable \( f: \mathcal{X} \to \mathbb{R} \), if \( y \) is a \( k \)-th-order stationary point (\( \textit{k-critical} \) for short) for (Q) then \( x = \varphi(y) \) is stationary for (P). We say \( \varphi \) satisfies the \textit{k} \( \Rightarrow \textit{1} \) property if it does so at all \( y \in \mathcal{M} \).

The precise definitions of each type of desirable points above is given in Section 3. We fully characterize these properties and explain how to check them on specific examples. We then apply our results to study lifts arising in applications ranging from low-rank matrices and tensors to neural networks.

Note that \( 2 \Rightarrow 1 \) at \( y \) implies \( 1 \Rightarrow 1 \) at \( y \) since 2-critical points are 1-critical, but no other implication between the different properties holds in general—see Remark 2.12. We also mention that \( C^\infty \) smoothness is not necessary for the above properties or for their characterizations. For example, it suffices for the manifold \( \mathcal{M} \) and the lift \( \varphi \) to be of class \( C^k \) for the definition of \( k \Rightarrow 1 \) and its characterization to apply. For \( \textit{local} \Rightarrow \textit{local} \) it suffices for \( \mathcal{M} \) to be a topological space satisfying certain properties (see Appendix A) and for \( \varphi \) to be continuous.

Our characterizations of \( \textit{local} \Rightarrow \textit{local} \) and \( \textit{1} \Rightarrow \textit{1} \) are easy to state as follows.

**Theorem 2.3.** The lift \( \varphi: \mathcal{M} \to \mathcal{X} \) satisfies \( \textit{local} \Rightarrow \textit{local} \) at \( y \in \mathcal{M} \) if and only if it is open at \( y \). If \( \varphi \) does not satisfy \( \textit{local} \Rightarrow \textit{local} \) at \( y \), there is a smooth cost \( f \) such that \( y \) is a local minimum for (Q) but \( \varphi(y) \) is not a local minimum for (P).

---

\(^2\)The converse question is simple: preimages of local minima are local minima by continuity of \( \varphi \) (Proposition 3.6), and preimages of stationary points are stationary by differentiability of \( \varphi \) (Proposition 3.13).
By definition, the map \( \varphi \) is open at \( y \in \mathcal{M} \) if it maps neighborhoods of \( y \) (that is, sets containing \( y \) in their interior) to neighborhoods of \( \varphi(y) \) (in the subspace topology on \( \mathcal{X} \) from \( \mathcal{E} \))—a purely topological property. Proving that openness is sufficient for “local \( \Rightarrow \) local” is easy. Proof of its necessity requires substantial work, deferred to Appendix A. Our proof in the appendix provides the result in a more general, topological setting without using smoothness. It also provides (possibly new) conditions which are equivalent to openness and may be easier to check for some lifts.

Our characterization for “1 \( \Rightarrow \) 1” involves the image of the differential of the lift map \( \varphi \), and is proved in Section 3.2.1.

**Theorem 2.4.** The lift \( \varphi: \mathcal{M} \rightarrow \mathcal{X} \) satisfies “1 \( \Rightarrow \) 1” at \( y \in \mathcal{M} \) if and only if \( \text{im } D\varphi(y) = T_y\mathcal{X} \), where \( x = \varphi(y) \). If \( \varphi \) does not satisfy “1 \( \Rightarrow \) 1”, there is a linear cost \( f \) such that \( y \) is 1-critical for (Q) but \( \varphi(y) \) is not stationary for (P).

Here \( \varphi \) is viewed as a smooth map between smooth manifolds \( \mathcal{M} \rightarrow \mathcal{E} \), and its differential \( D\varphi(y) \) maps the tangent space \( T_y\mathcal{M} \) to (in general, a subset of) the tangent cone \( T_y\mathcal{X} \), see Definition 3.1. Since \( D\varphi(y) \) is a linear map and \( T_y\mathcal{M} \) is a linear space, “1 \( \Rightarrow \) 1” is rarely satisfied: unless all tangent cones of \( \mathcal{X} \) are linear subspaces, for every smooth lift \( \varphi \), there exists a (linear) \( f \) such that some stationary point for (Q) does not map to a stationary point for (P).

Our characterization for “2 \( \Rightarrow \) 1” is more complicated, involving the second derivative of \( \varphi \) as well. We state an equivalent condition for “2 \( \Rightarrow \) 1”, as well as sufficient and necessary conditions for it that are easier to check in some applications, in Theorem 3.23. If “2 \( \Rightarrow \) 1” fails at \( y \), we show in Corollary 3.18 that there exists a convex quadratic cost \( f \) such that \( y \) is 2-critical for (Q) but \( \varphi(y) \) is not stationary for (P). Note also that if “1 \( \Rightarrow \) 1” holds at \( y \) then so does “2 \( \Rightarrow \) 1” by definition.

Understanding stationarity on \( \mathcal{X} \) requires knowledge of its tangent cones. These can be hard to characterize. We show that it is sometimes possible to obtain an explicit expression for the tangent cones simultaneously with proving “1 \( \Rightarrow \) 1” and “2 \( \Rightarrow \) 1” for some lift of \( \mathcal{X} \), see Sections 3.5 and 4. This is somewhat surprising since the tangent cones to \( \mathcal{X} \) are defined independently of any lift.

Given a set \( \mathcal{X} \), it is also natural to seek constructions of a smooth lift \( \varphi: \mathcal{M} \rightarrow \mathcal{X} \) satisfying desirable properties. We give a systematic construction of a map \( \varphi: \mathcal{M} \rightarrow \mathcal{X} \) in Section 4 which maps a set \( \mathcal{M} \) surjectively onto \( \mathcal{X} \). When the set \( \mathcal{M} \) constructed in this way is a smooth manifold, we obtain a smooth lift and give conditions under which \( \varphi \) satisfies each of the above properties.

We now proceed to apply our results to study various lifts arising in the literature.

### 2.1 The sphere-to-simplex Hadamard lift

There is growing interest in optimizing over the probability simplex \( \mathcal{X} = \Delta^{n-1} \subset \mathcal{E} = \mathbb{R}^n \) by lifting it to the sphere via the Hadamard lift

\[
\mathcal{M} = S^{n-1}, \quad \varphi(y) = y \odot y, \quad \text{(Had)}
\]

where \( \odot \) denotes the entrywise (Hadamard) product. Using this lift leads to fast algorithms for high-dimensional problems (Q), see [54, 39, 13]. This is also essentially the lift that appears in the eigenvalue example (scenario (b)) in Section 1, see Example 4.13. This lift is particularly natural in applications involving probabilities since the push-forward under \( \varphi \) of the standard metric on the sphere is the Fisher-Rao metric on the simplex [5, Prop. 2.1].

We can characterize precisely where each of our desirable properties holds.

**Proposition 2.5.** The lift (Had) satisfies “local \( \Rightarrow \) local” everywhere, “1 \( \Rightarrow \) 1” at \( y \) if and only if \( y_i \neq 0 \) for all \( i \) (i.e., at preimages of the relative interior of the simplex), and “2 \( \Rightarrow \) 1” everywhere.
We prove this proposition in Corollary 4.12 by showing that the lift (had) is a special case of our construction of lifts in Section 4 and can be analyzed using the general results we prove there.

The relation between desirable points for (Q) and for (P) have been previously studied in [39], where the authors show that 2-critical points for (Q) map to 2nd-order KKT points for (P), viewed as a nonlinear program, for any twice-differentiable cost. This is a strengthening of “2 ⇒ 1”.

2.2 Smooth semidefinite programs via Burer–Monteiro

Consider the domain $\mathcal{X}$ of a rank-constrained semidefinite program (SDP),

$$\mathcal{X} = \{ X \in \mathbb{S}^n_r : \text{rank}(X) \leq r, \langle A_i, X \rangle = b_i \text{ for } i = 1, \ldots, m \} \subseteq \mathcal{E} = \mathbb{S}^n,$$

where $\langle U, V \rangle = \text{Tr}(U^\top V)$ is the (Frobenius) inner product on $\mathcal{E}$. The Burer–Monteiro approach [10] to optimizing over $\mathcal{X}$ consists of optimizing over the following parametrization instead:

$$\mathcal{M} = \{ R \in \mathbb{R}^{n \times r} : h_i(R) := \langle A_i, R \rangle - b_i = 0 \text{ for } i = 1, \ldots, m \}, \quad \varphi(R) = RR^\top. \quad \text{(BM)}$$

Burer and Monteiro prove in [11, Prop. 2.3] that local minima for (Q) map under $\varphi$ to local minima for (P) for linear costs $f$.

Under some conditions on the $A_i, b_i$ (which are satisfied generically [15, Prop. 1] as well as for several applications of interest [9]), the constraints $h_i(R) = 0$ constitute (constant-rank) local defining functions (in the sense of [35, Thm. 5.12]) for $\mathcal{M}$, which is then an embedded submanifold of $\mathbb{R}^{n \times r}$. In that case, $\mathcal{M}$ and $\varphi$ constitute a smooth lift of $\mathcal{X}$. In [9], assuming $f$ is linear (as is typical for SDPs), the authors use the assumption that the $h_i$ are local defining functions to prove that (in our terminology) rank-deficient 2-critical points for (Q) map under $\varphi$ to stationary points for (P). This was also shown for nonlinear $f$ in [28], though under more restrictive conditions on $A_i, b_i$ (e.g., $A_iA_j = 0$ for $i \neq j$). In all cases, these results allow to capture benign non-convexity when $f$ is convex, as then stationary points for (P) are global minima.

Using our framework, we can generalize these results to any (twice-differentiable) costs and remove the restrictions on the rank of the 2-critical points.

**Proposition 2.6.** The Burer–Monteiro lift (BM) satisfies “local ⇒ local” everywhere. If $\mathcal{M}$ in (BM) is a smooth manifold with (constant-rank) local defining functions $\{h_i\}_{i=1}^m$, then this lift satisfies “1 ⇒ 1” at $R$ if and only if rank$(R) = r$, and “2 ⇒ 1” everywhere.

We prove this result too in Corollary 4.12 using general properties of our lift construction in Section 4. Our theory also yields explicit expressions for the tangent cones to $\mathcal{X}$ in (4.0.4), which (to our knowledge) have not previously appeared in the literature.

2.3 Low-rank matrices

Consider the set $\mathcal{X} = \mathbb{R}^{m \times n}_r$ of matrices in $\mathcal{E} = \mathbb{R}^{m \times n}$ with rank at most $r$. We study several natural lifts of this real algebraic variety. The first one we study is based on the rank factorization of a matrix:

$$\mathcal{M} = \mathbb{R}^{m \times n} \times \mathbb{R}^{n \times r}, \quad \varphi(L, R) = LR^\top. \quad \text{(LR)}$$

The authors of [23] showed (in our terminology) that “1 ⇒ 1” does not hold everywhere, but “2 ⇒ 1” does. We further proved in [36] that this lift does not satisfy “local ⇒ local” everywhere either. We strengthen these results here using our unified framework, by characterizing precisely where each of these properties hold. The proof of the following proposition is given in Section 5.1.
Proposition 2.7. The lift of \( X = \mathbb{R}_{\leq r}^{m \times n} \) given by (LR) satisfies:

- “local \( \Rightarrow \) local” at \( (L, R) \) if and only if \( \text{rank}(L) = \text{rank}(R) = \text{rank}(LR^\top) \),
- “1 \( \Rightarrow \) 1” at \( (L, R) \) if and only if \( \text{rank}(L) = \text{rank}(R) = r \),
- and “2 \( \Rightarrow \) 1” everywhere on \( \mathcal{M} \) [23].

The second lift we study for \( \mathbb{R}_{\leq r}^{m \times n} \) is the desingularization lift introduced in [29]. It is given by

\[
\mathcal{M} = \{ (X, S) \in \mathbb{R}^{m \times n} \times \text{Gr}(n, n-r) : S \subseteq \ker X \}, \quad \varphi(X, S) = X. \tag{Desing}
\]

Here \( \text{Gr}(n, n-r) \) is the Grassmann manifold of \( (n-r) \)-dimensional subspaces of \( \mathbb{R}^n \) [8, §9]. We proved in [36] that this lift too does not satisfy “local \( \Rightarrow \) local”. The following proposition parallels the one above and is proved in Section 5.2.

Proposition 2.8. The lift of \( X = \mathbb{R}_{\leq r}^{m \times n} \) given by (Desing) satisfies:

- “local \( \Rightarrow \) local” at \( (X, S) \) if and only if \( \text{rank}(X) = r \); the same is true for “1 \( \Rightarrow \) 1”.
- “2 \( \Rightarrow \) 1” everywhere on \( \mathcal{M} \).

A potential advantage of the desingularization lift over the matrix factorization lift is that the preimage of a matrix, \( \varphi^{-1}(X) \), is compact for the former but not for the latter.

Note that both lifts (LR) and (Desing) satisfy “1 \( \Rightarrow \) 1” and “local \( \Rightarrow \) local” at preimages of rank-\( r \) matrices, but the lift (LR) further satisfies “local \( \Rightarrow \) local” at “balanced” preimages of lower-rank matrices. We also mention that no smooth lift of \( \mathbb{R}_{\leq r}^{m \times n} \) can satisfy “1 \( \Rightarrow \) 1” at preimages of lower-rank matrices by Theorem 2.4, since the tangent cones to such matrices are not linear spaces.

In [41], the authors experiment with various SVD-type lifts for optimization over matrices of rank exactly \( r \). The following proposition, proved in Appendix D, gives some of the properties of these lifts, extended to parametrize all of \( \mathbb{R}_{\leq r}^{m \times n} \).

Proposition 2.9. The SVD lift and its modification from [41] satisfy the following.

- The SVD lift of \( \mathbb{R}_{\leq r}^{m \times n} \) given by

\[
\mathcal{M} = \text{St}(m, r) \times \mathbb{R}^r \times \text{St}(n, r), \quad \varphi(U, \sigma, V) = U \text{diag}(\sigma)V^\top, \tag{SVD}
\]

satisfies “local \( \Rightarrow \) local” at \( (U, \sigma, V) \) if and only if \( |\sigma_1|, \ldots, |\sigma_r| \) are nonzero and distinct; the same holds for “1 \( \Rightarrow \) 1”.

- The modified SVD lift

\[
\mathcal{M} = \text{St}(m, r) \times \mathbb{S}^r \times \text{St}(n, r), \quad \varphi(U, M, V) = U MV^\top, \tag{MSVD}
\]

satisfies “local \( \Rightarrow \) local” at \( (U, M, V) \) if and only if the eigenvalues of \( M \) satisfy \( \lambda_i(M) + \lambda_j(M) \neq 0 \) for all \( i, j \); the same holds for “1 \( \Rightarrow \) 1”.

In [41, §6.3], the authors observed that Riemannian gradient descent running on (Q) gets stuck in a suboptimal point for a certain matrix completion problem using (SVD) but not using (MSVD). We can use Proposition 2.9 to understand their observation. Their algorithm only generates iterates with strictly positive diagonals \( \sigma \) in (SVD) and strictly positive-definite middle factors \( M \) in (MSVD), and can only converge to such points. Proposition 2.9 shows that (MSVD) satisfies “local \( \Rightarrow \) local” and “1 \( \Rightarrow \) 1” in that region, while (SVD) does not.
2.4 Low-rank tensors

Tensor factorization formats correspond to lifts mapping factors to low-rank tensors, for various notions of tensor rank. For example, the canonical polyadic (CP) decomposition of rank at most 1 corresponds to the lift of \( X = \{ X \in \mathbb{R}^{n_1 \times \cdots \times n_d} : \text{CP-rank}(X) \leq 1 \} \) [32] via tensor product \( \otimes \) as:

\[
\mathcal{M} = \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_d}, \quad \varphi(v_1, \ldots, v_d) = v_1 \otimes \cdots \otimes v_d.
\]

Other examples include CP decompositions of higher rank, Tucker and Tensor Train (TT) decompositions, and more generally tensor networks [32, 14]. Surprisingly, none of these lifts satisfy “2 \( \Rightarrow \) 1”: we derive this from more general obstructions to “2 \( \Rightarrow \) 1” for multilinear lifts in Section 5.3. Here is one take-away: any stationarity guarantees for algorithms targeting second-order critical points over the factors in a tensor decomposition must exploit the structure of the cost function.

2.5 Neural networks

Training neural networks is done via lifts. Indeed, here \( \mathcal{M} \) is the manifold of weights and biases of a fixed neural network architecture (typically a linear space; sometimes a product of spheres if normalization constraints are present). The lift \( \varphi \) maps a choice of weights and biases to the function given by the corresponding neural network. The image \( X = \varphi(\mathcal{M}) \) of this lift is the set of functions that can be represented by the architecture, viewed as a subset of some linear space \( \mathcal{E} \) of functions (e.g., an \( L^p \) space).

The authors of [43] show that such \( \varphi \) is not open for any choice of (nonconstant) Lipschitz continuous activation functions. Our Theorem 2.3 then implies that “local \( \Rightarrow \) local” fails for all neural network lifts used in practice. Consequently, training such a neural network by optimizing over its weights and biases might yield a spurious local minimum that does not parametrize a local minimum in function space. In that case, a different parametrization of the same function might not be a local minimum for \( (Q) \).

When the neural network architecture is linear with three or more layers, the corresponding lift is multilinear, hence does not satisfy “2 \( \Rightarrow \) 1” by the same general obstructions from Section 5.3 we use for tensor decompositions. Similarly to the tensor case, this implies that proofs of guarantees for training algorithms must exploit the structure in specific cost functions (the loss). Additional study of lifts defined by linear neural networks was done in [51, 30], where the authors characterize (in our terminology) “1 \( \Rightarrow \) 1” for lifts defined by linear and linear convolutional architectures.

2.6 Submersions and higher order stationary points

All the sets \( \mathcal{X} \) we consider in this paper contain dense smooth submanifolds. Moreover, even though lifts of such sets \( \mathcal{X} \) do not satisfy “1 \( \Rightarrow \) 1” everywhere on the lift, they do so at preimages of points on this submanifold, allowing us to prove much stronger guarantees. More precisely, we define the following subset of \( \mathcal{X} \).

Definition 2.10. A point \( x \in \mathcal{X} \) is smooth if there is an open neighborhood \( U \subseteq \mathcal{E} \) containing \( x \) such that \( U \cap \mathcal{X} \) is a smooth embedded submanifold of \( \mathcal{E} \). It is called nonsmooth or singular otherwise. The smooth locus of \( \mathcal{X} \), denoted \( \mathcal{X}^{\text{smth}} \), is the set of smooth points of \( \mathcal{X} \).

For all constraint sets in practical optimization problems we are aware of, \( \mathcal{X}^{\text{smth}} \) is itself a smooth embedded submanifold of \( \mathcal{E} \). In general, it is a union of smooth embedded submanifolds, though possibly of different dimensions. For example, if \( \mathcal{X} = \mathbb{R}^{m \times n}_{\leq r} \) then \( \mathcal{X}^{\text{smth}} = \mathbb{R}^{m \times n}_{\leq r} \) and if
$X = \Delta^{n-1}$ then $X^{\text{smth}} = \Delta_0^{n-1}$ consisting of strictly positive simplex vectors. All the lifts we consider for these sets in Sections 2.1 and 2.3 indeed satisfy "1 $\Rightarrow$ 1" on the preimages of $X^{\text{smth}}$ (though that is not always the case).

If the lift satisfies "1 $\Rightarrow$ 1" at preimages of smooth points, then it is a submersion there and hence preserves not only local minima, but also stationary points of all orders. The following proposition, proved in Section 3.2.1, formalizes this.

**Proposition 2.11.** Let $y \in \varphi^{-1}(X^{\text{smth}}) \subseteq M$. If $\varphi$ satisfies "1 $\Rightarrow$ 1" at $y$, then it also satisfies "local $\Rightarrow$ local" and "k $\Rightarrow$ k" for all $k \geq 1$ at $y$.

Here "k $\Rightarrow$ k" is defined analogously to Definition 2.2, where kth-order stationarity (or "k-criticality" for short) of $x \in X^{\text{smth}}$ is defined using curves similarly to 1- and 2-criticality [12, §3.1.1]. This property can be used in proofs of benign nonconvexity.

**Remark 2.12** (Relations between lift properties). Aside from Proposition 2.11, the only relation between the three properties in Definition 2.2 is that "1 $\Rightarrow$ 1" at $y$ implies "2 $\Rightarrow$ 1" at $y$ (since 2-critical points are 1-critical). None of the other possible implications hold in general: The desingularization lift (Desing) shows that "2 $\Rightarrow$ 1" at $y$ implies neither "1 $\Rightarrow$ 1" nor "local $\Rightarrow$ local" at $y$ in general. The example $\varphi(x) = x^3$ viewed as a lift from $M = \mathbb{R}$ to $X = \mathbb{R}$ satisfies "local $\Rightarrow$ local" at the origin but neither "2 $\Rightarrow$ 1" nor "1 $\Rightarrow$ 1", hence "local $\Rightarrow$ local" does not imply the other two properties. Finally, the standard parametrization of the cochleoid curve [56] satisfies "1 $\Rightarrow$ 1" but not "local $\Rightarrow$ local" at all preimages of the origin, hence "1 $\Rightarrow$ 1" does not imply "local $\Rightarrow$ local".

Submersions between smooth manifolds, including quotients by group actions [8, §9.2] and several lifts arising in practice (Example 3.14), satisfy "local $\Rightarrow$ local" and "k $\Rightarrow$ k" for all $k \geq 1$ [8, Prop. 9.6]. Therefore, a lift $\varphi$ composed with a submersion $\psi$ as $\varphi \circ \psi$ inherits the properties of $\varphi$. We study such compositions of lifts in Section 3.3, and apply our results to a composition used in the robotics and computer vision literature [19] in Example 3.29.

### 3 Characterizations of lifts

In this section, we relate the landscapes of (P) and (Q) and prove the characterizations of our lift properties stated in Section 2. To this end, we formally define the different types of desirable points we consider. We first define the (contingent or Bouligand) tangent cone$^3$ [16, §2.7].

**Definition 3.1.** The tangent cone to $X$ at $x \in X$ is the set

$$T_xX = \left\{ v = \lim_{i \to \infty} \frac{x_i - x}{\tau_i} : x_i \in X, \tau_i > 0 \text{ for all } i, \tau_i \to 0 \right\} \subseteq E.$$

This is a closed (not necessarily convex) cone [47, Lem. 3.12].

In particular, if $\gamma$ is a differentiable curve in $X$ with $\gamma(0) = x$, then $\gamma'(0) \in T_xX$. If $x$ is a smooth point of $X$ (Definition 2.10), then $T_xX$ is the usual tangent space to $X$ at $x$ [46, Ex. 6.8].

**Definition 3.2** (Desirable points for (P)). A point $x \in X$ is a

- (a) global minimum for (P) if $f(x) = \min_{x' \in X} f(x')$.

---

$^3$In this paper, a cone is a set $K$ such that $x \in K \implies \alpha x \in K$ for all $\alpha > 0$. 
(b) local minimum for \((P)\) if there is a neighborhood \(U \subseteq \mathcal{X}\) of \(x\) such that \(f(x) = \min_{x' \in U} f(x')\).

(c) (first-order) stationary point for \((P)\) if \(Df(x)[v] \geq 0\) for all \(v \in T_x \mathcal{X}\), or equivalently, if \(\nabla f(x)\) is in the dual \((T_x \mathcal{X})^*\) of the tangent cone.

In words, \(x\) is stationary if the cost function is non-decreasing to first order along all tangent directions at \(x\). Local minima of \((P)\) are stationary [47, Thm. 3.24]. The dual of a cone \(K \subseteq \mathcal{E}\) contained in a Euclidean space \(\mathcal{E}\) is defined by

\[
K^* = \{x \in \mathcal{E} : \langle x, x' \rangle \geq 0 \text{ for all } x' \in K\}.
\]

The equivalence in part (c) then follows since \(Df(x)[v] = \langle \nabla f(x), v \rangle\) by definition of the (Euclidean) gradient \(\nabla f(x)\). We use the following properties of dual cones throughout (see [20, Prop. 4.5] for proofs):

- The dual cone is always a closed convex cone.
- If \(K_1 \subseteq K_2\), then \(K_2^* \subseteq K_1^*\).
- The bidual cone \(K^{**} = (K^*)^*\) of \(K\) is equal to the closure of its convex hull: \(K^{**} = \overline{\text{conv}(K)}\).
- In particular, \(K^{**} \supseteq K\).
- If \(K\) is a linear space, then its dual \(K^*\) is equal to its orthogonal complement \(K^\perp\).

Next, we define desirable points for \((Q)\).

**Definition 3.3 (Desirable points for \((Q)\)).**

(a)-(b) Global and local minima for \((Q)\) are defined exactly as for \((P)\).

(c) A point \(y \in \mathcal{M}\) is first-order stationary (or “1-critical”) for \((Q)\) if for each smooth curve \(c : \mathbb{R} \to \mathcal{M}\) satisfying \(c(0) = y\), we have \((g \circ c)'(0) \geq 0\), or equivalently, \(^4\) \((g \circ c)'(0) = 0\).

(d) A point \(y \in \mathcal{M}\) is second-order stationary (or “2-critical”) for \((Q)\) if it is 1-critical and \((g \circ c)''(0) \geq 0\) for all smooth curves \(c : \mathbb{R} \to \mathcal{M}\) satisfying \(c(0) = y\).

If \(\mathcal{M}\) is embedded in a linear space, first-order stationarity in Definition 3.3(c) coincides with Definition 3.2(c) by [46, Ex. 6.8]. Definition 3.3 can be rephrased in terms of the Riemannian gradient and Hessian of \(g\), as follows.

**Proposition 3.4 ([8, §4.2, §6.1]).** A point \(y \in \mathcal{M}\) is 1-critical for \((Q)\) if and only if \(\nabla g(y) = 0\). It is 2-critical if and only if \(\nabla^2 g(y) \succeq 0\) and \(\nabla^2 g(y) \succeq 0\).

We proceed to study the connections between desirable points for \((Q)\) and \((P)\). As mentioned in Section 2, the connection between global minima of \((Q)\) and \((P)\) is straightforward.

**Proposition 3.5.** A point \(y \in \mathcal{M}\) is a global minimum of \((Q)\) if and only if \(x = \varphi(y)\) is a global minimum of \((P)\).

**Proof.** Because \(\varphi(\mathcal{M}) = \mathcal{X}\), we have \(\inf_{y \in \mathcal{M}} g(y) = \inf_{y \in \mathcal{M}} f(\varphi(y)) = \inf_{x \in \mathcal{X}} f(x) =: p^*.\) Therefore, \(y\) is a global minimum for \((Q)\) iff \(g(y) = f(x) = p^*\) which happens iff \(x\) is a global minimum for \((P)\). \(\square\)

Since computing global minima is hard, the remainder of this section is devoted to characterizing the properties in Definition 2.2 that yield connections between the other types of points.

\(^4\)If \((g \circ c)'(0) > 0\), let \(\tilde{c}(t) = c(-t)\) and note that \((g \circ \tilde{c})'(0) < 0\).
3.1 Local minima

In this section, we investigate the relationship between the local minima of \((P)\) and those of \((Q)\). Preimages of local minima on \(X\) are always local minima on \(M\) merely because \(\varphi\) is continuous.

**Proposition 3.6.** Let \(x\) be a local minimum for \((P)\). Any \(y \in \varphi^{-1}(x)\) is a local minimum for \((Q)\).

**Proof.** There exists a neighborhood \(U\) of \(x\) in \(X\) such that \(f(x) \leq f(x')\) for all \(x' \in U\). Since \(\varphi: M \to X\) is continuous, the set \(U = \varphi^{-1}(U)\) is a neighborhood of \(y\) in \(M\). Pick an arbitrary \(y' \in U\): it satisfies \(\varphi(y') = x'\) for some \(x' \in U\). Hence, \(g(y) = f(x) \leq f(x') = g(y')\), i.e., \(y\) is a local minimum of \((Q)\).

Unfortunately, lifting can introduce spurious local minima, that is, local minima for \((Q)\) that exist only because of the lift and not because they were present in \((P)\) to begin with.

**Example 3.7** (Nodal cubic). Consider the nodal cubic
\[
X = \{x \in \mathbb{R}^2 : x_2 = x_1^2(x_1 + 1)\},
\]
and the following lift,\(^5\) as depicted in Figure 1:
\[
M = \{y \in \mathbb{R}^3 : y_1 = y_3^2 - 1, \ y_2 = y_1y_3\}, \quad \varphi(y_1, y_2, y_3) = (y_1, y_2).
\]

Let \(f(x) = -x_1 - x_2\). Then the point \(y = (0, 0, 1)\) is a local minimum for \(g = f \circ \varphi\) but \(\varphi(y) = (0, 0)\) is not even stationary for \(f\). Indeed, we have \((1, 1) \in T_{(0,0)}X\) and \(Df(0,0)[(1,1)] = -2 < 0\).

![Figure 1: Nodal cubic in \(\mathbb{R}^2\) as the shadow of its lift in \(\mathbb{R}^3\), colored by the value of the function \(f(x) = -x_1 - x_2\). The highlighted points are \(x = (0,0)\) (not stationary for \((P)\)), and its two preimages on the lift, including \(y = (0,0,1)\) (a spurious local minimum for \((Q)\)).](image)

To ensure that a lift does not introduce spurious local minima, we need to verify that it satisfies the “local \(\Rightarrow\) local” property (Definition 2.2(a)). We proceed to prove the easy direction of Theorem 2.3 stating that openness implies “local \(\Rightarrow\) local”. The converse is more involved and is deferred to Appendix A.

\(^5\)The curve \(M\) is obtained by blowing up \(X\) at the origin in the sense of algebraic geometry [24, Ch. 17].
Proof of Theorem 2.3. Assume $\varphi$ is open at $y$, and that $y$ is a local minimum for (Q). Then, there exists a neighborhood $\mathcal{U}$ of $y$ on $\mathcal{X}$ such that $g(y) \leq g(y')$ for all $y' \in \mathcal{U}$. The set $U = \varphi(\mathcal{U})$ is a neighborhood of $x = \varphi(y)$ in $\mathcal{X}$ by openness of $\varphi$ at $y$. Moreover, each $x' \in U$ is of the form $x' = \varphi(y')$ for some $y' \in \mathcal{U}$. Therefore, $f(x) = g(y) \leq g(y') = f(x')$ for all $x' \in U$, that is, $x$ is a local minimum of (P). For the converse, see Theorem A.2.

Not all lifts of interest are open. In particular, all lifts of low-rank matrices in Section 2.3 as well as the neural network lifts in Section 2.5 fail to be open.

Remark 3.8. In Appendix A, we introduce an equivalent condition for openness of $\varphi$ at $y$ that we call the Subsequence Lifting Property (SLP), see Definition A.1(3); we find that it is sometimes easier to check. For example, Burer and Monteiro prove that the lift (BM) satisfies “local $\Rightarrow$ local” in [11, Prop. 2.3] by (in our terminology) proving SLP holds.

We note in passing that all continuous, surjective, open maps are quotient maps, hence if $\varphi$ is a smooth lift of $\mathcal{X}$ satisfying “local $\Rightarrow$ local” then it is a quotient map from $\mathcal{M}$ to $\mathcal{X}$.

3.2 Stationary points

In this section, we investigate the relationship between the first- and second-order stationary points for (Q) and (first-order) stationary points for (P). To that end, we begin by relating the (Riemannian) gradient and Hessian of $g = f \circ \varphi$ to the (Euclidean) counterparts of $f$. This relation depends on the first and second derivatives of the lift $\varphi$.

Definition 3.9. Let $\varphi: \mathcal{M} \to \mathcal{X}$ be a smooth lift and fix $y \in \mathcal{M}$. For each $v \in T_y \mathcal{M}$, choose a curve $c_v$ on $\mathcal{M}$ satisfying $c(0) = y$ and $c'(0) = v$. Define maps $L_y, Q_y: T_y \mathcal{M} \to \mathcal{E}$ by

$$L_y(v) = (\varphi \circ c_v)'(0), \quad Q_y(v) = (\varphi \circ c_v)''(0).$$

We write $L_y^c$ and $Q_y^c$ when we wish to emphasize the lift.

As a point of notation: $\varphi \circ c_v$ is a curve in $\mathcal{E}$ hence $(\varphi \circ c_v)''$ denotes its Euclidean acceleration. In contrast, $c_v$ is a curve on $\mathcal{M}$ hence $c'_v$ denotes its Riemannian acceleration, see [8, §5.8, §8.12].

Of course, $L_y$ is simply the differential $D\varphi(y)$, and is therefore linear and independent of the choice of curves $c_v$. The map $Q_y$ will play an important role in characterizing “2 $\Rightarrow$ 1” in Section 3.2.2, where we also clarify its inconsequential dependence on the choice of curve $c_v$. We explain how to compute $L_y$ and $Q_y$ without explicitly choosing curves $c_v$ in Section 3.4.

The gradients and Hessians of $f$ and $g = f \circ \varphi$ are neatly related as follows in terms of $L_y$.

Definition 3.10. For any $w \in \mathcal{E}$, define $\varphi_w: \mathcal{M} \to \mathbb{R}$ by $\varphi_w(y) = \langle w, \varphi(y) \rangle$.

Lemma 3.11. For any twice differentiable cost $f: \mathcal{E} \to \mathbb{R}$, any $y \in \mathcal{M}$, and $x = \varphi(y)$, we have

$$\nabla g(y) = L^*_y(\nabla f(x)), \quad \nabla^2 g(y) = L^*_y \circ \nabla^2 f(x) \circ L_y + \nabla^2 \varphi \nabla f(x)(y),$$

where $L^*_y: \mathcal{E} \to T_y \mathcal{M}$ is the adjoint of $L_y$.

Proof. For any $v \in T_y \mathcal{M}$, let $c_v$ be a smooth curve on $\mathcal{M}$ satisfying $c_v(0) = y$, $c'_v(0) = v$ and $c''_v(0) = 0$ (e.g., let $c_v$ be a geodesic). Let $\gamma_v = \varphi \circ c_v$: it satisfies $\gamma_v(0) = x$ and $\gamma'_v(0) = L_y(v)$. Then

$$\langle \nabla g(y), v \rangle = (g \circ c_v)'(0) = (f \circ \gamma_v)'(0) = \langle \nabla f(x), L_y(v) \rangle = \langle L^*_y(\nabla f(x)), v \rangle.$$
Since this holds for all \( v \in T_yM \), we conclude that \( \nabla g(y) = L^*_y(\nabla f(x)) \). Next,

\[
\langle \nabla^2 g(y)[v], v \rangle = (g \circ c_v)'(0) = (f \circ c_v)'(0) = \langle \nabla^2 f(x)[L_y(v)], L_y(v) \rangle + \langle \nabla f(x), \gamma_v''(0) \rangle,
\]

where the first equality uses \( c_v''(0) = 0 \), see [8, §5.9]. On the other hand, with Definition 3.10,

\[
\langle \nabla^2 \varphi \nabla f(x)(y)[v], v \rangle = (\varphi \nabla f(x) \circ c_v)'(0) = \frac{d^2}{dt^2} \langle \nabla f(x), \gamma_v(t) \rangle \bigg|_{t=0} = \langle \nabla f(x), \gamma_v''(0) \rangle,
\]

hence

\[
\langle \nabla^2 g(y)[v], v \rangle = \langle \nabla^2 f(x)[L_y(v)], L_y(v) \rangle + \langle \nabla^2 \varphi \nabla f(x)(y)[v], v \rangle
\]

\[
= \langle (L^*_y \circ \nabla^2 f(x) \circ L_y + \nabla^2 \varphi \nabla f(x)(y))[v], v \rangle.
\]

Since this holds for all \( v \in T_yM \) and both \( \nabla^2 g(y) \) and \( L^*_y \circ \nabla^2 f(x) \circ L_y + \nabla^2 \varphi \nabla f(x)(y) \) are self-adjoint linear maps on \( T_yM \), we conclude that they are equal. \( \square \)

We turn to proving our characterizations of “\( k \Rightarrow 1 \)” for \( k = 1, 2 \) announced in Section 2.

### 3.2.1 “\( 1 \Rightarrow 1 \)” : Lifts preserving 1-critical points

Preimages of stationary points on \( X \) are always 1-critical on \( M \). We show this after a helpful lemma.

**Lemma 3.12.** Fix \( y \in M \) and let \( x = \varphi(y) \). Then \( \text{im} L_y \subseteq T_xX \). Moreover, \( y \) is 1-critical for \((Q)\) if and only if \( \nabla f(x) \in (\text{im} L_y)^\perp = (\text{im} L_y)^* \).

**Proof.** The first claim follows from Definition 3.1 for the tangent cone \( T_xX \) and the fact that \( L_y(v) = (\varphi \circ c_v)'(0) \) for a curve \( c_v \) as in Definition 3.9. For the second claim, \( y \) is 1-critical for \((Q)\) if and only if \( \nabla f(x) = L^*_y(\nabla f(x)) = 0 \), or equivalently, \( \nabla f(x) \in \ker(L^*_y) = (\text{im} L_y)^\perp \). \( \square \)

**Proposition 3.13.** If \( x \in X \) is stationary for \((P)\), then any \( y \in \varphi^{-1}(x) \) is 1-critical for \((Q)\).

**Proof.** If \( x \in X \) is stationary for \((P)\), then \( \nabla f(x) \in (T_xX)^* \). Since \( T_xX \supseteq \text{im} L_y \), taking duals on both sides we get that \( \nabla f(x) \in (T_xX)^* \subseteq (\text{im} L_y)^\perp \), hence \( y \) is 1-critical for \((Q)\) by Lemma 3.12. \( \square \)

The converse to Proposition 3.13 is false in general. In fact, Example 3.7 shows that a lift need not even map local minima to stationary points on \( X \). We therefore proceed to prove Theorem 2.4 characterizing the “\( 1 \Rightarrow 1 \)” property.

**Proof of Theorem 2.4.** Suppose \( \text{im} L_y = T_xX \), so \( (\text{im} L_y)^\perp = (\text{im} L_y)^* = (T_xX)^* \). If \( y \) is 1-critical for \((Q)\), then \( \nabla f(x) \in (\text{im} L_y)^\perp \) by Lemma 3.12. Therefore, \( \nabla f(x) \in (T_xX)^* \), which is the definition of \( x \) being stationary for \((P)\). Thus, “\( 1 \Rightarrow 1 \)” holds.

Now suppose \( \text{im} L_y \neq T_xX \). This implies \( (T_xX)^* \neq (\text{im} L_y)^\perp \). Indeed, otherwise we would have

\[
\text{im} L_y = (\text{im} L_y)^{\perp \perp} = (\text{im} L_y)^* = (T_xX)^* \supseteq T_xX \supseteq \text{im} L_y,
\]

which would imply \( \text{im} L_y = T_xX \). (The right-most inclusion above is by Lemma 3.12.) Using \( \text{im} L_y \subseteq T_xX \) again, we see that \( (T_xX)^* \subseteq (\text{im} L_y)^\perp \). Therefore, the above observations imply that \( (T_xX)^* \subseteq (\text{im} L_y)^\perp \). Pick \( w \in (\text{im} L_y)^\perp \setminus (T_xX)^* \) and define \( f(x') = \langle w, x' \rangle \) for \( x' \in \mathcal{E} \). Then \( \nabla f(x) = w \in (\text{im} L_y)^\perp \) so \( y \) is 1-critical for \((Q)\) by Lemma 3.12, but \( \nabla f(x) \notin (T_xX)^* \), so \( x \) is not stationary for \((P)\). Hence “\( 1 \Rightarrow 1 \)” is not satisfied at \( y \).

This argument also shows that if \( \varphi \) does not satisfy “\( 1 \Rightarrow 1 \)” at \( y \), then \( y \) is 1-critical for \((Q)\) but \( x \) is not stationary for \((P)\) if and only if \( \nabla f(x) \in (\text{im} L_y)^\perp \setminus (T_xX)^* \), showing that if “\( 1 \Rightarrow 1 \)” fails at \( y \) then this is witnessed by a linear cost \( f \). \( \square \)
As discussed in Section 2, Theorem 2.4 implies that “1 ⇒ 1” rarely holds on all of \( \mathcal{M} \). Nevertheless, “1 ⇒ 1” does usually hold at preimages of smooth points, that is, points around which \( \mathcal{X} \) is a smooth embedded submanifold of \( \mathcal{E} \) as in Definition 2.10. We now prove Proposition 2.11, stating that if “1 ⇒ 1” holds at such points then “local ⇒ local” and “\( k \Rightarrow k \)” hold there as well.

**Proof of Proposition 2.11.** Let \( U \subseteq \mathcal{E} \) be an open neighborhood of \( \varphi(y) \) in \( \mathcal{E} \) such that \( U \cap \mathcal{X} \) is a smooth embedded submanifold of \( \mathcal{E} \). Since \( \varphi(\mathcal{M}) = \mathcal{X} \), we have \( \varphi^{-1}(U \cap \mathcal{X}) = \varphi^{-1}(U) =: V \), which is open in \( \mathcal{M} \) by continuity of \( \varphi \). Therefore, \( V \) is also a smooth manifold, since it is an open subset of \( \mathcal{M} \), and \( \varphi|_V: V \to U \cap \mathcal{X} \) is a smooth map between smooth manifolds. By Theorem 2.4, \( \varphi \) satisfies “1 ⇒ 1” at \( y \) iff \( T_y \mathcal{X} = \text{im} \, D\varphi(y) = \text{im} \, D(\varphi|_V)(y) \), where \( \varphi|_V \) is viewed as a map \( V \to \mathcal{E} \). Since \( U \cap \mathcal{X} \) is an embedded submanifold of \( \mathcal{E} \), the differential of \( \varphi|_V \) viewed as a map \( V \to \mathcal{E} \) coincides with its differential viewed as a map \( V \to U \cap \mathcal{X} \), hence the latter is a submersion near \( y \) [35, Prop. 4.1]. By [35, Prop. 4.28], this implies \( \varphi \) is open at \( y \), hence it satisfies “local ⇒ local” at \( y \) by Theorem 2.3. To see that \( \varphi \) further satisfies “\( k \Rightarrow k \)” for all \( k \geq 1 \), note that any curve passing through \( \varphi(y) \) is the image under \( \varphi \) of a curve passing through \( y \) [35, Thm. 4.26], and apply Definition 3.3 for \( k = 1, 2 \) and [12, Eq. (3.11)] for \( k \geq 2 \).

The converse of Proposition 2.11 fails. For example, \( \varphi(y) = y^5 \) viewed as a map \( \mathbb{R} \to \mathbb{R} \) satisfies “local ⇒ local” at \( y = 0 \) but not “1 ⇒ 1” since \( T_y y^5 = 0 \) for this \( y \). That example also shows that “1 ⇒ 1” can fail at the preimage of a smooth point. Likewise, “1 ⇒ 1” can hold at the preimage of a nonsmooth point, as the standard parametrization of the cochleoid curve [56] shows. The only examples of lifts we know of that satisfy “1 ⇒ 1” everywhere are smooth maps between smooth manifolds that are submersions.

**Example 3.14 (Submersions).** Examples of submersions in optimization, that is, lifts of the form \( \varphi: \mathcal{M} \to \mathcal{X} \) where \( \mathcal{X} \) is an embedded submanifold of \( \mathcal{E} \) and \( \text{im} \, D\varphi(y) = T_{\varphi(y)} \mathcal{X} \) for all \( y \in \mathcal{M} \), include:

- Quotient maps by smooth, free, and proper Lie group actions [8, §9], [2, §3.4], used in particular to optimize over Grassmannians by lifting to Stiefel manifolds [21].

- The map \( \text{SO}(p) \to \text{St}(p,d) \) taking the first \( d \) columns of a rotation matrix, which is used in the rotation averaging algorithm of the robotics paper [19], see Example 3.29 below.

- Deep orthogonal linear networks, mapping \( \text{O}(p)^n \to \text{O}(p) \) by \( \varphi(Q_1, \ldots, Q_n) = Q_1 \cdots Q_n \), whose properties are studied in [1].

**Theorem 2.4 and Proposition 2.11** show that these lifts satisfy “1 ⇒ 1” and “local ⇒ local”, and hence also “\( k \Rightarrow k \)” for all \( k \geq 1 \).

Failure of a lift to satisfy “1 ⇒ 1” means that it may introduce spurious critical points. In the next section, we characterize the “2 ⇒ 1” property, which allows algorithms to avoid these spurious points by using second-order information.

### 3.2.2 “2 ⇒ 1”: Lifts mapping 2-critical points to 1-critical points

Since “1 ⇒ 1” fails on many sets of interest, we proceed to study “2 ⇒ 1”. As Section 2 demonstrates, this property is satisfied for many interesting lifts. We begin by stating an equivalent characterization for “2 ⇒ 1” involving the following set.
Definition 3.15. For \( y \in \mathcal{M} \) and \( x = \varphi(y) \in \mathcal{X} \), define
\[
W_y = \left\{ w \in \mathcal{E} : \text{there exists a twice differentiable function } f : \mathcal{E} \to \mathbb{R} \right. \\
\text{such that } \nabla f(x) = w \text{ and } y \text{ is 2-critical for (Q)} \right\}.
\]
We write \( W_y^\varphi \) when we wish to emphasize the lift.

Theorem 3.16. The lift \( \varphi : \mathcal{M} \to \mathcal{X} \) satisfies “\( 2 \Rightarrow 1 \)” at \( y \) if and only if \( W_y \subseteq (T_x\mathcal{X})^* \) where \( x = \varphi(y) \).

Proof. Say \( \varphi \) satisfies “\( 2 \Rightarrow 1 \)” at \( y \) and let \( w \in W_y \). Pick \( f \) such that \( y \) is 2-critical for (Q) and \( \nabla f(x) = w \). By “\( 2 \Rightarrow 1 \)”, we know \( x \) is stationary for \( f \), hence \( w = \nabla f(x) \in (T_x\mathcal{X})^* \). Conversely, say \( W_y \subseteq (T_x\mathcal{X})^* \) and let \( y \) be 2-critical for (Q) with some cost \( f \). Then \( \nabla f(x) \in W_y \subseteq (T_x\mathcal{X})^* \), hence \( x \) is stationary for (P). This shows “\( 2 \Rightarrow 1 \)”.

Since the 2-criticality of \( y \) for (Q) only depends on the first two derivatives of \( f \), we can restrict the functions \( f \) in Definition 3.15 to be of class \( C^\infty \) or even quadratic polynomials whose Hessians are a multiple of the identity, as the following proposition shows.

Proposition 3.17. For \( y \in \mathcal{M} \) and \( x = \varphi(y) \), the set \( W_y \) in Definition 3.15 satisfies:
\[
W_y = \left\{ w \in \mathcal{E} : \exists \alpha > 0 \text{ s.t. } y \text{ is 2-critical for (Q) with } f(x') = \langle x', w \rangle + \frac{\alpha}{2} \|x' - x\|^2 \right\}.
\]
In particular, \( W_y \) is a convex cone.

Proof. The inclusion \( \supseteq \) is clear from the definition of \( W_y \). Conversely, if \( w \) is in \( W_y \) then \( y \) is 2-critical for (Q) for some \( f \) with \( \nabla f(x) = w \). Let \( g = f \circ \varphi \) and \( \alpha = \lambda_{\max}(\nabla^2 f(x)) \), and define
\[
f_\alpha(x') = \langle w, x' \rangle + \frac{\alpha}{2} \|x - x'\|^2, \quad g_\alpha = f_\alpha \circ \varphi.
\]
Note that \( \nabla f_\alpha(x) = w \) and, by Lemma 3.11, we have \( \nabla g_\alpha(y) = L_y^*(w) = \nabla g(y) = 0 \) and
\[
\nabla^2 g_\alpha(y) = L_y^* \circ \nabla^2 f_\alpha(x) \circ L_y + \nabla^2 f_\alpha(x) \circ L_y = L_y^* \circ (\alpha I) \circ L_y + \nabla^2 \varphi_w(y) \\
\geq L_y^* \circ \nabla^2 f(x) \circ L_y + \nabla^2 \varphi_w(y) = \nabla^2 g(y) \geq 0.
\]
Thus, \( y \) is 2-critical for \( g_\alpha \). This shows the reverse inclusion.

\( W_y \) is a convex cone since if \( w_1, w_2 \) are in \( W_y \) as witnessed by functions \( f_1, f_2 \), then any \( w = \lambda_1 w_1 + \lambda_2 w_2 \) with \( \lambda_1, \lambda_2 \geq 0 \) is in \( W_y \) as witnessed by \( f = \lambda_1 f_1 + \lambda_2 f_2 \).

Proposition 3.17 shows that if “\( 2 \Rightarrow 1 \)” is not satisfied at \( y \), then there exists a simple strongly convex quadratic cost \( f \) for which \( y \) is 2-critical for (Q) but \( x = \varphi(y) \) is not stationary for (P).

Corollary 3.18. Suppose \( \varphi \) does not satisfy “\( 2 \Rightarrow 1 \)” at \( y \in \mathcal{M} \) and denote \( x = \varphi(y) \). Then \( W_y \setminus (T_x\mathcal{X})^* \neq \emptyset \) and for any \( w \) in that set there exists \( \alpha > 0 \) such that if \( f(x') = \langle w, x' \rangle + \frac{\alpha}{2} \|x' - x\|^2 \), then \( y \) is 2-critical for (Q) but \( x \) is not stationary for (P).

We conjecture that the reverse inclusion in Theorem 3.16 always holds (it does for all the lifts in Section 2). If this is indeed true, then \( \varphi \) satisfies the “\( 2 \Rightarrow 1 \)” property at \( y \) if and only if \( (T_x\mathcal{X})^* = W_y \), neatly echoing the condition for “\( 1 \Rightarrow 1 \)”, namely, \( (T_x\mathcal{X})^* = (\text{im} L_y)^\perp \).
Conjecture 3.19. It always holds that $(T_x\mathcal{X})^* \subseteq W_y$.

The description of $W_y$ can be complicated. It is therefore worthwhile to derive sufficient conditions for “$2 \Rightarrow 1$” that are easier to check. We do so by identifying two (admittedly technical) sets whose duals contain $\nabla f(x)$ if $x = \varphi(y)$ and $y$ is 2-critical for $(Q)$. The sufficient conditions then require the duals of these two subsets to be contained in $(T_x\mathcal{X})^*$.

Definition 3.20. For $y \in \mathcal{M}$, define

$$A_y = \{w \in \mathcal{E} : \exists c: \mathbb{R} \to \mathcal{M} \text{ smooth s.t. } c(0) = y, \ (\varphi \circ c)'(0) = 0, \ (\varphi \circ c)''(0) = w\},$$

$$B_y = \{w \in \mathcal{E} : \exists c_i: \mathbb{R} \to \mathcal{M} \text{ smooth s.t. } c_i(0) = y, \ \lim_{i \to \infty} (\varphi \circ c_i)'(0) = 0, \ \lim_{i \to \infty} (\varphi \circ c_i)''(0) = w\}.$$

We write $A_y^*, B_y^*$ when we wish to emphasize the lift.

The following are the basic properties these two sets satisfy. We give further expressions for $A_y, B_y$ and $W_y$ in Proposition 3.26 below.

Proposition 3.21. Fix $y \in \mathcal{M}$ and denote $x = \varphi(y)$.

(a) $A_y$ and $B_y$ are cones, and $B_y$ is closed.

(b) $A_y \subseteq T_x\mathcal{X}$ and $\text{im } L_y \subseteq A_y \subseteq B_y$. Moreover, $\text{im } L_y + A_y = A_y$ and $\text{im } L_y + B_y = B_y$.

(c) If $y$ is 2-critical for $g = f \circ \varphi$, then $\nabla f(x) \in B_y^*$.

(d) $W_y \subseteq B_y^* \subseteq A_y^* \subseteq (\text{im } L_y)^{-1}$.

Proof. The proofs of part (a) and the second half of part (b) are given in Appendix B.

(b) If $c: \mathbb{R} \to \mathcal{M}$ satisfies $c(0) = y$ and $(\varphi \circ c)'(0) = 0$, then $(\varphi \circ c)(t) \in \mathcal{X}$ for all $t$ and $(\varphi \circ c)(t) = x + (t^2/2)(\varphi \circ c)''(0) + \mathcal{O}(t^3)$, hence by Definition 3.1 we have

$$(\varphi \circ c)''(0) = \lim_{t \to 0} \frac{(\varphi \circ c)(t) - x}{t^2/2} \in T_x\mathcal{X}.$$

This shows $A_y \subseteq T_x\mathcal{X}$.

If $w \in \text{im } L_y$ so $w = L_y(v)$ for some $v \in T_y\mathcal{M}$, let $c: \mathbb{R} \to \mathcal{M}$ be a curve satisfying $c(0) = y$ and $c'(0) = v$. Define $\tilde{c}: \mathbb{R} \to \mathcal{M}$ by $\tilde{c}(t) = c(t^2/2)$, and note that $\tilde{c}(0) = y$, $(\varphi \circ \tilde{c})'(0) = 0$, and $(\varphi \circ \tilde{c})''(0) = (\varphi \circ c)'(0) = w$. Hence $w$ is in $A_y$. This shows $\text{im } L_y \subseteq A_y$. It is clear that $A_y \subseteq B_y$ from Definition 3.20.

(c) Suppose $y$ is 2-critical for $g = f \circ \varphi$ and $w \in B_y$. Let $c_i: \mathbb{R} \to \mathcal{M}$ witness $w \in B_y$. Because $y$ is 1-critical, $(g \circ c_i)'(0) = 0$ for all $i$. Because $y$ is 2-critical, for all $i$ we have

$$(g \circ c_i)''(0) = \langle \nabla f(x), (\varphi \circ c_i)''(0) \rangle + \langle \nabla^2 f(x)[(\varphi \circ c_i)'(0)], (\varphi \circ c_i)'(0) \rangle \geq 0.$$

Taking $i \to \infty$, we conclude that $\langle \nabla f(x), w \rangle \geq 0$ and hence $\nabla f(x) \in B_y^*$ as claimed.

(d) If $w \in W_y$ then there exists $f$ such that $\nabla f(x) = w$ and $y$ is 2-critical for $(Q)$, hence $w \in B_y^*$ by part (c). The other inclusions follow by taking duals in part (b).

We remark that neither the inclusion $B_y \subseteq T_x\mathcal{X}$ nor $T_x\mathcal{X} \subseteq B_y$ hold in general, see the end of Appendix B. Combining part (d) above with Theorem 3.16 yields the following sufficient conditions for “$2 \Rightarrow 1$”.

17
Corollary 3.22. Fix $y \in \mathcal{M}$ and denote $x = \varphi(y)$.

(a) If $A_y^* \subseteq (T_x\mathcal{X})^*$, and in particular if $A_y = T_x\mathcal{X}$, then “2 $\Rightarrow$ 1” holds at $y$.

(b) If $B_y^* \subseteq (T_x\mathcal{X})^*$, then “2 $\Rightarrow$ 1” holds at $y$.

The conditions in Corollary 3.22 yield simpler proofs of “2 $\Rightarrow$ 1” for many lifts. For example, the condition in Corollary 3.22(a) holds for the Burer–Monteiro lift of Section 2.2. While it does not hold for the lifts of low-rank matrices in Section 2.3, they do satisfy the stronger condition in Corollary 3.22(b). In fact, the condition in Corollary 3.22(b) holds in all the examples satisfying “2 $\Rightarrow$ 1” that we consider. It would be interesting to determine whether it is necessary as well.

We now state and prove the chain of implications we find the most useful for verifying or refuting “2 $\Rightarrow$ 1”, as well as for computing tangent cones (see Section 3.5).

Theorem 3.23. Let $\varphi\colon \mathcal{M} \to \mathcal{X}$ be a smooth lift and fix $y \in \mathcal{M}$. We have the following chain of implications for “2 $\Rightarrow$ 1”:

$$
T_x\mathcal{X} \subseteq A_y
\iff T_x\mathcal{X} = A_y
\implies B_y^* \subseteq (T_x\mathcal{X})^*
\implies W_y \subseteq (T_x\mathcal{X})^*
\iff \varphi \text{ satisfies } “2 \Rightarrow 1” \text{ at } y
\implies (\im L_y)^\perp \cap (Q_y(T_y\mathcal{M}))^* \subseteq (T_x\mathcal{X})^*.
$$

Proof. The equivalence of the first two conditions follows by Proposition 3.21(b). The second condition implies the third by Proposition 3.21(b) as well. The third condition implies the fourth by Proposition 3.21(d), which itself is equivalent to “2 $\Rightarrow$ 1” at $y$ by Theorem 3.16.

The last implication gives a necessary condition for “2 $\Rightarrow$ 1” to hold. Suppose there exists $w \in (\im L_y)^\perp \cap (Q_y(T_y\mathcal{M}))^* \setminus (T_x\mathcal{X})^*$. Define $f(x') = \langle w, x' \rangle$, whose gradient and Hessian at $x$ are $\nabla f(x) = w$ and $\nabla^2 f(x) = 0$. For any curve $c\colon I \to \mathcal{M}$ satisfying $c(0) = y$, denote $v = c'(0) \in T_y\mathcal{M}$. Let $g = f \circ \varphi$. Note that $(g \circ c)'(0) = \langle w, L_y(v) \rangle = 0$ since $w \in (\im L_y)^\perp$ and

$$
(g \circ c)''(0) = (f \circ \varphi \circ c)''(0) = \langle w, (\varphi \circ c)''(0) \rangle = \langle w, Q_y(v) \rangle \geq 0,
$$

where the second equality follows from the chain rule, the third equality follows from Lemma 3.24(a) below, and the inequality follows from $w \in (Q_y(T_y\mathcal{M}))^*$. Thus, $y$ is 2-critical for (Q). However, $\nabla f(x) = w \notin (T_x\mathcal{X})^*$ so $x$ is not stationary for (P), hence “2 $\Rightarrow$ 1” does not hold at $y$.

Our goal now is to derive more explicit expressions for the sets $A_y, B_y, W_y$ in terms of the maps $L_y$ and $Q_y$ from Definition 3.9. Such expressions allow us to compute these sets in specific examples. To do so, we first recall that the value of $Q_y(v)$ depends on the choice of curve $c_v$ in Definition 3.9. Before proceeding, we characterize the ambiguity in $Q_y(v)$ arising from different such choices, verifying that it causes no issues.

Lemma 3.24. For each $y \in \mathcal{M}$ and $v \in T_y\mathcal{M}$, let $c_v\colon I \to \mathcal{M}$ be a curve satisfying $c_v(0) = y$ and $c_v'(0) = v$, so we can set $Q_y(v) = (\varphi \circ c_v)''(0)$ according to Definition 3.9.

(a) For any other curve $c$ satisfying $c(0) = y$ and $c'(0) = v$, we have $(\varphi \circ c)''(0) - (\varphi \circ c_v)''(0) = L_y(c''(0) - c_v''(0)) \in \im L_y$. 

18
(b) For any \( u \in T_y \mathcal{M} \), there exists a curve \( c \) as in part (a) satisfying \( c''(0) - c''_v(0) = u \), hence 
\[(\varphi \circ c)^\prime\prime(0) - (\varphi \circ c_v)^\prime\prime(0) = L_y(u).\]

In particular, \( \{ (\varphi \circ c)^\prime\prime(0) : c(0) = y \text{ and } c'(0) = v \} = Q_y(v) + \text{im } L_y \).

Proof.  
(a) For any \( w \in \mathcal{E} \), recall the function \( \varphi_w(y) = \langle w, \varphi(y) \rangle \) from Definition 3.10. Let \( c : I \to \mathcal{M} \) be a curve satisfying \( c(0) = y \) and \( c'(0) = v \). Then, on the one hand,
\[
\frac{d^2}{dt^2} \varphi_w(c(t)) \bigg|_{t=0} = \frac{d^2}{dt^2} (w, (\varphi \circ c)(t)) \bigg|_{t=0} = \langle w, (\varphi \circ c)^\prime\prime(0) \rangle.
\]
On the other hand, using the Riemannian structure on \( \mathcal{M} \),
\[
\frac{d^2}{dt^2} \varphi_w(c(t)) \bigg|_{t=0} = \langle \nabla^2 \varphi_w(y)[c'(0)], c'(0) \rangle + \langle \nabla \varphi_w(y), c''(0) \rangle.
\]
By Lemma 3.11, we have \( \nabla \varphi_w(y) = L^*_y(w) \), so \( \langle \nabla \varphi_w(y), c''(0) \rangle = \langle w, L_y(c''(0)) \rangle \). We conclude that
\[
\langle w, (\varphi \circ c)^\prime\prime(0) \rangle = \langle \nabla^2 \varphi_w(y)[v], v \rangle + \langle w, L_y(c''(0)) \rangle.
\]  
(3.2.1)

The first term on the right-hand side is independent of \( c \). Thus, for any \( w \in \mathcal{E} \) we have
\[
\langle w, (\varphi \circ c)^\prime\prime(0) - (\varphi \circ c_v)^\prime\prime(0) \rangle = \langle w, L_y(c''(0) - c''_v(0)) \rangle,
\]
which proves the claim.

(b) For the first claim, set \( c(t) = \exp_y(tv + t^2(c''_v(0) - u)/2) \) where \( \exp \) is the exponential map of \( \mathcal{M} \) [8, Exer. 5.46]. The second claim follows from part (a). \( \square \)

Lemma 3.24 shows that the possible values of \( Q_y(v) \) (depending on the choice of curve \( c_v \) in Definition 3.9) differ by an element of \( \text{im } L_y \), and conversely, every element of \( Q_y(v) + \text{im } L_y \) can be obtained by an appropriate choice of \( c_v \). Consequently, if \( w \in (\text{im } L_y)^\perp \), then the inner product \( \langle w, Q_y(v) \rangle \) is independent of the choice of \( c_v \) in Definition 3.9. In fact, (3.2.1) shows that it is a quadratic form in \( v \in T_y \mathcal{M} \) given by:
\[
\langle w, Q_y(v) \rangle = \langle \nabla^2 \varphi_w(y)[v], v \rangle \quad \forall v \in T_y \mathcal{M}.
\]  
(3.2.2)

We stress that this identity requires \( w \in (\text{im } L_y)^\perp \) in general. It allows us to view \( \langle w, Q_y(v) \rangle \) interchangeably as either a quadratic form in \( v \) on \( T_y \mathcal{M} \) or a linear form in \( w \) on \( (\text{im } L_y)^\perp \).

Remark 3.25 (Disambiguation of \( Q_y \)). Given the above ambiguity in \( Q_y \), it would be natural to define the codomain of \( Q_y \) to be the quotient vector space \( \mathcal{E}/\text{im } L_y \). This would make it independent of the choice of \( c_v \). Subsets of the quotient \( \mathcal{E}/\text{im } L_y \) are in bijection with subsets of \( \mathcal{E} \) that are closed under addition with \( \text{im } L_y \) (i.e. subsets \( S \subseteq \mathcal{E} \) such that \( S + \text{im } L_y = S \)), which includes \( A_y \) and \( B_y \). Subsets of the dual vector space to \( \mathcal{E}/\text{im } L_y \) are in bijection with subsets of \( (\text{im } L_y)^\perp \), which includes \( W_y \). Hence we could equivalently phrase our conditions for “\( 2 \Rightarrow 1 \)” in terms of subsets of \( \mathcal{E}/\text{im } L_y \) and its dual. However, we have several techniques to obtain expressions for \( Q_y \) without explicitly choosing curves, see Section 3.4 below. Thus, Definition 3.9 mirrors the computations we do more closely. In practice, we view \( Q_y \) as taking values in \( \mathcal{E} \), and consider two maps differing by elements of \( \text{im } L_y \) as equivalent for the purpose of verifying “\( 2 \Rightarrow 1 \)” since they yield the same sets \( A_y, B_y, W_y \).
We now express the sets $A_y, B_y, W_y$ appearing in our conditions for $“2 \Rightarrow 1”$ in terms of the maps $L_y$ and $Q_y$. We explain how to compute $L_y$ and $Q_y$ in various settings in Section 3.4 below.

**Proposition 3.26.** For any $y \in \mathcal{M}$,

(a) $A_y = Q_y(\ker L_y) + \im L_y$.

(b) $B_y = \bigcup_{(v_i)_{i \geq 1} : L_y(v_i) \to 0} \lim_{i \to \infty} (Q_y(v_i) + \im L_y)$.\(^6\)

(c) $W_y = \{ w \in A^*_y : \forall v \in \ker L_y, (\nabla^2 \varphi_w(y)[v], v) = 0 \implies (\nabla^2 \varphi_w(y)[v] = 0 \}\}.

We always have

$$B_y \supseteq \{ \lim_i Q_y(v_i) : L_y(v_i) \to 0 \} + \im L_y,$$

(3.2.3)

but inclusion may be strict depending on the curves $c_v$ in Definition 3.9. In practice, the expressions for $Q_y$ we obtain using our techniques from Section 3.4 below are smooth, and the subset of $B_y$ in (3.2.3) is large enough to prove $“2 \Rightarrow 1”$ in every example where it holds.

**Proof.** (a) If $w \in A_y$, then $w = (\varphi \circ c)”(0)$ for some smooth curve $c$ on $\mathcal{M}$ such that $c(0) = y$ and $0 = (\varphi \circ c)'(0) = L_y(c'(0))$, so $c'(0) \in \ker L_y$. By Lemma 3.24(a), we have $w \in Q_y(c'(0)) + \im L_y$. Since $\varphi(c'(0)) = L_y(v) = 0$, this shows $Q_y(\ker L_y) + \im L_y \subseteq A_y$.

(b) If $w \in B_y$, then there are smooth curves $c_i$ such that $c_i(0) = y$, $\int L_y(c_i'(0)) \to 0$ and $(\varphi \circ c_i)”(0) = w$ exists, we conclude that $\lim_i Q_y(c_i'(0)) + \im L_y = w + \im L_y$ exists as well, and $w$ is contained in this limit. This shows the inclusion $\subseteq$ in the claim.

Conversely, suppose $w \in \lim_i (Q_y(v_i) + \im L_y)$ for some sequence $(v_i)_{i \geq 1} \subseteq T_y\mathcal{M}$ such that $L_y(v_i) \to 0$. Then $\varphi(c_i(0)) = \nabla^2 \varphi(w(y)[v_i])$. By Lemma 3.24(b), there exist curves $c_i$ satisfying $c_i(0) = y$, $c_i'(0) = v_i$ and $(\varphi \circ c_i)”(0) = Q_y(v_i) + L_y(u_i)$. Then $(\varphi \circ c_i)”(0) = L_y(v_i) \to 0$ and $(\varphi \circ c_i)”(0) \to w$, so $w \in B_y$ and hence the reverse inclusion in the claim also holds.

(c) Let $x = \varphi(y)$. By Proposition 3.17, a vector $w \in \mathcal{E}$ is contained in $W_y$ if and only if there exists $\alpha > 0$ such that $y$ is 2-critical for $(\mathcal{Q})$ with cost $g_\alpha = f_\alpha \circ \varphi$ where $f_\alpha(x') = \langle w, x' \rangle + \frac{\alpha}{2} \| x' - x \|^2$. By Lemma 3.11, this is equivalent to

$$\nabla g_\alpha(y) = L^*_y(w) = 0, \quad \nabla^2 g_\alpha(y) = \alpha L^*_y \circ L_y + \nabla^2 \varphi_w(y) \succeq 0.$$

In other words, $w \in W_y$ if $w \in (\im L_y)^\perp$ and $\nabla^2 \varphi_w(y) + \alpha L^*_y \circ L_y \succeq 0$ for some $\alpha > 0$. To understand when the second condition holds, we decompose $T_y\mathcal{M} = \ker L_y \oplus (\ker L_y)^\perp$ and express the relevant self-adjoint operators on $T_y\mathcal{M}$ in block matrix form with respect to a basis compatible with this decomposition. More explicitly, choose a basis as described above.

\(^6\)A sequence $(v_i + W)_{i \geq 1}$ of translates of a subspace $W$ of a (topological) vector space $V$ converges (necessarily to another translate of $W$) if there exist $w_i \in W$ such that $(v_i + w_i)_{i \geq 1} \subseteq V$ converges in $V$. 

20
and denote \( n = \dim \ker \mathbf{L}_y \) and \( m = \dim(ker \mathbf{L}_y)^\perp \). Assume first that \( m > 0 \). We represent \( \nabla^2 \varphi_w(y) \) and \( \alpha \mathbf{L}_y^* \circ \mathbf{L}_y \) in that basis as

\[
\nabla^2 \varphi_w(y) = \begin{bmatrix} \Phi_1 & \Phi_2 \\ \Phi_2^\top & \Phi_3 \end{bmatrix}, \quad \text{with} \quad \Phi_1 \in \mathbb{S}^n, \Phi_3 \in \mathbb{S}^m.
\]

Thus,

\[
\alpha \mathbf{L}_y^* \circ \mathbf{L}_y = \begin{bmatrix} 0 & 0 \\ 0 & \alpha \Psi \end{bmatrix}, \quad \text{with} \quad \Psi \in \mathbb{S}^m_{>0}.
\]

By the generalized Schur complement theorem \cite[Thm. 1.20]{57}, the block-matrix on the right-hand side is positive semidefinite exactly if

\[
\Phi_1 \succeq 0, \quad \text{im} \Phi_2 \subseteq \text{im} \Phi_1 \quad \text{and} \quad \Phi_3 + \alpha \Psi \succeq \Phi_2 \Phi_1^\top \Phi_2,
\]

where \( \Phi_1^\dagger \) is the Moore–Penrose pseudo-inverse of \( \Phi_1 \). The last condition holds upon choosing \( \alpha \geq \lambda_{\max}(\Phi_2 \Phi_1^\top \Phi_2 - \Phi_3)/\lambda_{\min}(\Psi) \). Thus, we deduce the following expression for \( W_y \):

\[
W_y = \{ w \in (im \mathbf{L}_y)^\perp : \Phi_1 \succeq 0 \ \text{and} \ \text{im} \Phi_2 \subseteq \text{im} \Phi_1 \},
\]

with \( \Phi_1 \) and \( \Phi_2 \) as defined above. We now work out basis-free characterizations of the properties \( \Phi_1 \succeq 0 \) and \( \text{im} \Phi_2 \subseteq \text{im} \Phi_1 \).

First, notice that \( \Phi_1 \succeq 0 \) iff

\[
\begin{bmatrix} v_1^\top & 0^\top_m \end{bmatrix} \begin{bmatrix} \Phi_1 & \Phi_2 \\ \Phi_2^\top & \Phi_3 \end{bmatrix} \begin{bmatrix} v_1 \\ 0^\top_m \end{bmatrix} \succeq 0 \quad \text{for all} \ v_1 \in \mathbb{R}^n,
\]

or in basis-free terms, \( \langle \nabla^2 \varphi_w(y)[v], v \rangle \succeq 0 \) for all \( v \in \ker \mathbf{L}_y \). If \( w \in (im \mathbf{L}_y)^\perp \) then (3.2.2) shows that this is also equivalent to \( \langle w, Q_y(v) \rangle \succeq 0 \) for all \( v \in \ker \mathbf{L}_y \), which is in turn equivalent to \( \langle w, Q_y(v) + L_y(u) \rangle \succeq 0 \) for all \( v \in \ker \mathbf{L}_y \) and \( u \in T_y M \). This last condition is just \( w \in A_y^* \) by part (a).

Second, we must understand for which vectors \( w \) it holds that \( \text{im} \Phi_2 \subseteq \text{im} \Phi_1 \), or equivalently, \( \ker \Phi_1 \subseteq \ker \Phi_2^\perp \) (recall that \( \Phi_1^\top = \Phi_1 \)). If \( \Phi_1 \succeq 0 \), then \( v_1 \in \ker \Phi_1 \) iff \( v_1^\top \Phi_1 v_1 = 0 \). Moreover, if \( v_1 \in \ker \Phi_1 \) then

\[
\begin{bmatrix} \Phi_1 & \Phi_2 \\ \Phi_2^\top & \Phi_3 \end{bmatrix} \begin{bmatrix} v_1 \\ 0^\top_m \end{bmatrix} = \begin{bmatrix} 0^\top_m \\ \Phi_2 v_1 \end{bmatrix},
\]

which vanishes iff \( v_1 \in \ker \Phi_2^\top \). Thus, assuming \( \Phi_1 \succeq 0 \), the inclusion \( \text{im} \Phi_2 \subseteq \text{im} \Phi_1 \) is equivalent to the implication

\[
\begin{bmatrix} v_1^\top & 0^\top_m \end{bmatrix} \begin{bmatrix} \Phi_1 & \Phi_2 \\ \Phi_2^\top & \Phi_3 \end{bmatrix} \begin{bmatrix} v_1 \\ 0^\top_m \end{bmatrix} = 0 \quad \Longrightarrow \quad \begin{bmatrix} \Phi_1 & \Phi_2 \\ \Phi_2^\top & \Phi_3 \end{bmatrix} \begin{bmatrix} v_1 \\ 0^\top_m \end{bmatrix} = 0, \quad \text{for all} \ v_1 \in \mathbb{R}^n.
\]
In basis-free terms, we have shown that, if $\Phi_1 \succeq 0$, then $\im(\Phi_2) \subseteq \im(\Phi_1)$ is equivalent to the implication $\langle \nabla^2 \varphi_w(y)[v], v \rangle = 0 \implies \nabla^2 \varphi_w(y)[v] = 0$ holding for all $v \in \ker L_y$. Putting everything together,

$$W_y = \{ w \in (\im L_y)\perp : \Phi_1 \succeq 0 \text{ and } \im(\Phi_2) \subseteq \im(\Phi_1) \}
= \{ w \in (\im L_y)\perp : w \in A_y^* \text{ and } \forall v \in \ker L_y, \langle \nabla^2 \varphi_w(y)[v], v \rangle = 0 \implies \nabla^2 \varphi_w(y)[v] = 0 \}
= \{ w \in A_y^* : \forall v \in \ker L_y, \langle \nabla^2 \varphi_w(y)[v], v \rangle = 0 \implies \nabla^2 \varphi_w(y)[v] = 0 \},$$

where the last equality holds because $A_y^* \subseteq (\im L_y)\perp$ by Proposition 3.21(b). This is the claimed expression for $W_y$.

If $m = 0$, or equivalently, if $L_y = 0$, then $w \in W_y$ iff $w \in (\im L_y)\perp = E$ and $\nabla^2 \varphi_w(y) \succeq 0$. This in turn is equivalent to $w \in A_y^* = (Q_y(\ker L_y))^\perp \cap (\im L_y)\perp$ by (3.2.2), so $W_y = A_y^*$ in this case. Conversely, if $w \in A_y^*$ and $L_y = 0$ then $\nabla^2 \varphi_w(y) \succeq 0$ so the condition in the claimed expression for $W_y$ is satisfied automatically: it also evaluates to $A_y^*$. This verifies that the claimed expression for $W_y$ holds for $m = 0$ as well.

### 3.3 Composition of lifts

In this section, we ask: when are lift properties preserved under composition? We use the following proposition both to compute $L_y$ and $Q_y$ in various settings, and to study some of the lifts appearing in the literature in Sections 4 and 5.

**Proposition 3.27.** Let $\varphi : \mathcal{M} \to \mathcal{X}$ be a smooth lift, and let $\psi : \mathcal{N} \to \mathcal{M}$ be a smooth map between smooth manifolds such that $\varphi \circ \psi : \mathcal{N} \to \mathcal{X}$ is surjective. Both $\varphi$ and $\varphi \circ \psi$ are smooth lifts for $\mathcal{X}$. For $z \in \mathcal{N}$ and $y = \psi(z) \in \mathcal{M}$, the following hold.

(a) If $\varphi \circ \psi$ satisfies “local $\implies$ local” at $z$, then $\varphi$ satisfies “local $\implies$ local” at $y$. If $\psi$ is open (in particular, if $\psi$ is a submersion) at $z$, and if $\varphi$ satisfies “local $\implies$ local” at $y$, then $\varphi \circ \psi$ satisfies “local $\implies$ local” at $z$.

(b) If $\varphi \circ \psi$ satisfies “1 $\implies$ 1” or “2 $\implies$ 1” at $z$, then $\varphi$ satisfies the corresponding property at $y$. If $\psi$ is a submersion at $z$ and $\varphi$ satisfies “1 $\implies$ 1” or “2 $\implies$ 1” at $y$, then $\varphi \circ \psi$ satisfies the corresponding property at $z$.

(c) If $\psi$ is a submersion at $z$, then

$$L_z^{\varphi \circ \psi} = L_y^{\varphi} \circ L_z^{\psi} \quad \text{ and } \quad Q_z^{\varphi \circ \psi} \equiv Q_z^{\varphi} \circ L_z^{\psi} \mod \im L_z^{\varphi \circ \psi}. $$

Moreover, $\im L_z^{\varphi \circ \psi} = \im L_y^{\varphi}$, $A_z^{\varphi \circ \psi} = A_y^{\varphi}$, $B_z^{\varphi \circ \psi} = B_y^{\varphi}$, and $W_z^{\varphi \circ \psi} = W_y^{\varphi}$.

The proof is straightforward but long, and is deferred to Appendix C.1. Here we denote $v = w \mod \im L_y$ to mean $v - w \in \im L_y$. By Lemma 3.24, equality of $Q_z^{\varphi \circ \psi}$ and $Q_y^{\varphi}$ modulo $\im L_y = \im L_z^{\varphi \circ \psi}$ means that either one can be used to verify “2 $\implies$ 1”.

Proposition 3.27 shows that, given a smooth lift $\varphi : \mathcal{M} \to \mathcal{X}$, there is no benefit to further lifting $\mathcal{M}$ to another smooth manifold through $\psi : \mathcal{N} \to \mathcal{M}$ in terms of our properties. Indeed, if $\varphi$ does not satisfy one of our properties, then neither does $\varphi \circ \psi$ for any smooth $\psi$ (we cannot ‘fix’ a bad lift by lifting it further). On the other hand, this proposition also tells us that our properties, as well as the sets involved in their characterization, are preserved under submersions. This notably means lift properties can be checked through charts of $\mathcal{M}$. Moreover, for lifts to a
manifold $\mathcal{M}$ which is a quotient of another manifold $\overline{\mathcal{M}}$ (these arise naturally when quotienting by group actions, see [8, §9]), Proposition 3.27 allows us to verify our properties on the total space $\overline{\mathcal{M}}$, which is often easier.

**Remark 3.28.** If $\psi: N \to \mathcal{M}$ is a submersion, for each $z \in N$, let $V_z = \ker D\psi(z)$ and $H_z = (\ker D\psi(z))^\perp$ be the so-called vertical and horizontal spaces at $z$, which satisfy $T_zN = V_z \oplus H_z$ and $H_z \cong T_{\psi(z)}\overline{\mathcal{M}}$. Proposition 3.27 implies that $L_z^{\psi\circ\psi} = L_z^{\psi\circ\psi} \circ \text{Proj}_{H_z}$ and $Q_z^{\psi\circ\psi} = Q_z^{\psi\circ\psi} \circ \text{Proj}_{H_z}$ where $\text{Proj}_{H_z}$ denotes orthogonal projection onto $H_z$, so it suffices to consider the restrictions of $L_z^{\psi\circ\psi}$ and $Q_z^{\psi\circ\psi}$ to the horizontal space at $z$. The latter is often simpler than $T_zN$, see [8, §9.4].

We end this section with an implicit application of Proposition 3.27 seen in the robotics and computer vision literature.

**Example 3.29** (Shohan rotation averaging). In [19], Dellaert et al. estimate a set of $n$ rotations of $\mathbb{R}^d$ from noisy measurements of pairs of relative rotations. Their algorithm involves the Burer-Monteiro lift (BM) with $\mathcal{M} = \text{St}(p,d)^n$ for appropriate $p \geq d$, composed with the submersion $\psi: \text{SO}(p)^n \to \text{St}(p,d)^n$ extracting the first $d$ columns of each matrix. Using our framework, we can analyze this composition and thereby strengthen the guarantees proved in [19]. Indeed, since the Burer-Monteiro lift satisfies “2 ⇒ 1” and “local ⇒ local” by Proposition 2.6 and $\psi$ is a submersion, Proposition 3.27 shows that the composed lift satisfies “local ⇒ local” and “2 ⇒ 1” as well. Furthermore, if every stationary point for the original low-rank SDP of [19] is globally optimal (e.g., if the conditions of [9] hold), then any 2-critical point for their lifted problem is globally optimal and therefore the lifted problem enjoys benign nonconvexity.

### 3.4 Computing $L_y$ and $Q_y$

Theorem 3.23 gives several conditions for “2 ⇒ 1” that (together with Proposition 3.26) can be checked using $L_y$ and $Q_y$ from Definition 3.9. We therefore consider various strategies for computing $L_y$ and $Q_y$ depending on how we can access $\mathcal{M}$. Since $Q_y$ is only defined modulo $\text{im}L_y$, different methods may yield different expressions, any of which can be used to verify “2 ⇒ 1”.

**$\mathcal{M}$ through charts:** Suppose we are given a chart $\psi: U \to \mathcal{M}$ on $\mathcal{M}$, which is a diffeomorphism from an open subset $U \subseteq \mathcal{E}'$ of some linear space $\mathcal{E}'$ onto its image, and let $y \in \psi(U)$. Then we can compose $\varphi$ with $\psi$ to obtain a lift to a linear space $\overline{\varphi} = \varphi \circ \psi$. By Proposition 3.27, the lift $\varphi$ satisfies “1 ⇒ 1” or “2 ⇒ 1” at $y \in \mathcal{M}$ if and only if $\overline{\varphi}$ satisfies the corresponding property at $z = \psi^{-1}(y)$. Thus, it suffices to compute $L_z^{\overline{\varphi}}$ and $Q_z^{\overline{\varphi}}$ and use them to check “2 ⇒ 1” at $z$. Since $U$ is an open subset of a linear space $\mathcal{E}'$, it is natural to compute $L_z^{\overline{\varphi}}$ and $Q_z^{\overline{\varphi}}$ directly from Definition 3.9 using curves $\overline{c}_\nu(t) = z + t\overline{v}$ which are straight lines through $z$ in direction $\overline{v} \in \mathcal{E}' = T_zU$. This choice yields the expressions

$$L_z^{\overline{\varphi}}(\overline{v}) = (\overline{\varphi} \circ \overline{c}_\nu)'(0) = D\overline{\varphi}(z)[\overline{v}], \quad \text{and} \quad Q_z^{\overline{\varphi}}(\overline{v}) = (\overline{\varphi} \circ \overline{c}_\nu)''(0) = D^2\overline{\varphi}(z)[\overline{v}, \overline{v}],$$

where $D\overline{\varphi}(z)$ and $D^2\overline{\varphi}(z)$ are the ordinary first- and second-order derivative maps of $\overline{\varphi}$ viewed as a map between linear spaces $\mathcal{E}' \to \mathcal{E}$. In particular, if $\mathcal{M}$ is itself a linear space (e.g., for the (LR) lift), we may take $U = \mathcal{E}' = \mathcal{M}$ and $\psi = \text{id}$ and use (3.4.1) with $\overline{\varphi} = \varphi$.

**$\mathcal{M}$ embedded in a linear space:** Suppose now that $\mathcal{M}$ is an embedded submanifold of another linear space $\mathcal{E}'$. By [8, Prop. 3.31], the lift $\varphi$ can be extended to a smooth map on a neighborhood $V$ of $\mathcal{M}$ in $\mathcal{E}'$, denoted by $\overline{\varphi}: V \to \mathcal{E}$. This means $\overline{\varphi}$ is a smooth map defined on an open subset $V \subseteq \mathcal{E}'$ containing $\mathcal{M}$ and it satisfies $\overline{\varphi}|_\mathcal{M} = \varphi$. If $c_\nu$ is a curve on $\mathcal{M}$ passing through $y \in \mathcal{M} \subseteq V$
with velocity \( v \in T_y\mathcal{M} \subseteq T_yV = \mathcal{E}' \), then \( \varphi \circ c_v = \overline{\varphi} \circ c_v \) because the curve is contained in \( \mathcal{M} \) where \( \overline{\varphi} \) agrees with \( \varphi \). Denote by \( u_v = \dot{c}_v(0) \) the ordinary (extrinsic) acceleration of \( c_v \) at \( t = 0 \), viewed as a curve in \( \mathcal{E}' \). Then Definition 3.9 and the chain rule give

\[
L_y(v) = (\overline{\varphi} \circ c_v)'(0) = D\overline{\varphi}(y)[v], \quad Q_y(v) = (\overline{\varphi} \circ c_v)''(0) = D^2\overline{\varphi}(y)[v, v] + D\overline{\varphi}(y)[w].
\]

To better understand \( u_v \), let \( h : \mathcal{E}' \to \mathbb{R}^k \) be a local defining function for \( \mathcal{M} \) around \( y \), that is, \( h \) is smooth, \( \mathcal{M} \) is locally its zero-set, and \( \text{rank} Dh(y') = k \) for all \( y' \) around \( y \). For a curve \( c_v \) as above, we have \( h(c_v(t)) \equiv 0 \) around \( t = 0 \), so in particular \( (h \circ c_v)'(0) = (h \circ c_v)''(0) = 0 \). By the chain rule, the latter equations can be written as

\[
Dh(y)[v] = 0 \quad \text{and} \quad D^2h(y)[v, v] + Dh(y)[w] = 0. \quad (3.4.2)
\]

Conversely, for any \( v, u_v \in \mathcal{E}' \) satisfying (3.4.2), there exists a curve on \( \mathcal{M} \) passing through \( y \) with velocity \( v \) and extrinsic acceleration \( u_v \) by [46, Prop. 13.13]. Thus, the expressions (3.4.2) describe all possible velocities and extrinsic accelerations of curve as they pass through \( y \). This set of all possible such accelerations of curves on \( \mathcal{M} \) passing through \( y \in \mathcal{M} \) with velocity \( v \in T_y\mathcal{M} \) is the second-order tangent set to \( \mathcal{M} \) at \( y \) for \( v \), and is denoted by \( T^2_y\mathcal{M} \) [46, Def. 13.11]. The above discussion shows that

\[
T_y\mathcal{M} = \ker Dh(y) \quad \text{and} \quad T^2_{y,v}\mathcal{M} = \{ u \in \mathcal{E}' : Dh(y)[u] = -D^2h(y)[v, v] \}. \quad (3.4.3)
\]

As a result, for any extension \( \overline{\varphi} \) of \( \varphi \) and all \( v \in T_y\mathcal{M} \), we have

\[
L_y(v) = D\overline{\varphi}(y)[v] \quad \text{and} \quad Q_y(v) = D^2\overline{\varphi}(y)[v, v] + D\overline{\varphi}(y)[w] \quad \text{for some } u_v \in T^2_{y,v}\mathcal{M}. \quad (3.4.4)
\]

Note that \( T^2_{y,v}\mathcal{M} \) is an affine subspace of \( \mathcal{E}' \) which is a translate of the subspace \( T_y\mathcal{M} \), as can be seen from (3.4.3). Therefore, while different choices of \( u_v \) lead to different expressions for \( Q_y \), they are all equal modulo \( \text{im} L_y \).

\( \mathcal{M} \) as a quotient manifold: Suppose next that \( \mathcal{M} \) is a quotient manifold of \( \overline{\mathcal{M}} \) with quotient map \( \pi : \overline{\mathcal{M}} \to \mathcal{M} \) [8, §9]. Then \( \overline{\varphi} = \varphi \circ \pi \) gives a smooth lift of \( \mathcal{X} \) to \( \overline{\mathcal{M}} \). Since \( \pi \) is a submersion, Proposition 3.27 and Remark 3.28 imply that to check “2 ⇒ 1”, we need only compute \( L_z \) and \( Q_z \) for \( \overline{\varphi} \) restricted to the horizontal spaces (\( \ker D\pi(z) \))\(^\perp \) using the preceding two methods.

Computing \( \nabla^2\varphi_w \): To check the equivalent condition in Theorem 3.23 for any presentation of \( \mathcal{M} \), we need to compute \( W_y \). If we use Proposition 3.26 to do so, we need an expression for the Riemannian Hessian \( \nabla^2\varphi_w(y) \) where \( \varphi_w(y) = \langle w, \varphi(y) \rangle \) for \( w \in (\text{im} L_y)\(^\perp \). Given \( Q_y \), we can obtain \( \nabla^2\varphi_w(y) \) as the unique self-adjoint operator on \( T_y\mathcal{M} \) that defines the quadratic form (3.2.2). Conversely, if we compute \( \nabla^2\varphi_w(y) \) for all \( w \in (\text{im} L_y)\(^\perp \), e.g., using the techniques from [8, §5.5], we can set \( Q_y(v) \) to be the unique element of \((\text{im} L_y)\(^\perp \) satisfying (3.2.2), providing another way to compute \( Q_y \).

We now illustrate the above techniques for computing \( L_y \) and \( Q_y \). The following example uses charts.

Example 3.30 (Desingularization of \( \mathbb{R}^{m \times n}_{\leq r} \)). Consider the desingularization lift (Desing) of bounded rank matrices \( \mathcal{X} = \mathbb{R}^{m \times n}_{\leq r} \). We compute \( L \) and \( Q \) using charts. For \( (X_0, S_0) \in \mathcal{M} \), let \( Y_0 \in \mathbb{R}^{n \times (n-r)} \) be a matrix satisfying \( \text{col}(Y_0) = S_0 \), so \( X_0 Y_0 = 0 \). Since \( \text{rank}(Y_0) = n - r \), we can find \( n - r \) linearly independent rows in \( Y_0 \). Let \( J \in \mathbb{R}^{(n-r) \times (n-r)} \) be the invertible submatrix of \( Y_0 \) obtained by
extracting these rows, and let $\Pi \in \mathbb{R}^{n \times n}$ be a permutation matrix sending those $n - r$ rows to the first rows. Then
\[
\Pi Y_0 J^{-1} = \begin{bmatrix} I_{n-r} \\ W_0 \end{bmatrix} \quad \text{and} \quad X_0 = \begin{bmatrix} -Z_0 W_0, & Z_0 \end{bmatrix} \Pi \quad \text{for some } Z_0 \in \mathbb{R}^{m \times r} \text{ and } W_0 \in \mathbb{R}^{r \times (n-r)},
\]
where the second identity is implied by $0 = X_0 Y_0 J^{-1} = X_0 \Pi^T (\Pi Y_0 J^{-1})$. Accordingly, a chart $\psi: \mathbb{R}^{m \times r} \times \mathbb{R}^{r \times (n-r)} \to M$ containing $(X_0, S_0)$ is given by
\[
\psi(Z, W) = \left( [-ZW, \ Z] \Pi, \ col \left( \Pi^T \begin{bmatrix} I_{n-r} \\ W \end{bmatrix} \right) \right).
\]

Composing with $\varphi$, we obtain the lift $\tilde{\varphi}(Z, W) = \varphi(\psi(Z, W)) = [-ZW, \ Z] \Pi$, and by (3.4.1),
\[
\begin{aligned}
L_{\tilde{\varphi}}(Z, W) &= D\tilde{\varphi}(Z, W)(\dot{\tilde{\varphi}}(Z, W)) = [-\dot{Z}W - Z\ddot{W}, \ \dot{Z}] \Pi, \\
Q_{\tilde{\varphi}}(Z, W) &= D^2\tilde{\varphi}(Z, W)(\dot{\tilde{\varphi}}(Z, W)) = [-2\dot{Z}W, \ 0] \Pi.
\end{aligned}
\]

For $V = [V_1, \ V_2] \Pi \in \mathbb{R}^{m \times n}$ where $V_2$ is $m \times r$, the Hessian $\nabla^2 \tilde{\varphi}_V(Z, W)$ is the ordinary Euclidean Hessian of $\tilde{\varphi}_V(Z, W) = \langle V, \tilde{\varphi}(Z, W) \rangle$, given by
\[
\nabla^2 \tilde{\varphi}_V(Z, W)(\dot{Z}, \ddot{W}) = [-V_1 W^T, \ Z^T V_1].
\]

We use the above expressions to show that this lift satisfies “$2 \Rightarrow 1$” everywhere on $M$ in Section 5.2.

Next, we illustrate the embedded submanifold case on the following low-dimensional example, which shows that the necessary condition for “$2 \Rightarrow 1$” in the last implication of Theorem 3.23 is not sufficient.

**Example 3.31.** Consider the lift of the unit disk $X$ in $\mathbb{R}^2$ to
\[
M = \{ y \in \mathbb{R}^3 : y_1^2 + y_2^2 + y_3^4 = 1 \}, \quad \varphi(y) = (y_1, y_2).
\]

Note that $M$ is an embedded submanifold of $\mathbb{R}^3$ with defining function $h(y) = y_1^2 + y_2^2 + y_3^4 - 1$, since $\nabla h(y) \neq 0$ for all $y \in M$. The first two derivatives of $h$ are
\[
Dh(y)[\dot{y}] = 2y_1 \dot{y}_1 + 2y_2 \dot{y}_2 + 4y_3^3 \dot{y}_3, \quad D^2h(y)[\dot{y}, \ddot{y}] = 2\dot{y}_1^2 + 2\dot{y}_2^2 + 12y_3^2 \dot{y}_3^2.
\]

Let $y = (1, 0, 0)$ and $x = \varphi(y) = (1, 0)$. We get from (3.4.3) that
\[
T_y M = \{ \dot{y} \in \mathbb{R}^3 : \dot{y}_1 = 0 \}, \quad T^2_{y, \dot{y}} M = (-\dot{y}_2^2, 0, 0) + T_y M.
\]

Because $\varphi$ extends to a linear map $\overline{\varphi}(y) = (y_1, y_2)$ on all of $\mathbb{R}^3$, whose first two derivatives are $D\overline{\varphi}(y)[\dot{y}] = (\dot{y}_1, \dot{y}_2)$ and $D^2\overline{\varphi}(y)[\dot{y}, \ddot{y}] = 0$, we have from (3.4.4) that
\[
L_y(\dot{y}) = (0, \dot{y}_2) \quad \text{and} \quad Q_y(\dot{y}) = (-\dot{y}_2^2, 0).
\]

On the other hand,
\[
T_x X = \{ \dot{x} \in \mathbb{R}^2 : \dot{x}_1 \leq 0 \} \supset \text{im} L_y = \{ \dot{x} \in \mathbb{R}^2 : \dot{x}_1 = 0 \}.
\]
Since \( \text{im} \ L_y \neq T_x \mathcal{X} \), “\( 1 \Rightarrow 1' \)" does not hold at \( y \). Checking our sufficient conditions, Proposition 3.26 gives

\[
A_y = Q_y(\ker(L_y)) + \text{im} \ L_y = \text{im} \ L_y, \quad B_y = \bigcup_{(\hat{y}_i)_1 = 0} \lim_{L_y(\hat{y}_i) \rightarrow 0} (Q_y(\hat{y}_i) + \text{im} \ L_y) = \text{im} \ L_y,
\]

where the last equality follows since \( L_y(\hat{y}_i) = (0, (\hat{y}_i)_2) \rightarrow 0 \) implies \( Q_y(\hat{y}_i) = (-(\hat{y}_i)_2^2, 0) \rightarrow 0 \). Thus, all sufficient conditions in Theorem 3.23 fail.

For the equivalent condition in Theorem 3.23, if \( w \in A_y^* \), or equivalently, \( w = (w_1, 0) \), then \( \nabla^2 \varphi_w(y) \) is the unique self-adjoint operator on \( T_y \mathcal{M} \) satisfying

\[
(\nabla^2 \varphi_w(y)[\hat{y}], \hat{y} = \langle w, Q_y(\hat{y}) \rangle = -w_1 \hat{y}_2^2, \quad \text{hence} \quad \nabla^2 \varphi_w(y)[\hat{y}] = (0, -w_1 \hat{y}_2, 0).
\]

For \( u \in \ker L_y \), or equivalently, \( u = (0, 0, u_3) \), we get \( (\nabla^2 \varphi_w(y)[u], u) = 0 \) and \( \nabla^2 \varphi_w(y)[u] = 0 \). Proposition 3.26 then shows that \( W_y = A^*_y \subset (T_x \mathcal{X})^* \), hence “\( 2 \Rightarrow 1 \)" does not hold.

Nevertheless, the necessary condition in Theorem 3.23 does hold. Indeed,

\[
Q_y(T_y \mathcal{M}) + \text{im} \ L_y = \{(-\hat{y}_2^2, 0) : \hat{y}_2 \in \mathbb{R}\} + \{(0, \hat{y}_2) : \hat{y}_2 \in \mathbb{R}\} = T_x \mathcal{X},
\]

so taking duals on both sides yields the desired condition.

### 3.5 Low-rank PSD matrices, and computing tangent cones

As an application of the theory we developed so far, we consider the set of bounded-rank PSD matrices

\[
\mathcal{X} = \{X \in \mathbb{R}^{n \times n} : X^T = X, \ X \succeq 0, \ \text{rank}(X) \leq r\} = S^{n}_{\geq 0} \cap \mathbb{R}^{n \times n}_{\leq r},
\]

with \( r < n \) together with its lift to \( \mathcal{M} = \mathbb{R}^{n \times r} \) via the factorization map \( \varphi(R) = RR^\top \). This is a special case of the Burer–Monteiro lift for SDPs (BM) without constraints \((m = 0)\). Constructions in the next section enable us to deduce the properties of the general lift from this special case.

**Proposition 3.32.** The lift \( \varphi(R) = RR^\top \) from \( \mathcal{M} = \mathbb{R}^{n \times r} \) to \( \mathcal{X} = \mathbb{R}^{n \times n}_{\leq r} \cap S^{n}_{\geq 0} \) satisfies

- “local ⇒ local” and “\( 2 \Rightarrow 1 \)" everywhere, and
- “\( 1 \Rightarrow 1' \)" at \( R \in \mathcal{M} \) if and only if \( \text{rank}(R) = r \).

Moreover, with \( X = RR^\top \), the sufficient condition \( A_R = T_X \mathcal{X} \) for “\( 2 \Rightarrow 1 \)" holds everywhere, and

\[
T_X \mathcal{X} = T_X \mathbb{R}^{n \times n}_{\leq r} \cap T_X S^{n}_{\geq 0}
\]

\[
= \{V \in S^n : V_\perp \succeq 0, \ \text{rank}(V_\perp) \leq r - \text{rank}(X) \ \text{where} \ V_\perp = \text{Proj}_{\text{col}(X)^\perp} V \text{Proj}_{\text{col}(X)^\perp}\}
\]

\[
= \left\{ U \begin{pmatrix} V_1 & V_2 \\ V_2^\top & V_3 \end{pmatrix} U^\top : V_1 \in S^{\text{rank}(X)}, \ V_3 \succeq 0, \ \text{rank}(V_3) \leq r - \text{rank}(X) \right\},
\]

where in the last line \( U \in O(n) \) is an eigenmatrix for \( X \) satisfying \( X = U \Sigma U^\top \) with \( \Sigma \in \mathbb{R}^{n \times n} \) diagonal containing the eigenvalues of \( X \) in descending order.
Proof. For “local ⇒ local”, we sketch the proof from [11] of (in our terminology) SLP, the condition from Remark 3.8 that is equivalent to “local ⇒ local” by Theorem A.2. Fix \( R \in \mathcal{M} \) and \( X = RR^\top \). For any sequence \( (X_i)_{i \geq 1} \subseteq \mathcal{X} \) converging to \( X \), let \( X_i = U_i \Sigma_i U_i^\top \) be a size-\( r \) eigendecomposition for each \( X_i \) (so \( \Sigma_i \in \mathbb{R}^{r \times r} \) is a diagonal matrix of eigenvalues of \( X_i \), possibly including zeros). Let \( R_i = U_i \Sigma_i \) and note that \( \|R_i\| = \|X_i\|^{1/2} \) which is bounded, hence after passing to a subsequence we may assume that \( \lim_i R_i = \tilde{R} \) exists. By continuity of \( \varphi \), we have \( X = \lim_i X_i = \lim_i \varphi(R_i) = \varphi(\lim_i R_i) = \varphi(\tilde{R}) = \tilde{R} \tilde{R}^\top \), hence \( \tilde{R} \tilde{R}^\top = RR^\top \). By [11, Lemma 2.5], there is an orthogonal \( Q \in O(r) \) satisfying \( R = \tilde{R}Q \), so \( (R_iQ)_{i \geq 1} \subseteq \mathcal{M} \) is a sequence converging to \( R \) satisfying \( \varphi(R_iQ) = X_i \). This proves SLP holds.

For \( V \in \mathbb{R}^{n \times n} \) and \( X = RR^\top \in \mathcal{X} \), define

\[
V \perp := \text{Proj}_{\ker(X)} V \text{Proj}_{\ker(X)} = \text{Proj}_{\col(R)} V \text{Proj}_{\col(R)} \perp,
\]

where we used the fact that \( \col(R) = \col(X) = \ker(X)^\perp \). The tangent cone at \( X \) to \( \mathbb{S}^n_{\geq 0} \) is given by [25, Eq. (9)]

\[
T_X \mathbb{S}^n_{\geq 0} = \left\{ V \in \mathbb{S}^n : (V u, u) \geq 0, \text{ for all } u \in \ker(X) \right\} = \left\{ V \in \mathbb{S}^n : V \perp \geq 0 \right\} = \left\{ U \begin{pmatrix} V_1 & V_2 \\ V_2 & V_3 \end{pmatrix} U^\top : V_1 \in \mathbb{R}^{\text{rank}(X) \times \text{rank}(X)}, \ V_3 \geq 0 \right\}, \tag{3.5.1}
\]

and the tangent cone at \( X \) to \( \mathbb{R}^{n \times n}_{\leq r} \) is given by [48, Thm. 3.2]

\[
T_X \mathbb{R}^{n \times n}_{\leq r} = \left\{ V \in \mathbb{R}^{n \times n} : \text{rank}(V \perp) \leq r - \text{rank}(X) \right\} = \left\{ U \begin{pmatrix} V_1 & V_2 \\ V_2 & V_3 \end{pmatrix} U^\top : V_1 \in \mathbb{R}^{\text{rank}(X) \times \text{rank}(X)}, \ \text{rank}(V_3) \leq r - \text{rank}(X) \right\}.
\]

Hence the intersection \( T_X \mathbb{R}^{n \times n}_{\leq r} \cap T_X \mathbb{S}^n_{\geq 0} \) is given by the claimed expression. Furthermore, the tangent cone to an intersection is always included in the intersection of the tangent cones, which follows easily from Definition 3.1. Hence \( T_X \mathcal{X} \subseteq T_X \mathbb{R}^{n \times n}_{\leq r} \cap T_X \mathbb{S}^n_{\geq 0} \) and it suffices to show the reverse inclusion. We do so simultaneously with proving “2 ⇒ 1”.

Since \( \mathcal{M} \) is a linear space, the expressions (3.4.1) with the identity chart give

\[
\begin{align*}
L_R(\tilde{R}) &= D\varphi(R)[\tilde{R}] = RR^\top + \tilde{R} \tilde{R}^\top, \\
Q_R(\tilde{R}) &= D^2\varphi(R)[\tilde{R}, \tilde{R}] = 2\tilde{R} \tilde{R}^\top.
\end{align*}
\]

Therefore,

\[
\text{im} L_R = \{ V \in \mathbb{S}^n : V \perp = 0 \}. \tag{3.5.2}
\]

Indeed, if \( V \in \text{im} L_R \) then \( V = RR^\top + \tilde{R} \tilde{R}^\top \) for some \( \tilde{R} \in \mathbb{R}^{n \times r} \) and hence \( V \perp = 0 \) since \( \text{Proj}_{\col(R)}^\perp R = 0 \), while if \( V \in \mathbb{S}^n \) satisfies \( V \perp = 0 \) then

\[
V = \text{Proj}_{\col(R)} V + V \text{Proj}_{\col(R)} - \text{Proj}_{\col(R)} V \text{Proj}_{\col(R)} = L_R \left( VR^\top - \frac{1}{2} RR^\dagger VR \right) \in \text{im} L_R,
\]

where \( RR^\dagger = \text{Proj}_{\col(R)} \). Furthermore, we have

\[
Q_R(\ker L_R) \supseteq \{ V \in \mathbb{S}^n_{\geq 0} : \text{rank}(V) \leq r - \text{rank}(X) \}.
\]
Indeed, if \( V \in \mathbb{S}^n_{\geq 0} \) with \( \text{rank}(V) \leq r - \text{rank}(X) \), let \( V = 2\hat{R}_0\hat{R}_0^\top \) be a (rescaled) Cholesky decomposition with \( \hat{R}_0 \in \mathbb{R}^{n \times r} \), so \( \text{rank}(\hat{R}_0) = \text{rank}(V) \). Since

\[
\dim \text{col}(R^\top) = \text{rank}(R^\top) = \text{rank}(R) = \text{rank}(X) \leq r - \text{rank}(\hat{R}_0) = \dim \ker(\hat{R}_0),
\]

there is an orthogonal matrix \( Q \in O(r) \) satisfying \( Q\text{col}(R^\top) \subseteq \ker(\hat{R}_0) \). Let \( \tilde{R} = \hat{R}_0Q \) and note that \( V = 2\tilde{R}\tilde{R}^\top \) so \( Q_R(\tilde{R}) = V \), and \( \tilde{R}R^\top = 0 \) so \( \tilde{R} \in \ker L_R \). Thus, with Proposition 3.26(a),

\[
A_R = Q_R(\ker L_R) + \text{im} L_R \supseteq T_X\mathbb{R}^{n \times n}_{\leq r} \cap T_X\mathbb{S}^n_{\geq 0}.
\]

On the other hand, by Proposition 3.21(b), we have \( A_R \subseteq T_X X \). Thus, we have the chain of inclusions

\[
T_X\mathbb{R}^{n \times n}_{\leq r} \cap T_X\mathbb{S}^n_{\geq 0} \subseteq Q_R(\ker L_R) + \text{im} L_R \subseteq T_X X \subseteq T_X\mathbb{R}^{n \times n}_{\leq r} \cap T_X\mathbb{S}^n_{\geq 0},
\]

so all the above inclusions are equalities. In particular, we obtain the claimed expression for \( T_X X \) and “\( 2 \Rightarrow 1 \)” everywhere on \( \mathcal{M} \) by Theorem 3.23. Our claims about “\( 1 \Rightarrow 1 \)” follow from (3.5.2) and Theorem 2.4.

Finding an explicit expression for tangent cones can be difficult in general. In Proposition 3.32, the set \( \mathcal{X} \) was an intersection of two sets whose tangent cones are known, namely \( \mathbb{R}^{n \times n}_{\leq r} \) and \( \mathbb{S}^n_{\geq 0} \), which gave us an inclusion \( T_X \mathcal{X} \subseteq T_X\mathbb{R}^{n \times n}_{\leq r} \cap T_X\mathbb{S}^n_{\geq 0} \). However, the tangent cone to an intersection can be strictly contained in the intersection of the tangent cones.\(^7\) The proof of “\( 2 \Rightarrow 1 \)” in Proposition 3.32 proceeds by showing \( A_R = T_X\mathbb{R}^{n \times n}_{\leq r} \cap T_X\mathbb{S}^n_{\geq 0} \), which gives \( T_X \mathcal{X} = T_X\mathbb{R}^{n \times n}_{\leq r} \cap T_X\mathbb{S}^n_{\geq 0} = A_R \) because \( A_R \subseteq T_X \mathcal{X} \) by Proposition 3.21(b). This simultaneously gives us “\( 2 \Rightarrow 1 \)” and an expression for the tangent cone.

This illustrates a more general and, as far as we know, novel technique of getting expressions for the tangent cones using lifts. Generalizing the above discussion, if we have an inclusion \( T_x \mathcal{X} \subseteq S \) for some set \( S \) and we are able to prove \( A_y \supseteq S \) for some \( y \in \varphi^{-1}(x) \), then we must have \( T_x \mathcal{X} = S \) by Proposition 3.21(b). In this case, we also conclude that “\( 2 \Rightarrow 1 \)” holds at \( y \) by Theorem 3.23. In Section 4, we shall see another setting in which we naturally have a superset for \( T_x \mathcal{X} \) (see Lemma 4.8), and which allows us to derive expressions for \( T_x \mathcal{X} \) from lifts satisfying “\( 1 \Rightarrow 1 \)” and “\( 2 \Rightarrow 1 \)”.

A general condition implying that the tangent cone to an intersection is the intersection of the tangent cones is given in [46, Thm. 6.42]. That condition does not apply to \( \mathcal{X} = \mathbb{R}^{n \times n}_{\leq r} \cap \mathbb{S}^n_{\geq 0} \) because \( \mathbb{R}^{n \times n}_{\leq r} \) is not Clarke-regular in the sense of [46, Def. 6.4]. Our approach circumvents Clarke regularity, exploiting the existence of an appropriate lift instead.

### 4 Constructing lifts via fiber products

In this section, we give a systematic construction of lifts for a large class of sets \( \mathcal{X} \). If the resulting lifted space is a smooth manifold, we also give conditions under which the lift satisfies our desirable properties. Moreover, under these conditions we can obtain expressions for the tangent cones to \( \mathcal{X} \). We shall see that several natural lifts, including the Hadamard and Burer–Monteiro lifts from Section 2, are special cases of this construction.

\(^7\)For example, consider intersecting the circle in the plane with one of its tangent lines.
Suppose the set $X$ is presented in the form

$$X = \{ x \in E : F(x) \in Z \} = F^{-1}(Z),$$

where $Z \subseteq E'$ is some subset of a linear space and $F : E \rightarrow E'$ is smooth. This form is general—any set $X$ can be written in this form by letting $F$ be the identity and $Z = X$. However, we shall see that our framework is most useful when $Z$ is a product of simple sets for which we have smooth lifts satisfying desirable properties. For example, any set defined by $k$ smooth equalities $g_i(x) = 0$ and $\ell$ smooth inequalities $h_j(x) \geq 0$ can be written in this form by letting $F(x) = (g_1(x), \ldots, g_k(x), h_1(x), \ldots, h_\ell(x))$ and $Z = \{0\}^k \times \mathbb{R}_{\geq 0}^\ell$. We can also incorporate semidefiniteness and rank constraints of smooth functions of $x$ by taking Cartesian products of $Z$ with $\mathbb{R}^{m \times n}$ or $\mathbb{S}^n_{\geq 0}$.

Suppose now that we have a smooth lift $\psi : N \rightarrow Z$. We can use this lift of $Z$ to construct a lift of $X$ by taking the fiber product of $F$ and $\psi$.

**Definition 4.1.** Let $X$ be a subset of $E$ defined by a smooth map $F : E \rightarrow E'$ and a set $Z \subseteq E'$ as $X = F^{-1}(Z)$. Suppose $\psi : N \rightarrow Z$ is a smooth lift of $Z$ to the smooth manifold $N$. Then the fiber product lift of $X$ with respect to $F$ and $\psi$ is $\varphi : M_{F, \psi} \rightarrow X$ where

$$M_{F, \psi} = \{(x, y) \in E \times N : F(x) = \psi(y)\} \quad \text{and} \quad \varphi(x, y) = x.$$ 

Here $M_{F, \psi}$ is the (set-theoretic) fiber product of the maps $F : E \rightarrow E'$ and $\psi : N \rightarrow E'$.

The following commutative diagram illustrates Definition 4.1. Its top horizontal arrow is the coordinate projection $\pi(x, y) = y$.

$$\begin{array}{ccc}
M_{F, \psi} & \xrightarrow{\pi} & N \\
\downarrow \varphi & & \downarrow \psi \\
E \supseteq X & \xrightarrow{F} & Z \subseteq E'
\end{array}$$

The fiber product $M_{F, \psi}$ need not be a smooth manifold even when both $F$ and $\psi$ are smooth maps between smooth manifolds. Accordingly, we make the following assumption:

**Assumption 4.2.** The differential of $(x, y) \mapsto F(x) - \psi(y)$ has constant rank in a neighborhood of $M_{F, \psi}$ in $E \times N$.

Assumption 4.2 not only implies $M_{F, \psi}$ is a smooth embedded submanifold of $E \times N$, but also that $F(x) - \psi(y) = 0$ is a (constant-rank) defining function for it (in the sense of [35, Thm. 5.12]). Under this assumption, the tangent space to $M_{F, \psi}$ is given by

$$T_{(x, y)}M_{F, \psi} = \{(\dot{x}, \dot{y}) \in E \times T_yN : DF(x)[\dot{x}] = D\psi(y)[\dot{y}]\}. \quad (4.0.1)$$

We proceed to give some examples of the above construction. We then study fiber product lifts in general and instantiate our results on these examples.

**Example 4.3** (Sphere to ball). Let $X = \{ x \in \mathbb{R}^n : \|x\|_2 \leq 1 \}$ be the unit Euclidean ball. Let $Z = \mathbb{R}_{\geq 0}$ and $F(x) = 1 - x^\top x$ so $X = F^{-1}(Z)$. Let $\psi : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ be the smooth lift $\psi(y) = y^2$. Then $M_{F, \psi} = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : 1 - x^\top x = y^2\}$, which is just the unit sphere in $\mathbb{R}^{n+1}$, and $\varphi(x, y) = x$ projects onto the first $n$ coordinates. This lift is used in [44, §2.7] to apply a solver for quadratic programming over the sphere (Q) to quadratic programs over the ball (P).
Example 4.4 (Sphere to simplex). The Hadamard lift (Had) from Section 2.1 can be obtained as a special case of Definition 4.1. Indeed, let $\mathcal{X} = \Delta^{n-1} = \{x \in \mathbb{R}^n : x \geq 0, \sum_{i=1}^n x_i = 1\}$ be the standard simplex, let $Z = \mathbb{R}_{\geq 0}^n \times \{0\}$ and $F(x) = (x, \sum_{i=1}^n x_i - 1)$ so $\mathcal{X} = F^{-1}(Z)$. Let $\psi : \mathbb{R}^n \to Z$ be $\psi(y) = (y^{\odot 2}, 0)$ where superscript $\odot 2$ denotes entrywise squaring. Then

$$M_{F,\psi} = \left\{ (x, y) \in \mathbb{R}^n \times \mathbb{R}^n : x = y^{\odot 2}, \sum_{i=1}^n x_i = 1 \right\} = \left\{ (x, y) \in \mathbb{R}^n \times \mathbb{R}^n : x = y^{\odot 2}, \|y\|_2 = 1 \right\},$$

and $\varphi(x, y) = x$. It is easy to check that the coordinate projection $\pi$ defines a diffeomorphism of $M_{F,\psi}$ with $S^{n-1}$, and that the composition $\varphi \circ \pi^{-1} : S^{n-1} \to \mathcal{X}$ yields the Hadamard lift (Had) from Section 2.1. By Proposition 3.27, the fiber product lift of the simplex is equivalent (for the purposes of checking our desirable properties) to the lift (Had).

Example 4.5 (Torus to annulus). Let $\mathcal{X} = \{x \in \mathbb{R}^n : r_1 \leq \|x\|_2 \leq r_2\}$, where we assume $0 < r_1 < r_2$. Let $Z = \mathbb{R}_\geq^2$ and $F(x) = (x^T x - r_1^2, r_2^2 - x^T x)$. Let $\psi : \mathbb{R}^2 \to Z$ be $\psi(y) = y^{\odot 2}$ so

$$M_{F,\psi} = \left\{ (x, y) \in \mathbb{R}^n \times \mathbb{R}^2 : x^T x - r_1^2 = y_1^2, r_2^2 - x^T x = y_2^2 \right\},$$

$$= \left\{ (x, y) \in \mathbb{R}^n \times \mathbb{R}^2 : \|x\|_2 = \sqrt{r_1^2 + y_1^2}, \|y\|_2 = \sqrt{r_2^2 - r_1^2} \right\}.$$

This is an $n$-dimensional manifold diffeomorphic to $S^{n-1} \times S^1$, with diffeomorphism

$$\Phi(x, y) = [(r_1^2 + y_1^2)^{-1/2} x, (r_2^2 - r_1^2)^{-1/2} y].$$

Viewed differently, the equivalent (by Proposition 3.27) lift $\varphi \circ \Phi^{-1}$ is the composition

$$S^{n-1} \times S^1 \to S^{n-1} \times \Delta^1 \to \mathcal{X},$$

where the first map is the Hadamard lift from the sphere to the simplex from the preceding example, and the second map is $(y, \theta) \mapsto \sqrt{\theta r_1^2 + \theta^2 r_2^2 y}$. If $n = 2$, then $\mathcal{X}$ is an annulus and $M$ is a torus.

Example 4.6 (Smooth SDPs). The Burer–Monteiro lift (BM) from Section 2.2 is also a special case of Definition 4.1. To see this, in the notation of Section 2.2, let

$$Z = (S^m_{\geq 0} \cap \mathbb{R}_{\geq 0}^{n \times m}) \times \{0\}^m, \quad F(X) = (X, \langle A_1, X \rangle - b_1, \ldots, \langle A_m, X \rangle - b_m),$$

and $\psi(R) = (RR^T, 0, \ldots, 0)$ defined on $\mathcal{N} = \mathbb{R}^{n \times r}$. In this case,

$$M_{F,\psi} = \{(X, R) \in \mathbb{R}^n \times \mathbb{R}^{n \times r} : X = RR^T, \langle A_i R, R \rangle = b_i\},$$

which the projection $\pi$ maps diffeomorphically onto the set $M$ in (BM). Furthermore, Assumption 4.2 is equivalent in this case to the assumption in Proposition 2.6 that $h_i(R) = \langle A_i R, R \rangle - b_i$ are local defining functions. Thus, proving Proposition 2.6 is equivalent to proving the corresponding properties for the fiber product lift above under Assumption 4.2.

Now that we have seen several examples of fiber product lifts, we ask: when do desirable properties of the lift $\psi : \mathcal{N} \to Z$ imply the corresponding properties for the fiber product lift $\varphi : M_{F,\psi} \to \mathcal{X}$? This is answered by the next few propositions.

Proposition 4.7. Under Assumption 4.2, if $\psi : \mathcal{N} \to Z$ satisfies “local ⇒ local” at $y \in \mathcal{N}$, then $\varphi : M_{F,\psi} \to \mathcal{X}$ satisfies “local ⇒ local” at $(x, y) \in M_{F,\psi}$ for any $x \in \pi^{-1}(y)$.
Proof. By Theorem 2.3, it is equivalent to show that openness of ψ at (x, y) implies openness of ϕ at (x, y). Assumption 4.2 implies that $\mathcal{M}_{\psi, \phi}$ is an embedded submanifold of $\mathcal{E} \times \mathcal{N}$, hence its manifold topology coincides with the subspace topology induced from $\mathcal{E} \times \mathcal{N}$. Thus, to show ϕ is open at (x, y), it suffices to show $\varphi((U \times V) \cap \mathcal{M}_{\psi, \phi})$ is open for any open $U \subseteq \mathcal{E}$ containing x and open $V \subseteq \mathcal{N}$ containing y, since such sets form a basis for the subspace topology on $\mathcal{M}_{\psi, \phi}$. Since ψ is open at y, we get that $\psi(V) \subseteq Z$ is open. Since $F$ is continuous, $F^{-1}(\psi(V)) \subseteq \mathcal{X}$ is open. Since $\varphi(x, y) = x$, we have

$$
\varphi((U \times V) \cap \mathcal{M}_{\psi, \phi}) = \{x \in U \cap \mathcal{X} : \exists y \in V \text{ s.t. } F(x) = \psi(y)\} = \{x \in U \cap \mathcal{X} : F(x) \in \psi(V)\} = (U \cap \mathcal{X}) \cap F^{-1}(\psi(V)),
$$

which is open in $\mathcal{X}$ as the intersection of two open sets. Thus, $\varphi$ is open at (x, y).

Note that the above proof, and hence the conclusion of Proposition 4.7, apply more generally whenever $\mathcal{M}_{\psi, \phi}$ is endowed with the subspace topology induced from $\mathcal{E} \times \mathcal{N}$ (but is not necessarily a smooth manifold) and when all maps involved are continuous (but not necessarily smooth).

We now turn to studying “1 ⇒ 1” and “2 ⇒ 1”. Along the way, we give another instance of the technique for finding tangent cones via lifts outlined in Section 3.5. To do so, we begin by giving a superset of the tangent cone, obtained from the fact that $\mathcal{X}$ is an inverse image [46, Thm. 6.31]. As usual, $DF(x)^{-1}$ denotes the preimage under the differential (which may not be invertible).

Lemma 4.8. The following inclusion holds for all $x \in \mathcal{X}$:

$$
T_x\mathcal{X} \subseteq \{\dot{x} \in \mathcal{E} : DF(x)[\dot{x}] \in TF(x)Z\} = DF(x)^{-1}(TF(x)Z).
$$

Proof. If $v \in T_x\mathcal{X}$ then by Definition 3.1 there exist sequences $(x_i)_{i \geq 1} \subseteq \mathcal{X}$ converging to $x$ and $(\tau_i)_{i \geq 1} \subseteq \mathbb{R}_{>0}$ converging to zero satisfying $v = \lim_{i \to \infty} \frac{x-x_i}{\tau_i}$. Because $F$ is differentiable at $x$, we have $F(x_i) = F(x) + DF(x)[x_i - x] + o(\|x_i - x\|)$, so

$$
DF(x)[v] = \lim_{i \to \infty} \frac{F(x_i) - F(x)}{\tau_i}.
$$

Since $F(x_i) \in Z$ for all $i$, we conclude that $DF(x)[v] \in TF(x)Z$ by Definition 3.1.

Proposition 4.9. Let $(x, y) \in \mathcal{M}_{\psi, \phi}$. Under Assumption 4.2, if $\psi$ satisfies “1 ⇒ 1” at $y \in \mathcal{N}$, then $\varphi$ satisfies “1 ⇒ 1” at $(x, y)$, and equality holds in Lemma 4.8.

Proof. Since $\psi$ satisfies “1 ⇒ 1” at $y$, Theorem 2.4 yields $\text{im} L_y^{\psi} = T_{\psi(y)}Z = TF(x)Z$. Assumption 4.2 implies that $\mathcal{M}_{\psi, \phi}$ is an embedded submanifold of $\mathcal{E} \times \mathcal{N}$. Since $\varphi$ extends to $\varphi(x, y) = x$ defined on all of $\mathcal{E} \times \mathcal{N}$, we get from (3.4.4) that $L_{(x,y)}^{\varphi}(\dot{x}, \dot{y}) = \dot{x}$ for all $(\dot{x}, \dot{y}) \in T_{(x,y)}\mathcal{M}_{\psi, \phi}$. By (4.0.1),

$$
\text{im} L_{(x,y)}^{\varphi} = DF(x)^{-1}(\text{im} L_y^{\psi}) = DF(x)^{-1}(TF(x)Z).
$$

Using Lemma 4.8 and Proposition 3.21(b), we get the chain of inclusions

$$
T_x\mathcal{X} \subseteq DF(x)^{-1}(TF(x)Z) = \text{im} L_{(x,y)}^{\varphi} \subseteq T_x\mathcal{X}.
$$

We conclude that all these sets are equal and hence that “1 ⇒ 1” holds for $\varphi$ at $(x, y)$.
Proposition 4.10. Under Assumption 4.2, if $\psi$ satisfies the sufficient condition $A^\psi_y = T_{\psi(y)}Z$ for \textquotedblleft 2 $\Rightarrow$ 1\textquotedblright at $y \in N$ (recall Theorem 3.23), then $\varphi$ satisfies the sufficient condition $A^{\varphi}_{(x,y)} = T_xX$ for \textquotedblleft 2 $\Rightarrow$ 1\textquotedblright at $(x,y) \in M_{F,\psi}$, and equality holds in Lemma 4.8.

Proof. By Lemma 4.8, we always have $T_xX \subseteq DF(x)^{-1}(T_F(x)Z)$. For the reverse inclusion and the desired sufficient condition for \textquotedblleft 2 $\Rightarrow$ 1\textquotedblright, it suffices to prove that $DF(x)^{-1}(T_F(x)Z) \subseteq A^\psi_{(x,y)}$ since $A^\psi_{(x,y)} \subseteq T_xX$ by Proposition 3.21(b).

Suppose $DF(x)[\dot{x}] \in T_F(x)Z$. By hypothesis, $T_F(x)Z = T_{\psi(y)}Z = A^\psi_y$, so by Proposition 3.26(a):

$$DF(x)[\dot{x}] = Q^{\psi}_y(v) + L^{\psi}_y(u), \quad \text{for some } v \in \ker L^{\psi}_y \text{ and } u \in T_yN. \tag{4.0.2}$$

Because $v \in \ker L^{\psi}_y$, we have $(0,v) \in T_{(x,y)}M_{F,\psi}$ by (4.0.1). Let $t \mapsto c(t) = (c_x(t),c_y(t))$ be a smooth curve on $M_{F,\psi}$ passing through $(x,y)$ with velocity $(0,v)$ at $t = 0$. Because $F(c_x(t)) = \psi(c_y(t))$ for all $t$ near 0, differentiating this expression twice we get

$$D^2F(x)[c''_x(0),c'_x(0)] + DF(x)[c''_x(0)] = (\psi \circ c_y)''(0).$$

The first term vanishes since $c'_x(0) = 0$. Using Definition 3.9 for $Q^{\psi}_y$ with Lemma 3.24, we obtain

$$DF(x)[c''_x(0)] = Q^{\psi}_y(v) + L^{\psi}_y(u'), \quad \text{for some } u' \in T_yN. \tag{4.0.3}$$

Subtracting (4.0.3) from (4.0.2) yields (using Definition 3.9 for $L^{\psi}_y$ and (4.0.1) for $T_{(x,y)}M_{F,\psi}$)

$$DF(x)[\dot{x} - c''_x(0)] = L^{\psi}_y(u - u'), \quad \text{hence } (\dot{x} - c''_x(0), u - u') \in T_{(x,y)}M_{F,\psi}.$$

Since $D\varphi(x,y)[\dot{x},\dot{y}] = \dot{x}$, it follows that $\dot{x} - c''_x(0) \in \im L^{\varphi}_{(x,y)}$. By Definition 3.9 and Lemma 3.24 again, there exists $w \in T_{(x,y)}M_{F,\psi}$ satisfying

$$c''_x(0) + L^{\varphi}_{(x,y)}(w) = Q^{\varphi}_{(x,y)}(0,v) \in Q^{\varphi}_{(x,y)}(\ker L^{\varphi}_{(x,y)}),$$

from which we see $\dot{x} \in Q^{\varphi}_{(x,y)}(\ker L^{\varphi}_{(x,y)}) + \im L^{\varphi}_{(x,y)} = A^\varphi_{(x,y)}$ (again with Proposition 3.26(a)). \qed

We remark that other sufficient conditions for equality in Lemma 4.8 to be achieved are given in [46, Exer. 6.7, Thm. 6.31]. However, they do not apply to Example 4.6 ($Z$ is not Clarke-regular and $DF(X)$ may not be surjective). In contrast, our approach via lifts does apply to this example, and gives \textquotedblleft 2 $\Rightarrow$ 1\textquotedblright and an expression for the tangent cones simultaneously, see Corollary 4.12 below.

As the examples in the beginning of this section illustrate, $Z$ is often a product of sets. It is therefore useful to note that a product of lifts satisfying desirable properties also satisfies those properties:

Proposition 4.11. Suppose $Z_i \subseteq E_i$ for $i = 1, \ldots, k$ are subsets admitting smooth lifts $\psi_i: N_i \rightarrow Z_i$. Let $Z = Z_1 \times \cdots \times Z_k$ and $\psi = \psi_1 \times \cdots \times \psi_k: N_1 \times \cdots \times N_k \rightarrow Z$, which is a smooth lift of $Z$. Then the following hold.

(a) $T_zZ \subseteq T_{z_1Z_1} \times \cdots \times T_{z_kZ_k}$ (the inclusion may be strict, see [46, Prop. 6.41]).

(b) $\psi$ satisfies \textquotedblleft local $\Rightarrow$ local\textquotedblright at $y$ if and only if $\psi_i$ satisfies \textquotedblleft local $\Rightarrow$ local\textquotedblright at $y_i$, for all $i$.

(c) We have $\im L^{\psi}_y = \im L^{\psi_{y_1}}_{y_1} \times \cdots \times \im L^{\psi_{y_k}}_{y_k}$. In particular, $\psi$ satisfies \textquotedblleft 1 $\Rightarrow$ 1\textquotedblright at $y$ if $\psi_i$ satisfies \textquotedblleft 1 $\Rightarrow$ 1\textquotedblright at $y_i$ for all $i$, in which case equality in (a) holds.

32
(d) We have \(Q_y^\psi \equiv Q_{y_1}^\psi \times \cdots \times Q_{y_k}^\psi \mod \ker L_y\). Moreover, \(A_y^\psi = A_{y_1}^\psi \times \cdots \times A_{y_k}^\psi\) and likewise for \(B_y^\psi\) and \(W_y^\psi\). In particular, \(\psi\) satisfies “2 \(\Rightarrow\) 1” at \(y\) if \(\psi_i\) satisfies “2 \(\Rightarrow\) 1” at \(y_i\) for all \(i\).

(e) \(\psi\) satisfies the sufficient condition \(A_y^\psi = T_yZ\) for “2 \(\Rightarrow\) 1” if \(\psi_i\) satisfies the corresponding conditions \(A_{y_i}^\psi = T_{y_i}Z_i\) for all \(i\).

The proof is given in Appendix C.2. Equality in Proposition 4.11(a) is achieved when each \(Z_i\) is Clarke-regular at \(z_i\) [46, Prop. 6.41]. By Remark 3.25, equality of \(Q_y^\psi\) and \(Q_{y_1}^\psi \times \cdots \times Q_{y_k}^\psi\) modulo \(\ker L_y\) means that either one can be used to verify “2 \(\Rightarrow\) 1”.

We can now revisit the examples from the beginning of this section.

**Corollary 4.12.** The lifts in Examples 4.3, 4.4, 4.5 and 4.6 satisfy the following.

- The sphere to ball lift in Example 4.3 satisfies “local \(\Rightarrow\) local” everywhere, “1 \(\Rightarrow\) 1” at \(y\) if and only if \(y \neq 0\) (i.e., at preimages of the interior of the ball), and “2 \(\Rightarrow\) 1” everywhere.

- The sphere to simplex lift (Had) satisfies “local \(\Rightarrow\) local” everywhere, “1 \(\Rightarrow\) 1” at \(y\) if and only if \(y_i \neq 0\) for all \(i\) (i.e., at preimages of the relative interior of the simplex), and “2 \(\Rightarrow\) 1” everywhere.

- The lift of the annulus in Example 4.5 satisfies “local \(\Rightarrow\) local” everywhere, “1 \(\Rightarrow\) 1” at \(y\) if and only if \(y_1, y_2 \neq 0\) (i.e., at preimages of the interior of the annulus), and “2 \(\Rightarrow\) 1” everywhere.

- The Burer–Monteiro lift (BM) under the smoothness assumption satisfies “local \(\Rightarrow\) local” everywhere, “1 \(\Rightarrow\) 1” at \(R\) if and only if \(\text{rank}(\text{BM}) = r\) (i.e., at preimages of points of rank \(r\)), and “2 \(\Rightarrow\) 1” everywhere. Moreover, we get an expression for the tangent cones to \(X\):

\[
T_X X = \{V \in \mathbb{R}^n : V \in T_X S^n_0 \cap T_X \mathbb{R}^{n \times n}_{\leq r} \text{ and } \langle A_i, V \rangle = 0 \text{ for all } i\}. \tag{4.0.4}
\]

In particular, this proves Propositions 2.5 and 2.6. Note that an expression for \(T_X S^n_0 \cap T_X \mathbb{R}^{n \times n}_{\leq r}\) is derived in Proposition 2.6 (incidentally, also as a consequence of the sufficient condition for “2 \(\Rightarrow\) 1” used in Proposition 4.10).

**Proof.** For the first three bullet points, consider the lift \(\psi(y) = y^2\) from \(N = \mathbb{R}\) to \(Z = \mathbb{R}_{\geq 0}\). Observe that it satisfies “local \(\Rightarrow\) local” everywhere, “1 \(\Rightarrow\) 1” at \(y \neq 0\), and the sufficient condition \(A_y = T_{\psi(y)} Z\) for “2 \(\Rightarrow\) 1” at \(y = 0\). Indeed, at \(y \neq 0\) we have \(L_y(y) = 2y\psi\) which is an isomorphism of \(T_y N = \mathbb{R}\) and \(T_y Z = \mathbb{R}\); and at \(y = 0\) we have \(L_y = 0\) and \(Q_y(y) = 2y^2\) by (3.4.1) so \(A_y = Q_y(\ker L_y) + \ker L_y = \mathbb{R}_{\geq 0} = T_0 Z\). Propositions 4.7 and 4.9–4.11 imply that the first three lifts satisfy “local \(\Rightarrow\) local” and “2 \(\Rightarrow\) 1” everywhere, and give the claimed “if” directions for “1 \(\Rightarrow\) 1”. The “only if” directions follow from Theorem 2.4.

For the Burer–Monteiro lift, consider the lift \(\psi(R) = RR^\top\) from \(N = \mathbb{R}^{n \times r}\) to \(Z = S^n_{\geq 0} \cap \mathbb{R}^{n \times n}_{\leq r}\). Proposition 3.32 shows that \(\psi\) satisfies “local \(\Rightarrow\) local” and the sufficient condition \(A_R = T_{RR^\top} Z\) for “2 \(\Rightarrow\) 1” everywhere, as well as “1 \(\Rightarrow\) 1” at points \(R\) of rank \(r\). Therefore, Propositions 4.7 and 4.9–4.11 imply that the Burer–Monteiro lift satisfies “local \(\Rightarrow\) local” and “2 \(\Rightarrow\) 1” everywhere and “1 \(\Rightarrow\) 1” at \(R\) if \(\text{rank}(R) = r\). The “1 \(\Rightarrow\) 1” property does not hold at other points by Theorem 2.4. Proposition 4.10 gives the claimed expression for the tangent cones to \(X\). □
Example 4.13. We can now revisit the example from Section 1 about computing the smallest eigenvalue of a symmetric matrix $A = U \text{diag}(\lambda) U^\top$ with $U$ orthogonal. There,

$$X = \Delta^{d-1}, \quad M = S^{d-1} \quad \text{and} \quad \varphi(y) = \text{diag}(U^\top y y^\top U).$$

Observe that $\varphi(y) = (U^\top y)^\odot 2$, which is the composition of the diffeomorphism $y \mapsto U^\top y$ from the sphere to itself and the Hadamard lift from Example 4.4. We conclude that this lift satisfies “2 ⇒ 1” everywhere on $M$ by Proposition 3.27 and Corollary 4.12. Therefore, any 2-critical point for $(Q)$ maps to a stationary point for $(P)$, for any cost $f$. If $f$ is convex, then since $X$ is also convex any stationary point for $(P)$ is globally optimal. Thus, in this case any 2-critical point for $(Q)$ is globally optimal and its nonconvexity is benign. This is well-known for the eigenvalue problem, which corresponds to the case of linear $f$.

5 Analysis of low rank lifts

In this section, we use our theory to prove the remaining results from Section 2 concerning lifts of low rank matrices and tensors.

5.1 Proof of Proposition 2.7 ($LR^\top$ lift)

For “local ⇒ local”, we follow [36, Prop. 2.34] and show SLP holds at every “balanced” $(L,R)$ factorizations, i.e., one satisfying $\text{rank}(L) = \text{rank}(R) = \text{rank}(LR^\top)$.\(^8\) Let $X = LR^\top$ and suppose $(X_i)_{i \in \mathbb{N}} \subseteq \mathcal{X}$ converges to $X$. Let $X_i = U_i \Sigma_i V_i^\top$ be a size-$r$ SVD of $X_i$ where $\Sigma_i \in \mathbb{R}^{r \times r}$ is diagonal with the first $r$ singular values of $X_i$ (possibly including zeros) on the diagonal. Let $L_i = U_i \Sigma_i^{1/2}$ and $R_i = V_i \Sigma_i^{1/2}$. These satisfy $\varphi(L_i, R_i) = L_i R_i^\top = X_i$ and $\|L_i\| = \|R_i\| = \|X_i\|^{1/2}$. Since $\|X_i\|$ are bounded, after passing to a subsequence we may assume that the limit $(L_\infty, R_\infty) = \lim_i (L_i, R_i)$ exists. By continuity of $\varphi$, we must have $L_\infty R_\infty^\top = X$. Since $L_i^\top L_i = R_i^\top R_i = \Sigma_i$ for all $i$, we also have $L_\infty^\top L_\infty = R_\infty^\top R_\infty$. This implies $\text{rank}(L_\infty) = \text{rank}(R_\infty) = \text{rank}(X)$ by considering the polar decompositions of $L_\infty, R_\infty$, see [36, Lem. 2.33]. By [36, Lem. 2.32], there exists $J \in \text{GL}(r)$ satisfying $L = L_\infty J$ and $R = R_\infty J^{-\top}$. Therefore, $(L_i J, R_i J^{-\top})$ converges to $(L, R)$ and is a lift of $X_i$, showing that SLP holds. We conclude that $\varphi$ satisfies “local ⇒ local” at $(L,R)$ such that $\text{rank}(L) = \text{rank}(R) = \text{rank}(LR^\top)$ by Theorem A.2.

Conversely, we show that if $\text{rank}(L), \text{rank}(R)$ and $\text{rank}(LR^\top)$ are not all equal then “local ⇒ local” does not hold at $(L,R)$ by constructing a sequence $(X_i)$ converging to $X = LR^\top$ no subsequence of which can be lifted to a sequence converging to $(L,R)$. It always holds that $\text{rank}(LR^\top) \leq \min\{\text{rank}(L), \text{rank}(R)\}$. Assume $\text{rank}(X) < \text{rank}(L)$ (in particular, $\text{rank}(X) < r$). The case $\text{rank}(X) < \text{rank}(R)$ is similar. Define

$$L_i = \text{Proj}_{\text{col}(X)} L + i^{-1} L_\perp, \quad \text{and} \quad R_i = \text{Proj}_{\text{row}(X)} R + i^{-1} R_\perp,$$

for $L_\perp \in \mathbb{R}^{m \times r}$ satisfying

(a) $\text{col}(L_\perp) \perp \text{col}(X)$;

(b) $\text{rank}(L_\perp) = r - \text{rank}(X)$.

\(^8\)Compare this argument to the proof of Proposition 2.6. In both cases, the proof relies on a characterization of the fibers of $\varphi$ or a subset of them.
(c) \( \text{col}(L_\perp) \neq \text{col}(L) \cap \text{col}(X)^\perp \). 

The matrix \( R_\perp \in \mathbb{R}^{n \times r} \) satisfies the analogous conditions. Such \( L_\perp \) can be obtained by letting its first column be an arbitrary nonzero vector \( v_1 \in \text{col}(L)^\perp \) (which exists since \( \text{rank}(L) \leq r < \min\{m, n\} \)), letting the next \( r - \text{rank}(X) - 1 \) columns be linearly independent vectors in \( \text{col}(X)^\perp \cap \text{span}\{v_1\}^\perp \) (which exist because \( r - \text{rank}(X) - 1 \leq \min\{m, n\} - \text{rank}(X) - 1 \leq \dim \text{col}(X)^\perp \cap \text{span}\{v_1\}^\perp \)), and the remaining rank(\( X_i \)) columns to be zero.

Note that \( \text{rank}(L_i) = \text{rank}(R_i) = r \) for all \( i \) by conditions (a) and (b). Define \( X_i = L_iR_i^\top \) which converges to \( X \) as \( i \to \infty \), and suppose \((\tilde{L}_{ij}, \tilde{R}_{ij})\) is a lift of a subsequence of \((X_i)\) converging to \((L, R)\). Because \( \text{rank}(L_i) = \text{rank}(R_i) = r \), we also have \( \text{rank}(X_i) = r \), and there exist \( J_{ij} \in \text{GL}(r) \) such that \((\tilde{L}_{ij}, \tilde{R}_{ij}) = (L_iJ_{ij}, R_iJ_{ij}^{-\top})\), see [36, Lem. 2.32]. However, we have \( \text{col}(\tilde{L}_{ij}) = \text{col}(L_{ij}) = \text{col}(X) \cap \text{col}(L_\perp) \) is constant and not equal to \( \text{col}(L) \) by condition (c), contradicting \( \tilde{L}_{ij} \to L \). Thus, no subsequence of \((X_i)\) can be lifted, showing that \( \varphi \) does not satisfy “local ⇒ local” at \((L, R)\) by Theorem A.2. For an explicit example of a cost \( f \) and point \((L, R)\) which is a local minimum for \((Q)\) but such that \( LR^\top \) is not a local minimum for \((P)\), see [36, Prop. 2.30].

For “1 ⇒ 1” and “2 ⇒ 1”, note that \( \mathcal{M} \) is a linear space, hence (3.4.1) gives

\[
L_{(L,R)}(\hat{L}, \hat{R}) = \hat{L}R^\top + L\hat{R}^\top, \quad Q_{(L,R)}(\hat{L}, \hat{R}) = 2\hat{L}\hat{R}^\top.
\]

If \( \text{rank}(LR^\top) = r \), then [36, Prop. 2.15] (which is a slight generalization of the proof of “1 ⇒ 1” in Proposition 2.6) shows \( \text{im} L_{(L,R)} = T_{LR^\top}\mathcal{X} \) hence “1 ⇒ 1” holds. If \( \text{rank}(LR^\top) < r \) then \( T_{LR^\top}\mathcal{X} \) is not a linear space, hence “1 ⇒ 1” does not hold at \((L, R)\) by Theorem 2.4. To show “2 ⇒ 1” holds everywhere on \( \mathcal{M} \), it suffices to show \( B^*_{(L,R)} \subseteq (TX\mathcal{X})^* \) whenever \( \text{rank}(LR^\top) < r \) by Theorem 3.23. Since \( \text{rank}(LR^\top) < r \) we must have either \( \text{rank}(L) < r \) or \( \text{rank}(R) < r \), assume the former (the case \( \text{rank}(R) < r \) is similar). Then there exists \( w \in \mathbb{R}^r \) such that \( Lw = 0 \) and \( \|w\|^2 = 1 \). For any \( u \in \mathbb{R}^m \), \( v \in \mathbb{R}^n \), and \( i \in \mathbb{N} \), let \( \hat{L}_i = i^{-1}uw^\top \) and \( \hat{R}_i = (i/2)vw^\top \). Then

\[
L_{(L,R)}(\hat{L}_i, \hat{R}_i) = i^{-1}uw^\top R^\top \xrightarrow{i \to \infty} 0, \quad Q_{(L,R)}(\hat{L}_i, \hat{R}_i) = uv^\top,
\]

showing that \( uv^\top \in B_{(L,R)} \). Thus, \( B_{(L,R)} \) contains all rank-1 matrices, showing that \( B^*_g = \{0\} = (TX\mathcal{X})^* \).

5.2 Proof of Proposition 2.8 (desingularization lift)

We begin by showing “local ⇒ local” does not hold at \((X, S) \in \mathcal{M} \) if \( \text{rank}(X) < r \). Since \((X, S) \in \mathcal{M} \), the subspace \( S \subseteq \mathbb{R}^n \) has dimension \( n - r \) and \( S \subseteq \ker(X) \). We construct a sequence converging to \( X \) such that no subsequence of \( S \) can be lifted to a sequence converging to \((X, S)\), demonstrating that SLP does not hold at \((X, S)\). Let \( X = USV^\top \) be a thin SVD for \( X \) where \( U \in \text{St}(m, \text{rank}(X)) \), \( V \in \text{St}(n, \text{rank}(X)) \). Since \( S \subseteq \ker(X) \), we have \( \text{col}(V) \subseteq S^\perp \). Let \( U_\perp \in \text{St}(m, r - \text{rank}(X)) \) have columns orthogonal to those of \( U \), and \( V_\perp \in \text{St}(n, r - \text{rank}(X)) \) have columns orthogonal to those of \( V \), such that some column of \( V_\perp \) is contained in \( S \). Define

\[
X_i = X + i^{-1}U_\perp V_\perp^\top \in \mathbb{R}^{m \times n}.
\]

Note that \( X_i \to X \), that \( \text{rank}(X_i) = r \) for all \( i \), and that \( \ker(X_i) = \ker(X) \cap \text{col}(V_\perp)^\perp = S' \neq S \) for all \( i \). Suppose \((X_{ij}, S_{ij}) \in \mathcal{M} \) is a lift of a subsequence converging to \((X, S)\). Since \( \text{rank}(X_i) = r \), we must have \( S_{ij} = \ker(X_{ij}) = S' \) for all \( j \). Thus, \( \lim_j S_{ij} = S' \neq S \), a contradiction. Therefore,
no subsequence of \((X_i)\) can be lifted, so we conclude that \(\varphi\) does not satisfy “local \(\Rightarrow\) local” at \((X, \mathcal{S})\) by Theorem A.2. For an explicit example of a cost \(f\) and point \((X, \mathcal{S}) \in \mathcal{M}\) that is a local minimum for \((Q)\), but such that \(X\) is not a local minimum for \((P)\), see [36, Prop. 2.37].

We show “local \(\Rightarrow\) local” does hold at \((X, \mathcal{S}) \in \mathcal{M}\) if \(\operatorname{rank}(X) = r\) by showing that “1 \(\Rightarrow\) 1” holds there, which suffices by Proposition 2.11.

For “1 \(\Rightarrow\) 1” and “2 \(\Rightarrow\) 1”, we use the results of Example 3.30. Recall from that example that every \((X, \mathcal{S}) \in \mathcal{M}\) is in the image of a chart of the form

\[
\psi(Z, W) = \begin{pmatrix} -ZW, & Z \end{pmatrix} \Pi, \quad \operatorname{col}(\Pi^T \begin{bmatrix} I_{n-r}^\top \\ W \end{bmatrix}), \quad (Z, W) \in \mathbb{R}^{m \times r} \times \mathbb{R}^{r \times (n-r)}
\]

for some permutation matrix \(\Pi \in \mathbb{R}^{n \times n}\). Proposition 3.27 implies that our desirable properties hold at \((Z, W)\) if they hold at \((X, \mathcal{S})\), so it suffices to consider the composed lift

\[
\tilde{\varphi}(Z, W) = [-ZW, \ Z] \Pi =: X,
\]

defined on the linear space \(\mathbb{R}^{m \times r} \times \mathbb{R}^{r \times (n-r)}\). For this lift, we computed \(L_{(Z,W)}\) and \(Q_{(Z,W)}\) in (3.4.5):

\[
L_{(Z,W)}(\hat{Z}, \hat{W}) = [-\hat{Z}W - Z\hat{W}, \ \hat{Z}] \Pi, \quad Q_{(Z,W)}(\hat{Z}, \hat{W}) = [-2\hat{Z}W, \ 0] \Pi.
\]

Suppose \(\operatorname{rank}(X) = r\) and \((X, \mathcal{S}) = \psi(Z, W)\). Note that \(\operatorname{col}(X) = \operatorname{col}(Z)\), so \(\operatorname{rank}(Z) = r\). If \(L_{(Z,W)}(\hat{Z}, \hat{W}) = 0\), then \(\hat{Z} = 0\) and \(Z\hat{W} = 0\). This implies \(\hat{W} = 0\) since \(Z\) has full column rank. Thus, \(L_{(Z,W)}\) is injective, but since its domain has dimension \((m+n-r)r = \dim \mathbb{R}^{m \times n}\), we conclude that it is an isomorphism. Thus, “1 \(\Rightarrow\) 1” holds at \((Z, W)\) by Theorem 2.4. If \(\operatorname{rank}(X) < r\) then \(T_X \mathcal{X}\) is not a linear space [26, Thm. 2.2], hence “1 \(\Rightarrow\) 1” cannot hold for any lift by Theorem 2.4.

Suppose \(\operatorname{rank}(X) < r\). We show “2 \(\Rightarrow\) 1” holds at \((Z, W)\) by showing that \(B_{(Z,W)}^* \subseteq (T_X \mathcal{X})^*\). To that end, note that \(\operatorname{rank}(Z) = \operatorname{rank}(X) < r\), so there is a unit vector \(w \in \mathbb{R}^r\) satisfying \(Zw = 0\). Let \(u \in \mathbb{R}^m\) and \(v \in \mathbb{R}^{n-r}\) be arbitrary. For any \(i \in \mathbb{N}\), let \(\hat{Z}_i = i^{-1}uw^\top\) and \(\hat{W}_i = iwv^\top\). Then

\[
L_{(Z,W)}(\hat{Z}_i, \hat{W}_i) = [-i^{-1}uw^\top W - i(Zw)v^\top, \ i^{-1}uw^\top] \Pi \xrightarrow{i \to \infty} 0,
\]

\[
Q_{(Z,W)}(\hat{Z}_i, \hat{W}_i) \equiv [-2uv^\top, \ 0] \Pi.
\]

We conclude that

\[
B_{(Z,W)} \supseteq (\mathbb{R}^{m \times (n-r)} \times \{0\}) \Pi + \operatorname{im} L_{(Z,W)} \implies B_{(Z,W)}^* \subseteq (\{0\} \times \mathbb{R}^{m \times r}) \Pi \cap (\operatorname{im} L_{(Z,W)})^\perp.
\]

To characterize \((\operatorname{im} L_{(Z,W)})^\perp\), observe that \(V = \begin{bmatrix} V_1, & V_2 \end{bmatrix} \Pi \in \mathbb{R}^{m \times n}\) with \(V_1 \in \mathbb{R}^{m \times (n-r)}\) satisfies \(V \in (\operatorname{im} L_{(Z,W)})^\perp\) iff the following holds for all \((\hat{Z}, \hat{W}) \in \mathbb{R}^{m \times r} \times \mathbb{R}^{r \times (n-r)}\):

\[
\langle V, L_{(Z,W)}(\hat{Z}, \hat{W}) \rangle = \langle V_1, -\hat{Z}W - Z\hat{W} \rangle + \langle V_2, \hat{Z} \rangle = \langle \hat{Z}, V_2 - V_1 W^\top \rangle - \langle \hat{W}, Z^\top V_1 \rangle = 0.
\]

This is equivalent to \(V_2 = V_1 W^\top\) and \(Z^\top V_1 = 0\). Thus, if \(V_1 = 0\) then \(V = 0\), hence \(B_{(Z,W)}^* = \{0\} = (T_X \mathcal{X})^*\). This shows “2 \(\Rightarrow\) 1” holds at \((Z, W)\), and hence also at \((X, \mathcal{S})\) by Proposition 3.27.

### 5.3 Multilinear lifts, and tensors

In this section, we prove two obstructions to “2 \(\Rightarrow\) 1” for multilinear lifts, which apply in particular to lifts defined by tensor factorizations and linear neural networks as discussed in Sections 2.4-2.5.
Proposition 5.1. Suppose φ: M → X ⊆ E is a smooth lift where M ⊆ E = E_1 ×⋯× E_d is a smooth embedded submanifold of a product of Euclidean spaces E_i, and φ is defined on all of E' and is multilinear in its d arguments. If M contains a point (y_1,..,y_d) such that y_i = 0 for three indices i, and 0 = φ(y_1,..,y_d) is not an isolated point of X, then φ does not satisfy “2 ⇒ 1” at (y_1,..,y_d).

Proof. Let (y_1,..,y_d) ∈ E'. Note that if y_i = 0 for some i, then φ(y_1,..,y_d) = 0 by the multilinearity of φ. Also, since φ is multilinear and defined on all of E', its differential in the ambient Euclidean space is

\[ Dφ(y_1,..,y_d)[(\dot{y}_1),..,\dot{y}_d] = \frac{d}{dt} \bigg|_{t=0} φ(y_1 + t\dot{y}_1,..,y_d + t\dot{y}_d) = \sum_{i=1}^{d} φ_{y_i}[(y_1,..,y_{i-1},\dot{y}_i,y_{i+1},..,y_d)]. \]

Similarly,

\[ D^2φ(y_1,..,y_d)[(\dot{y}_1),..,\dot{y}_d,\ddot{y}_1,..,\ddot{y}_d] = \frac{d}{dt} Dφ(y_1 + t\dot{y}_1,..,y_d + t\dot{y}_d)[\ddot{y}_1,..,\ddot{y}_d] = 2 \sum_{1≤ i,j≤ d, i≠j} φ_{y_i,y_j}[(y_1,..,\dot{y}_{i-1},\dot{y}_i,\dot{y}_{i+1},..,\dot{y}_{j-1},\ddot{y}_j,\ddot{y}_{j+1},..,\ddot{y}_d)]. \]

In particular, if y_i = 0 for at least three indices i, then Dφ(y_1,..,y_d) = D^2φ(y_1,..,y_d) = 0. Hence (3.4.4) gives L_{(y_1,..,y_d)} = 0 and Q_{(y_1,..,y_d)} = 0. This implies

\[ (\text{im } L_{(y_1,..,y_d)})^⊥ \cap (Q_{(y_1,..,y_d)}(T_{(y_1,..,y_d)}M))^\ast = E. \]

The necessary condition for “2 ⇒ 1” given by the last implication in Theorem 3.23 is satisfied iff (T_0X)^\ast = E, or equivalently T_0X = \{0\}. This holds if and only if 0 is an isolated point of X, since if (x_i) ⊆ X \ {0} is a sequence converging to 0, then after passing to a subsequence (x_i/\|x_i\|) converges and gives a nonzero element of T_0X.

Proposition 5.1 implies that the lifts corresponding to linear neural networks, as well as standard tensor decompositions such as CPD, Tensor Train (TT), and Tucket, all do not satisfy “2 ⇒ 1” as points with at least three zero factors.

Proposition 5.1 might suggest that failure of “2 ⇒ 1” can be avoided by normalizing the arguments of the lift to have unit norm. Specifically, by multilinearity of φ we have

\[ φ(y_1,..,y_d) = \left( \prod_{i=1}^{d} \|y_i\| \right) φ\left( \frac{y_1}{\|y_1\|},..,\frac{y_d}{\|y_d\|} \right), \text{ whenever } y_i ≠ 0 \text{ for all i}. \]

Using this observation, one could replace a lift φ: R^{n_1} ×⋯× R^{n_d} → X to a product of Euclidean spaces by a lift ψ: R × S^{n_1-1} ×⋯× S^{n_d-1} to a product of R and several spheres, satisfying ψ(λ, x_1,..,x_d) = λφ(x_1,..,x_d). Only one factor can be zero in this new lift, so Proposition 5.1 does not apply and we might hope that “2 ⇒ 1” holds. Unfortunately, this may not resolve the problem as there is another obstruction to “2 ⇒ 1” for the following specific form of a lift.

Proposition 5.2. Suppose φ: M → X is a smooth lift of the form

\[ φ(λ,Y_1,..,Y_d) = \sum_{i=1}^{r} λ_i \cdot (Y_1)_{:,i} \otimes⋯\otimes(Y_d)_{:,i}, \]
where $\mathcal{M} \subseteq \mathbb{R}^r \times \mathbb{R}^{n_1 \times r} \times \cdots \times \mathbb{R}^{n_d \times r}$. Denote $X = \varphi(\lambda, Y_1, \ldots, Y_d)$. If $d \geq 3$ and
\[
\text{col}(Y_1) \perp \cdots \perp \text{col}(Y_d) \perp \not\subseteq (T_X \mathcal{X})^*,
\]  
then $\varphi$ does not satisfy “$2 \Rightarrow 1$” at $(\lambda, Y_1, \ldots, Y_d)$ for any $\lambda \in \mathbb{R}^r$. If $d = 2$ and (5.3.1) holds, then $\varphi$ does not satisfy “$2 \Rightarrow 1$” at $(0, Y_1, Y_2)$.

Proof of Proposition 5.2. For any $W \in \text{col}(Y_1) \perp \cdots \perp \text{col}(Y_d) \perp$, we have
\[
\langle W, D\varphi(\lambda, Y_1, \ldots, Y_d)[\dot{\lambda}, \dot{Y}_1, \ldots, \dot{Y}_d] \rangle = \langle W, D^2\varphi(\lambda, Y_1, \ldots, Y_d)[(\dot{\lambda}, \dot{Y}_1, \ldots, \dot{Y}_d), (\dot{\lambda}, \dot{Y}_1, \ldots, \dot{Y}_d)] \rangle = 0,
\]
for all $(\dot{Y}_1, \ldots, \dot{Y}_d)$ if $d \geq 3$ or $d = 2$ and $\lambda = 0$, by multilinearity. Since $L_{(\lambda, Y_1, \ldots, Y_d)}$ is the restriction of $D\varphi(\lambda, Y_1, \ldots, Y_d)$ to $T_{(\lambda, Y_1, \ldots, Y_d)} \mathcal{M}$ and $Q_{(\lambda, Y_1, \ldots, Y_d)}$ is given by (3.4.4), we get
\[
\text{col}(Y_1) \perp \cdots \perp \text{col}(Y_d) \perp \subseteq (\text{im} L_{(\lambda, Y_1, \ldots, Y_d)})^+ \cap (Q_{(\lambda, Y_1, \ldots, Y_d)}(T_{(\lambda, Y_1, \ldots, Y_d)} \mathcal{M}))^*,
\]
if either $d \geq 3$ or $d = 2$ and $\lambda = 0$. Thus, if $\text{col}(Y_1) \perp \cdots \perp \text{col}(Y_d) \perp \not\subseteq (T_X \mathcal{X})^*$ then the necessary condition for “$2 \Rightarrow 1$” from Theorem 3.23 does not hold.

Proposition 5.2 applies in particular to lifts corresponding to symmetric and normalized CP decompositions and ODECO tensors [45], as well as the SVD lift (SVD). As discussed in Section 2, these obstructions to “$2 \Rightarrow 1$” imply that guarantees for second-order optimization algorithms running on $(Q)$ must use the structure in the particular cost function involved. This is particularly significant since our obstructions apply to a broad class of lifts arising naturally in several applications.

6 Conclusions and future work

For the pair of problems $(Q)$ and $(P)$, we characterized the properties the lift $\varphi: \mathcal{M} \rightarrow \mathcal{X}$ needs to satisfy in order to map desirable points of $(Q)$ to desirable points of $(P)$. We noted that global minima for $(Q)$ always map to global minima for $(P)$ (Proposition 3.5), and showed that local minima for $(Q)$ map to local minima for $(P)$ if and only if $\varphi$ is open (Theorem 2.3). We also showed that 1-critical points for $(Q)$ map to stationary points for $(P)$ if and only if the differential of $\varphi$, viewed as a map from tangent spaces of $\mathcal{M}$ to tangent cones of $\mathcal{X}$, is surjective (Theorem 2.4). This requires the tangent cones of $\mathcal{X}$ to be linear spaces. We then characterized when 2-critical points for $(Q)$ map to stationary points for $(P)$, and gave two sufficient conditions and a necessary condition that may be easier to check for some examples (Theorem 3.23). We explained several techniques to compute all quantities involved in these conditions in Section 3.4.

Using our theory, we studied the above properties for a variety of lifts, including several lifts of low-rank matrices and tensors (Section 5) and the Burer–Monteiro lift for smooth SDPs (Corollary 4.12). We also proposed a systematic construction of lifts using fiber products that applies when $\mathcal{X}$ is the preimage of a smooth function (Section 4). We gave conditions under which it satisfies our desirable properties. In some cases, we can also obtain an expression for the tangent cones of $\mathcal{X}$ simultaneously with “$2 \Rightarrow 1$”, as explained in Section 3.5.

We end by listing several future directions suggested by this work.

(a) “$k \Rightarrow 1$” for general $k$: Several lifts of interest, notably tensor factorizations with more than two factors, do not satisfy “$2 \Rightarrow 1$”. It would therefore be interesting to characterize “$k \Rightarrow 1$” for general $k$, i.e., when do $k$-critical points for $(Q)$ map to stationary points for $(P)$
for any \( k \) times differentiable cost \( f \)? Do lifts that are multilinear in \( k \) arguments, such as order-\( k \) tensor lifts, satisfy “\( k \Rightarrow 1 \)”?

What can be said about “\( k \Rightarrow \ell \)” for \( \ell > 1 \)? Already for \( \ell = 2 \), the second-order optimality conditions on \( \mathcal{X} \) can be involved [47, Thm. 3.45]. On the positive side, if “\( 1 \Rightarrow 1 \)” holds at a preimage of a smooth point, then “\( k \Rightarrow k \)” holds there for all \( k \geq 1 \) by Proposition 2.11.

(b) **Robust “\( k \Rightarrow 1 \)”**: Algorithms run for finitely many iterations in practice, hence can only find approximate \( k \)-critical points for (Q). It is therefore important to characterize “robust” versions of “\( k \Rightarrow 1 \)”, guaranteeing that approximate \( k \)-critical points for (Q) map to approximate stationary points for (P). Note that if \( \mathcal{X} \) lacks regularity, care is needed when defining approximate stationarity for (P), see [37].

(c) **Obstructions to “local \( \Rightarrow \) local” and “\( k \Rightarrow 1 \)”**: For some sets \( \mathcal{X} \), we are not aware of any lifts which satisfy, say, “local \( \Rightarrow \) local” or “2 \( \Rightarrow 1 \)”.

(d) **Regularization on the lift**: It is common to modify (Q) by adding a regularizer to \( g = f \circ \varphi \), see [31, 53, 50]. For example, with the lift \((L, R) \mapsto LR^\top\), we may regularize (Q) by adding \( \frac{1}{2} \left( \|L\|_F^2 + \|R\|_F^2 \right) \), motivated by the fact that its minimum over a fiber \( \{(L, R) : LR^\top = X\} \) is the nuclear norm \( \|X\|_* \) [50]. Our framework does not directly apply in this case (because the regularizer may not be constant over fibers, hence may not factor through \( \varphi \)). Can it be extended to relate the landscape of the regularized (Q) to that of (P)?

(e) **Bypassing tangent cones via lifts**: To verify “\( 1 \Rightarrow 1 \)” and “\( 2 \Rightarrow 1 \)” on concrete examples of \( \mathcal{X} \) using the theory in this paper, we need to understand the tangent cones to \( \mathcal{X} \), which is often challenging. Many sets \( \mathcal{X} \) encountered in applications are only defined implicitly via a lift \( \varphi : \mathcal{M} \to \mathcal{E} \). Examples include the set of tensors admitting a certain type of factorization, the set of functions parametrized by a given neural network architecture, and the set of positions and orientations attainable by a robotic arm with a given joint configuration. Are there sufficient conditions for “\( k \Rightarrow 1 \)” that can be checked using \( \varphi \) and \( \mathcal{M} \) alone, without an explicit expression for the tangent cones to \( \mathcal{X} \)?

(f) **Dynamical systems on \( \mathcal{M} \) and their image on \( \mathcal{X} \)**: This paper is focused on comparing properties of points on \( \mathcal{M} \) and their images on \( \mathcal{X} \). In contrast, several applications are concerned with properties of entire trajectories of dynamical systems on \( \mathcal{M} \), and it may be interesting to compare these properties with their counterparts for the images of the trajectories on \( \mathcal{X} \). Examples of such comparisons include relating gradient flow on the weights of a neural network to gradient flow in function or measure spaces [7, 6, 27], and the “algorithmic equivalence” technique used in [3, 22, 40] to study mirror descent by showing that its continuous-time analogue is equivalent to gradient flow on a reparametrized problem.

**Acknowledgements**

We thank Christopher Criscitiello and Quentin Rebjock for numerous conversations and comments on drafts of this paper. NB was supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number MB22.00027.
References

[1] P. Ablin. Deep orthogonal linear networks are shallow. arXiv preprint arXiv:2011.13831, 2020.

[2] P.-A. Absil, R. Mahony, and R. Sepulchre. Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton, NJ, 2008.

[3] E. Amid and M. K. K. Warmuth. Reparameterizing mirror descent as gradient descent. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, Advances in Neural Information Processing Systems, volume 33, pages 8430–8439. Curran Associates, Inc., 2020.

[4] K. Antonakopoulos, P. Mertikopoulos, G. Piliouras, and X. Wang. Adagrad avoids saddle points. In International Conference on Machine Learning, pages 731–771. PMLR, 2022.

[5] N. Ay, J. Jost, H. Vân Lê, and L. Schwachhöfer. Information geometry, volume 64. Springer, 2017.

[6] F. Bach and L. Chizat. Gradient descent on infinitely wide neural networks: Global convergence and generalization. arXiv preprint arXiv:2110.08084, 2021.

[7] B. Bah, H. Rauhut, U. Terstiege, and M. Westdickenberg. Learning deep linear neural networks: Riemannian gradient flows and convergence to global minimizers. Information and Inference: A Journal of the IMA, 02 2021.

[8] N. Boumal. An introduction to optimization on smooth manifolds. Cambridge University Press, 2023.

[9] N. Boumal, V. Voroninski, and A. S. Bandeira. Deterministic guarantees for Burer–Monteiro factorizations of smooth semidefinite programs. Communications on Pure and Applied Mathematics, 73(3):581–608, 2020.

[10] S. Burer and R. D. Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. Mathematical Programming, 95(2):329–357, 2003.

[11] S. Burer and R. D. Monteiro. Local minima and convergence in low-rank semidefinite programming. Mathematical Programming, 103(3):427–444, 2005.

[12] C. Cartis, N. I. Gould, and P. L. Toint. Second-order optimality and beyond: Characterization and evaluation complexity in convexly constrained nonlinear optimization. Foundations of Computational Mathematics, 18:1073–1107, 2018.

[13] J. Chok and G. M. Vasil. Convex optimization over a probability simplex. arXiv preprint arXiv:2305.09046, 2023.

[14] A. Cichocki. Tensor networks for big data analytics and large-scale optimization problems. arXiv preprint arXiv:1407.3124, 2014.

[15] D. Cifuentes. On the Burer–Monteiro method for general semidefinite programs. Optimization Letters, pages 1–11, 2021.

[16] F. H. Clarke, Y. S. Ledyaev, R. J. Stern, and P. R. Wolenski. Nonsmooth Analysis and Control Theory, volume 178. Springer, 2008.
[17] D. Cox, J. Little, and D. O’Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer, 2013.

[18] F. Curtis, Z. Lubberts, and D. Robinson. Concise complexity analyses for trust region methods. *Optimization Letters*, 12(8):1713–1724, June 2018.

[19] F. Dellaert, D. M. Rosen, J. Wu, R. Mahony, and L. Carlone. Shonan rotation averaging: Global optimality by surfing \( SO(p)^n \). In A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, editors, *Computer Vision – ECCV 2020*, pages 292–308, Cham, 2020. Springer International Publishing.

[20] F. R. Deutsch. *Best Approximation in Inner Product Spaces*. Springer, 2012.

[21] A. Edelman, T. Arias, and S. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.

[22] U. Ghai, Z. Lu, and E. Hazan. Non-convex online learning via algorithmic equivalence. *arXiv preprint arXiv:2205.15235*, 2022.

[23] W. Ha, H. Liu, and R. Foygel Barber. An equivalence between critical points for rank constraints versus low-rank factorizations. *SIAM Journal on Optimization*, 30(4):2927–2955, Jan. 2020.

[24] J. Harris. *Algebraic Geometry: A First Course*. Springer, 1992.

[25] J.-B. Hiriart-Urruty and J. Malick. A fresh variational-analysis look at the positive semidefinite matrices world. *Journal of Optimization Theory and Applications*, 153(3):551–577, 2012.

[26] S. Hosseini, D. R. Luke, and A. Uschmajew. *Tangent and Normal Cones for Low-Rank Matrices*, pages 45–53. Springer International Publishing, Cham, 2019.

[27] A. Jacot, F. Gabriel, and C. Hongler. Neural tangent kernel: Convergence and generalization in neural networks. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[28] M. Journé, F. Bach, P.-A. Absil, and R. Sepulchre. Low-rank optimization on the cone of positive semidefinite matrices. *SIAM Journal on Optimization*, 20(5):2327–2351, 2010.

[29] V. Khrulkov and I. Oseledets. Desingularization of bounded-rank matrix sets. *SIAM Journal on Matrix Analysis and Applications*, 39(1):451–471, 2018.

[30] K. Kohn, T. Merkh, G. Montúfar, and M. Trager. Geometry of linear convolutional networks. *arXiv preprint arXiv:2108.01538*, 2021.

[31] C. Kolb, C. L. Müller, B. Bischl, and D. Rügamer. Smoothing the edges: A general framework for smooth optimization in sparse regularization using Hadamard overparametrization. *arXiv preprint arXiv:2307.03571*, 2023.

[32] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.

[33] J. Leake and N. K. Vishnoi. Optimization and sampling under continuous symmetry: Examples and Lie theory. *arXiv preprint arXiv:2109.01080*, 2021.
[34] J. D. Lee, I. Panageas, G. Piliouras, M. Simchowitz, M. I. Jordan, and B. Recht. First-order methods almost always avoid strict saddle points. *Mathematical programming*, 176(1):311–337, 2019.

[35] J. M. Lee. *Introduction to Smooth Manifolds*. Springer, 2012.

[36] E. Levin. Towards optimization on varieties. Undergraduate senior thesis, Princeton University, 2020.

[37] E. Levin, J. Kileel, and N. Boumal. Finding stationary points on bounded-rank matrices: a geometric hurdle and a smooth remedy. *Mathematical Programming*, 2022.

[38] M. Lezcano Casado. *Geometric optimisation on manifolds with applications to deep learning*. PhD thesis, University of Oxford, 2021.

[39] Q. Li, D. McKenzie, and W. Yin. From the simplex to the sphere: faster constrained optimization using the Hadamard parametrization. *Information and Inference: A Journal of the IMA*, 12(3):iaad017, 06 2023.

[40] Z. Li, T. Wang, J. D. Lee, and S. Arora. Implicit bias of gradient descent on reparametrized models: On equivalence to mirror descent. *arXiv preprint arXiv:2207.04036*, 2022.

[41] B. Mishra, G. Meyer, S. Bonnabel, and R. Sepulchre. Fixed-rank matrix factorizations and Riemannian low-rank optimization. *Computational Statistics*, 29(3-4):591–621, 2014.

[42] G. Olikier, K. A. Gallivan, and P.-A. Absil. An apocalypse-free first-order low-rank optimization algorithm. *arXiv preprint arXiv:2201.03962*, 2022.

[43] P. Petersen, M. Raslan, and F. Voigtlaender. Topological properties of the set of functions generated by neural networks of fixed size. *Foundations of Computational Mathematics*, 21(2):375–444, 04 2021.

[44] A.-H. Phan, M. Yamagishi, D. Mandic, and A. Cichocki. Quadratic programming over ellipsoids with applications to constrained linear regression and tensor decomposition. *Neural Computing and Applications*, 32(11):7097–7120, 2019.

[45] E. Robeva. Orthogonal decomposition of symmetric tensors. *SIAM Journal on Matrix Analysis and Applications*, 37(1):86–102, 2016.

[46] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317. Springer, 2009.

[47] A. Ruszczyński. *Nonlinear Optimization*. Princeton University Press, 2006.

[48] R. Schneider and A. Uschmajew. Convergence results for projected line-search methods on varieties of low-rank matrices via Łojasiewicz inequality. *SIAM Journal on Optimization*, 25(1):622–646, 2015.

[49] B. Siciliano, L. Sciavicco, L. Villani, and G. Oriolo. *Robotics: Modelling, Planning and Control*. Springer, 2009.

[50] N. Srebro, J. Rennie, and T. Jaakkola. Maximum-margin matrix factorization. In L. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems*, volume 17. MIT Press, 2004.
A Lifts preserving local minima

We characterize the lifts that map local minima of (Q) to local minima of (P). To this end, we introduce a number of properties related to preservation of local minima and then prove that they are all equivalent. Recall that $\overline{S}$ is our notation for the closure of a set $S$.

**Definition A.1.** Let $\varphi : \mathcal{M} \to \mathcal{X}$ be a continuous, surjective map from a topological space $\mathcal{M}$ to a metric space $\mathcal{X}$ with distance $\text{dist}$, and let $x = \varphi(y)$.

1. $\varphi$ is open at $y$ if $\varphi(U)$ is a neighborhood of $x$ in $\mathcal{X}$ for all neighborhoods $U$ of $y$ in $\mathcal{M}$.
2. $\varphi$ is approximately open at $y$ if $\overline{\varphi(U)}$ is a neighborhood of $x$ in $\mathcal{X}$ for all neighborhoods $U$ of $y$ in $\mathcal{M}$.
3. $\varphi$ satisfies the Subsequence Lifting Property (SLP) at $y$ if for every sequence $(x_i)_{i \geq 1} \subseteq \mathcal{X}$ converging to $x$ there exists a subsequence indexed by $(i_j)_{j \geq 1}$ and a sequence $(y_{i_j})_{j \geq 1} \subseteq \mathcal{M}$ converging to $y$ such that $\varphi(y_{i_j}) = x_{i_j}$ for all $j \geq 1$.
4. $\varphi$ satisfies the Approximate Subsequence Lifting Property (ASLP) at $y$ if for every sequence $(x_i)_{i \geq 1} \subseteq \mathcal{X}$ converging to $x$ and every sequence $(\epsilon_i)_{i \geq 1} \subseteq \mathbb{R}_{>0}$ converging to 0 there exists a subsequence indexed by $(i_j)_{j \geq 1}$ and a sequence $(y_{i_j})_{j \geq 1} \subseteq \mathcal{M}$ converging to $y$ such that $\text{dist}(\varphi(y_{i_j}), x_{i_j}) \leq \epsilon_{i_j}$ for all $j \geq 1$.

**Theorem A.2.** If $\mathcal{M}$ is Hausdorff, second-countable and locally compact (all of which hold if $\mathcal{M}$ is a topological manifold), then the four properties of $\varphi$ at $y \in \mathcal{M}$ in Definition A.1 are equivalent to each other and to the “local $\Rightarrow$ local” property at $y$ (Definition 2.2(a)).
Proof. We show that ASLP $\implies$ approximate openness $\implies$ openness $\implies$ SLP $\implies$ “local $\implies$ local” $\implies$ ASLP.

Suppose $\varphi$ satisfies ASLP at $y$. Suppose there exists a neighborhood $U$ of $y$ such that $\varphi(U)$ is not a neighborhood of $x = \varphi(y)$. Then we can find a sequence $(x_i)_{i \geq 1} \subseteq X$ such that $x_i \to x$ but $x_i \notin \varphi(U)$ for all $i$. Set $\epsilon_i = \frac{1}{2} \text{dist}(x_i, \varphi(U)) > 0$ and apply ASLP to find a sequence $(y_j)_{j \geq 1} \subseteq M$ such that $y_j \to y$ and $\text{dist}(\varphi(y_j), x_i) \leq \epsilon_i$. Because $\text{dist}(\varphi(y_j), x_i) < \text{dist}(x_i, \varphi(U))$, we have $\varphi(y_j) \notin \varphi(U)$ for all $i$. However, because $U$ is a neighborhood of $y$ and $y_j \to y$, we must have $y_j \in U$ for all large $i$, a contradiction. Thus, $\varphi(U)$ is a neighborhood of $x$, so $\varphi$ is approximately open at $y$.

Suppose $\varphi$ is approximately open at $y$, and let $U$ be a neighborhood of $y$ in $M$. Because $M$ is locally compact, we can find a compact neighborhood $V \subseteq U$ of $x$. Since $\varphi$ is continuous and $V$ is compact, we have that $\varphi(V)$ is compact; since $X$ is Hausdorff (it is a metric space), it follows that $\varphi(V)$ is closed. Combining with the fact that $\varphi$ is approximately open at $y$, we deduce that $\varphi(V)$ is a neighborhood of $x$. Since $\varphi(U) \supseteq \varphi(V)$, we conclude that $\varphi(U)$ is a neighborhood of $x$ as well. Thus, $\varphi$ is open at $y$.

Suppose $\varphi$ is open at $y$, and $(x_j)_{j \geq 1} \subseteq X$ converges to $x = \varphi(y)$. Owing to the topological properties of $M$, there is a sequence of open neighborhoods $U_i$ of $y$ with compact closures such that $U_i \supseteq \overline{U_{i+1}}$ and $\bigcap_{i=1}^{\infty} U_i = \{y\}$, see Lemma A.3 following this proof. Because $\varphi$ is open, each $\varphi(U_i)$ is an open neighborhood of $x$ such that $\varphi(U_i) \supseteq \varphi(U_{i+1})$ and $x \in \bigcap_{i=1}^{\infty} \varphi(U_i)$. Moreover, because $\varphi(U_i)$ is a neighborhood of $x$ and $x_j \to x$, there exists index $J(i)$ such that $x_j \in \varphi(U_i)$ for all $j \geq J(i)$. After passing to a subsequence of $(x_j)$, we may assume $x_j \in \varphi(U_i)$ and pick $y_j \in U_j$ satisfying $x_j = \varphi(y_j)$. Because $(y_j)$ is an infinite sequence contained in the compact set $\overline{U_1}$, after passing to a subsequence again we may assume that $\lim_j y_j$ exists. With $i$ arbitrary, we have for all $j > i$ that $y_j \in U_j \subseteq U_{i+1}$, hence that $\lim_j y_j \in \overline{U_{i+1}} \subseteq U_i$. This holds for all $i$, hence $\lim_j y_j \in \bigcup_i U_i = \{y\}$. Thus, $y = \lim_i y_i$ and $\varphi(y_i) = x_i$, so $\varphi$ satisfies SLP.

Suppose $\varphi$ satisfies SLP at $y$. Let $f : X \to \mathbb{R}$ be a cost function on $X$ and $g = f \circ \varphi$. Suppose $x = \varphi(y)$ is not a local minimum for $f$ on $X$, that is, there exists a sequence $(x_i)_{i \geq 1} \subseteq X$ converging to $x$ such that $f(x_i) < f(x)$ for all $i$. Applying SLP, after passing to a subsequence we can find a sequence $(y_i)_{i \geq 1} \subseteq M$ converging to $y$ such that $\varphi(y_i) = x_i$. Since $g(y_i) = f(x_i) < f(x) = g(y)$ and $y_i \to y$, we conclude that $y$ is not a local minimum for $g$. By contrapositive, this shows that $\varphi$ satisfies the “local $\Rightarrow$ local” property at $y$.

For the last implication, we proceed by contrapositive once again. Suppose $\varphi$ does not satisfy ASLP at $y$. Then, we can find sequences $(x_i)_{i \geq 1} \subseteq X$ converging to $x$ and $(\epsilon_i)_{i \geq 1} \subseteq \mathbb{R}_{>0}$ converging to $0$ such that no subsequence of $(x_i)$ can be approximately lifted to $M$ in the sense of ASLP. Let $\tilde{B}(x, \epsilon) = \{x' \in X : \text{dist}(x, x') \leq \epsilon\}$. Notice that $x = \varphi(y) \notin \tilde{B}(x_i, \epsilon_i)$ for all but finitely many indices $i$, as otherwise the constant sequence $y_i \equiv y$ would give an approximate lift of a subsequence. Since $x_i \to x$ and $\epsilon_i \to 0$, after passing to a subsequence we may assume that the closed balls $\tilde{B}(x_i, \epsilon_i)$ are pairwise disjoint and none contain $x$. Define the following sum of smooth bump functions centered at the $x_i$

$$f(x') = \begin{cases} \exp \left( \frac{1}{1-(\text{dist}(x_i, x')/\epsilon_i)^2} \right) & \text{if } x' \in \tilde{B}(x_i, \epsilon_i) \text{ for some } i, \\ 0 & \text{otherwise.} \end{cases}$$

This is well defined because the balls $\tilde{B}(x_i, \epsilon_i)$ are disjoint. (As a side note, we remark that if $X$ is a metric subspace of a Euclidean space $\mathcal{E}$ as in our general treatment, then $f$ extends to a smooth function on $\mathcal{E}$.) Note that $x$ is not a local minimum for $f$ since $x_i \to x$ and $f(x_i) = -1 < 0 = f(x)$. However, $y$ is a local minimum for $g = f \circ \varphi$. Indeed, if there was a sequence $(y_i)$ converging to $y$ such that $g(y_i) < g(y) = 0$, then we would have $\varphi(y_i) \in \tilde{B}(x_{n_i}, \epsilon_{n_i})$ for an infinite subsequence $(n_i)$ with $n_i \to \infty$ (since we must have $\varphi(y_i) \to x$ by continuity of $\varphi$), showing that $(y_i)$ is an
approximate lift of the subsequence \((x_{n_i})\): a contradiction to our assumptions about \((x_i), (\epsilon_i)\).

Thus, \(\varphi\) does not satisfy the “local \(\Rightarrow\) local” property at \(y\).

\[\square\]

**Lemma A.3.** Suppose \(\mathcal{M}\) is Hausdorff, second-countable, and locally compact. Then for any \(y \in \mathcal{M}\) there is a sequence of open neighborhoods \(U_i\) of \(y\) with compact closures such that \(U_i \supseteq \overline{U}_{i+1}\) and \(\bigcap_{i=1}^{\infty} U_i = \{y\}\).

**Proof.** Because \(\mathcal{M}\) is second-countable and locally compact, we can find a countable basis of open neighborhoods with compact closures \(\{V_j\}_{j \geq 1}\) for \(y\). Since \(\mathcal{M}\) is Hausdorff and \(\{V_j\}\) is a basis for \(y\), we have \(\bigcap_{j=1}^{\infty} V_j = \{y\}\). Indeed, if \(y' \neq y\), then there exists a neighborhood of \(y\) not containing \(y'\), and this neighborhood contains \(V_i\) for some \(i\) by definition of a local basis. By replacing \(V_i\) by \(\bigcap_{j=1}^{i} V_j\) (which preserves their intersection), we may assume \(V_i \supseteq V_{i+1}\). We construct \(\{U_i\}_{i \geq 1}\) inductively. Set \(U_1 = V_1\), which is an open neighborhood of \(y\) by assumption. Having constructed \(U_1, \ldots, U_i\), use local compactness to find a compact neighborhood \(K_{i+1} \subseteq V_{i+1} \cap U_i\) of \(y\) and let \(U_{i+1}\) be the interior of \(K_{i+1}\). Then \(U_{i+1}\) is an open neighborhood of \(y\) by construction, and \(\overline{U}_{i+1} \subseteq K_{i+1} \subseteq U_i\) which also shows \(\overline{U}_{i+1}\) is compact as a closed subset of the compact set \(K_{i+1}\). Finally, we have \(\{y\} \subseteq \bigcap_{i=1}^{\infty} U_i \subseteq \bigcap_{i=1}^{\infty} V_i = \{y\}\) hence \(\bigcap_{i=1}^{\infty} U_i = \{y\}\). \[\square\]

**B Basic properties of \(A_y\) and \(B_y\)**

**Proof of Proposition 3.21.** (a) For \(w \in A_y\), let \(c: \mathbb{R} \to \mathcal{M}\) satisfy \(c(0) = y\), \((\varphi \circ c)'(0) = 0\) and \((\varphi \circ c)'(0) = w\). For any \(\alpha \geq 0\) define \(\varphi(t) = c(\sqrt{\alpha t})\). Note that \(\varphi(0) = y\), \((\varphi \circ c)'(0) = \sqrt{\alpha(\varphi \circ c)'(0)} = 0\) and \((\varphi \circ c)'(0) \in A_y\). Thus, \(\alpha w \in A_y\) so \(A_y\) is a cone. The proof that \(B_y\) is a cone is analogous.

Suppose \(w_j \in B_y\) and \(w_j \to w\). Let \(c_{j,i}: \mathbb{R} \to \mathcal{M}\) satisfy \((\varphi \circ c_{j,i})'(0) \xrightarrow{i \to \infty} 0\) and \((\varphi \circ c_{j,i})'(0) \xrightarrow{i \to \infty} w_j\). For each \(j\), pick \(i_j\) large enough so that \(\|((\varphi \circ c_{j,i})'(0)) \leq 1/j\) and \(\|((\varphi \circ c_{j,i})'(0) - w_j\| \leq 1/j\). Set \(\tilde{c}_j = c_{j,i_j}\) and observe that \((\varphi \circ \tilde{c}_j)'(0) \to 0\) and \((\varphi \circ \tilde{c}_j)'(0) \to w\), showing \(w \in B_y\) by construction.

(b) We prove the last claim in part (b). Since \(0 \in \text{im} L_y\), we trivially have \(A_y + \text{im} L_y \supseteq A_y\) and similarly for \(B_y\). Conversely, if \(w \in A_y\), let \(c: \mathbb{R} \to \mathcal{M}\) satisfy \(c(0) = y\), \((\varphi \circ c)'(0) = 0\) and \((\varphi \circ c)'(0) = w\). By Lemma 3.24(b), for any \(v \in T_y \mathcal{M}\) there exists a curve \(\tilde{c}: \mathbb{R} \to \mathcal{M}\) satisfying \(\tilde{c}(0) = y\), \(\tilde{c}'(0) = c'(0)\), and \(\tilde{c}''(0) = c''(0) + v\), the sum being taken inside \(T_y \mathcal{M}\). Then

\[((\varphi \circ \tilde{c})'(0) = D\varphi(y)[\tilde{c}'(0)] = D\varphi(y)[c'(0)] = (\varphi \circ c)'(0) = 0,\]

and by Lemma 3.24(a), \((\varphi \circ \tilde{c})''(0) = (\varphi \circ c)''(0) + L_y(v) = w + L_y(v)\). This shows \(A_y + \text{im} L_y \subseteq A_y\). The proof that \(B_y + \text{im} L_y \subseteq B_y\) is analogous.

The proofs of the first half of part (b) and parts (c)-(d) are given in the main body. \[\square\]

To see that \(B_y \not\subseteq T_x \mathcal{X}\) in general, consider \(B_y\) for the lift \(\varphi: \mathcal{M} = \mathbb{R} \times (\mathbb{R}^2)^2 \to \mathcal{X} = \mathbb{R}^{2 \times 2}_{\leq 1}\) given by \(\varphi(\lambda, u, v) = \lambda uv^\top\) at \(y = (0, e_1, e_1)\) where \(e_1 = [1, 0]^\top\). We remark that “2 \(\Rightarrow 1\)” does not hold for this lift by Proposition 5.2. To see that \(T_x \mathcal{X} \not\subseteq B_y\) in general, note that if \(T_x \mathcal{X} \not\subseteq B_y\) then \(B_y^* \subseteq (T_x \mathcal{X})^*\) and hence “2 \(\Rightarrow 1\)” holds at \(y\) by Corollary 3.22(b). In particular, we have \(T_x \mathcal{X} \not\subseteq B_y\) for the above example as well.

45
C Compositions and products of lifts

C.1 Compositions

Proof of Proposition 3.27. Recall that “local ⇒ local” is equivalent to openness by Theorem 2.3.

(a) For the first statement, suppose \( \varphi \circ \psi \) is open at \( z \) and let \( V \subseteq M \) be a neighborhood of \( y \).
Because \( \psi \) is continuous, \( \psi^{-1}(V) \) is a neighborhood of \( z \), so \((\varphi \circ \psi)(\psi^{-1}(V))\) is a neighborhood of \( \varphi(y) \), hence \( \varphi(V) \supseteq (\varphi \circ \psi)(\psi^{-1}(V)) \) is a neighborhood of \( \varphi(y) \) as well. This shows \( \varphi \) is open at \( y \), and hence satisfies “local ⇒ local” there by Theorem 2.3.

For the second statement, if \( U \subseteq N \) is a neighborhood of \( z \), then \( \psi(U) \subseteq M \) is a neighborhood of \( y \) since \( \psi \) is open at \( z \). If \( \varphi \) satisfies “local ⇒ local”, or equivalently, is open at \( y \), then \( \varphi(\psi(U)) \subseteq X \) is a neighborhood of \( \varphi(y) \), hence \( \varphi \circ \psi \) is open at \( z \) and satisfies “local ⇒ local” there by Theorem 2.3.

(b) For “1 ⇒ 1”, the chain rule gives \( D(\varphi \circ \psi)(z) = D\varphi(y) \circ D\psi(z) \) for any \( z \in N \). If \( \varphi \circ \psi \) satisfies “1 ⇒ 1”, then Theorem 2.4 shows that \( \text{im } D(\varphi \circ \psi)(z) = T_xX \), hence \( \text{im } D\varphi(y) \supseteq T_xX \). Since \( \text{im } D\varphi(y) \subseteq T_xX \) by Lemma 3.12, we conclude that \( \text{im } D\varphi(y) = T_xX \) and hence \( \varphi \) satisfies “1 ⇒ 1” at \( y \) by Theorem 2.4. Conversely, suppose \( \psi \) is a submersion at \( z \), so \( D\psi(z) \) is surjective. Then \( \text{im } D\varphi(y) = \text{im } D(\varphi \circ \psi)(z) \). By Theorem 2.4, we conclude that \( \varphi \) satisfies “1 ⇒ 1" at \( y \) if and only if \( \varphi \circ \psi \) does so at \( z \).

For “2 ⇒ 1”, suppose first that \( \varphi \circ \psi \) satisfies “2 ⇒ 1" at \( z \) and \( y \) is 2-critical for \( f \circ \varphi \) where \( f \) is some cost function on \( X \). For any curve \( \tilde{c} : R \to N \) such that \( \tilde{c}(0) = z \), the curve \( c = \psi \circ \tilde{c} : R \to M \) satisfies \( c(0) = y \). Because \( y \) is 2-critical, we have

\[
(f \circ \varphi \circ c)'(0) = (f \circ \varphi \circ \psi \circ \tilde{c})'(0) = 0, \quad \text{and } (f \circ \varphi \circ c)''(0) = (f \circ \varphi \circ \psi \circ \tilde{c})''(0) \geq 0.
\]

This shows \( z \) is 2-critical for \( f \circ \varphi \circ \psi \) on \( N \). Because \( \varphi \circ \psi \) satisfies “2 ⇒ 1”, we conclude that \( \varphi(\psi(z)) = \varphi(y) \) is stationary for \( f \) on \( X \), hence \( \varphi \) satisfies “2 ⇒ 1” at \( y \).

Conversely, suppose \( \psi \) is a submersion at \( z \) and \( \varphi \) satisfies “2 ⇒ 1" at \( y \). Suppose \( z \) is 2-critical for \( f \circ \varphi \circ \psi \). Because \( \psi \) is a submersion at \( z \), for any curve \( c : I \to M \) satisfying \( c(0) = y \), there exists a potentially smaller interval \( I' \subseteq R \) containing \( t = 0 \) in its interior and a curve \( \tilde{c} : I' \to N \) such that \( \psi \circ \tilde{c} = c \). For example, by [35, Thm. 4.26] there exists a smooth local section \( \sigma \) of \( \psi \) defined near \( y \) satisfying \( \sigma(y) = z \), in which case we can set \( \tilde{c} = \sigma \circ c \). Because \( z \) is 2-critical, it follows that \( y \) is 2-critical for \( f \circ \varphi \), by (C.1.1). Since \( \varphi \) satisfies “2 ⇒ 1" at \( y \), we conclude that \( \varphi(y) = \varphi(\psi(z)) \) is stationary for \( f \), which shows \( \varphi \circ \psi \) satisfies “2 ⇒ 1" at \( z \).

(c) The first claimed equality is just the chain rule \( D(\varphi \circ \psi)(z) = D\varphi(y) \circ D\psi(z) \). For the second claimed equality, let \( v = L^\varphi_z(u) \in T_yM \). Recall that \( Q^\varphi \circ \psi \)(u) = (\varphi \circ \psi \circ \tilde{c}_u)'(0) \) where \( \tilde{c}_u : R \to N \) is some curve satisfying \( \tilde{c}_u(0) = z \) and \( \tilde{c}'_u(0) = u \), and \( Q^\varphi \psi(v) = (\varphi \circ c_v)'(0) \) for some curve \( c_v \) on \( M \) satisfying \( c_v(0) = y \) and \( c'_v(0) = v \). Since \( \psi \circ \tilde{c}_u \) is another curve on \( M \) satisfying \( (\psi \circ \tilde{c}_u)(0) = y \) and \( (\psi \circ \tilde{c}_u)'(0) = L^\varphi_z(u) = v \), Lemma 3.24 shows that \( Q^\varphi \psi(L^\varphi_z(u)) - Q^\varphi \circ \psi(u) \in \text{im } L^\varphi_y = \text{im } L^\varphi \psi \) for all \( u \in T_yN \) as claimed, where we used the fact that \( L^\varphi \circ \psi = L^\varphi_y \circ L^\varphi_z \) and the surjectivity of \( L^\varphi_z \).
Since $L^\psi_z = L^\psi_y \circ L^\psi_x$, we have $\ker L^\psi_z = (L^\psi_y)^{-1}(\ker L^\psi_y)$. Because $L^\psi_y$ is surjective, we have $L^\psi_z((L^\psi_y)^{-1}(\ker L^\psi_y)) = \ker L^\psi_y$. Using the results of the preceding paragraph,

$$A^\psi_z = Q^\psi_z(\ker L^\psi_z) + \im L^\psi_z = Q^\psi_y(\ker L^\psi_y) + \im L^\psi_y = \im L^\psi_y = A^\psi_y.$$

If $w \in B^\varphi_y$ then there exist $v_i \in T_y M$ such that $L^\varphi_y(v_i) \to 0$ and $w \in \lim_i(Q^\varphi_y(v_i) + \im L^\varphi_y)$. Since $L^\varphi_x$ is surjective, there exist $u_i \in T_z N$ satisfying $L^\varphi_x(u_i) = v_i$. We then have $L^\varphi_z(u_i) = L^\varphi_y(v_i) \to 0$, and $Q^\varphi_z(u_i) + \im L^\varphi_z = Q^\varphi_y(v_i) + \im L^\varphi_y$. Taking $i \to \infty$, we conclude that $w \in B^\varphi_z$, and $B^\varphi_y \subseteq B^\varphi_z$.

Conversely, if $w \in B^\varphi_z$ then there exist $u_i \in T_z N$ such that $L^\varphi_z(u_i) \to 0$ and $w \in \lim_i(Q^\varphi_z(u_i) + \im L^\varphi_z)$. Let $v_i = L^\varphi_y(u_i)$ to conclude that $w \in B^\varphi_y$, and hence $B^\varphi_z = B^\varphi_y$.

The argument for part (b) shows that $y$ is 2-critical for $f \circ \varphi$ if and only if $z$ is 2-critical for $f \circ \varphi \circ \psi$, which implies $W^\varphi_y = W^\varphi_z$ by Definition 3.15.

\[\square\]

### C.2 Products

**Proof of Proposition 4.11.** (a) If sequences $((y^{(j)}_1, \ldots, y^{(j)}_k))_{j \geq 1} \subseteq Z$ and $\tau_j \to 0$ such that $\tau_j > 0$ satisfy

$$(v_1, \ldots, v_k) = \lim_{j \to \infty} \frac{(y^{(j)}_1, \ldots, y^{(j)}_k) - (y_1, \ldots, y_k)}{\tau_j},$$

then $v_i = \lim_{j \to \infty} \frac{y^{(j)}_i - y_i}{\tau_j}$ for all $i = 1, \ldots, k$, which shows that if $(v_1, \ldots, v_k) \in T_{(z_1, \ldots, z_k)}Z$ then $v_i \in T_{z_i}Z$ for all $i$. For an example where the inclusion is strict, see [46, Prop. 6.41].

(b) Since products of open sets form a basis for the (product) topology on $Z$, we conclude that $\psi$ is open at $y = (y_1, \ldots, y_k)$ iff $\psi(U_1 \times \cdots \times U_k) = \psi_1(U_1) \times \cdots \times \psi_k(U_k)$ is a neighborhood of $y$ whenever $U_i \subseteq M_i$ is a neighborhood of $y_i$ for all $i$. Because the interior of a product of sets in the product topology is the product of their interiors, we conclude that $\psi_1(U_1) \times \cdots \times \psi_k(U_k)$ is a neighborhood of $y$ if and only if $\psi_i(U_i)$ is a neighborhood of $y_i$ for all $i$. This proves part (b) by Theorem 2.3.

(c) We have $L^\psi_{(y_1, \ldots, y_k)} = L^\psi_{y_1} \times \cdots \times L^\psi_{y_k}$, which is defined on $T_{(y_1, \ldots, y_k)}N = T_{y_1}N_1 \times \cdots \times T_{y_k}N_k$.

Therefore, $\im L^\psi_{(y_1, \ldots, y_k)} = \im L^\psi_{y_1} \times \cdots \times \im L^\psi_{y_k}$. If $\psi_i$ satisfies $\{1 \Rightarrow 1\}$ at $y_i$ for all $i$, then $\im L^\psi_{y_i} = T_{z_i}Z_i$ for all $i$ by Theorem 2.4, so $\im L^\psi_{(y_1, \ldots, y_k)} = T_{z_1}Z_1 \times \cdots \times T_{z_k}Z_k$ where $z_i = \psi_i(y_i)$. Since $\im L^\psi_{(y_1, \ldots, y_k)}$ is always contained in $T_{(z_1, \ldots, z_k)}Z$ by Proposition 3.21(b), we get the chain of inclusions

$$T_{z_1}Z_1 \times \cdots \times T_{z_k}Z_k = \im L^\psi_{(y_1, \ldots, y_k)} \subseteq T_{(z_1, \ldots, z_k)}Z \subseteq T_{z_1}Z_1 \times \cdots \times T_{z_k}Z_k,$$

where the last inclusion is part (a). We conclude that equality in part (a) holds and $\im L^\psi_{(y_1, \ldots, y_k)} = T_{(z_1, \ldots, z_k)}Z$ so $\{1 \Rightarrow 1\}$ holds.
(d) For any \(v = (v_1, \ldots, v_k) \in T_{(y_1, \ldots, y_k)} M\), let \(c_{v_i}\) be a smooth curve on \(M\) satisfying \(Q_{y_i}^\psi (v_i) = (\psi_i \circ c_{v_i})''(0)\) as in Definition 3.9. Then \(c(t) = (c_{v_1}(t), \ldots, c_{v_k}(t))\) is a smooth curve on \(M\) passing through \((y_1, \ldots, y_k)\) with velocity \(v\) and

\[
(\psi \circ c)''(0) = ((\psi_1 \circ c_{v_1})''(0), \ldots, (\psi_k \circ c_{v_k})''(0)) = \left( Q_{y_1}^\psi \times \cdots \times Q_{y_k}^\psi \right)(v),
\]

showing that \(Q_{(y_1, \ldots, y_k)}^\psi \equiv Q_{y_1}^\psi \times \cdots \times Q_{y_k}^\psi \mod \text{im}(L_{(y_1, \ldots, y_k)}^\psi)\) by Lemma 3.24. By part (c) above, we also have

\[
\ker L_{(y_1, \ldots, y_k)}^\psi = \ker(L_{y_1}^\psi \times \cdots \times L_{y_k}^\psi) = \ker L_{y_1}^\psi \times \cdots \times \ker L_{y_k}^\psi,
\]

so Proposition 3.26(a) gives

\[
A_{y_1, \ldots, y_k}^\psi = Q_{y_1, \ldots, y_k}^\psi (\ker L_{y_1, \ldots, y_k}^\psi) + \text{im} L_{y_1, \ldots, y_k}^\psi = Q_{y_1}^\psi (\ker L_{y_1}^\psi) \times \cdots \times Q_{y_k}^\psi (\ker L_{y_k}^\psi) + \text{im} L_{y_1}^\psi \times \cdots \times \text{im} L_{y_k}^\psi = \left( Q_{y_1}^\psi (\ker L_{y_1}^\psi) + \text{im} L_{y_1}^\psi \right) \times \cdots \times \left( Q_{y_k}^\psi (\ker L_{y_k}^\psi) + \text{im} L_{y_k}^\psi \right) = A_{y_1}^\psi \times \cdots \times A_{y_k}^\psi.
\]

The proof for \(B_{(y_1, \ldots, y_k)}\) similarly follows from Proposition 3.26(b). Now recall Definition 3.10. For any \((w_1, \ldots, w_k) \in \mathcal{E}_1 \times \cdots \times \mathcal{E}_k\) we have

\[
\psi_{(w_1, \ldots, w_k)}(y_1, \ldots, y_k) = \langle (w_1, \ldots, w_k), \psi(y_1, \ldots, y_k) \rangle = \sum_{i=1}^k \langle w_i, \psi_i(y_i) \rangle = \sum_{i=1}^k (\psi_i)_{w_i}(y_i),
\]

where the first and last equalities are the definitions of \(\psi_{(w_1, \ldots, w_k)}(y_1, \ldots, y_k), (\psi_i)_{w_i}(y_i)\) from Definition 3.10. Therefore,

\[
\nabla^2 \psi_{(w_1, \ldots, w_k)}(y_1, \ldots, y_k)[\tilde{y}_1, \ldots, \tilde{y}_k] = \left( \nabla^2 (\psi_1)_{w_1}(y_1)[\tilde{y}_1], \ldots, \nabla^2 (\psi_k)_{w_k}(y_k)[\tilde{y}_k] \right).
\]

The claimed expression for \(W_{(y_1, \ldots, y_k)}^\psi\) follows from Proposition 3.26(c) together with (C.2.1).

Taking duals in part (a), we get \((T_{(z_1, \ldots, z_k)}Z)^* \supseteq (T_{z_1}Z_1)^* \times \cdots \times (T_{z_k}Z_k)^*\) where we used the fact that the dual of a product of cones is the product of their duals. If \(\psi_i\) satisfies “2 \Rightarrow 1” for all \(i\), then \(W_{y_i}^\psi \subseteq (T_{z_i}Z_i)^*\) by Theorem 3.16, in which case

\[
W_{(y_1, \ldots, y_k)}^\psi = W_{y_1}^\psi \times \cdots \times W_{y_k}^\psi \subseteq (T_{z_1}Z_1)^* \times \cdots \times (T_{z_k}Z_k)^* \subseteq (T_{(z_1, \ldots, z_k)}Z)^*.
\]

Thus, \(\psi\) satisfies “2 \Rightarrow 1” at \((y_1, \ldots, y_k) \in \mathcal{N}\).

(e) This follows from (a) and (d). 

\[\square\]

D Proof of Proposition 2.9 (SVD lifts)

We first analyze the SVD lift, and then its modification.
SVD lift: Considering the lift (SVD). Fix \((U, \sigma, V) \in \mathcal{M}\) such that the entries of \(|\sigma|\) are all nonzero but not distinct. Choose \(k < \ell\) such that \(|\sigma_k| = |\sigma_\ell|\) and let \(X = \varphi(U, \sigma, V) = U\text{diag}(\sigma)V^\top\). We show “local ⇒ local” does not hold at \((U, \sigma, V)\) by constructing a sequence converging to \(X\) such that no subsequence of it can be lifted to a sequence converging to \((U, \sigma, V)\). Choose \(\alpha_i \in \mathbb{R}^r\), converging to zero such that \(|\sigma + \alpha_i|\) has distinct entries which are all nonzero. Let \(Q^{(k, \ell)} \in O(r)\) be the Givens rotation matrix rotating by \(\pi/4\) in \((k, \ell)\) plane, given explicitly by

\[
Q_{r,s}^{(k, \ell)} = \begin{cases} 
1/\sqrt{2} & \text{if } r = s = k \text{ or } r = s = \ell, \\
1/\sqrt{2} & \text{if } r = \ell, s = k, \\
-1/\sqrt{2} & \text{if } r = k, s = \ell, \\
1 & \text{if } r = s \notin \{k, \ell\}, \\
0 & \text{otherwise}, 
\end{cases}
\]

and let \(U^{(k, \ell)} = U\text{sign}(|\sigma|)Q^{(k, \ell)}\text{sign}(\sigma) \in \text{St}(m, r)\) and \(V^{(k, \ell)} = VQ^{(k, \ell)} \in \text{St}(n, r)\). Define

\[
X_i = \varphi(U^{(k, \ell)}, \sigma + \alpha_i, V^{(k, \ell)}) \in \mathcal{M},
\]

whose limit is

\[
\lim_{i \to \infty} X_i = U^{(k, \ell)}\text{diag}(\sigma)(V^{(k, \ell)})^\top = U\text{sign}(\sigma)\left[Q^{(k, \ell)}\text{diag}(|\sigma|)(Q^{(k, \ell)})^\top\right]V^\top = U\text{diag}(\sigma)V^\top = X.
\]

The third equality above follows since \(|\sigma| = |\sigma_\ell|\) so the corresponding submatrix of \(\text{diag}(\sigma)\) is a multiple of the identity, and \(Q^{(k, \ell)}\) is orthogonal and acts by the identity outside of that submatrix. Suppose \((U_{ij}, \sigma_{ij}, V_{ij})\) is a lift of some subsequence of \((X_{ij})\) converging to \((U, \sigma, V)\). The singular values of \(X_{ij}\) are the entries of \(|\sigma_{ij}|\), which are also the entries of \(|\sigma + \alpha_{ij}|\). Since all these singular values are distinct, \(U_{ij}\) and \(V_{ij}\) must contain the \(k\)th and \(\ell\)th columns of \(U^{(k, \ell)}\) and \(V^{(k, \ell)}\) up to sign, since the singular vectors are unique up to sign [52, Thm. 4.1]. But then it cannot happen that \((U_{ij}, V_{ij}) \to (U, V)\) by construction of \(Q^{(k, \ell)}\), a contradiction. Thus, no subsequence of \((X_{ij})\) can be lifted to a sequence on \(\mathcal{M}\) converging to \((U, \sigma, V)\), showing that “local ⇒ local” does not hold there. Proposition 2.11 together with the fact that \(X^\text{smth} = \mathbb{R}^m_{r \times n}\) implies that “\(1 \Rightarrow 1\)” does not hold at such \((U, \sigma, V)\) either.

Next, fix \((U, \sigma, V) \in \mathcal{M}\) such that \(\sigma_k = 0\) for some \(k\), and let \(X = \varphi(U, \sigma, V)\). Since \(\text{rank}(X) < r < \min(m, n)\), there exist unit vectors \(u_k' \in \mathbb{R}^m\) and \(v_k' \in \mathbb{R}^n\) such that \(U^\top u_k' = 0\) and \(V^\top v_k' = 0\). Let \(U^{(k)}, V^{(k)}\) be obtained from \(U, V\) by replacing their \(k\)th columns with \(u_k', v_k'\), respectively, and let \(\alpha_i \in \mathbb{R}^r\) be a sequence converging to zero such that \(|\sigma + \alpha_i|\) has distinct entries which are all nonzero. Define \(X_i = \varphi(U^{(k)}, \sigma + \alpha_i, V^{(k)})\) which converge to \(X\) as \(i \to \infty\). If \((U_{ij}, \sigma_{ij}, V_{ij})\) is a lift of a subsequence of \((X_i)\) converging to \((U, \sigma, V)\), then one of the columns of \(U_{ij}, V_{ij}\) must be \(u_k', v_k'\), up to sign because \(|\sigma_{ij}|\) has distinct entries which are the singular values of \(X_{ij}\). This contradicts \(U_{ij} \to U\) and \(V_{ij} \to V\). Thus, “local ⇒ local” does not hold at such \((U, \sigma, V)\). Theorem 2.4 shows that “\(1 \Rightarrow 1\)” does not hold there either since \(T_X \mathcal{X}\) is not a linear space.

Finally, fix \((U, \sigma, V) \in \mathcal{M}\) such that all the entries of \(|\sigma|\) are nonzero and distinct. We verify that “\(1 \Rightarrow 1\)” holds there using Theorem 2.4 by showing \(\text{im} L_{(U, \sigma, V)} = T_X \mathcal{X}\). Since \(\mathcal{M}\) is a product of embedded submanifolds of linear spaces, we have from (3.4.4) that

\[
T_{(U, \sigma, V)} = T_U \text{St}(m, r) \times T_\sigma \mathbb{R}^r \times T_V \text{St}(n, r) = \{(\hat{U}, \hat{\sigma}, \hat{V}) : U^\top \hat{U} + \hat{U}^\top U = V^\top \hat{V} + \hat{V}^\top V = 0\},
\]

and

\[
L_{(U, \sigma, V)}(\hat{U}, \hat{\sigma}, \hat{V}) = D\varphi(U, \sigma, V)[\hat{U}, \hat{\sigma}, \hat{V}] = \hat{U}\text{diag}(\sigma)V^\top + \hat{U}\text{diag}(\hat{\sigma})V^\top + U\text{diag}(\sigma)\hat{V}^\top,
\]
where \((\hat{U}, \hat{\sigma}, \hat{V}) \in T_{(U,\sigma,V)} M\). Let \(U_\perp \in \text{St}(m, m - r)\) and \(V_\perp \in \text{St}(n, n - r)\) satisfy \(U_\perp^T U_\perp = 0\) and \(V_\perp^T V_\perp = 0\). By [8, §7.3], we have \(\hat{U} \in T_U \text{St}(m, r)\) and \(\hat{V} \in T_V \text{St}(n, r)\) if and only if

\[
\hat{U} = U\Omega_u + U_\perp B_u, \quad \hat{V} = V\Omega_v + V_\perp B_v, \tag{D.0.1}
\]

where \(\Omega_u, \Omega_v \in \text{Skew}(r) := \{\Omega \in \mathbb{R}^{r \times r} : \Omega^T + \Omega = 0\}\) and \(B_u \in \mathbb{R}^{(m-r) \times r}\), \(B_v \in \mathbb{R}^{(n-r) \times r}\) are arbitrary. Using this parametrization,

\[
L_{(U,\sigma,V)}(\hat{U}, \hat{\sigma}, \hat{V}) = [U \quad U_\perp] \begin{bmatrix} \Omega_u \text{diag}(\sigma) - \text{diag}(\sigma)\Omega_v + \text{diag}(\hat{\sigma}) \sigma - \text{diag}(\sigma)B_v^T & 0 \\ B_u \text{diag}(\sigma) & \sigma \end{bmatrix} \begin{bmatrix} V \\ V_\perp \end{bmatrix}^T. \tag{D.0.2}
\]

Since \(T_X Y\) is a linear space and \(\dim T_{(U,\sigma,V)} M = \dim T_X Y\), we have \(\text{im} L_{(U,\sigma,V)} = T_X Y\) if and only if \(L_{(U,\sigma,V)}\) is injective. Suppose therefore that \(L_{(U,\sigma,V)}(\hat{U}, \hat{\sigma}, \hat{V}) = 0\). By (D.0.2) this is equivalent to

\[
B_u = 0, \quad B_v = 0, \quad \hat{\sigma} = 0, \quad \Omega_u \text{diag}(\sigma) - \text{diag}(\sigma)\Omega_v = 0, \tag{D.0.3}
\]

where \(\hat{\sigma} = 0\) follows by considering the diagonal of the top left block in (D.0.2). The fourth equality in (D.0.3) together with the skew-symmetry of \(\Omega_u, \Omega_v\) gives for all \(i, j\) that

\[
(\Omega_u)_{i,j} = \sigma_i \sigma_j \implies \frac{-\sigma_i}{\sigma_j} (\Omega_u)_{i,j} = (\Omega_u)_{j,i} = -\frac{-\sigma_j}{\sigma_i} (\Omega_u)_{i,j} \implies \frac{\sigma_i^2 - \sigma_j^2}{\sigma_i \sigma_j} (\Omega_u)_{i,j} = 0.
\]

Since \(|\sigma_i| \neq |\sigma_j|\) whenever \(i \neq j\), we get \((\Omega_u)_{i,j} = 0\) and \((\Omega_u)_{i,j} = 0\) for all \(i \neq j\), hence \(\Omega_u = \Omega_v = 0\).

We conclude that \((\hat{U}, \hat{\sigma}, \hat{V}) = 0\) so \(L_{(U,\sigma,V)}\) is injective and “1 \Rightarrow 1” holds. By Proposition 2.11, “local \Rightarrow local” holds there as well.

All cases have been checked, so the first bullet in Proposition 2.9 is proved.

**Modified SVD lift:** We now consider the modified SVD lift \(\text{St}(m, r) \times S^r \times \text{St}(n, r) \rightarrow \mathbb{R}_{\leq r}^{m \times n}\) defined by \(\varphi(U, M, V) = U M V^T\). Fix \((U, M, V) \in M\) such that \(\text{rank}(M) = r\) and \(\lambda_k(M) + \lambda_\ell(M) = 0\) for some \(k < \ell\). We show “local \Rightarrow local” does not hold at \((U, M, V)\) by constructing a sequence converging to \(X = \varphi(U, M, V)\) such that no subsequence of it can be lifted to a sequence converging to \((U, M, V)\). Let \(\alpha_i \in \mathbb{R}^r\) be a sequence converging to zero such that \(|\lambda(M) + \alpha_i|\) has distinct entries that are all nonzero. Let \(M = W \text{diag}(\lambda(M)) W^T\) be an eigendecomposition of \(M\), where \(W \in \text{O}(r)\). Define

\[
U^{(k,\ell)} = U W T^{(k,\ell)} S^{(k,\ell)} W^T \in \text{St}(m, r), \quad V^{(k,\ell)} = V W T^{(k,\ell)} W^T \in \text{St}(n, r), \tag{D.0.4}
\]

where \(T^{(k,\ell)} \in \text{O}(r)\) is the permutation interchanging the \(k\)th and \(\ell\)th entries of a vector while fixing all the others, and \(S^{(k,\ell)} \in \text{O}(r)\) flips the signs of these entries (so it’s a diagonal matrix with all 1’s on the diagonal except for the \(k\)th and \(\ell\)th entries which are \(-1\)). Note that \(T^{(k,\ell)}\) and \(S^{(k,\ell)}\) are symmetric and commute. Let

\[
X_i = \varphi(U^{(k,\ell)}, M + W \text{diag}(\alpha_i) W^T, V^{(k,\ell)}),
\]

whose singular values are \(|\lambda(M) + \alpha_i|\) and are all distinct. Note that \(X_i \rightarrow X\), because

\[
\lim_{i \to \infty} X_i = U^{(k,\ell)} M (V^{(k,\ell)})^T = (U W) T^{(k,\ell)} (S^{(k,\ell)} \text{diag}(\lambda(M)))(T^{(k,\ell)})^T (V W)^T = (U W) \text{diag}(\lambda(M))(V W)^T = U M V^T = X,
\]

50
where the first equality on the second line follows because $S^{(k,\ell)}$ flips the signs of $\lambda_k(M)$ and of $\lambda_{\ell}(M) = -\lambda_k(M)$, and conjugation by $T^{(k,\ell)}$ interchanges them again.

Suppose $(U_{ij}, M_{ij}, V_{ij})$ is a lift of a subsequence of $(X_i)$ converging to $(U, M, V)$. The singular values of $X_{ij}$ are the entries $|\lambda(M_{ij})|$, which are also the entries of $|\lambda(M) + \alpha_{ij}|$ (possibly permuted). Combining this with $\lambda(M_{ij}) \to \lambda(M)$ (since $M_{ij} \to M$), $\alpha_{ij} \to 0$, and the fact that $\lambda(M)$ has no zero entries, we further get that the entries of $\lambda(M_{ij})$ are equal to the entries of $\lambda(M) + \alpha_{ij}$ for all large $j$. After passing to a subsequence, we may assume that this equality holds for all $j$. Let $M_{ij} = W_{ij} \diag(\lambda(M) + \alpha_{ij}) W_{ij}^\top$ be an eigendecomposition of $M_{ij}$. Since $O(r)$ is compact, we may also assume after further passing to a subsequence that

$$\lim_{j \to \infty} W_{ij} = \tilde{W} \in O(r),$$

exists. At this point, we have two SVDs of $X_{ij}$, namely

$$X_{ij} = [U^{(k,\ell)}W \sign(\lambda(M) + \alpha_{ij})] \diag(|\lambda(M) + \alpha_{ij}|) [V^{(k,\ell)}W]^\top = [U_{ij} W_{ij} \sign(\lambda(M) + \alpha_{ij})] \diag(|\lambda(M) + \alpha_{ij}|) [V_{ij} W_{ij}]^\top.$$

Because the singular values $|\lambda(M) + \alpha_{ij}|$ of $X_{ij}$ are distinct, its singular vectors are unique up to sign [52, Thm. 4.1]. Specifically, there exists $S_{ij} \in \diag(\{\pm 1\}^r)$ satisfying

$$U^{(k,\ell)}W \sign(\lambda(M) + \alpha_{ij}) = U_{ij} W_{ij} \sign(\lambda(M) + \alpha_{ij}) S_{ij} \text{ implying } U^{(k,\ell)}W = U_{ij} W_{ij} S_{ij},$$

$$V^{(k,\ell)}W = V_{ij} W_{ij} S_{ij},$$

where in the first line we used the fact that diagonal matrices commute. Because $S_{ij}$ takes values in a finite set, after passing to a subsequence again we may assume $S_{ij} = S$ is fixed. Then $U^{(k,\ell)} = U_{ij} W_{ij} S W^\top$ and $V^{(k,\ell)} = V_{ij} W_{ij} S W^\top$ for all $j$, and taking $j \to \infty$ we conclude that $U^{(k,\ell)} = U\tilde{W} SW^\top$ and $V^{(k,\ell)} = \tilde{V}\tilde{W} SW^\top$. Equating this to (D.0.4), we obtain

$$U\tilde{W} SW^\top = UWT^{(k,\ell)} S^{(k,\ell)} W^\top \text{ and } \tilde{V}\tilde{W} SW^\top = VWT^{(k,\ell)} W^\top.$$

Rearranging gives $W^\top \tilde{W} S = T^{(k,\ell)} S^{(k,\ell)}$ and $W^\top \tilde{W} S = T^{(k,\ell)}$ so in fact, $T^{(k,\ell)} S^{(k,\ell)} = T^{(k,\ell)}$, a contradiction. Thus, no subsequence of $(X_i)$ can be lifted, so the lift does not satisfy “local ⇒ local” at $(U, M, V)$. By Proposition 2.11 and the fact that $X \in \mathbb{R}^{m \times n} = X^{\text{smth}}$, we conclude that “1 ⇒ 1” is not satisfied at $(U, M, V)$ either.

Now fix $(U, M, V)$ such that $\lambda_k(M) = 0$ for some $k$, and let $X, W$ be as above. Since $r < \min\{m, n, n\}$, there exist unit vectors $u'_k \in \mathbb{R}^{m}$ and $v'_k \in \mathbb{R}^{n}$ such that $(UW)^\top u'_k = 0$ and $(VW)^\top v'_k = 0$. Let $Y \in O(m)$ and $Z \in O(n)$ send the $k$th columns of $UW$ and $VW$ to $u'_k$ and $v'_k$, respectively, and act by the identity on their orthogonal complements. Let $\alpha_i \in \mathbb{R}^r$ converge to zero such that $|\lambda(M) + \alpha_i|$ are distinct and nonzero. Define $X_i = \varphi(YU, M + \alpha_i, ZV)$, which converge to $YUM(ZV)^\top = X$. Suppose $(U_{ij}, M_{ij}, V_{ij})$ is a lift of a subsequence of $(X_i)$ converging to $(U, M, V)$. Let $M_{ij} = W_{ij} \diag(\lambda(M) + \alpha_i) W_{ij}^\top$ be an eigendecomposition. After passing to a subsequence, we may assume $W_{ij} \to \tilde{W}$ in $O(r)$. Because the singular vectors of $X_{ij}$ are unique up to sign, there exists $S_{ij} \in \diag(\{\pm 1\}^r)$ satisfying

$$YU = U_{ij} W_{ij} S_{ij} W_{ij}^\top,$$

and similarly for $ZV$. Because $S_{ij}$ takes values in a finite set, after passing to a subsequence again we may assume $S_{ij} = S$ is fixed for all $j$, in which case we get

$$YU = U_{ij} W_{ij} S W^\top \to U\tilde{W} SW^\top.$$
This is a contradiction since \( \text{col}(YU) \neq \text{col}(U) = \text{col}(U\tilde{W}SW^\top) \) by construction of \( Y \). Thus, no subsequence of \( (X_i) \) can be lifted so “local \( \Rightarrow \) local” does not hold at such \( (U, M, V) \). Since \( T_X\mathcal{X} \) is not a linear space, Theorem 2.4 shows that “1 \( \Rightarrow \) 1” does not hold there either.

Finally, fix \( (U, M, V) \in \mathcal{M} \) such that \( \lambda_i(M) + \lambda_j(M) \neq 0 \) for all \( i, j \) and let \( X = \varphi(U, M, V) \). We show that “1 \( \Rightarrow \) 1” holds at \( (U, M, V) \) by showing \( \text{im} L_{(U, M, V)} = T_X\mathcal{X} \) and appealing to Theorem 2.4. Note that \( T_{(U, M, V)}\mathcal{M} = T_{U\text{St}}(m, r) \times S^r \times T_{V\text{St}}(n, r) \). Then (3.4.4) gives

\[
L_{(U, M, V)}(\dot{U}, \dot{M}, \dot{V}) = \dot{U}MV^\top + U\dot{M}V^\top + UM\dot{V}^\top.
\]

Writing \( \dot{U}, \dot{V} \) as in (D.0.1), we get similarly to (D.0.2) that

\[
L_{(U, M, V)}(\dot{U}, \dot{M}, \dot{V}) = [U \ U_\perp] \begin{bmatrix} \Omega_uM - M\Omega_v + \dot{M} & MB_u^\top \\ B_uM & 0 \end{bmatrix} [V \ V_\perp]^\top.
\] (D.0.5)

Since \( \text{rank}(M) = r \), we have \( \text{col}(U) = \text{col}(X) \) and \( \text{col}(V) = \text{col}(X^\top) \). By [48, Thm. 3.1], we have

\[
T_X\mathbb{R}^{m \times n}_{\leq r} = \left\{ \dot{X} = [U \ U_\perp] \begin{bmatrix} \dot{X}_1 & \dot{X}_2 \\ \hat{X}_3 & 0 \end{bmatrix} [V \ V_\perp]^\top : \dot{X}_1 \in \mathbb{R}^{r \times r}, \dot{X}_2 \in \mathbb{R}^{r \times (n-r)}, \hat{X}_3 \in \mathbb{R}^{(m-r) \times r} \right\}.
\]

Since \( \lambda_i(M) + \lambda_j(M) \neq 0 \) for all \( i, j \), for any \( \dot{X}_1 \in \mathbb{R}^{r \times r} \) we can pick \( \Omega \in \text{Skew}(r) \) such that \( \Omega M + M\Omega = \text{skew}(\dot{X}_1) = \frac{\dot{X}_1 - \dot{X}_1^\top}{2} \). Indeed, if \( M = W\Lambda W^\top \) is an eigendecomposition of \( M \), define \( \Omega \) by setting

\[
(W^\top \Omega W)_{i,j} = \frac{(W^\top \text{skew}(\dot{X}_1) W)_{i,j}}{\lambda_i(M) + \lambda_j(M)},
\]

which is clearly skew-symmetric and solves the above equation. We set \( \Omega_u = -\Omega_v = \Omega \) and \( \dot{M} = \text{sym}(\dot{X}_1) = \frac{\dot{X}_1 + \dot{X}_1^\top}{2} \in \mathbb{S}^r \). Finally, we set

\[
B_u = \dot{X}_2^\top M^{-\top}, \quad B_v = \hat{X}_3 M^{-1}.
\]

With these choices, we get \( L_{(U, M, V)}(\dot{U}, \dot{M}, \dot{V}) = \dot{X} \), showing that \( L_{(U, M, V)} \) is surjective and “1 \( \Rightarrow \) 1” holds. By Proposition 2.11, the lift satisfies “local \( \Rightarrow \) local” at \( (U, M, V) \) as well.

**E  Notation and basic definitions**

Here we collect notation and standard definitions used throughout the paper.

- \( \mathcal{E} \) is a linear space endowed with an inner product \( \langle \cdot, \cdot \rangle \) and induced norm \( \| \cdot \| \).
- \( \mathcal{M} \) is a smooth Riemannian manifold endowed with a Riemannian metric, also denoted \( \langle \cdot, \cdot \rangle \), with its induced norm \( \| \cdot \| \). The tangent space to \( \mathcal{M} \) at \( y \in \mathcal{M} \) is denoted \( T_y\mathcal{M} \).
- If \( S \subseteq \mathcal{E} \) is a subspace of an inner product space \( \mathcal{E} \), we denote by \( \text{Projs}(x) \) the orthogonal projection of \( x \in \mathcal{E} \) onto \( S \).
- \( \mathcal{X} \) is endowed with its subspace topology from \( \mathcal{E} \), and \( \mathcal{M} \) is endowed with its manifold topology.
- A neighborhood of a point \( x \) is a set that contains \( x \) in its interior. (We do not require neighborhoods to be open.)
• A map \( \varphi \) between two topological spaces is open at \( y \) if it maps neighborhoods of \( y \) in \( M \) to neighborhoods of \( \varphi(y) \) in \( X \). Globally, a map is open if it is open at all points, or equivalently if it maps open subsets of \( M \) to open subsets of \( X \).

• A smooth curve on \( M \) passing through \( y \in M \) with velocity \( v \in T_yM \) is a smooth map \( c: \mathbb{R} \to M \) satisfying \( c(0) = y \) and \( c'(0) = v \).

• If \( c: \mathbb{R} \to M \) is a smooth curve on a Riemannian manifold \( M \) such that \( c(0) = y \), then \( c''(0) \in T_yM \) denotes its intrinsic (Riemannian) acceleration. Accordingly, if \( \gamma = \varphi \circ c \) then \( \gamma''(0) \) denotes its (standard) acceleration in the Euclidean space \( \mathcal{E} \).

• A cone is a set \( K \subseteq \mathcal{E} \) such that \( v \in K \) implies \( \alpha v \in K \) for all \( \alpha > 0 \).

• The dual cone \( K^* \) of a cone \( K \) is \( K^* = \{ u \in \mathcal{E} : \langle u, v \rangle \geq 0 \text{ for all } v \in K \} \). We use the following properties throughout (see [20, Prop. 4.5] for proofs):
  - The dual cone is always a closed convex cone.
  - If \( K_1 \subseteq K_2 \), then \( K_2^* \subseteq K_1^* \).
  - The bidual cone \( K^{**} \) of \( K \) is equal to the closure of its convex hull \( K^{**} = \overline{\text{conv}}(K) \). In particular, \( K^{**} \supseteq K \).
  - If \( K \) is a linear space, then its dual \( K^* \) is equal to its orthogonal complement \( K^\perp \).

• If \( \psi: \mathcal{N} \to M \) is a smooth map between smooth manifolds, its differential at \( z \in \mathcal{N} \) is denoted \( D\psi(z) : T_z\mathcal{N} \to T_{\psi(z)}M \).

• If \( g: M \to \mathbb{R} \) is a smooth function, its Riemannian gradient at \( y \in M \) is denoted \( \nabla g(y) \in T_yM \). It is the unique element of \( T_yM \) satisfying \( \langle \nabla g(y), v \rangle = Dg(y)[v] \) for all \( v \in T_yM \). The Riemannian Hessian of \( g \) at \( y \) is a self-adjoint linear map denoted \( \nabla^2 g(y) : T_yM \to T_yM \), see [8, §8.11] for the definition.

If \( f: \mathcal{E} \to \mathbb{R} \) is a smooth function on a linear space, the usual definitions of the directional derivative \( Df(x)[v] \), the gradient \( \nabla f(x) \), and the Hessian \( \nabla^2 f(x) \) coincide with the above definitions specialized to \( \mathcal{M} = \mathcal{E} \).

• We write \( S \succeq 0 \) to denote a positive-semidefinite (PSD) matrix or a PSD self-adjoint linear map \( S \), and we write \( S \succ 0 \) to indicate \( S \) is positive-definite.