



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/94926/>

Version: Accepted Version

---

**Article:**

Halupczok, I and Yin, Y (2018) Lipschitz stratifications in power-bounded o-minimal fields. Journal of the European Mathematical Society, 20 (11). pp. 2717-2767. ISSN: 1435-9855

<https://doi.org/10.4171/JEMS/823>

---

© 2018 EMS Publishing House. This is an author produced version of a paper published in Journal of the European Mathematical Society. Uploaded in accordance with the publisher's self-archiving policy.

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

# Lipschitz stratifications in power-bounded $o$ -minimal fields

Immanuel Halupczok      Yimu Yin

9th Feb. 2016

## Abstract

We propose to grok Lipschitz stratifications from a non-archimedean point of view and thereby show that they exist for closed definable sets in any power-bounded  $o$ -minimal structure on a real closed field. Unlike the previous approaches in the literature, our method bypasses resolution of singularities and Weierstraß preparation altogether; it transfers the situation to a non-archimedean model, where the quantitative estimates appearing in Lipschitz stratifications are sharpened into valuation-theoretic inequalities. Applied to a uniform family of sets, this approach automatically yields a family of stratifications which satisfy the Lipschitz conditions in a uniform way.

**Keywords.** Lipschitz stratifications, polynomially bounded fields, power-bounded fields

In this paper we prove the existence of Lipschitz stratifications for any closed definable set in a polynomially bounded  $o$ -minimal structure on  $\mathbb{R}$ , and, in fact, even more generally, in a power-bounded  $o$ -minimal structure on a real closed field  $\mathcal{R}$ . The notion of a Lipschitz stratification was introduced by Mostowski in his dissertation [Mos]. It is much stronger than Whitney's conditions or Verdier's condition (w) formulated in [Ver]; it imposes a global condition and ensures that the Lipschitz type of the stratified set is locally constant along each stratum.

Throughout this paper,  $\mathcal{R}$  is a power-bounded real closed field. A classical example of such a structure is  $\mathbb{R}_{\text{an}}$ : the reals with restricted analytic functions as described in [DMM]; beyond this (subanalytic) level, there is e.g. the class of quasianalytic structures; see [Ro]. If the field  $\mathcal{R}$  is just  $\mathbb{R}$ , then power-bounded is equivalent to polynomially bounded. In

---

I. Halupczok (corresponding author): School of Mathematics, University of Leeds, Leeds, LS2 9JT, UK; e-mail: math@karimmi.de

Y. Yin: 607 Xi Chang Hall, Department of Philosophy, Sun Yat-sen University, 135 Xingang Road West, Guangzhou, China, 510275; e-mail: yimu.yin@hotmail.com

*Mathematics Subject Classification (2010):* 03C64; 03C60, 32C07, 32B20, 32S60, 58A35

other real closed fields, power-boundedness is more general and more natural; we recall that notion in Definition 1.1.1.

Here is a first version of our main result.

**Theorem 1** (Lipschitz stratifications). *Let  $X \subseteq \mathcal{R}^n$  be a closed definable subset in a power-bounded real closed field  $\mathcal{R}$ . Then there exists a definable Lipschitz stratification of  $X$ .*

The notion of Lipschitz stratification is recalled in Definition 1.2.4, and Subsection 1.1 clarifies how the terminology should be adapted in the case  $\mathcal{R} \neq \mathbb{R}$ . For compact sets  $X$  and in the case  $\mathcal{R} = \mathbb{R}$ , the semi-analytic case of this theorem was established in [Par1] and the subanalytic case in [Par2]. Recently, Nguyen and Valette [NV] generalized Parusinski's proof to polynomially bounded structures on  $\mathbb{R}$ . (In [NV] the result is stated for compact  $X$ , but their proof also goes through for arbitrary closed  $X$ ; see [Ngu].)

A main motivation for Lipschitz stratifications is that one has local bilipschitz triviality along strata, which in turn implies that any two points within the same stratum have neighborhoods which are in bilipschitz bijection. The proof of this result is rather easy in  $\mathbb{R}$ , but the argument uses integration along vector fields; this is highly non-definable, and it does not generalize to other real closed fields. We believe that local bilipschitz triviality (along strata of a Lipschitz stratification) can also be obtained in  $\mathcal{R} \neq \mathbb{R}$ , but the argument might be much more involved. More precisely, a proof of the existence of *definable* local bilipschitz trivializations within  $\mathbb{R}$  would probably directly generalize to  $\mathcal{R}$ . Some results in that direction exist. For example, Valette [Val] proved the existence of definable bilipschitz trivializations in polynomially bounded  $\mathcal{o}$ -minimal structures, but using certain triangulations instead of Lipschitz stratifications.

Using the existence of Skolem functions and the Compactness Theorem, one easily deduces that Theorem 1 also works uniformly in families, in the sense that given a uniformly definable family of sets, one finds a uniformly definable family of Lipschitz stratifications. However, the notion of Lipschitz stratifications involves a constant  $C$  (a stratification is Lipschitz if some conditions hold for sufficiently big  $C$ ), and a natural question is whether that  $C$  can be chosen to be the same for an entire family. In this paper, we obtain uniform Lipschitz stratifications in families in this strong sense; the precise statement is Theorem 1.3.5.

Our approach to the construction of Lipschitz stratifications is quite different from all previous ones. The main difference is that we use the technique from non-standard analysis of replacing  $\mathcal{R}$  by a bigger real closed field  $\mathcal{R}'$  (an elementary extension). The infinite and infinitesimal elements in  $\mathcal{R}'$  make it possible to simplify the formulation of statements involving limits. In particular, we obtain simpler characterizations of Lipschitz stratifications: Whereas the original definition of a Lipschitz stratification uses subtle inequalities

depending on two different constants  $c$  and  $C$ , we obtain an equivalent definition, formulated using  $\mathcal{R}'$ , which needs neither  $c$  nor  $C$  (see Definition 1.6.5 and Proposition 1.6.11). The aforementioned strong uniformity in families is obtained as a side effect of using that approach: We prove that Lipschitz stratifications exist in families within  $\mathcal{R}'$ . The fact that the parameters of the family are allowed to run over the bigger field  $\mathcal{R}'$  allows us to deduce the strong uniformity result within  $\mathcal{R}$ .

On our way, we also obtain various other equivalent characterizations of Lipschitz stratifications: Proposition 1.2.5 provides some characterizations purely within the standard model, where  $c$  and  $C$  are used in a less subtle way, and Proposition 1.8.3 provides a new characterization of Lipschitz stratifications in terms of partial flags, which is invariant under  $\mathrm{GL}_n$ . (To our knowledge, the only previously known  $\mathrm{GL}_n$ -invariant characterization was the one in terms of vector bundles given e.g. in [Par1, Proposition 1.5].)

Typically, proofs carried out using non-standard analysis in an elementary extension  $\mathcal{R}'$  can be translated back to “classical” proofs within  $\mathcal{R}$  (at the cost of making them much less readable). However, for one key ingredient to our proof – a precise estimate of the gradient of functions near a singular locus; cf. Corollary 2.2.2 and Remark 2.2.3 – we use some deeper model theoretic results. More precisely,  $\mathcal{R}'$  naturally carries a valuation, which specifies the order of magnitude of elements. The proof of our estimate builds on model theory of  $\mathcal{R}'$  as a *valued* field, i.e., we consider definable sets in a language including the valuation. This setting has been studied by van den Dries and Lewenberg [DL, Dri1] under the name of “ $T$ -convex fields”. In that setting, the second author of the present paper obtained a result which is somewhat related to Weierstraß Preparation in valued fields (Proposition 2.2.1) and that in turn implies the above-mentioned Corollary 2.2.2.

In Section 1 we recall the notion of Lipschitz stratifications and prove the equivalence of its various characterizations. We also give an overview of the proof of existence of Lipschitz stratifications (in Subsection 1.9). The entire remainder of the paper is devoted to the details of that proof. Section 2 discusses the various ingredients and Section 3 contains the proof itself.

## 0.1 Acknowledgement

The research reported in this paper has received financial support from the ERC Advanced Grant 246903 NMNAG, from the IHES, from the European Research Council under the European Community’s Seventh Framework Programme (FP7/2007-2013) with ERC Grant Agreement no. 615722 MOTMELSUM, from the Labex CEMPI (ANR-11-LABX-0007-01), from the Fund for Scientific Research of Flanders, Belgium (grant G.0415.10) and from the laboratoire de mathématiques de l’université de Savoie Mont Blanc.

# 1 Characterizations of Lipschitz Stratifications

In this section, we recall the definition of Lipschitz stratifications, we formulate several alternative definitions and we prove that all those definitions are equivalent. This does not yet use any deep model theory; the only model theoretic ingredient we use is the notion of elementary extensions (and their existence).

## 1.1 Basic Notation

We fix some notation which will be used throughout this paper.

Recall that an  $o$ -minimal structure on  $\mathbb{R}$  is polynomially bounded if every definable function  $\mathbb{R} \rightarrow \mathbb{R}$  is ultimately bounded by a polynomial. One essential aspect of this notion is the dichotomy obtained by Miller [Mil2]: In any structure that is not polynomially bounded, one can already define exponentiation. To obtain a similar dichotomy for other real closed fields  $\mathcal{R}$ , one needs a generalization of polynomially bounded [Mil1]: a definable function only needs to be bounded by a kind of generalized power function. Here is the precise definition.

**Definition 1.1.1** (Power bounded). Suppose that  $\mathcal{R}$  is an  $o$ -minimal real closed field. A *power function* in  $\mathcal{R}$  is a definable endomorphism of the multiplicative group  $\mathcal{R}^\times$ . We call  $\mathcal{R}$  *power bounded* if for every definable function  $f: \mathcal{R} \rightarrow \mathcal{R}$ , there exists a power function  $g$  such that  $|f(x)| \leq g(x)$  for all sufficiently big  $x$ .

There is a precise sense in which a power function is of the form  $x \mapsto x^\lambda$ , where  $\lambda$  is an element of a certain subfield of  $\mathcal{R}$ . Since we will use power-boundedness only indirectly, we do not elaborate on this; see [Mil1] for details.

**Notation 1.1.2** (Structures and language). Throughout this paper, we fix a power-bounded  $o$ -minimal real closed field  $\mathcal{R}$  in a language  $\mathcal{L}$  expanding the ring language. (At some point, we will impose that  $\mathcal{R}$  is, without loss, sufficiently big).

By *definable* we mean definable with arbitrary parameters; in contrast,  *$\mathcal{L}$ -definable* means definable without parameters (apart from those which are constants in the language).

*Remark 1.1.3.* It is somewhat customary, in  $o$ -minimal geometry, to not specify a language and only work with the notion of definable sets. However, specifying a language allows us to keep track of the parameters needed to define sets, and this will be needed for some model theoretic arguments. For the moment, the reader unfamiliar with our approach may assume that  $\mathcal{L}$  contains a constant for each element of  $\mathcal{R}$ , so that  $\mathcal{L}$ -definable means the same as definable.

**Notation 1.1.4** (Coordinate projections). Given  $d \leq n$ , we write  $\text{pr}_d: \mathcal{R}^n \rightarrow \mathcal{R}$  for the projection to the  $d$ -th coordinate,  $\text{pr}_{\leq d}: \mathcal{R}^n \rightarrow \mathcal{R}^d$  for the projection to the first  $d$  coordinates and  $\text{pr}_{> d}: \mathcal{R}^n \rightarrow \mathcal{R}^{n-d}$  for the projection to the last  $n - d$  coordinates.

We use the usual notation and conventions for  $\mathcal{o}$ -minimal expansions of real closed fields; see e.g. [Dri2]. We quickly recall the most important ones.

**Notation 1.1.5** (Infima and suprema). By  $\mathcal{o}$ -minimality, any definable subset  $X \subseteq \mathcal{R}$  has an infimum and a supremum (which may be  $\pm\infty$ ); we denote them by  $\inf(X)$  and  $\sup(X)$ .

**Notation 1.1.6** (Norms and distances). We write  $|\cdot|$  for the absolute value on  $\mathcal{R}$ ,  $\|a\|$  for the Euclidean Norm of  $a \in \mathcal{R}^n$  ( $\|a\|$  is an element of  $\mathcal{R}_{\geq 0}$ ) and  $\|M\|$  for the operator norm of a matrix  $M$ , i.e.,  $\|M\| = \sup\{\|Ma\| : \|a\| = 1\}$ . Given a point  $a \in \mathcal{R}^n$  and a definable set  $X \subseteq \mathcal{R}^n$ , we write  $\text{dist}(a, X) := \inf\{\|a - x\| : x \in X\}$  for the distance from  $a$  to  $X$ ; we define that distance to be  $\infty$  if  $X$  is empty.

**Notation 1.1.7** (Topology). The real closed field  $\mathcal{R}$  comes with a natural topology induced by the order on  $\mathcal{R}$ ; this also induces a topology on  $\mathcal{R}^n$ . Given a definable set  $X \subseteq \mathcal{R}^n$ , we write  $\text{cl}(X)$  for its topological closure,  $\text{int}(X)$  for its interior, and  $\partial X := \text{cl}(X) \setminus X$  for its frontier (not to be mixed up with the boundary, which is also sometimes denoted by  $\partial X$ ). We call  $X$  *definably connected* if  $X$  is not the disjoint union of two relatively closed (in  $X$ ) definable subsets. The *definable connected components* of  $X$  are defined accordingly. (Any definable set in an  $\mathcal{o}$ -minimal structure has finitely many definable connected components.)

The topology on  $\mathcal{R}$  might be totally disconnected, so the usual notion of connectedness does not behave as desired. However, in the case  $\mathcal{R} = \mathbb{R}$ , definably connected is the same as connected.

**Notation 1.1.8** (Derivatives). For an open set  $X \subseteq \mathcal{R}^n$ , derivatives of functions  $f: X \rightarrow \mathcal{R}^m$  are defined as the usual limits. By  $\mathcal{o}$ -minimality, derivatives exist almost everywhere. For functions  $f: X \rightarrow \mathcal{R}$ , we write  $\partial_i f$  for the derivative with respect to the  $i$ -th variable ( $1 \leq i \leq n$ ), and for  $f = (f_1, \dots, f_m): X \rightarrow \mathcal{R}^m$  and  $a \in X$ , we write

$$\text{Jac}_a f := \begin{pmatrix} \partial_1 f_1(a) & \cdots & \partial_n f_1(a) \\ \vdots & & \vdots \\ \partial_1 f_m(a) & \cdots & \partial_n f_m(a) \end{pmatrix}$$

for the Jacobian matrix of  $f$  at the point  $a$ . In the case  $m = 1$ , we also write  $\nabla f(a)$  instead of  $\text{Jac}_a f$ . We define the class  $C^p$  of  $p$ -fold continuously differentiable functions in the usual way.

The notion of manifolds makes sense over ( $\mathcal{o}$ -minimal) fields  $\mathcal{R} \neq \mathbb{R}$  only if one restricts to definable manifolds. All manifolds we will encounter will moreover be embedded.

**Notation 1.1.9** (Manifolds and tangent spaces). A  $d$ -dimensional definable  $C^p$  submanifold of  $\mathcal{R}^n$  (for  $d \leq n$  and  $p \geq 1$ ) is a definable set  $X \subseteq \mathcal{R}^n$  such that there exists a finite definable open cover of  $X$  by sets  $U_i$ , each of which is in definable  $C^p$ -bijection with an open set  $V_i \subseteq \mathcal{R}^d$ . The tangent space of  $X$  at some  $a \in X$  is denoted by  $T_a(X)$ . (We consider  $T_a(X)$  as a subspace of  $\mathcal{R}^n$ .)

Note that  $T_a(X)$  is definable uniformly in  $a$ .

## 1.2 Various definitions of Lipschitz stratifications

We use the following notation and conventions for stratifications:

**Definition 1.2.1** (Stratifications). Let  $X \subseteq \mathcal{R}^n$  be a definable subset of dimension  $d$ . A *definable stratification* of  $X$  is a family  $\mathcal{X} = (X^0 \subseteq X^1 \subseteq \dots \subseteq X^d = X)$  of closed definable subsets of  $X$  satisfying the properties below. We set  $X^{-1} := \emptyset$ . For  $0 \leq i \leq d$ , the set  $\overset{\circ}{X}^i := X^i \setminus X^{i-1}$  is called the  *$i$ -th skeleton*, and each definably connected component of each skeleton is called a *stratum*. We call  $\mathcal{X}$  a stratification if the following conditions hold.

- For each  $i$ ,  $\dim X^i \leq i$ ;
- for each  $i$ ,  $\overset{\circ}{X}^i$  is either empty or a definable  $C^1$  submanifold of  $\mathcal{R}^n$  of dimension  $i$  (not necessarily connected);
- for each stratum  $S$ , the topological closure  $\text{cl}(S)$  is a union of strata.

(Note that in the generality of power-bounded  $\mathcal{o}$ -minimal structures, one cannot expect to obtain smooth strata.)

Mostowski's original definition of when a stratification is a Lipschitz stratification uses the notion of a chain: a sequence of points  $(a^\ell)_{0 \leq \ell \leq m}$  that starts with an arbitrary point  $a^0 \in X$ , and where the remaining points lie in lower dimensional skeletons, but "not too far from  $a^0$ ", and only in "those skeletons  $\overset{\circ}{X}^i$  which are much closer to  $a^0$  than  $X^{i-1}$ ". The precise inequalities specifying these distances are quite subtle. There exists an equivalent definition involving Lipschitz vector fields [Par1, Proposition 1.5], which avoids the subtleties of bounding the aforementioned distances. However, that definition quantifies over vector fields, which makes it less suitable for our model theoretic approach. Therefore, in this paper, we use the original definition in terms of chains. (More precisely, we use the simplified variant of that original definition given in [Par1].)

As already mentioned in the introduction, we will use methods from non-standard analysis to simplify the definition of Lipschitz stratifications: After having replaced  $\mathcal{R}$  by an elementary extension, we will define a valuation on  $\mathcal{R}$ , which will allow us to replace the subtle bounds on distances by simple valuative inequalities. However, that valuative definition is *not* a straight-forward translation of Mostowski's definition in the usual non-standard analysis way. To make such a translation possible, one needs to first modify Mostowski's definition in such a way that certain quantifiers become simpler.

To prove that our new definition is equivalent to the old one, our strategy is as follows. We introduce two new variants of Mostowski's definition: one of them *a priori* weaker and one of them *a priori* stronger. Both variants have simpler quantifiers, so that they can be translated to valuative versions. For those valuative versions, it will not be very hard to prove that the weak one implies the strong one, hence implying that all definitions are equivalent.

In the following, we start by giving all those definitions of Lipschitz stratifications which do not use the valuation. The valuative versions are stated in Subsection 1.6, and the proofs of the equivalences are given in Subsection 1.7.

Lipschitz stratifications are defined in terms of projections to the tangent spaces of the skeletons  $\mathring{X}^i$ ; we first fix notation for those projection maps.

**Definition 1.2.2.** Given a definable stratification  $\mathcal{X}$  of a definable subset  $X \subseteq \mathcal{R}^n$  and a point  $a \in \mathring{X}^i$ , let

$$P_a : \mathcal{R}^n \longrightarrow T_a \mathring{X}^i$$

be the orthogonal projection onto the tangent space of  $\mathring{X}^i$  at  $a$ , considered as a map  $\mathcal{R}^n \longrightarrow \mathcal{R}^n$ .

The various definitions of Lipschitz stratifications only differ in the way that certain constants are treated. To avoid writing almost the same definition three times (and to make it clear how exactly the definitions differ), we introduce a general notion of a stratification  $\mathcal{X}$  "satisfying the Mostowski Conditions for given constants". For readers who just want to understand one single definition of Lipschitz stratifications, one possible definition is encoded in the notation used for the constants: Increasing lowercase constants and decreasing uppercase constants both makes the Mostowski Conditions more restrictive; and  $\mathcal{X}$  is a Lipschitz stratification if no matter how big the lowercase constants are chosen, one can find values for the uppercase constants such that the Mostowski Conditions are satisfied (see Proposition 1.2.5 (2)).

Note: The Mostowski Conditions impose conditions on all chains, so a more restrictive notion of chains yields a less restrictive notion of Mostowski Conditions.

**Definition 1.2.3** (Chains and Mostowski Conditions). Let  $\mathcal{X} = (X^i)_i$  be a definable stratification (of a definable set  $X \subseteq \mathcal{R}^n$ ), and let  $c, c', C', C'', C''' \in \mathcal{R}$  be given.

A *plain chain* (in  $\mathcal{X}$ ) is a sequence of points  $a^0, a^1, \dots, a^m$  ( $m \geq 0$ ) with  $a^\ell \in \mathring{X}^{e_\ell}$ ,  $e_0 > e_1 > \dots > e_m$  satisfying the following conditions.

1. For  $\ell = 1, \dots, m$ , we have:

$$\|a^0 - a^\ell\| < c \cdot \text{dist}(a^0, X^{e_\ell}).$$

2. For each  $i$  with  $e_m \leq i < e_0$ , we have one of two different conditions (which should be considered as specifying which  $i$  should be among the  $e_\ell$  and which should not):

$$\begin{cases} \text{dist}(a^0, X^{i-1}) \geq C' \cdot \text{dist}(a^0, X^i) & \text{if } i \in \{e_1, \dots, e_m\} \\ \text{dist}(a^0, X^{i-1}) < c' \cdot \text{dist}(a^0, X^i) & \text{if } i \notin \{e_1, \dots, e_m\}. \end{cases}$$

An *augmented chain* (in  $\mathcal{X}$ ) consists of a plain chain  $a^0, a^2, a^3, \dots, a^m$  ( $m \geq 1$ ,  $a^\ell \in \mathring{X}^{e_\ell}$ ) together with an additional point  $a^1 \in \mathring{X}^{e_1}$ , where  $e_1 := e_0$ , satisfying

$$\|a^0 - a^1\| \leq \frac{\text{dist}(a^0, X^{e_1-1})}{C''}. \quad (1.1)$$

We say that  $\mathcal{X}$  *satisfies the Mostowski Conditions* for  $(c, c', C', C'', C''')$  if the following two conditions hold:

For every plain chain  $(a^i)_{0 \leq i \leq m}$  with  $m \geq 1$ , we have

$$\|(1 - P_{a^0})P_{a^1}P_{a^2} \dots P_{a^m}\| < \frac{C''' \|a^0 - a^1\|}{\text{dist}(a^0, X^{e_m-1})}, \quad (\text{m1})$$

For every augmented chain  $(a^i)_{0 \leq i \leq m}$  (with  $m \geq 1$ ), we have

$$\|(P_{a^0} - P_{a^1})P_{a^2}P_{a^3} \dots P_{a^m}\| < \frac{C''' \|a^0 - a^1\|}{\text{dist}(a^0, X^{e_m-1})}. \quad (\text{m2})$$

We use the convention that if  $X^{e_m-1}$  is empty, then in (m1) and (m2), we require the left hand side to be equal to 0.

Concerning nomenclature, note that what Parusinski calls a  $c$ -chain in [Par1] is what we would call a plain chain of maximal length, using the same constant  $c$  and  $c' := C' := 2c^2$ . Also, we use different conventions regarding the case when  $X^{e_m-1}$  is empty. (Our convention seems more natural to us, though it almost implies  $X^0 \neq \emptyset$ .)

Parusinski's version of the definition of a Lipschitz stratification is the following:

**Definition 1.2.4** (Lipschitz stratifications). Let  $c > 1$  ( $c \in \mathcal{R}$ ) be given. A definable stratification  $\mathcal{X}$  is a *Lipschitz stratification* if there exists a  $C \in \mathcal{R}$  such that  $\mathcal{X}$  satisfies the Mostowski Conditions for  $(c, 2c^2, 2c^2, 2c, C)$ .

A priori, this notion seems to depend on the choice of  $c$ . However, it turns out that different choices of  $c$  yield equivalent notions; this follows e.g. from [Par1, Proposition 1.5].

That we have  $c' = C'$  in Definition 1.2.4 means that Condition (2) in Definition 1.2.3 uniquely specifies the set  $\{e_1, e_2, \dots, e_m\}$  in terms of the initial point  $a^0$  and the length  $m$  of the chain. However, a side effect of identifying  $c'$  with  $C'$  is that the strength of the condition is not a monotone function in the value of the constant: a chain for some  $c' = C'$  might neither stay a chain when making  $c' = C'$  bigger, nor when making them smaller. This has various disadvantages, the main one for us being that only monotone conditions can nicely be simplified by reformulating them in an elementary extension. Another consequence is that one has to be quite precise about the relations between the various constants:  $c$  vs.  $2c^2$  vs.  $2c$ .

In contrast, the following two equivalent characterizations are monotone in  $c$  and  $C$  in the above sense, and they are much more robust with respect to small modifications of Definition 1.2.3.

**Proposition 1.2.5** (Characterizations of Lipschitz stratifications). *The following conditions on an  $\mathcal{L}$ -definable stratification  $\mathcal{X}$  are equivalent:*

1.  $\mathcal{X}$  is a Lipschitz stratification (in the sense of Definition 1.2.4).
2. For every  $c \in \mathcal{R}$ , there exists a  $C \in \mathcal{R}$  such that  $\mathcal{X}$  satisfies the Mostowski Conditions for  $(c, c, C, C, C)$ .
3. For every  $c \in \mathcal{R}$ , there exists a  $C \in \mathcal{R}$  such that  $\mathcal{X}$  satisfies the Mostowski Conditions for  $(c, c, 1, \frac{1}{c}, C)$ .

The monotonicity in  $c$  and  $C$  means that both (2) and (3) in the proposition can be considered as statements about big  $c$  and  $C$ , namely: “No matter how big  $c$  is, the Mostowski Conditions hold for all sufficiently big  $C$ .”

Characterization (2) imposes conditions only on very few chains: since  $C$  can be assumed to be big compared to  $c$ , for most points  $a^0 \in X$ , neither of the two inequalities in Definition 1.2.3 (2) holds, hence forbidding those  $a^0$  as starting points of chains. In contrast, *every* sequence of points in decreasing skeletons is relevant in (3) for some  $c$ . (Note that putting  $C' = 1$  makes the first condition of Definition 1.2.3 (2) trivially true.) For these reasons, the implications (3)  $\Rightarrow$  (1)  $\Rightarrow$  (2) are very easy to prove, assuming that we read Definition 1.2.4 as “for every  $c > 1$  there exists  $C$ ” (which we can, using the result that it is independent of  $c$ ). The proof (2)  $\Rightarrow$  (3) is harder; this will follow from Proposition 1.6.11. In fact, this is a good example of a proof which becomes much easier after translating the statements to valuative ones in an elementary extension.

*Proof of Proposition 1.2.5 (3)  $\Rightarrow$  (1).* Let  $c > 1$  be given (from Definition 1.2.4). Then (3) yields a  $C$  such that the Mostowski Conditions hold for  $(2c^2, 2c^2, 1, \frac{1}{2c^2}, C)$ . Thus they also hold for  $(c, 2c^2, 2c^2, 2c, C)$ .  $\square$

*Proof of Proposition 1.2.5 (1)  $\Rightarrow$  (2).* Let  $c$  be given (from (2)); without loss,  $c > 1$ . By Definition 1.2.4, there exists  $C$  such that the Mostowski Conditions hold for  $(c, 2c^2, 2c^2, 2c, C)$ . Hence they also hold for  $(c, c, C', C', C')$ , where  $C' := \max\{C, 2c^2\}$ .  $\square$

*Remark 1.2.6.* It is possible to translate the valuative proof of (2)  $\Rightarrow$  (3) given in Subsection 1.7 into a “conventional” proof within the original field  $\mathcal{R}$ ; we leave the details of this to the interested reader as an exercise. Such a translation in particular yields how, given a function  $f_{(2)}: c \mapsto C$  witnessing (2), one obtains a function  $f_{(3)}: c \mapsto C$  witnessing (3). Roughly,  $f_{(3)} = \underbrace{(f_{(2)} \circ g) \circ \cdots \circ (f_{(2)} \circ g)}_{\dim X \text{ times}}$  for some simple function  $g$ .

### 1.3 Uniform families of Lipschitz stratifications

As mentioned in the introduction, we will obtain Lipschitz stratifications uniformly in families, in a very strong sense. We now make this precise.

**Notation 1.3.1** (Definable families). For the whole subsection, we fix a definable set  $Q$  (say, a subset of  $\mathcal{R}^N$ ); all definable families are parametrized by  $Q$ : A *definable family* of subsets of  $\mathcal{R}^n$  is simply a definable subset  $X \subseteq \mathcal{R}^n \times Q$ , where we write

$$X_q := \{x \in \mathcal{R}^n : (x, q) \in X\}$$

for the fiber at  $q \in Q$ .

We also define families of stratifications in the obvious way:

**Definition 1.3.2.** Suppose that  $X$  is a definable family of  $d$ -dimensional subsets of  $\mathcal{R}^n$  (for some fixed  $d \leq n$ ). A *definable family of stratifications* of  $X$  is a tuple  $\mathcal{X} = (X^i)_{0 \leq i \leq d}$  of families of definable sets such that for each  $q \in Q$ ,  $\mathcal{X}_q := (X_q^i)_{0 \leq i \leq d}$  is a stratification of  $X_q$ ;  $\mathcal{X}$  is a *definable family of Lipschitz stratifications* if each  $\mathcal{X}_q$  is a Lipschitz stratification.

The more interesting concept is that of a family of stratifications that are *uniformly* Lipschitz; this says that the constant  $C$  appearing in the definition of Lipschitz stratifications can be chosen uniformly for the entire family. Here is the precise definition.

**Definition 1.3.3** (Uniformly Lipschitz stratifications). A definable family  $\mathcal{X}$  of stratifications (of a definable family  $X$  of sets) is a family of *uniformly Lipschitz stratifications* if one of the following equivalent conditions holds:

1. For every  $c \in \mathcal{R}$  there exists a  $C \in \mathcal{R}$  such that for every  $q \in Q$ ,  $\mathcal{X}_q$  satisfies the Mostowski Conditions for  $(c, 2c^2, 2c^2, 2c, C)$ .
2. For every  $c \in \mathcal{R}$  there exists a  $C \in \mathcal{R}$  such that for every  $q \in Q$ ,  $\mathcal{X}_q$  satisfies the Mostowski Conditions for  $(c, c, C, C, C)$ .
3. For every  $c \in \mathcal{R}$  there exists a  $C \in \mathcal{R}$  such that for every  $q \in Q$ ,  $\mathcal{X}_q$  satisfies the Mostowski Conditions for  $(c, c, 1, \frac{1}{c}, C)$ .

The above proofs of the non-uniform implications  $(3) \Rightarrow (1) \Rightarrow (2)$  (of Proposition 1.2.5) also work without modification in the uniform case. The implication  $(2) \Rightarrow (3)$  is re-stated as (a part of) Proposition 1.6.11 and will be proved in Subsection 1.7.

*Remark 1.3.4.* The reader may have noticed that in Definition 1.3.3 (1), we wrote “for every  $c$ ”, instead of fixing a  $c > 1$ , as in Definition 1.2.4. We believe that also the *a priori* weaker versions with fixed  $c$  are equivalent, but we didn’t check that carefully.

Now we can finally state the full version of the main result of this paper.

**Theorem 1.3.5** (Uniformly Lipschitz stratifications). *Fix a power-bounded real closed field  $\mathcal{R}$  in a language  $\mathcal{L}$ . Suppose that  $X$  is an  $\mathcal{L}$ -definable family of closed,  $d$ -dimensional subsets of  $\mathcal{R}^n$  (i.e.,  $X$  is an  $\mathcal{L}$ -definable subset of  $\mathcal{R}^n \times Q$ , whose fibers  $X_q \subseteq \mathcal{R}^n$  are closed and  $d$ -dimensional, for  $q \in Q$ ). Then there exists an  $\mathcal{L}$ -definable family  $\mathcal{X} = (X^i)_{0 \leq i \leq d}$  of uniformly Lipschitz stratifications of  $X$  (in the sense of Definition 1.3.3).*

## 1.4 Enlarging the model

The conditions in Definition 1.3.3 are clearly first order properties. Therefore, when proving the implication  $(2) \Rightarrow (3)$  and the existence of uniformly Lipschitz stratifications, we may work in an elementary extension. More precisely, we will take the point of view that without loss,  $\mathcal{R}$  itself is already large, so that in particular, it possesses an elementary substructure  $\mathcal{R}_0 \not\cong \mathcal{R}$ . It is not difficult to check that the convex closure of  $\mathcal{R}_0$  within  $\mathcal{R}$  is a (non-trivial) valuation ring of  $\mathcal{R}$ ; we denote it by  $\mathcal{O}_{\mathcal{R}}$ . Intuitively, elements of  $\mathcal{R} \setminus \mathcal{O}_{\mathcal{R}}$  may be regarded as “infinite” and elements in the maximal ideal of  $\mathcal{O}_{\mathcal{R}}$  as “infinitesimal”; more generally, bigger valuation means smaller order of magnitude, where two elements are considered as having the same order of magnitude if they differ at most by a factor from  $\mathcal{R}_0^\times$ . (Note that even if  $\mathcal{R}_0$  is non-archimedean, we consider all its elements as having the same order of magnitude.)

It is a standard technique to study  $\mathbb{R}$  by passing to an elementary extension. This implicitly uses the above valuation, but one usually considers definability only in the original language  $\mathcal{L}$ . In contrast, in this paper, we will explicitly consider  $\mathcal{R}$  as a structure in the language expanded by a predicate for  $\mathcal{O}_{\mathcal{R}}$ . The model theory of such structures has

been studied by van den Dries and Lewenberg [DL, Dri1], and a key ingredient to our proof of existence of Lipschitz stratifications builds on those results.

**Notation 1.4.1** (Valuation). For the remainder of Section 1, we suppose that we have two  $\mathcal{L}$ -structures  $\mathcal{R}_0 \preceq \mathcal{R}$ . We write  $\mathcal{O}_{\mathcal{R}} \subseteq \mathcal{R}$  for the valuation ring obtained as the convex closure of  $\mathcal{R}_0$  in  $\mathcal{R}$ , i.e.,

$$\mathcal{O}_{\mathcal{R}} = \{a \in \mathcal{R} : -b < a < b \text{ for some } b \in \mathcal{R}_0\}.$$

We write  $\Gamma := \mathcal{R}^{\times} / \mathcal{O}_{\mathcal{R}}^{\times}$  for the value group and  $\text{val}: \mathcal{R} \rightarrow \Gamma \cup \{\infty\}$  for the valuation. Let  $\mathcal{L}_{\text{val}}$  be the expansion of the language  $\mathcal{L}$  by a predicate for  $\mathcal{O}_{\mathcal{R}}$ .

In [DL, Dri1], the language  $\mathcal{L}_{\text{val}}$  is denoted by  $\mathcal{L}_{\text{convex}}$  and an  $\mathcal{L}_{\text{val}}$ -structure obtained from  $o$ -minimal structures  $\mathcal{R}_0 \preceq \mathcal{R}$  as in Notation 1.4.1 is called “ $T$ -convex”, where  $T$  is the theory of  $\mathcal{R}$  as an  $\mathcal{L}$ -structure. It has been proved in [DL] that being  $T$ -convex is an elementary property, i.e., that for any  $\mathcal{L}_{\text{val}}$ -structure  $\mathcal{R}'$  which is elementarily equivalent to  $\mathcal{R}$ , the valuation ring  $\mathcal{O}_{\mathcal{R}'} \subseteq \mathcal{R}'$  is also the convex closure of an  $\mathcal{L}$ -elementary substructure  $\mathcal{R}'_0 \preceq \mathcal{R}'$ . In particular, we can assume that  $\mathcal{R}$  is sufficiently saturated as an  $\mathcal{L}_{\text{val}}$ -structure (by possibly further enlarging both,  $\mathcal{R}_0$  and  $\mathcal{R}$ ); this will be useful for (model theoretic) compactness arguments.

**Assumption 1.4.2.** For the remainder of the paper, we assume that  $\mathcal{R}$  is sufficiently saturated, as a structure in the language  $\mathcal{L}_{\text{val}}$ .

(To be precise, we will need  $\mathcal{R}$  to be  $|\mathcal{L}_{\text{val}}|^{+}$ -saturated.)

*Remark 1.4.3.* The result that being  $T$ -convex is an elementary property is only used for convenience, to be able to fix  $\mathcal{R}$  once and for all. In reality, in those parts of the paper where we do need to consider elementary extensions of  $\mathcal{R}$  as an  $\mathcal{L}_{\text{val}}$ -structure (namely Theorem 1.6.7 and its proof), we do not need  $\mathcal{O}_{\mathcal{R}}$  to be the convex closure of an elementary substructure.

## 1.5 Valuative Notation

We fix some notation related to the newly introduced valuation. First of all, note that even when working with the language  $\mathcal{L}_{\text{val}}$ , all stratifications we consider are  $\mathcal{L}$ -definable (instead of  $\mathcal{L}_{\text{val}}$ -definable), and the notions of definable connectedness and definable manifolds still refer to the language  $\mathcal{L}$ .

Now that we have a valuation, it is useful to also have valuative versions of norms and distances; we use the following notation. Note that by [Dri1, Proposition 4.3], the value group  $\Gamma$  (with the induced structure) is  $o$ -minimal. In particular, suprema and infima of definable subsets of  $\Gamma$  exist.

**Notation 1.5.1** (Valuative norms and distances). For  $a = (a_1, \dots, a_n) \in \mathcal{R}^n$ , we set  $\text{val}(a) := \min_i \text{val}(a_i) = \text{val}(\|a\|)$ . If in addition, we have a definable set  $X \subseteq \mathcal{R}^n$ , we set  $\text{valdist}(a, X) := \sup_{x \in X} \text{val}(a - x) = \text{val}(\text{dist}(a, X))$ , where  $\text{valdist}(a, \emptyset) := -\infty$ . For a matrix  $M = (m_{ij})_{ij}$ , we set  $\text{val}(M) := \min_{i,j} \text{val}(m_{ij})$ .

We recall some facts about those definitions.

**Lemma 1.5.2.** *Let  $M$  and  $N$  be matrices with coefficients on  $\mathcal{R}$ . Then we have the following (where some statements implicitly impose conditions on the numbers of rows/columns of  $M$  and  $N$ ):*

1. We have  $\text{val}(MN) \geq \text{val}(M) + \text{val}(N)$  (and in particular  $\text{val}(Ma) \geq \text{val}(M) + \text{val}(a)$  for  $a \in \mathcal{R}^n$ ).
2. The matrix  $M$  lies in  $\text{GL}_n(\mathcal{O}_{\mathcal{R}})$  iff  $M \in \text{GL}_n(\mathcal{R})$  and we have both  $\text{val}(M) \geq 0$  and  $\text{val}(M^{-1}) \geq 0$ .
3. If  $M \in \text{GL}_n(\mathcal{O}_{\mathcal{R}})$ , then  $\text{val}(MN) = \text{val}(N)$  (and in particular  $\text{val}(Ma) = \text{val}(a)$  for  $a \in \mathcal{R}^n$ ).
4. We have  $\text{val}(M) = \text{val}(\|M\|)$ , where  $\|M\|$  is the operator norm of  $M$  (or, in fact, any other of the usual norms).

*Proof.* (1) Easy computation.

(2) Clear.

(3) Follows from (1) and (2).

(4) We have

$$\text{val}(\|M\|) + \text{val}(\|a\|) \stackrel{(\star)}{\leq} \text{val}(\|Ma\|) \stackrel{(\star\star)}{\geq} \text{val}(M) + \text{val}(\|a\|) \quad (1.2)$$

(using the definition of the operator norm to get  $(\star)$ , and using (1) to get  $(\star\star)$ ). By choosing  $a$  such that  $\|Ma\| = \|M\| \cdot \|a\|$ , we obtain an equality at  $(\star)$  and hence  $\text{val}(\|M\|) \geq \text{val}(M)$ . To obtain  $\text{val}(\|M\|) \leq \text{val}(M)$ , we choose an  $a$  which yields an equality at  $(\star\star)$ : if the  $j$ -th column of  $M$  has an entry  $m_{ij}$  satisfying  $\text{val}(M) = \text{val}(m_{ij})$ , then we can take  $a$  to be the  $j$ -th standard basis vector.  $\square$

All balls we consider in this paper are valuative balls. We use the following notation.

**Notation 1.5.3** (Balls). Given  $a \in \mathcal{R}^n$  and  $\lambda \in \Gamma$ , we write

$$B_{>\lambda}(a) := \{x \in \mathcal{R}^n : \text{val}(x - a) > \lambda\} \quad \text{and} \\ B_{\geq\lambda}(a) := \{x \in \mathcal{R}^n : \text{val}(x - a) \geq \lambda\}$$

for the open and closed ball of valuative radius  $\lambda$ .

## 1.6 Valuative Lipschitz Stratifications

The valuation allows us to simplify Conditions (2) and (3) of Definition 1.3.3 in the “usual non-standard analysis way”. This leads to a valuative version of chains and Lipschitz stratifications, which we now introduce.

**Definition 1.6.1** (val-chains). Fix a definable stratification  $\mathcal{X} = (X^i)_i$  of a definable set  $X \subseteq \mathcal{R}^n$ . A *plain val-chain* (in  $\mathcal{X}$ ) is a sequence of points  $a^0, \dots, a^m$  ( $m \geq 0$ ) with  $a^\ell \in \mathring{X}^{e_\ell}$ ,  $e_0 > e_1 > \dots > e_m$  such that for all  $1 \leq \ell \leq m$ , we have

$$\lambda_\ell := \text{val}(a^0 - a^\ell) > \text{valdist}(a^0, X^{e_{\ell-1}}) \quad \text{and} \quad (1.3)$$

$$\text{val}(a^0 - a^\ell) = \text{valdist}(a^0, X^{e_{\ell-1}-1}). \quad (1.4)$$

An *augmented val-chain* (in  $\mathcal{X}$ ) is a sequence of points  $a^0, \dots, a^m$  ( $m \geq 1$ ) with  $a^\ell \in \mathring{X}^{e_\ell}$ ,  $e_0 = e_1 > \dots > e_m$  such that (1.3) holds for  $1 \leq \ell \leq m$  and (1.4) holds for  $2 \leq \ell \leq m$ . By a *val-chain*, we mean either a plain or an augmented one.

The numbers  $e_\ell$  (for  $0 \leq \ell \leq m$ ) are the *dimensions* of the val-chain, and its *distances* are the valuations  $\lambda_\ell$  ( $1 \leq \ell \leq m$ ) together with  $\lambda_{m+1} := \text{valdist}(a^0, X^{e_m-1})$  (which might be  $-\infty$ ).

*Remark 1.6.2.* An equivalent way of characterizing a plain val-chain is the following. Choose any point  $a^0$  in any skeleton  $\mathring{X}^{e_0}$ . Then choose the remaining points  $a^\ell$  ( $1 \leq \ell \leq m$ ) in skeletons  $\mathring{X}^{e_\ell}$  as close as possible to  $a^0$  in the valuative sense, where  $\{e_1, \dots, e_m\}$  consists of the  $m$  biggest elements of the set  $\{j \leq e_0 : \text{valdist}(a^0, X^j) > \text{valdist}(a^0, X^{j-1})\}$ .

*Remark 1.6.3.* By (1.3), we have  $\lambda_1 > \dots > \lambda_{m+1}$ . This implies

$$\begin{aligned} \text{val}(a^k - a^\ell) &= \text{val}(a^0 - a^\ell) && \text{for } 0 \leq k < \ell \leq m \text{ and} \\ \text{valdist}(a^k, X^j) &= \text{valdist}(a^0, X^j) && \text{for } 0 \leq k \leq m \text{ and } j < e_k. \end{aligned}$$

In particular, if  $(a^\ell)_{0 \leq \ell \leq m}$  is a val-chain, then any sub-sequence of the form  $(a^\ell)_{k \leq \ell \leq m'}$  for  $0 \leq k \leq m' \leq m$  is also a val-chain (which is always plain if  $k \geq 1$ ). Moreover, if  $(a^\ell)_{0 \leq \ell \leq m}$  is an augmented val-chain, then  $a^0, a^2, a^3, \dots, a^{m'}$  is a plain val-chain (for  $1 \leq m' \leq m$ ).

**Definition 1.6.4** (valuative Mostowski Conditions). Let  $\mathcal{X}$  be a definable stratification and  $(a^\ell)_{0 \leq \ell \leq m}$  a val-chain with distances  $\lambda_\ell$ . By the *valuative Mostowski Condition at  $(a^\ell)_\ell$* , we mean one of the following two properties of  $\mathcal{X}$ . If  $(a^\ell)_\ell$  is a plain val-chain, the condition is

$$\text{val}((1 - P_{a^0})P_{a^1} \cdots P_{a^m}) \geq \lambda_1 - \lambda_{m+1}; \quad (\text{vm1})$$

if  $(a^\ell)_\ell$  is an augmented val-chain, the condition is

$$\text{val}((P_{a^0} - P_{a^1})P_{a^2} \cdots P_{a^m}) \geq \lambda_1 - \lambda_{m+1}. \quad (\text{vm2})$$

In the case  $\lambda_{m+1} = -\infty$ , the conditions are supposed to be read as “ $\text{val}(\dots) = \infty$ ”, i.e., the composition of the maps is 0. If  $X^0 \neq \emptyset$ , then  $\lambda_{m+1} = -\infty$  implies  $a^m \in X^0$ , and we anyway have  $P_{a^m} = 0$ . However, if  $X^0 = \emptyset$ , then this is a very strong condition, so as for classical Lipschitz stratifications, one can almost never have  $X^0 = \emptyset$ .

**Definition 1.6.5** (valuative Lipschitz stratifications). A definable stratification  $\mathcal{X} = (X^i)_i$  (of a definable set  $X \subseteq \mathcal{R}^n$ ) is a *valuative Lipschitz stratification* if it satisfies the valuative Mostowski conditions at every val-chain.

*Remark 1.6.6.* Whether or not in Definition 1.6.5 one considers val-chains consisting of a single point (i.e., with  $m = 0$ ) does not make a difference, since in that case, (vm1) is trivially true (since the right hand side is 0).

This is the notion of stratification we will use in the main proof in this paper, i.e., we will prove the existence of valuative Lipschitz stratifications. We will do this not only for  $\mathcal{L}$ -definable sets  $X$ , but also for sets definable with additional parameters from  $\mathcal{R}$ . By usual compactness arguments, this implies a family version of the result, and that in turn implies Theorem 1.3.5 about the existence of uniformly Lipschitz stratifications. The details of these implications are given at the end of this subsection.

**Theorem 1.6.7** (valuative Lipschitz stratifications). *Suppose that  $\mathcal{R}$  is a real closed field which is o-minimal and power-bounded as a structure in a language  $\mathcal{L}$ , and suppose that  $\mathcal{L}_{\text{val}}$  is an expansion of  $\mathcal{L}$  by a predicate for the convex closure of an elementary substructure  $\mathcal{R}_0 \not\cong \mathcal{R}$  (so  $\mathcal{R}$  is  $T$ -convex in the sense of [DL]). Suppose that  $X \subseteq \mathcal{R}^n$  is a closed,  $\mathcal{L}(A)$ -definable set for some parameter set  $A \subseteq \mathcal{R}$ . Then there exists an  $\mathcal{L}(A)$ -definable valuative Lipschitz stratification of  $X$ .*

The notion of a valuative Lipschitz stratification is just a reformulation of Proposition 1.2.5 (2) using the valuation, as we shall see below. To provide a similar reformulation of Proposition 1.2.5 (3), we introduce “weak val-chains”. (Those are only used here and in the next subsection.) Roughly, a weak val-chain is the same as a val-chain, except that additional intermediate points in skeletons of intermediate dimensions are allowed.

**Definition 1.6.8.** A *weak val-chain* (plain or augmented) is the same as a val-chain, except that the (strict) inequality (1.3) is replaced by a weak one:

$$\text{val}(a^0 - a^\ell) \geq \text{valdist}(a^0, X^{e_\ell-1}). \quad (1.5)$$

The val-chains from Definition 1.6.1 will sometimes be called *strict val-chains*, to emphasize the difference. The dimensions  $e_\ell$ , the distances  $\lambda_\ell$  and the valuative Mostowski Conditions are defined in the same way as for strict val-chains.

*Remark 1.6.9.* In fact, imposing (1.5) is necessary only for  $\ell = 1$  in augmented val-chains; in all other cases, (1.5) follows from (1.4) and  $X^{e_{\ell-1}-1} \supseteq X^{e_\ell-1}$ .

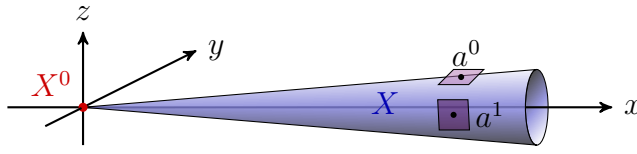


Figure 1.1: This is a Lipschitz stratification but not a valuative Lipschitz stratification, since the two tangent spaces of the augmented val-chain  $a^0 = (1, 0, r)$ ,  $a^1 = (1, r, 0)$  are too far apart from each other; see Example 1.6.12.

*Remark 1.6.10.* For weak val-chains, we only have weak inequalities  $\lambda_1 \geq \dots \geq \lambda_{m+1}$ , and a weak val-chain is strict iff all those inequalities between the  $\lambda_i$  are strict.

**Proposition 1.6.11.** *Suppose that  $\mathcal{X}$  is an  $\mathcal{L}$ -definable family of stratifications (of an  $\mathcal{L}$ -definable family  $X$  of subsets of  $\mathcal{R}^n$ ), parametrized by  $q \in Q$  for some  $\mathcal{L}$ -definable  $Q \subseteq \mathcal{R}^N$ . Then following conditions are equivalent:*

- (2) *Condition (2) of Definition 1.3.3.*
- (3) *Condition (3) of Definition 1.3.3.*
- (2') *For each  $q$ ,  $\mathcal{X}_q$  is a valuative Lipschitz stratification (in the sense of Definition 1.6.5).*
- (3') *For each  $q$ ,  $\mathcal{X}_q$  satisfies the valuative Mostowski Condition at every weak val-chain.*

Note that for the implications  $(x) \Rightarrow (x')$  to hold ( $x = 2, 3$ ), it is essential that  $\mathcal{X}$  is  $\mathcal{L}$ -definable without parameters outside of  $\mathcal{R}_0$ ; cf. Remark 1.7.2 below. However, the implications  $(x') \Rightarrow (x)$  seem to hold even for  $\mathcal{L}(A)$ -definable  $\mathcal{X}$ , where  $A \subseteq \mathcal{R}$ . (We did not check the details.)

*Example 1.6.12.* If  $X \subseteq \mathcal{R}^3$  is the cone defined by  $r^2x^2 = y^2 + z^2$  for some  $r \in \mathcal{R}$  of strictly positive valuation, then  $X^0 = X^1 = \{(0, 0, 0)\}$  defines a Lipschitz stratification of  $X$ , but not a valuative Lipschitz stratification; see Figure 1.1.

As promised, here is the precise argument on how to deduce Theorem 1.3.5 from Theorem 1.6.7 and Proposition 1.6.11.

*Proof of Theorem 1.3.5.* Let an  $\mathcal{L}$ -definable family  $X$  of closed  $d$ -dimensional subsets of  $\mathcal{R}^n$  be given (parametrized by  $q \in Q$ ); we would like to find a family  $\mathcal{X}$  of uniformly Lipschitz stratifications (Definition 1.3.3) of  $X$ . By Proposition 1.6.11, this is equivalent to  $\mathcal{X}_q$  being a valuative Lipschitz stratification for each  $q \in Q$ .

For each  $q \in Q$ , Theorem 1.6.7 provides an  $\mathcal{L}(q)$ -definable valuative Lipschitz stratification  $\mathcal{X}_q$  of  $X_q$ . By a standard compactness argument, we may assume that those  $\mathcal{X}_q$  are definable uniformly in  $q$ , i.e., that they are the fibers of an  $\mathcal{L}$ -definable family  $\mathcal{X}$  of stratifications, as desired.

The details of the compactness argument are as follows. For each  $\hat{q} \in Q$ , there exist  $\mathcal{L}$ -formulas  $\phi_{\hat{q}}^i(x, y)$  ( $0 \leq i \leq d$ ) such that the  $\phi_{\hat{q}}^i(x, \hat{q})$  define a valuative Lipschitz stratification of  $X_{\hat{q}}$ . Fix one  $\hat{q}$  and consider the set  $U_{\hat{q}}$  of those  $q \in Q$  such that  $(\phi_{\hat{q}}^i(x, q))_i$  defines a valuative Lipschitz stratification of  $X_q$ . Since being a valuative Lipschitz stratification is a first order property,  $U_{\hat{q}}$  is  $\mathcal{L}$ -definable. Finitely many sets  $U_{\hat{q}_1}, \dots, U_{\hat{q}_\ell}$  suffice to cover  $Q$ , since otherwise, the complements  $Q \setminus U_{\hat{q}}$  would form a partial type, which is satisfied by some  $q_0 \in Q$  (since  $\mathcal{R}$  is sufficiently saturated by Assumption 1.4.2), contradicting  $q_0 \in U_{q_0}$ . Now use the formulas  $\phi_{\hat{q}_1}^i(x, y), \dots, \phi_{\hat{q}_\ell}^i(x, y)$  to define  $\mathcal{X}$ ; more precisely, given  $q \in Q$ , let  $X_q^i$  be defined by  $\phi_{\hat{q}_j}^i(x, q)$ , where  $j$  is minimal with  $q \in U_{\hat{q}_j}$ .  $\square$

## 1.7 Equivalence of various definitions

We will now prove Proposition 1.6.11. More precisely, we will prove the following implications:

$$\begin{array}{ccc} (2) & \iff & (2') \\ & & \Downarrow \\ (3) & \iff & (3') \end{array} \quad (1.6)$$

Note that the right hand  $\uparrow$  is trivial, and anyway, we already proved  $\uparrow$  on the left hand side. Both horizontal  $\iff$  are simple applications of a standard method from non-standard analysis which we recall now:

**Lemma 1.7.1** (Translating: with/without valuation). *Suppose that  $Z$  is  $\mathcal{L}$ -definable and that  $f, g: Z \rightarrow \mathcal{R}_{\geq 0}$  are two  $\mathcal{L}$ -definable functions. Then the following are equivalent:*

1. *For every  $c \in \mathcal{R}_{\geq 0}$ , there exists  $C \in \mathcal{R}_{\geq 0}$  such that for every  $z \in Z$ ,  $f(z) \leq c$  implies  $g(z) \leq C$ .*
2. *For every  $z \in Z$ ,  $\text{val}(f(z)) \geq 0$  implies  $\text{val}(g(z)) \geq 0$ .*

*Proof.* Statement (1) is an  $\mathcal{L}$ -sentence in  $\mathcal{R}$ , so it is equivalent to the same sentence in  $\mathcal{R}_0$ ; we will use this version of (1). For the proof of this lemma, we assume without loss that all elements of  $\mathcal{R}_0$  are constants of  $\mathcal{L}$ .

(1)  $\Rightarrow$  (2): Let  $z_0 \in Z$  be given such that  $\text{val}(f(z_0)) \geq 0$ . Then  $f(z_0) \leq c$  for some  $c \in \mathcal{R}_0$  (by definition of the valuation). By (1) in  $\mathcal{R}_0$ , there exists a  $C \in (\mathcal{R}_0)_{\geq 0}$  such that  $\mathcal{R}_0 \models \forall z \in Z : (f(z) \leq c \rightarrow g(z) \leq C)$ . This sentence also holds in  $\mathcal{R}$  (where  $c, C$  are considered as constants from  $\mathcal{L}$ ), hence  $f(z_0) \leq c$  implies  $g(z_0) \leq C$ . This in turn implies  $\text{val}(g(z_0)) \geq 0$ .

(2)  $\Rightarrow$  (1): Let  $c \in (\mathcal{R}_0)_{\geq 0}$  be given. We consider “ $\exists C : \forall z \in Z : (f(z) \leq c \rightarrow g(z) \leq C)$ ” as a sentence where  $c$  is a constant from  $\mathcal{L}$ ; it suffices to prove that this sentence holds in  $\mathcal{R}$ . But indeed: since  $f(z) \leq c$  implies  $\text{val}(f(z)) \geq 0$ , we have  $\text{val}(g(z)) \geq 0$ , so we can take any  $C \in \mathcal{R}_{\geq 0}$  of negative valuation.  $\square$

*Remark 1.7.2.* For this lemma to be true, it is important that  $Z$ ,  $f$  and  $g$  are definable using parameters only from  $\mathcal{R}_0$ .

*Remark 1.7.3.* An easy special case of Lemma 1.7.1 is the one with  $f = 0$ : An  $\mathcal{L}$ -definable function  $g: Z \rightarrow \mathcal{R}$  is bounded iff it satisfies  $\text{val}(g(z)) \geq 0$  for all  $z \in Z$ .

*Proof of Proposition 1.6.11, (2)  $\iff$  (2').* This is just a straight-forward application of Lemma 1.7.1. The details are as follows.

Let the family  $\mathcal{X} = (X^i)_i$  of stratifications be fixed (parametrized by  $q \in Q$ ), and let  $Z$  be the set of all tuples  $z$  of the form  $(q, (a^\ell)_{0 \leq \ell \leq m})$ , with  $q \in Q$ ,  $a^\ell \in \mathring{X}_q^{e_\ell}$ ,  $e_0 \geq e_1 > e_2 > \dots > e_m$ , and  $m \geq 1$ . (We consider  $Z$  as an  $\mathcal{L}$ -definable set.) Given  $c, C \in \mathcal{R}$ , such a  $z \in Z$  witnesses that our family  $\mathcal{X}$  violates the Mostowski conditions for  $(c, c, C, C, C)$  if

- $(a^\ell)_{0 \leq \ell \leq m}$  is a chain in  $\mathcal{X}_q$  (either plain or augmented), i.e.:

$$\frac{\|a^0 - a^\ell\|}{\text{dist}(a^0, X_q^{e_\ell})} < c \quad \text{for } \ell = \begin{cases} 1, \dots, m & \text{if } e_0 > e_1 \\ 2, \dots, m & \text{if } e_0 = e_1 \end{cases} \quad (1.7)$$

$$\frac{\text{dist}(a^0, X_q^{i-1})}{\text{dist}(a^0, X_q^i)} \geq C \quad \text{for } i \in \begin{cases} \{e_1, \dots, e_m\} & \text{if } e_0 > e_1 \\ \{e_2, \dots, e_m\} & \text{if } e_0 = e_1 \end{cases} \quad (1.8)$$

$$\frac{\text{dist}(a^0, X_q^{i-1})}{\text{dist}(a^0, X_q^i)} < c \quad \text{for } e_m \leq i \leq e_0, i \notin \{e_0, \dots, e_m\} \quad (1.9)$$

$$\frac{\text{dist}(a^0, X_q^{e_1-1})}{\|a^0 - a^1\|} \geq C \quad \text{in the case } e_0 = e_1 \quad (1.10)$$

- and either (m1) or (m2) is violated:

$$\frac{\|(1 - P_{a^0})P_{a^1} \dots P_{a^m}\| \text{dist}(a^0, X_q^{e_m-1})}{\|a^0 - a^1\|} \geq C \quad \text{in the case } e_0 > e_1 \quad (1.11)$$

$$\frac{\|(P_{a^0} - P_{a^1})P_{a^2} \dots P_{a^m}\| \text{dist}(a^0, X_q^{e_m-1})}{\|a^0 - a^1\|} \geq C \quad \text{in the case } e_0 = e_1. \quad (1.12)$$

Define  $f(z)$  to be the maximum of all the left hand sides of (1.7) and (1.9) (for all  $\ell$  and  $i$ ) and  $g(z)$  to be the minimum of all the (relevant) left hand sides of (1.8), (1.10), (1.11), (1.12). Then Condition (2) of Definition 1.3.3 is exactly (1) of Lemma 1.7.1, and (2) of Lemma 1.7.1 says that there is no  $z \in Z$  satisfying the following modification of (1.7) – (1.12): replace “ $\star < c$ ” by “ $\text{val}(\star) \geq 0$ ” and “ $\star \geq C$ ” by “ $\text{val}(\star) < 0$ ”.

In this modified version, (1.7) – (1.10) state that  $(a^i)_i$  is a val-chain and (1.11), (1.12) state that the corresponding valuative Mostowski Condition is violated. Thus Lemma 1.7.1 (2) expresses that  $\mathcal{X}$  is a valuative Lipschitz stratification.  $\square$

*Proof of Proposition 1.6.11, (3)  $\iff$  (3').* The proof is almost the same as for (2)  $\iff$  (2'). The only differences are that (1.8) disappears and that (1.10) is replaced by

$$\frac{\|a^0 - a^1\|}{\text{dist}(a^0, X_q^{e_1-1})} \leq c. \quad (1.13)$$

Lemma 1.7.1 turns (1.13) into (1.5) for  $\ell = 1$ , so we obtain exactly weak val-chains (see also Remark 1.6.9).  $\square$

*Proof of Proposition 1.6.11, (2')  $\implies$  (3').* We assume that every strict val-chain satisfies the valuative Mostowski Conditions, and we have to prove the same for weak val-chains. Let  $a^0, \dots, a^m$  be a weak val-chain with dimensions  $e_i$  and distances  $\lambda_i$ . We do an induction over  $m$ . If this is already a strict val-chain, there is nothing to prove. Otherwise, choose any  $\ell$  such that  $\lambda_\ell = \lambda_{\ell+1}$  ( $1 \leq \ell \leq m$ ). Let us first suppose that  $(a^i)_i$  is a plain (weak) val-chain. Set

$$\begin{aligned} Q &:= (1 - P_{a^0})P_{a^1} \cdots P_{a^{\ell-1}} & \text{and} \\ Q' &:= P_{a^{\ell+1}} \cdots P_{a^m}; \end{aligned}$$

we need to show that

$$\text{val}(QP_{a^\ell}Q') \geq \lambda_1 - \lambda_{m+1}. \quad (1.14)$$

The sub-sequence  $a^0, \dots, a^{\ell-1}, a^{\ell+1}, \dots, a^m$  is still a weak val-chain, and by induction, it satisfies the Mostowski Conditions, i.e.:

$$\text{val}(QQ') \geq \lambda_1 - \lambda_{m+1}.$$

Moreover, we have  $\text{val}(Q) \geq \lambda_1 - \lambda_\ell$  (by the inductive hypothesis for  $a^0, \dots, a^{\ell-1}$ ) and  $\text{val}((1 - P_{a^\ell})Q') \geq \lambda_{\ell+1} - \lambda_{m+1}$  (by the inductive hypothesis for  $a^\ell, \dots, a^m$ ). Combining these three inequalities (and using  $\lambda_\ell = \lambda_{\ell+1}$ ) yields (1.14), since  $QP_{a^\ell}Q' = Q(1 - P_{a^\ell})Q' - QQ'$ .

Now suppose that  $a^0, \dots, a^m$  is an augmented val-chain. If  $\ell \geq 2$ , then the argument is exactly the same as for plain val-chains, with

$$Q = (P_{a^0} - P_{a^1})P_{a^2} \cdots P_{a^{\ell-1}}.$$

In the case  $\ell = 1$ , define  $Q' := P_{a^2} \cdots P_{a^m}$  (as before). The Mostowski conditions for  $a^0, a^2, \dots, a^m$  and  $a^1, a^2, \dots, a^m$  imply  $\text{val}((1 - P_{a^0})Q') \geq \lambda_1 - \lambda_{m+1}$  and  $\text{val}((1 - P_{a^1})Q') \geq \lambda_1 - \lambda_{m+1}$ ; this implies

$$\text{val}((P_{a^0} - P_{a^1})Q') \geq \lambda_1 - \lambda_{m+1},$$

which is what we had to show.  $\square$

## 1.8 A $\mathrm{GL}_n$ -invariant definition

To prove the existence of valuative Lipschitz stratifications, we will use yet another (equivalent) definition, which is more natural in the sense that it is clearly invariant under  $\mathrm{GL}_n(\mathcal{O}_{\mathcal{R}})$ . Note that Definition 1.6.5 (the definition of valuative Lipschitz stratifications) is already pretty close to being  $\mathrm{GL}_n(\mathcal{O}_{\mathcal{R}})$ -invariant, since  $\mathrm{GL}_n(\mathcal{O}_{\mathcal{R}})$  preserves valuations (by Lemma 1.5.2). To make it entirely  $\mathrm{GL}_n(\mathcal{O}_{\mathcal{R}})$ -invariant, one only needs to get rid of the orthogonal projections used to express that certain tangent spaces are close to each other; this is what we will do now.

That valuative Lipschitz stratifications are  $\mathrm{GL}_n(\mathcal{O}_{\mathcal{R}})$ -invariant directly implies that classical Lipschitz stratifications are  $\mathrm{GL}_n(\mathcal{R})$ -invariant; even though this is not a new result, we formulate it as Corollary 1.8.7.

There exists a natural valuative metric on the Grassmannians. It can be defined in many equivalent ways, some of which use orthogonal projections, and others being clearly  $\mathrm{GL}_n(\mathcal{O}_{\mathcal{R}})$ -invariant. We leave the proof of the equivalences to the reader.

**Definition 1.8.1.** For subspaces  $W_1, W_2 \subseteq \mathcal{R}^n$  of the same dimension, set  $\Delta(W_1, W_2) := \mathrm{val}(P_1 - P_2)$ , where  $P_i$  is the orthogonal projection onto  $W_i$ .

**Lemma 1.8.2.** For subspaces  $W_1, W_2 \subseteq \mathcal{R}^n$ , both of dimension  $d$  and for any  $\lambda \in \Gamma$ , the following are equivalent:

1.  $\Delta(W_1, W_2) \geq \lambda$
2. There exist  $\phi_1, \phi_2 \in \mathrm{Hom}(\mathcal{R}^d, \mathcal{R}^n)$  with  $\mathrm{val}(\phi_1 - \phi_2) \geq \lambda$  and  $\mathrm{im} \phi_i = W_i$ .
3. For every  $w_1 \in W_1$  there exists  $w_2 \in W_2$  such that  $\mathrm{val}(w_2 - w_1) \geq \mathrm{val}(w_1) + \lambda$ .

The Mostowski Condition bounding  $\mathrm{val}((1 - P_{a^0})P_{a^1})$  can be considered as the statement that  $\mathbf{T}_{a^0} \overset{\circ}{X}^0$  contains a subspace which is a good approximation of  $\mathbf{T}_{a^1} \overset{\circ}{X}^1$ . The following characterization of valuative Lipschitz stratifications is a generalization of this point of view to arbitrary val-chains; see Figure 1.2 for an overview over all sub-spaces.

**Proposition 1.8.3** (Valuative Lipschitz stratifications using flags). *The following conditions on a definable stratification  $\mathcal{X} = (X^i)_i$  are equivalent:*

1.  $\mathcal{X}$  is a valuative Lipschitz stratification (in the sense of Definition 1.6.5).
2. For every val-chain  $(a^i)_{i \leq m}$  (plain or augmented) with dimensions  $e_i$  and distances  $\lambda_i$ , there exist vector spaces  $V_{k,\ell}$  for  $0 \leq k \leq \ell \leq m$  with the following properties:

$$V_{k,m} \subseteq V_{k,m-1} \subseteq \dots \subseteq V_{k,k+1} \subseteq V_{k,k} = \mathbf{T}_{a^k} \overset{\circ}{X}^{e_k} \quad \text{for } 0 \leq k \leq m \quad (1.15)$$

$$\dim V_{k,\ell} = e_\ell \quad \text{for } 0 \leq k \leq \ell \leq m \quad (1.16)$$

$$\Delta(V_{k,\ell}, V_{k+1,\ell}) \geq \lambda_{k+1} - \lambda_{\ell+1} \quad \text{for } 0 \leq k < \ell \leq m, \quad (1.17)$$

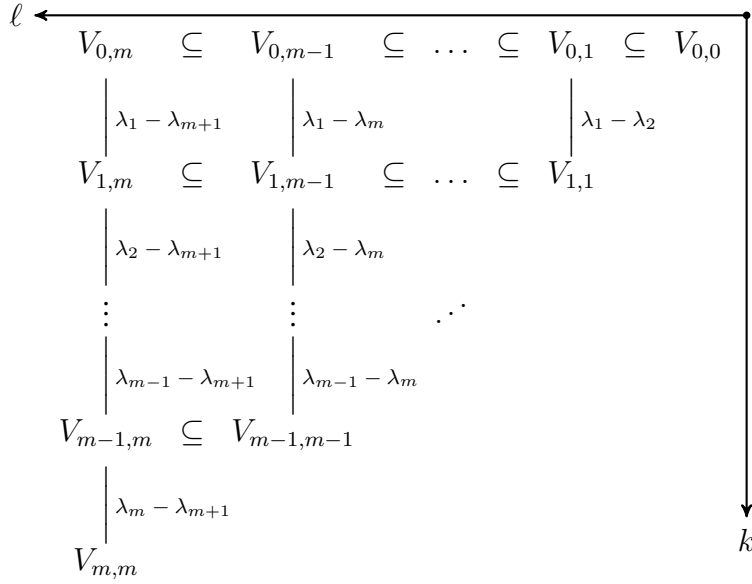


Figure 1.2: Diagrammatic representation of the vector spaces appearing in Proposition 1.8.3 (and Lemma 1.8.5); the labels of the vertical lines indicate the distance between the two corresponding spaces.

*Remark 1.8.4.* As in Definition 1.6.5, the above Condition (2) is trivial for val-chains consisting of a single point.

The proof of (1)  $\Rightarrow$  (2) is easy:

*Proof of Proposition 1.8.3, (1)  $\Rightarrow$  (2).* Given a val-chain  $(a^i)_i$ , we set  $V_{k,\ell} := \text{im}(Q_{k,\ell})$ , where

$$Q_{k,\ell} := \begin{cases} P_{a^k} P_{a^{k+1}} \cdots P_{a^\ell} & \text{if } (a^i)_{k \leq i \leq \ell} \text{ is a plain val-chain} \\ P_{a^k} P_{a^{k+2}} P_{a^{k+3}} \cdots P_{a^\ell} & \text{if } (a^i)_{k \leq i \leq \ell} \text{ is an augmented val-chain.} \end{cases} \quad (1.18)$$

Note that  $(a^i)_{k \leq i \leq \ell}$  is an augmented val-chain iff the entire sequence  $(a^i)_{0 \leq i \leq m}$  is augmented,  $k = 0$  and  $\ell \geq 1$ . In particular, if  $(a^i)_{0 \leq i \leq m}$  is augmented, then  $Q_{0,1} = P_{a^0}$ .

Condition (1.15) follows directly from this definition of  $V_{k,\ell}$ . Now fix  $0 \leq k < \ell \leq m$ . Then the valuative Mostowski Conditions for the subchain  $(a^i)_{k \leq i \leq \ell}$  imply

$$\text{val}(Q_{k,\ell} - Q_{k+1,\ell}) \geq \lambda_{k+1} - \lambda_{\ell+1}; \quad (1.19)$$

indeed, we have

$$Q_{k,\ell} - Q_{k+1,\ell} = \begin{cases} -(1 - P_{a^k}) P_{a^{k+1}} \cdots P_{a^\ell} & \text{if } (a^i)_{k \leq i \leq \ell} \text{ is plain} \\ (P_{a^k} - P_{a^{k+1}}) P_{a^{k+2}} \cdots P_{a^\ell} & \text{if } (a^i)_{k \leq i \leq \ell} \text{ is augmented.} \end{cases}$$

From (1.19), we first deduce (1.16): On the one hand, (1.18) directly implies  $\dim V_{k,\ell} \leq e_\ell$ . (Note that in the above case where  $Q_{0,1} = P_{a^0}$ , we have  $e_0 = e_1$ .) On the other hand,

repeatedly using (1.19) yields  $\text{val}(Q_{k,\ell} - Q_{\ell,\ell}) \geq \min_{k \leq i < \ell} (\lambda_{i+1} - \lambda_{\ell+1}) > 0$ , so since  $Q_{\ell,\ell} = P_{a^\ell}$  is the identity on  $V_{\ell,\ell} = \mathbf{T}_{a^\ell} \dot{X}^\ell$ , we have  $\ker Q_{k,\ell} \cap V_{\ell,\ell} = 0$  and hence  $\dim V_{k,\ell} = \text{rk } Q_{k,\ell} \geq e_\ell$ .

Finally, (1.19) implies (1.17) using Lemma 1.8.2 (2)  $\Rightarrow$  (1).  $\square$

We formulate the main part of the proof of the other direction as a general lemma about flags.

**Lemma 1.8.5.** *Fix  $m \geq 1$  and  $\lambda_1 \geq \dots \geq \lambda_{m+1} \in \Gamma$ . Suppose that for each  $0 \leq k \leq m$ , we have a (partial) flag*

$$V_{k,m} \subseteq V_{k,m-1} \subseteq \dots \subseteq V_{k,k+1} \subseteq V_{k,k} \subseteq \mathcal{R}^n \quad (1.20)$$

satisfying

$$\Delta(V_{k,\ell}, V_{k+1,\ell}) \geq \lambda_{k+1} - \lambda_{\ell+1} \quad \text{for } 0 \leq k < \ell \leq m. \quad (1.21)$$

(In particular, we assume  $\dim V_{k,\ell} = \dim V_{k+1,\ell}$ .) Let  $P_{k,\ell}: \mathcal{R}^n \rightarrow \mathcal{R}^n$  denote the orthogonal projection onto  $V_{k,\ell}$ . Under those assumptions, we have

$$\text{val}((1 - P_{0,0})P_{1,1}P_{2,2} \cdots P_{m,m}) \geq \lambda_1 - \lambda_{m+1}. \quad (1.22)$$

If moreover  $\dim V_{1,1} = \dim V_{0,0}$  (which in particular implies  $\dim V_{0,1} = \dim V_{0,0}$  and hence  $V_{0,1} = V_{0,0}$ ), then we moreover have

$$\text{val}((P_{0,0} - P_{1,1})P_{2,2}P_{3,3} \cdots P_{m,m}) \geq \lambda_1 - \lambda_{m+1}. \quad (1.23)$$

Before proving that lemma, we quickly check that it indeed implies the other direction of the proposition.

*Proof of Proposition 1.8.3, (2)  $\Rightarrow$  (1).* Let  $(a^i)_i$  be a val-chain. By (2) of the proposition, we have vector spaces  $V_{k,\ell}$  for  $0 \leq k \leq \ell \leq m$  satisfying the prerequisites of Lemma 1.8.5. If  $(a^i)_i$  is plain, then the Mostowski Condition (vm1) is (1.22); if  $(a^i)_i$  is augmented, then  $\dim V_{1,1} = \dim V_{0,0}$  and the Mostowski Condition (vm2) is (1.23).  $\square$

*Proof of Lemma 1.8.5.* We will prove the following two inequalities by downwards induction on  $k$ :

$$\text{val}((1 - P_{k,k}) \cdot P_{k+1,k+1} \cdots P_{m,m}) \geq \lambda_{k+1} - \lambda_{m+1} \quad \text{for } 0 \leq k \leq m \quad (1.24)$$

$$\text{val}((P_{k,i} - P_{k,i+1}) \cdot P_{k+1,k+1} \cdots P_{m,m}) \geq \lambda_{i+1} - \lambda_{m+1} \quad \text{for } 0 \leq k \leq i < m. \quad (1.25)$$

Note that (1.25) will be needed in the inductive proof of (1.24). Before we carry out this induction, let us already check that (1.24) implies the lemma: (1.22) is just (1.24) for  $k = 0$ . To get (1.23), we plug in

$$P_{0,0} - P_{1,1} = P_{0,0} \cdot (1 - P_{1,1}) + (P_{0,0} - 1) \cdot P_{1,1}. \quad (1.26)$$

The second summand obtained in this way is just (1.22) (up to sign) and hence has valuation as required. In the first summand, we repeat the factor  $(1 - P_{1,1})$  twice (which we may, since it is a projection), so it is equal to

$$\underbrace{P_{0,0} \cdot (1 - P_{1,1})}_{(a)} \cdot \underbrace{(1 - P_{1,1})P_{2,2}P_{3,3} \cdots P_{m,m}}_{(b)}. \quad (1.27)$$

By (1.24), (b) has valuation at least  $\lambda_2 - \lambda_{m+1}$  so it suffices to show that (a) has valuation at least  $\lambda_1 - \lambda_2$ . But indeed,  $\text{val}(P_{1,1} - P_{0,1}) \geq \lambda_1 - \lambda_2$  and  $P_{0,0} \cdot (1 - P_{0,1}) = 0$  since  $V_{0,1} = V_{0,0}$  (by the assumption  $\dim V_{1,1} = \dim V_{0,0}$ ). Thus it remains to prove (1.24) and (1.25).

For  $k = m$ , (1.24) is trivial (since  $\lambda_{m+1} - \lambda_{m+1} = 0$ ) and (1.25) is void, so suppose  $k < m$ . We give the details for (1.25); the proof of (1.24) works analogously; see below.

We will prove

$$\text{val}((P_{k,i} - P_{k,i+1}) \cdot Q \cdot P_{k+2,k+2} \cdots P_{m,m}) \geq \lambda_{i+1} - \lambda_{m+1} \quad (1.28)$$

for  $Q = P_{k+1,j} - P_{k+1,j+1}$  ( $j = k+1, \dots, m-1$ ) and for  $Q = P_{k+1,m}$ . The sum of all those  $Q$  is equal to  $P_{k+1,k+1}$ , so taking the sum of (1.28) for all those  $Q$  then yields (1.25).

Case  $Q = P_{k+1,m}$ : Since  $\text{val}(P_{k,m} - P_{k+1,m}) \geq \lambda_{k+1} - \lambda_{m+1} \geq \lambda_{i+1} - \lambda_{m+1}$ , we can replace  $Q$  by  $P_{k,m}$  in (1.28). Now (1.28) follows, since  $(P_{k,i} - P_{k,i+1})P_{k,m} = P_{k,m} - P_{k,m} = 0$ .

Case  $Q = P_{k+1,j} - P_{k+1,j+1}$ : By induction, we have

$$\text{val}(Q \cdot P_{k+2,k+2} \cdots P_{m,m}) \geq \lambda_{j+1} - \lambda_{m+1}. \quad (1.29)$$

If  $j \leq i$ , we are done, since  $\lambda_{j+1} - \lambda_m \geq \lambda_{i+1} - \lambda_{m+1}$ , so suppose now  $j > i$ . In that case, we have the following (“ $\approx$ ” explained below):

$$\begin{aligned} & (P_{k,i} - P_{k,i+1}) \cdot Q \cdot P_{k+2,k+2} \cdots P_{m,m} \\ &= (P_{k,i} - P_{k,i+1}) \cdot Q \cdot Q \cdot P_{k+2,k+2} \cdots P_{m,m} \\ &\approx (P_{k,i} - P_{k,i+1}) \cdot (P_{k,j} - P_{k,j+1}) \cdot Q \cdot P_{k+2,k+2} \cdots P_{m,m} \end{aligned}$$

Since  $i \neq j$ , we have  $(P_{k,i} - P_{k,i+1})(P_{k,j} - P_{k,j+1}) = 0$ , so to obtain (1.28), it remains to verify that the difference between the two sides of “ $\approx$ ” has valuation at least  $\lambda_{i+1} - \lambda_{m+1}$ .

This follows from (1.29) and the following:

$$\begin{aligned} \text{val}(Q - (P_{k,j} - P_{k,j+1})) &\geq \min\{\text{val}(P_{k+1,j} - P_{k,j}), \text{val}(P_{k+1,j+1} - P_{k,j+1})\} \\ &\geq \min\{\lambda_{k+1} - \lambda_{j+1}, \lambda_{k+1} - \lambda_{j+2}\} \geq \lambda_{i+1} - \lambda_{j+1}. \end{aligned}$$

This finishes the proof of (1.25). The proof of (1.24) is exactly the same: just replace  $(P_{k,i} - P_{k,i+1})$  by  $(1 - P_{k,k})$  everywhere in the proof and then plug in  $k$  for the remaining  $i$ 's in the proof. (Concerning the case  $Q = P_{k+1,j} - P_{k+1,j+1}$ , note that one then automatically has  $j > i = k$ .)  $\square$

From Proposition 1.8.3, one can easily deduce that the notion of Lipschitz stratifications is invariant under  $\mathrm{GL}_n$ . More precisely, we obtain the following.

**Corollary 1.8.6** ( $\mathrm{GL}_n(\mathcal{O}_{\mathcal{R}})$ -invariance). *If  $\mathcal{X} = (X^i)_i$  is a valutive Lipschitz stratification of a definable set  $X \subseteq \mathcal{R}^n$  and  $M \in \mathrm{GL}_n(\mathcal{O}_{\mathcal{R}})$ , then  $M(\mathcal{X}) := (M(X^i))_i$  is a valutive Lipschitz stratification of  $M(X)$ .*

*Proof.* We use the characterization of valutive Lipschitz stratification from Proposition 1.8.3 (2). By Lemma 1.5.2,  $M$  preserves valuations, so applying  $M$  to a val-chain  $(a^i)_i$  for  $\mathcal{X}$  yields a val-chain for  $M(\mathcal{X})$ . Moreover, if  $V_{k,\ell}$  are vector spaces satisfying the conditions (1.15) – (1.17) with  $V_{k,k} = \mathbf{T}_{a^k} \overset{\circ}{X}^{e_k}$ , then the spaces  $M(V_{k,\ell})$  satisfy the same conditions with  $M(V_{k,k}) = \mathbf{T}_{M(a^k)} M(\overset{\circ}{X}^{e_k})$ .  $\square$

**Corollary 1.8.7.** *If  $\mathcal{X} = (X^i)_i$  is a Lipschitz stratification of a definable set  $X \subseteq \mathcal{R}^n$  and  $M \in \mathrm{GL}_n(\mathcal{R})$ , then  $M(\mathcal{X}) := (M(X^i))_i$  is a Lipschitz stratification of  $M(X)$ .*

*Proof.* We may assume that  $\mathcal{X}$  and  $M$  are  $\mathcal{L}$ -definable. (This works in the same way as in the proof of Theorem 1 from Theorem 1.6.7: We first choose a language containing all the constants we need, and then we choose the models  $\mathcal{R}_0 \not\cong \mathcal{R}$ .)

By Proposition 1.6.11,  $\mathcal{X}$  is a valutive Lipschitz stratification; by Corollary 1.8.6,  $M(\mathcal{X})$  is a valutive Lipschitz stratification (which is still  $\mathcal{L}$ -definable, since  $M$  is), and finally, using Proposition 1.6.11 again, we deduce that  $M(\mathcal{X})$  is a Lipschitz stratification.  $\square$

## 1.9 Overview of the main proof

Here is an overview of the proof of Theorem 1.6.7, describing the main ideas in an informal way. Several technicalities are omitted.

We can easily stratify the given set  $X \subseteq \mathcal{R}^n$  in such a way that each stratum is the graph of a function. More precisely, given a  $d$ -dimensional stratum  $S$ , after a suitable coordinate transformation,  $S$  is the graph of some function  $\rho: \bar{S} \rightarrow \mathcal{R}^{n-d}$ , where  $\bar{S} := \mathrm{pr}_{\leq d}(S) \subseteq \mathcal{R}^d$ . Our final goal is to obtain bounds on valutive distances  $\Delta(V_1, V_2)$  (see Definition 1.8.1) between certain subspaces  $V_i$  of tangent spaces. To be able to easily express those distances in terms of the functions  $\rho$ , we need that  $\rho$  satisfies

$$\mathrm{val}(\mathrm{Jac}_{\bar{a}} \rho) \geq 0 \quad \text{for every } \bar{a} \in \bar{S}. \quad (1.30)$$

Indeed, for instance, under this assumption, we have, for  $a_1, a_2 \in S$ :

$$\Delta(\mathbf{T}_{a_1} S, \mathbf{T}_{a_2} S) = \mathrm{val}(\mathrm{Jac}_{\bar{a}_1} \rho - \mathrm{Jac}_{\bar{a}_2} \rho),$$

where  $\bar{a}_i = \mathrm{pr}_{\leq d}(a_i)$ .

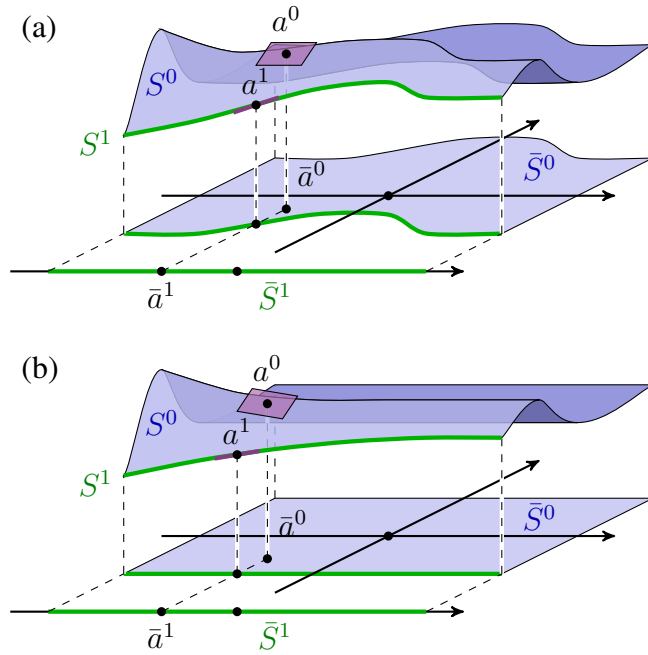


Figure 1.3: In this example, we consider a plain val-chain  $a^0, a^1$  with dimensions  $e_0 = 2$  and  $e_1 = 1$ . We assume that  $\text{pr}_{\leq e_1}(a^0) = \text{pr}_{\leq e_1}(a^1)$ . (a) We need to find a subspace of the tangent plane  $T_{a^0}(S^0)$  which is close to the tangent line  $T_{a^1}(S^1)$ . (b) To simplify this, we first deform the whole picture in such a way that the projection  $\text{pr}_{\leq e_0}(S^1)$  becomes a straight line.

Most of the bounds of the form  $\Delta(V_1, V_2)$  we aim for involve more than one stratum: Given a val-chain  $a^0, \dots, a^m$  with  $a^\ell \in S^\ell$ , we need to relate the tangent spaces of all those strata  $S^0, \dots, S^m$ . To be able to apply our above approach (of considering strata as graphs of functions and expressing distances of spaces in terms of Jacobians), we need to find a single coordinate transformation such that afterwards, each  $S^\ell$  is the graph of a function  $\rho^\ell$  satisfying (1.30). (We call such a coordinate transformation an “aligner” of  $S^0, \dots, S^m$ .) In Subsection 2.1 (Proposition 2.1.5), we obtain stratifications admitting aligners for any choice of  $n + 2$  strata. This is enough, since a val-chain consists of at most  $n + 2$  points.

To illustrate the remainder of the proof, we start by considering a plain val-chain consisting of only two points  $a^0 \in S^0, a^1 \in S^1$ . We suppose that we have already applied an aligner, so that  $S^\ell, \ell = 1, 2$ , is the graph of some function  $\rho^\ell$  on  $\bar{S}^\ell := \text{pr}_{\leq e_\ell} S^\ell$ , where  $e_\ell := \dim S^\ell$ .

The next step in the proof consists in reducing the case of arbitrary plain val-chains (of length 2) to plain val-chains satisfying  $\text{pr}_{\leq e_1}(a^0) = \text{pr}_{\leq e_1}(a^1)$ ; in the following, we assume this. In particular, this means that  $a^0$  determines  $a^1$  (assuming that  $S^0$  and  $S^1$  are fixed).

To establish the conditions from Proposition 1.8.3 concerning the val-chain  $a^0, a^1$ , we need to find a subspace  $V \subseteq \mathbf{T}_{a^0}S^0$  that is sufficiently close to  $\mathbf{T}_{a^1}S^1$ ; see Figure 1.3 (a). We choose  $V$  to be the subspace of  $\mathbf{T}_{a^0}S^0$  satisfying  $\text{pr}_{\leq e_0}(V) = \text{pr}_{\leq e_0}(\mathbf{T}_{a^1}S^1)$ . (From (1.30), one can deduce that this is a best possible approximation to  $\mathbf{T}_{a^1}S^1$ .) The distance  $\Delta(V, \mathbf{T}_{a^1}S^1)$  can directly be expressed in terms of Jacobians of the functions  $\rho^0$  and  $\rho^1$ , but this becomes simpler if we first apply a ‘‘rectilinearization’’: a transformation which translates the coordinates  $e_1 + 1, \dots, e_0$  in such a way that  $\text{pr}_{\leq e_0}(S^1)$  is sent to a subset of  $\mathcal{R}^{e_1} \times \{0\}^{e_0 - e_1}$  and which preserves all the other coordinates; see Figure 1.3 (b). After the rectilinearization has been applied,  $V$  is determined by the first  $e_1$  derivatives of  $\rho^0$ , and we obtain

$$\Delta(V, \mathbf{T}_{a^1}S^1) = \text{val}((\text{Jac}_{\bar{a}^0} \delta) \upharpoonright \mathcal{R}^{e_1} \times \{0\}^{e_0 - e_1}) = \min_{1 \leq i \leq e_1} \text{val}(\partial_i \delta(\bar{a}^0)),$$

where  $\bar{a}^0 := \text{pr}_{\leq e_0}(a^0)$  and  $\delta: \bar{S}^0 \rightarrow \mathcal{R}^{n - e_0}$  is the difference of  $\rho^0$  and the last  $n - e_0$  coordinates of  $\rho^1$ .

The desired bound on  $\Delta(V, \mathbf{T}_{a^1}S^1)$  depends on the valuative distance of  $a^1$  to a lower-dimensional stratum. We ensure that this bound holds by removing a lower-dimensional subset from  $\bar{S}^1$ . In terms of the function  $\delta$  defined above, this means that we need to find a set  $Z \subseteq \mathcal{R}^{e_1}$  of dimension less than  $e_1$  such that the first  $e_1$  partial derivatives of  $\delta$  at  $x \in \bar{S}^0$  are bounded in terms of the distance of  $\text{pr}_{\leq e_1}(x)$  to  $Z$ . More precisely, the bound we end up needing is

$$\text{val}(\partial_i \delta(x)) \geq \underbrace{\min\{\text{val}(\text{pr}_{> e_1}(x)), \text{val}(\delta(x))\}}_{=\text{val}(a^0 - a^1)} - \text{val} \text{dist}(\text{pr}_{\leq e_1}(x), Z) \quad \text{for } 1 \leq i \leq e_1. \quad (1.31)$$

The heart of the construction of valuative Lipschitz stratifications is Proposition 2.3.6, which provides such a lower-dimensional set  $Z$  for arbitrary functions  $\delta$ .

For longer val-chains, the arguments are similar: Given a plain val-chain  $a^0 \in S^0, \dots, a^m \in S^m$  (with  $\dim S^\ell = e_\ell$ ), we rectilinearize with respect to some of the coordinates of  $S^\ell$  for  $\ell = 1, \dots, m$  and we obtain a function  $\delta$  on (a certain subset of)  $\bar{S}^0$  whose first  $e_m$  derivatives need to be bounded by removing a lower-dimensional subset  $Z$  from  $\bar{S}^m$ . Together with an inductive assumption that everything already works well for the sub-chain  $a^0, \dots, a^{m-1}$ , we obtain the subspaces  $V_{k, \ell}$  needed by Proposition 1.8.3.

For augmented val-chains, the outline of the argument is the same; the biggest differences arise when the two first points  $a^0, a^1$  lie in the same stratum  $S^0$ , which, say, is the graph of  $\rho^0$ . In that case, instead of bounding first derivatives, we need to bound the second derivatives of  $\rho^0$  to obtain a bound  $\Delta(\mathbf{T}_{a^0}S^0, \mathbf{T}_{a^1}S^0)$ . Those bounds are obtained in essentially the same way as (1.31), namely by applying Proposition 2.3.6 to all first derivatives of  $\rho$ .

We end this overview by mentioning an issue related to aligners (i.e., the coordinate transformation ensuring (1.30)). Given a sequence  $S^0, \dots, S^m$  of strata, the set  $Z$  to be removed from  $S^m$  according to the above procedure may depend on the chosen aligner. When inductively assuming that this has already been done for  $S^0, \dots, S^{m-1}$ , we need that it has been done using the same aligner as the one we use for  $S^0, \dots, S^m$ . However, an aligner for  $S^0, \dots, S^{m-1}$  might not be suitable for  $S^m$ . The solution is that Proposition 2.1.5 states that all aligners can be found in a *finite* set  $\mathcal{C}_n$  of coordinate transformations (depending only on the ambient dimension  $n$ ). By applying the above procedure to every possible aligner of  $S^0, \dots, S^{m-1}$  in  $\mathcal{C}_n$ , we in particular ensure that we included aligners working for  $S^0, \dots, S^m$ .

## 2 Ingredients to the main proof

The entire remainder of the article is devoted to the proof of Theorem 1.6.7. We continue to use the notation introduced in Subsections 1.1, 1.4 and 1.5, though we make a slight change concerning the language: to avoid having to mention the parameters  $A$  from Theorem 1.6.7 everywhere, we now allow  $\mathcal{L}$  to contain additional constants from  $\mathcal{R}$ . Thus the general assumptions for the remainder of the paper are the following.

**Assumption 2.0.1.** For the remainder of the paper, we assume that  $\mathcal{R}$  is a real closed field which is power-bounded and  $o$ -minimal as an  $\mathcal{L}^0$ -structure and  $T$ -convex as an  $\mathcal{L}_{\text{val}}^0$ -structure. Moreover, we set  $\mathcal{L} := \mathcal{L}^0(A)$  and  $\mathcal{L}_{\text{val}} := \mathcal{L}_{\text{val}}^0(A)$  for some finite set of parameters  $A \subseteq \mathcal{R}$ , and we assume (without loss) that  $\mathcal{R}$  is sufficiently saturated.

Note that there is a hidden quantifier here: We will prove everything for *every* finite set  $A$  of parameters. This in particular means that we can use previously proved results for different  $A$ .

### 2.1 Alignable Bradycell Decompositions

The first step in the construction of a valutive Lipschitz stratification of a set  $X$  consists in partitioning  $X$  into pieces that can be “aligned”: After a suitable transformation of the coordinate system, they are graphs of functions whose derivatives have non-negative valuation.

**Definition 2.1.1** (Aligned sets). Let  $S$  be an  $\mathcal{L}$ -definable subset of  $\mathcal{R}^n$ . We say that  $S$  is *aligned* if, for  $d := \dim S$ , the set  $\bar{S} := \text{pr}_{\leq d}(S)$  is open in  $\mathcal{R}^d$  and  $S$  is the graph of an  $\mathcal{L}$ -definable  $C^1$  function  $f: \bar{S} \rightarrow \mathcal{R}^{n-d}$  satisfying

$$\text{val}(\text{Jac}_a f) \geq 0 \quad \text{for all } a \in \bar{S}. \quad (2.1)$$

The open set  $\bar{S}$  is referred to as the *base* of  $S$ . We say that  $\kappa \in \text{GL}_n(\mathcal{O}_{\mathcal{R}})$  is an *aligner* of an  $\mathcal{L}$ -definable set  $S \subseteq \mathcal{R}^n$  if  $\kappa(S)$  is aligned.

*Remark 2.1.2.* If such an aligned set  $S$  is a lowest dimensional stratum of a stratification of a closed definable set  $X \subseteq \mathcal{R}^n$ , then  $\bar{S}$  is both, open and closed and hence  $\bar{S} = \mathcal{R}^{\dim S}$ . This fits together with the fact that Lipschitz stratifications almost never have  $X^0 = \emptyset$  and it will also fit together with the convention that in val-chains, we set  $\lambda_{m+1} = -\infty$  if  $X^{e_m-1} = \emptyset$ .

*Remark 2.1.3.* These aligned sets are somewhat related to the  $L$ -regular cells of [KP, §1], to the regular  $M$ -cells of [Paw, §1], and to the  $\Lambda^m$ -regular cells of [Fis, Definition 1.2], though the latter are more sophisticated and control more derivatives, and all of these notions impose additional conditions on the base.

It will not be enough to partition our given set  $X$  into sets which can be aligned using some arbitrary  $\kappa \in \text{GL}_n(\mathcal{O}_{\mathcal{R}})$ ; we will also need some good control of these  $\kappa$ :

1. We need to find a *finite* set  $\mathcal{C}_n \subseteq \text{GL}_n(\mathcal{O}_{\mathcal{R}})$ , depending only on  $n$ , such that all aligners  $\kappa$  can be taken from  $\mathcal{C}_n$ . The precise set  $\mathcal{C}_n$  does not matter, so we postpone choosing it to Definition 2.1.10.
2. We need all  $\kappa$  to be  $\mathcal{L}$ -definable without additional parameters. To ensure this, we will choose  $\mathcal{C}_n \subseteq \text{GL}_n(\mathbb{Q})$ .
3. We need that a single  $\kappa$  works for several (given) pieces of the partition at once. More precisely, any  $n+2$  pieces should have a common aligner in  $\mathcal{C}_n$ . (This number  $n+2$  is what we need for the proof of existence of Lipschitz stratifications. The proofs in Subsection 2.1 would work equally well for any other fixed number.)

It is problematic that Item (3) above is not a condition on individual pieces, but on the partition as a whole: This makes it unclear whether, given a partition  $\mathcal{S} = (S_i)_i$  satisfying (3), one can refine parts of  $\mathcal{S}$  in a way that (3) is preserved without modifying the remainder of  $\mathcal{S}$ . To solve this problem, we will introduce the notion of “bradycells” (Definition 2.1.11). Bradycells will have the property that  $n+2$  of them always have a common aligner. Using that notion, we can state the main result of this subsection, which provides the desired partitions. Since the precise notion of bradycells is irrelevant for the remainder of the article, we postpone it.

**Definition 2.1.4.** A *bradycell decomposition* is a partition of  $\mathcal{R}^n$  into bradycells (see Definition 2.1.11).

**Proposition 2.1.5** (Bradycell decompositions). *1. Every finite partition of  $\mathcal{R}^n$  into  $\mathcal{L}$ -definable sets can be refined to a bradycell decomposition.*

2. For any set of at most  $n + 2$  bradycells  $S_1, \dots, S_k \subseteq \mathcal{R}^n$ , there exists a common aligner  $\kappa \in \mathcal{C}_n$ .

*Remark 2.1.6.* This entire subsection would become much simpler if all  $S_1, \dots, S_k$  in Proposition 2.1.5 (2) could be assumed to have different dimension; in particular, one could then choose  $\mathcal{C}_n$  to consist only of the coordinate permutations. In applications of the proposition, this will almost be the case: at most two of the bradycells will have the same dimension. It would probably be possible to also get rid of this (an approach like this has been used in [Hal]), but this would require considerably more work.

For the remainder of this subsection, we fix the following notation.

**Notation 2.1.7.** For  $d \leq n$ , we write  $\text{Gr}_{n,d}(\mathcal{R})$  for the Grassmannian variety, i.e., for the space of  $d$ -dimensional sub-vector spaces of  $\mathcal{R}^n$ .

**Definition 2.1.8.** For  $d \leq n$ , let  $\text{Gr}_{n,d}^\circ(\mathcal{R}) \subseteq \text{Gr}_{n,d}(\mathcal{R})$  be the open subset of those  $V \subseteq \mathcal{R}^n$  such that  $\text{pr}_{\leq d}(V) = \mathcal{R}^d$ , i.e., which project surjectively onto the first  $d$  coordinates. Such a  $V \in \text{Gr}_{n,d}^\circ(\mathcal{R})$  can be considered as the graph of a linear map  $M_V: \mathcal{R}^d \rightarrow \mathcal{R}^{n-d}$ ; we set  $J(V) := \|M_V\|$  (the operator norm of the matrix); for  $V \in \text{Gr}_{n,d}(\mathcal{R}) \setminus \text{Gr}_{n,d}^\circ(\mathcal{R})$ , we set  $J(V) := \infty$ .

The following lemma is the main tool to find the finitely many transformations  $\kappa \in \text{GL}_n(\mathbb{Q})$ . It is a purely geometrical-combinatorial result, closely related to [KP, Lemma 1.8]. Even though formulated in  $\mathcal{R}$ , it is just a statement about  $\mathbb{R}$  (as will become visible in the proof). Note also that the  $\kappa$  provided by the lemma are even elements of  $\text{O}_n(\mathbb{Q})$  (and not just in  $\text{GL}_n(\mathbb{Q})$ ).

**Lemma 2.1.9.** Fix arbitrary natural numbers  $n$  and  $\ell$ . Then we can find, for each  $d \leq n$ , a finite open covering of  $\text{Gr}_{n,d}(\mathcal{R})$  by  $\mathcal{L}$ -definable sets  $\Theta_\nu^d$  such that for any choice of  $\ell$  many of these sets  $\Theta_{\nu_1}^{d_1}, \dots, \Theta_{\nu_\ell}^{d_\ell}$ , there exists an orthogonal transformation  $\kappa \in \text{O}_n(\mathbb{Q})$  such that for every  $i \leq \ell$ , we have  $\text{val}(J(\kappa\Theta_{\nu_i}^{d_i})) \geq 0$ .

(Here,  $\text{val}(\infty) := -\infty$  and  $\text{val}(J(\kappa\Theta_{\nu_i}^{d_i})) \geq 0$  is a short hand notation for:  $\text{val}(J(\kappa V)) \geq 0$  for every  $V \in \Theta_{\nu_i}^{d_i}$ .)

*Proof of Lemma 2.1.9.* We will prove the stronger claim that the sets  $\Theta_\nu^d$  can be taken definable in the pure ring language. In that case, to get  $\text{val}(J(\kappa\Theta_{\nu_i}^{d_i})) \geq 0$ , it suffices to prove that the map  $V \mapsto J(V)$  is bounded on  $\kappa\Theta_{\nu_i}^{d_i}$  (by Remark 1.7.3). This boundedness is also a statement in the ring language, so we may as well assume  $\mathcal{R} = \mathbb{R}$ . In particular, any closed subset of  $\text{Gr}_{n,d_i}(\mathbb{R})$  is compact, so we can obtain boundedness of  $V \mapsto J(V)$  by proving

$$\text{cl}(\kappa\Theta_{\nu_i}^{d_i}) \subseteq \text{Gr}_{n,d_i}^\circ(\mathbb{R}).$$

Given a subset  $\Xi \subseteq \text{Gr}_{n,d}(\mathbb{R})$  (for any  $d \leq n$ ), we write  $Fo(\Xi)$  for the set of orthogonal transformations  $\kappa$  “forbidden by a space in  $\text{cl}(\Xi)$ ”, i.e.:

$$Fo(\Xi) := \{\kappa \in O_n(\mathbb{R}) : \text{cl}(\kappa\Xi) \not\subseteq \text{Gr}_{n,d}^\circ(\mathbb{R})\}.$$

Intuitively, we just need to choose the sets  $\Theta_\nu^d$  so small that no  $\ell$  of the sets  $Fo_\nu^d := Fo(\Theta_\nu^d)$  cover all of  $O_n(\mathbb{Q})$ . To make this argument precise, let  $\mu$  be the Haar measure on the compact group  $O_n(\mathbb{R})$ , normalized such that  $\mu(O_n(\mathbb{R})) = 1$ . It is enough to ensure that  $\mu(Fo_\nu^d) < 1/\ell$  for each  $\nu$  and  $d$ . (Then  $O_n(\mathbb{R}) \setminus \bigcup_{i=1}^\ell Fo_{\nu_i}^{d_i}$  is non-empty and open, and hence contains a  $\kappa \in O_n(\mathbb{Q})$ , as desired.)

To find finitely many sets  $\Theta_\nu^d$  with that property covering  $\text{Gr}_{n,d}(\mathbb{R})$ , we fix any definable metric on  $\text{Gr}_{n,d}(\mathbb{R})$  inducing the usual topology. Moreover, we fix any element  $V_0 \in \text{Gr}_{n,d}(\mathbb{Q})$ . Since the set  $Fo(\{V_0\}) \subseteq O_n(\mathbb{R})$  is a compact subset of lower dimension, we can find an open ball  $\Xi \subseteq \text{Gr}_{n,d}(\mathbb{R})$  around  $V_0$  such that  $\mu(Fo(\Xi)) < 1/\ell$ . (First choose any open set  $\mathcal{U} \supseteq Fo(\{V_0\})$  with  $\mu(\mathcal{U}) < 1/\ell$ , and then, using compactness of  $Fo(\{V_0\})$ , choose the radius of  $\Xi$  small enough to ensure  $Fo(\Xi) \subseteq \mathcal{U}$ .) We may moreover assume that  $\Xi$  has rational radius.

Now choose finitely many  $\kappa_\nu \in O_n(\mathbb{Q})$  such that the sets  $\Theta_\nu^d := \kappa_\nu(\Xi)$  cover  $\text{Gr}_{n,d}(\mathbb{R})$ . Then indeed,  $\mu(Fo_\nu^d) = \mu(Fo(\kappa_\nu(\Xi))) < 1/\ell$ .  $\square$

Using Lemma 2.1.9, we can now choose our finite set  $\mathcal{C}_n \subseteq \text{GL}_n(\mathbb{Q})$  and introduce the notion of bradycells.

**Definition 2.1.10** (The set  $\mathcal{C}_n$ ). For the remainder of this subsection, fix subsets  $\Theta_\nu^d \subseteq \text{Gr}_{n,d}(\mathbb{R})$  as provided by Lemma 2.1.9 using  $\ell = n + 2$ . Moreover, let  $\mathcal{C}_n \subseteq \text{GL}_n(\mathbb{Q})$  be a finite subset containing, for each choice of  $n + 2$  many sets  $\Theta_{\nu_1}^{d_1}, \dots, \Theta_{\nu_{n+2}}^{d_{n+2}}$ , an element  $\kappa$  satisfying  $\text{val}(J(\kappa\Theta_{\nu_i}^{d_i})) \geq 0$  ( $i = 1, \dots, n + 2$ ). (For any  $n + 2$  of the sets, the existence of such  $\kappa \in \text{GL}_n(\mathbb{Q})$  is asserted by the lemma, and there are only finitely many choices of  $n + 2$  sets.)

**Definition 2.1.11** (Bradycells). A *bradycell* is an  $\mathcal{L}$ -definable set  $S \subseteq \mathcal{R}^n$  such that for (at least) one of the sets  $\Theta_\nu^d$  chosen in Definition 2.1.10 (where  $d = \dim S$ ), we have the following:

1. For every  $x \in S$ , the tangent space  $\mathbf{T}_x S$  is an element of  $\Theta_\nu^d$ .
2. For every  $\kappa \in \mathcal{C}_n$  satisfying  $\text{val}(J(\kappa\Theta_\nu^d)) \geq 0$ ,  $\kappa(S)$  is aligned.

The content of Condition (2) is just that the projection  $\text{pr}_{\leq d}(\kappa(S))$  is open and that  $\kappa(S)$  is the graph of a function on that projection; the bound on the derivatives of the function is automatic by  $\text{val}(J(\kappa\Theta_\nu^d)) \geq 0$  and Condition (1).

Now that  $\mathcal{C}_n$  and bradycells are defined, we can finally prove the main result of this subsection.

*Proof of Proposition 2.1.5.* (1) We repeatedly refine the partition, ensuring that each piece of dimension  $d$  becomes a bradycell, proceeding from  $d = n$  downwards to  $d = 0$ . Thus fix  $d \leq n$ , and fix any  $d$ -dimensional piece  $S$ . It suffices to check that we can subdivide  $S$  into (finitely many)  $d$ -dimensional bradycells and an arbitrary lower-dimensional set.

After a first partitioning, we may assume that  $S$  is a definable  $C^1$  manifold and satisfies Condition (1) from Definition 2.1.11 for some set  $\Theta_\nu^d$ . To obtain Condition (2), we further partition  $S$  for each of those  $\kappa \in \mathcal{C}_n$  for which  $\text{val}(J(\kappa\Theta_\nu^d)) \geq 0$ : By a first partition, we ensure that  $\kappa(S)$  is the graph of a function  $f: \text{pr}_{\leq d}(\kappa(S)) \rightarrow \mathcal{R}^{n-d}$ . Then we remove a lower-dimensional set to ensure that  $\text{pr}_{\leq d}(\kappa(S))$  is open and that  $f$  is  $C^1$ .

(2) Consider bradycells  $S_1, \dots, S_k$  for some  $k \leq n + 2$ ; for each  $i \leq k$ , let  $\Theta_{\nu_i}^{d_i}$  be a corresponding set provided by Definition 2.1.11. By our choice of  $\mathcal{C}_n$  (Definition 2.1.10), there exists a  $\kappa \in \mathcal{C}_n$  such that for each  $i$ , we have  $\text{val}(J(\kappa\Theta_{\nu_i}^{d_i})) \geq 0$ . By Definition 2.1.11 (2),  $\kappa(S_i)$  is aligned.  $\square$

We end this subsection by proving a useful property of aligned sets.

**Lemma 2.1.12.** *Let  $S \subseteq \mathcal{R}^n$  be a  $d$ -dimensional aligned set, and suppose that  $B \subseteq \mathcal{R}^n$  is a valuative ball (open or closed) with  $B \cap S \neq \emptyset$  but  $B \cap \partial S = \emptyset$ . Then  $\bar{B} := \text{pr}_{\leq d}(B)$  is a subset of the base  $\bar{S} = \text{pr}_{\leq d}(S)$  of  $S$ .*

*Proof.* Suppose that  $\bar{B} \not\subseteq \bar{S}$ . Choose  $a \in B \cap S$ , set  $\bar{a} := \text{pr}_{\leq d}(a) \in \bar{B} \cap \bar{S}$  and choose  $\bar{b} \in \bar{B} \setminus \bar{S}$ . Let  $L := \{(1-t)\bar{a} + t\bar{b} \mid 0 < t < 1\}$  be the open line segment connecting  $\bar{a}$  and  $\bar{b}$ . We may assume  $L \subseteq \bar{S}$ ; otherwise, replace  $\bar{b}$  by the point of  $L \cap \partial \bar{S}$  which is closest to  $\bar{a}$ . (Such a point exists by  $o$ -minimality, and using that  $\partial \bar{S}$  is  $\mathcal{L}$ -definable.)

Let  $f$  be the function whose graph is  $S$ , and consider the function  $g: [0, 1] \rightarrow S$  sending  $t$  to  $f((1-t)\bar{a} + t\bar{b})$ . Using  $\text{val}(\text{Jac } f) \geq 0$ , we obtain  $\text{val}(g'(t)) \geq \text{val}(\bar{b} - \bar{a})$ , so using the Mean Value Theorem, we deduce, for any  $t_1, t_2 \in [0, 1]$ :

$$\text{val}(g(t_2) - g(t_1)) \geq \text{val}(t_2 - t_1) + \text{val}(\bar{b} - \bar{a}) \geq \text{rad}(B),$$

where the last inequality is strict if  $B$  is an open ball. This implies that  $b' := \lim_{t \rightarrow 1} g(t)$  exists and that  $b := (\bar{b}, b')$  satisfies  $\text{val}(b - a) \geq \text{val}(\bar{b} - \bar{a})$ . In particular,  $b \in B \cap \partial S$ , contradicting the assumption that this intersection is empty.  $\square$

*Remark 2.1.13.* Given an  $\mathcal{L}$ -definable  $C^1$  function  $f: X \rightarrow \mathcal{R}^{n-d}$  on an  $\mathcal{L}$ -definable set  $X \subseteq \mathcal{R}^d$ , a similar kind of Mean Value Theorem argument on a line segment allows us to bound  $\text{val}(f(a_1) - f(a_2))$  by  $\text{val}(a_1 - a_2) + \text{val}(\text{Jac } f)$  under suitable assumptions: If  $a_1$  and  $a_2$  both lie in a valuative ball  $B$  that is entirely contained in  $X$ , and  $\text{val}(\text{Jac}_a f) \geq \lambda$  for all  $a \in B$ , then

$$\text{val}(f(a_1) - f(a_2)) \geq \text{val}(a_1 - a_2) + \lambda.$$

In particular, for  $S$  and  $B$  as in Lemma 2.1.12, the entire preimage  $\text{pr}_{\leq d}^{-1}(\bar{B}) \cap S$  is contained in  $B$ .

## 2.2 Bounding derivatives using power-boundedness

A key ingredient to our proof of the existence of Lipschitz stratifications is the following proposition, which has been proved in [Yin]. This is the only (but crucial) place in the present paper where power-boundedness is used.

**Proposition 2.2.1** ([Yin, Corollary 2.17]). *Suppose that  $f: \mathcal{R}^n \rightarrow \mathcal{R}$  is an  $\mathcal{L}$ -definable function. Then there exists a finite  $\mathcal{L}$ -definable partition of  $\mathcal{R}^n$  into sets  $Y_\nu$  such that if  $B$  is an open valutive ball entirely contained in one of the sets  $Y_\nu$ , then either  $f(B) = \{0\}$  or  $f(B)$  is an open valutive ball not containing 0.*

Note that if  $f(B)$  is an open valutive ball not containing 0, then for any  $y_1, y_2 \in B$ , we have  $\text{val}(f(y_1)) = \text{val}(f(y_2))$ , and even  $\text{val}(f(y_1) - f(y_2)) > \text{val}(f(y_1))$ .

Here, we have rewritten Proposition 2.2.1 in the language of the present paper; the map  $\text{rv}$  appearing in [Yin] is defined in such a way that  $\text{rv}(a) = \text{rv}(a')$  iff either  $a = a' = 0$  or  $\text{val}(a - a') > \text{val}(a)$  for  $a, a' \in \mathcal{R}$  (and a valutive polydisc is just a product of valutive balls of possibly different radii). Note that the language used in [Yin] is, up to interdefinability, the same as ours; see [Yin, Definition 1.2 and Convention 1.11].

Instead of using Proposition 2.2.1 directly, we will use the following corollary:

**Corollary 2.2.2.** *Suppose that  $f: \mathcal{R}^n \rightarrow \mathcal{R}$  is an  $\mathcal{L}$ -definable function. Then there exists an  $\mathcal{L}$ -definable set  $Z \subseteq \mathcal{R}^n$  of dimension less than  $n$  such that for every  $y \in \mathcal{R}^n \setminus Z$ ,  $\partial_i f(y)$  exists and we have*

$$\text{val}(\partial_i f(y)) \geq \text{val}(f(y)) - \text{valdist}(y, Z) \quad (2.2)$$

for  $i = 1, \dots, n$ .

*Proof.* Apply Proposition 2.2.1 to  $f, \partial_1 f, \dots, \partial_n f$ , where the partial derivatives are extended by 0 to those points of  $\mathcal{R}^n$  where they do not exist; then set  $Z := Z_0 \cup \bigcup_{i,\nu} \partial Y_{i,\nu}$ , where  $(Y_{i,\nu})_\nu$  is the partition obtained for the  $i$ th of the above functions ( $i = 1, \dots, n+1$ ) and  $Z_0$  is the set of points where  $f$  is not differentiable; we claim that this set  $Z$  works.

Fix a  $y \in \mathcal{R}^n \setminus Z$  and set  $\zeta := \text{valdist}(y, Z)$  and  $B := B_{>\zeta}(y)$ . Then for each  $i$ , there exists a  $\nu$  such that  $B \subseteq Y_{i,\nu}$  for some  $\nu$ ; in particular,  $\text{val}(f(B))$  and  $\text{val}(\partial_i f(B))$  are singletons.

To prove (2.2), we use an Mean Value Theorem argument similar to the one in Remark 2.1.13, but in the opposite direction: Suppose for contradiction that  $y$  is a witness to the failure of (2.2), i.e.,  $\text{val}(f(B)) - \text{val}(\partial_i f(B)) > \zeta$  for some  $i$ . We choose  $y_1, y_2 \in B$  differing only in the  $i$ -th coordinate with  $\text{val}(y_1 - y_2) = \text{val}(f(B)) - \text{val}(\partial_i f(B))$ . The Mean Value Theorem yields a  $y_3 \in B$  such that

$$f(y_1) - f(y_2) = (y_1 - y_2) \cdot \partial_i f(y_3).$$

This leads to a contradiction: On the one hand, we have  $\text{val}(f(y_1) - f(y_2)) > \text{val}(f(B))$  (by our application of Proposition 2.2.1 to  $f$ ); on the other hand,

$$\text{val}((y_1 - y_2) \cdot \partial_i f(y_3)) = \text{val}(f(B)) - \text{val}(\partial_i f(B)) + \text{val}(\partial_i f(B)) = \text{val}(f(B)). \quad \square$$

*Remark 2.2.3.* Using Remark 1.7.3, Corollary 2.2.2 may be reformulated without making reference to the valuation. Since Remark 1.7.3 only applies to functions defined without parameters outside of  $\mathcal{R}_0$  (but we have made the change at the beginning of this section so that  $\mathcal{L}$  now might contain such parameters), one first needs to formulate the corollary for families of functions. In this way, one obtains that Corollary 2.2.2 is equivalent to the following statement: For any  $\mathcal{L}$ -definable family of functions  $f_q : \mathcal{R}^n \rightarrow \mathcal{R}$  (where  $q$  runs over some  $\mathcal{L}$ -definable set  $Q$ ), there exists a constant  $c \in \mathcal{R}$  (not depending on  $q$ ) and an  $\mathcal{L}$ -definable family of sets  $Z_q \subseteq \mathcal{R}^n$  of dimension less than  $n$  such that

$$|\partial_i f_q(y)| \leq \frac{c|f_q(y)|}{\text{dist}(y, Z_q)} \quad \text{for all } i \leq n, \text{ all } q \in Q \text{ and all } y \in \mathcal{R}^n \setminus Z_q. \quad (2.3)$$

Note that this bears some similarities to the  $\Lambda_L^1$ -regular functions in [Fis, Definition 1.1] (though (2.3) is false in, e.g., structures with exponential function). One has the feeling that there should be a more direct proof of (2.3), avoiding the machinery of  $T$ -convexity. For  $n = 1$  and when  $Q$  is a singleton, it is not too difficult to deduce it from power-boundedness. However, we do not know how to prove the general case more directly.

Here is another lemma, which does not really have anything to do with the previous results of this subsection, but which will be useful in conjunction with them.

**Lemma 2.2.4.** *Suppose that  $X \subseteq \mathcal{R}^n$  is a non-empty  $\mathcal{L}$ -definable set and  $f : X \rightarrow \mathcal{R}$  is an  $\mathcal{L}$ -definable function such that  $|f|$  is bounded (by an element of  $\mathcal{R}$ ). Then there exists an  $\mathcal{L}$ -definable element  $x_0 \in X$  such that  $\text{val}(f(x_0)) = \min\{\text{val}(f(x)) : x \in X\}$ . In particular, that minimum exists.*

*Proof.* Set  $s := \sup_x |f(x)|$ . The set  $X' := \{x \in X : |f(x)| \geq \frac{1}{2}s\}$  is  $\mathcal{L}$ -definable and non-empty, and every  $x \in X'$  satisfies  $\text{val}(x) = \text{val}(s)$ . Using definable choice (in the  $\mathcal{o}$ -minimal language  $\mathcal{L}$ ), we find an  $\mathcal{L}$ -definable  $x_0 \in X'$ .  $\square$

## 2.3 Sedating functions

To construct Lipschitz stratifications, we will need precise bounds on the valuations of the first derivatives of certain functions. The goal of this subsection is to prove the key tool for this: Proposition 2.3.6, which will allow us to obtain the desired bounds for any definable function after refining our stratification. We will also need bounds on second derivatives; those will be obtained in Corollary 2.3.10, by applying Proposition 2.3.6 to the first derivatives. Functions satisfying the desired bounds will be called “sedated”.

Before going into the details, here is an informal explanation. Given an  $\mathcal{L}$ -definable function  $f: X \rightarrow \mathcal{R}$  on an  $\mathcal{L}$ -definable set  $X \subseteq \mathcal{R}^n$ , we can remove a lower-dimensional subset from  $X$  using Corollary 2.2.2 to obtain a bound on  $\text{val}(\nabla f(x))$  which is good whenever  $x$  is not too close to the boundary of  $X$ :

$$\text{val}(\nabla f(x)) \geq \text{val}(f(x)) - \text{val}\text{dist}(x, \mathcal{R}^n \setminus X). \quad (2.4)$$

As a bound on  $\nabla f(x)$ , this is in some sense optimal, but it is often possible to get better bounds on individual partial derivatives: Very roughly, even near the boundary of  $X$ , one should be able to obtain good bounds on the partial derivatives in those directions which do not point towards the boundary; see Figure 2.1 (a). To construct stratifications, we will need such better bounds.

It is not so clear how to make this precise in general. Instead, the result in this subsection will provide the better bounds only in the rather specific situation we are in after the rectilinearization explained in Subsection 1.9: We only need a bound on the partial derivatives parallel to  $W := \mathcal{R}^{n'} \times \{0\}^{n-n'}$ , and that bound should not be affected by  $\text{dist}(x, W)$  being small even if  $W$  contains a boundary segment of  $X$ . (Such a bound makes most sense if  $X$  indeed has a boundary segment in  $W$ ; however, we will also prove and use the result when it doesn't.) The precise statement is that after removing a lower-dimensional subset from  $X$ , we obtain the estimate

$$\text{val}(\partial_i f(x)) \geq \text{val}(f(x)) - \text{val}\text{dist}(\text{pr}_{\leq n'}(x), \mathcal{R}^{n'} \setminus \text{pr}_{\leq n'}(X)) \quad \text{for } i = 1, \dots, n' \quad (2.5)$$

for points  $x \in X$  satisfying

$$\text{val}\text{dist}(x, \mathcal{R}^n \setminus X) \leq \text{val}(\text{pr}_{> n'}(x)). \quad (2.6)$$

Condition (2.6) ensures that  $x$  is not too close to a border of  $X$  different from  $W$ ; indeed, (2.5) cannot be expected for points close to a “diagonal border” like  $x_2$  in Figure 2.1.

Since it is  $\text{pr}_{\leq n'}(X)$  which appears in (2.5) and not  $X$  itself, the only lower-dimensional sets it makes sense to remove from  $X$  are sets of the form  $\text{pr}_{\leq n'}^{-1}(Z)$  for some  $Z \subseteq \text{pr}_{\leq n'}(X)$  (see Figure 2.1 (b)). This is how Proposition 2.3.6 is stated, and it is this set  $Z$  which will be used in the strategy outlined in Subsection 1.9 to shrink the  $n'$ -dimensional stratum.

The bound (2.5) is the one we will need to treat those augmented val-chains whose first two points  $a^0, a^1$  lie in two different strata; functions satisfying this bound will be called (a)-sedated. Proposition 2.3.6 also provides two variants of this, which are needed for other kinds of val-chains: to treat plain val-chains, we will need (b)-sedated functions, which satisfy a bound like (1.31), and to treat augmented val-chains whose first two points lie in the same stratum, we will need functions whose derivatives are (c)-sedated (see below).

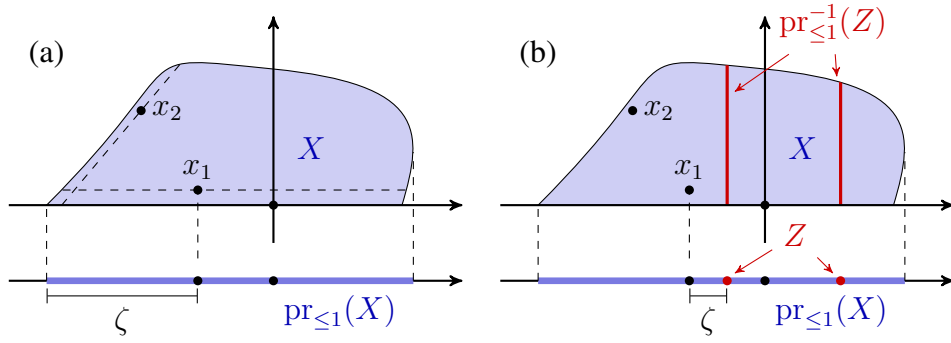


Figure 2.1: (a) At  $x_1$  and  $x_2$ , one can expect good bounds on the partial derivatives of  $f$  in the dashed directions, but not in the directions perpendicular to that. Proposition 2.3.6 provides good bounds on horizontal derivatives at points close to the  $x$ -axis: the bound on  $\partial_1 f(x_1)$  is computed using the distance  $\zeta$  in the projection  $\text{pr}_{\leq 1}(X)$ . Such a bound cannot be expected for  $\partial_1(x_2)$ , since  $x_2$  is close to a border of  $X$  different from the  $x$ -axis. (b) To obtain the bounds, it might be necessary to remove a lower-dimensional subset  $Z$  from  $\text{pr}_{\leq 1}(X)$ . This in effect weakens the condition on  $\partial_1(x_1)$ , since  $\zeta$  becomes smaller.

Everything described so far is what we need for short val-chains. For longer val-chains, say, living in strata of dimensions  $e_1 > \dots > e_m$ , we still need to bound the partial derivatives  $\partial_1 f(x), \dots, \partial_{e_m} f(x)$  of a function  $f$  with domain  $X \subseteq \mathcal{R}^{e_1}$ , but all the intermediate dimensions  $e_\ell$  also play a role, namely for the conditions specifying to which boundaries of  $X$  the point  $x$  is allowed to be close. To make this precise, we start by fixing some notation. In the whole subsection, we assume the following.

**Assumption 2.3.1.** Let the following be given:

- an integer  $m \geq 1$ ;
- integers  $0 < e_m < \dots < e_1$ ;
- an open  $\mathcal{L}$ -definable set  $X \subseteq \mathcal{R}^{e_1}$ .

**Notation 2.3.2.** We set  $Y := \text{pr}_{\leq e_m}(X)$ . For  $x \in X$ , we define:

- $\zeta_\ell := \zeta_\ell(x) := \text{dist}(\text{pr}_{\leq e_\ell}(x), \mathcal{R}^{e_\ell} \setminus \text{pr}_{\leq e_\ell}(X))$  for  $1 \leq \ell \leq m$
- $\sigma_\ell := \sigma_\ell(x) := \max\{1, \|\text{pr}_{> e_\ell}(x)\| \cdot \zeta_{\ell-1}(x)^{-1}\}$  for  $2 \leq \ell \leq m$ .

The shorter notation  $\zeta_\ell, \sigma_\ell$  will implicitly refer to a given point  $x \in X$  in context. Note that  $\zeta_\ell$  and  $\sigma_\ell$  implicitly also depend on  $X$ .

Some of this notation is illustrated in Figure 2.2. The purpose of  $\sigma_\ell$  is the following. One can only expect to obtain the best bounds on  $\partial_1 f(x), \dots, \partial_{e_m} f(x)$  at those  $x$  satisfying  $\text{val}(\sigma_\ell) = 0$  for all  $\ell$ . (Note that in the case  $m = 2$ , the condition  $\text{val}(\sigma_2) = 0$  is

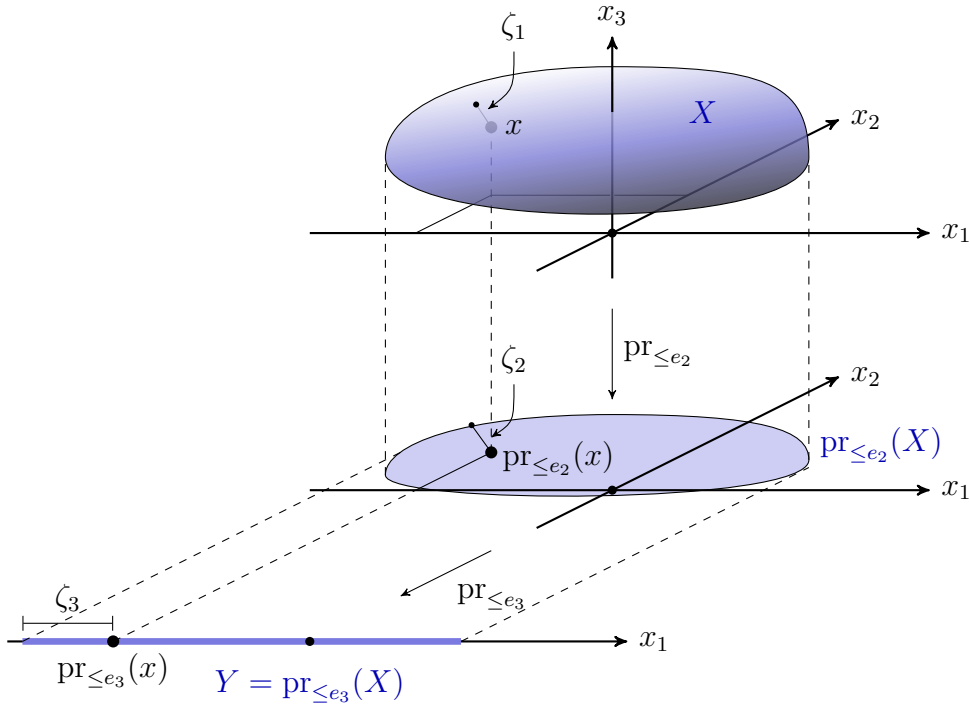


Figure 2.2: A picture illustrating some of Notation 2.3.2, in the case  $m = 3$ ,  $e_1 = 3$ ,  $e_2 = 2$ ,  $e_3 = 1$ .

exactly equivalent to (2.6).) However, even for  $x \in X$  not satisfying those conditions, it is possible to obtain a weakened bound, where the weakening is expressed in terms of the valuations of the  $\sigma_\ell$ . This leads to the following definition of sedated functions.

**Definition 2.3.3** (Sedated functions). Suppose that  $m$ ,  $e_\ell$  and  $X$  are given as in Assumption 2.3.1, and suppose that  $f: X \rightarrow \mathcal{R}$  is an  $\mathcal{L}$ -definable function. We consider three different versions:  $(v) \in \{(a), (b), (c)\}$ . In Version (b), we additionally assume  $m \geq 2$ . We call  $f$   $e_{[1,m]}-(v)$ -sedated (on  $X$ ) if it is  $C^1$  and if, for every  $x \in X$  and every  $1 \leq i \leq e_m$ , we have

$$\text{val}(\partial_i f(x)) \geq \text{val}(u_{(v)}(x)) - \text{val}(\zeta_m(x)) + \sum_{\ell=2}^m \text{val}(\sigma_\ell(x)), \quad \text{where} \quad (2.7)$$

$$u_{(a)}(x) = f(x), \quad u_{(b)}(x) = \max\{|f(x)|, \|\text{pr}_{>e_2}(x)\|\}, \quad u_{(c)}(x) = 1. \quad (2.8)$$

We call an  $\mathcal{L}$ -definable function  $X \rightarrow \mathcal{R}^n$   $e_{[1,m]}-(v)$ -sedated if each of its coordinate functions is  $e_{[1,m]}-(v)$ -sedated.

In this notation, “ $e_{[1,m]}$ ” is supposed to be considered as a short hand notation for the tuple  $(e_1, \dots, e_m)$ . In particular, for  $1 \leq k \leq \ell \leq m$  and  $f$  a function on a subset of  $\mathcal{R}^{e_k}$ , we also have a notion of being  $e_{[k,\ell]}-(v)$ -sedated.

*Remark 2.3.4.* Equation (2.7) depends on the domain  $X$ , since  $\zeta_\ell$  does. Nevertheless, if  $f$  is  $e_{[1,m]}(\mathbf{v})$ -sedated, then so is the restriction of  $f$  to any subset of  $X$ . Indeed, by shrinking  $X$ ,  $\zeta_\ell$  can only become smaller and  $\sigma_\ell$  can only become bigger, both of which make (2.7) easier to be satisfied.

*Remark 2.3.5.* If, in (b)-sedation, one allows  $e_1 = e_2$ , then (a) can be considered as a special case of (b) via some renumbering. However, for clarity, we wrote down the two cases separately.

**Proposition 2.3.6** (Sedating functions). *Fix  $(\mathbf{v}) \in \{(a), (b), (c)\}$ . Let  $m, e_\ell, X, Y$  be as in Assumption 2.3.1 and Notation 2.3.2 (with  $m \geq 2$  in Version (b)) and suppose that  $f: X \rightarrow \mathcal{R}$  is an  $\mathcal{L}$ -definable function which is  $e_{[1,m']}(\mathbf{v})$ -sedated for  $1 \leq m' < m$  (or  $2 \leq m' < m$ , in Version (b)). Suppose moreover that*

$$\begin{cases} \text{(a)} & \text{(no additional condition)} \\ \text{(b)} & f \text{ is } C^1 \text{ and } \text{val}(\nabla f(x)) \geq 0 \\ \text{(c)} & \text{val}(f(x)) \geq 0. \end{cases} \quad (2.9)$$

*Then there exists an  $\mathcal{L}$ -definable set  $Z \subseteq Y$  of dimension less than  $e_m$  such that the restriction of  $f$  to  $X \setminus \text{pr}_{\leq e_m}^{-1}(Z)$  is  $e_{[1,m]}(\mathbf{v})$ -sedated.*

*Remark 2.3.7.* The proposition directly implies the corresponding result for functions with range  $\mathcal{R}^n$ , by applying it to each of the coordinate functions.

*Remark 2.3.8.* In our application of this proposition, the bound (2.7) will only be needed on the subset  $X' := \{x \in X : \text{val}(\sigma_2) = \dots = \text{val}(\sigma_m) = 0\}$ , i.e., where the sum disappears and the bound is “best possible”. Nevertheless, we need to work with a notion of sedated functions imposing a bound on all of  $X$  for the following somewhat strange reason. The proof of Proposition 2.3.6 only works if  $X$  and  $f$  both are  $\mathcal{L}$ -definable; in particular, the “induction hypothesis” (that  $f$  is  $e_{[1,m']}(\mathbf{v})$ -sedated for  $m' < m$ ) is needed on an  $\mathcal{L}$ -definable set, so we need a formulation of that hypothesis which we can prove on all of  $X$ , and not just on  $X'$ .

The strategy of the proof of Proposition 2.3.6 is as follows. We will use Lemma 2.2.4 to choose, for each  $y \in Y$ , an element  $x = \tau(y) \in X_y := \{x \in X : \text{pr}_{\leq e_m}(x) = y\}$  where the difference between the two sides of (2.7) is worst, i.e., where the left hand side minus the right hand side is minimal. In particular, it suffices to prove that (2.7) holds for those  $x$ . Corollary 2.2.2 allows us to shrink  $Y$  in such a way that we obtain good bounds on the derivatives of  $f(\tau(y))$  in terms of  $\text{val}(\text{dist}(y, \mathcal{R}^{e_m} \setminus Y)) = \text{val}(\zeta_m)$ . We then obtain (2.7) by combining these bounds with the assumption about  $e_{[1,m']}(\mathbf{v})$ -sedation for  $m' < m$ .

To be able to apply Lemma 2.2.4 as described above, we need the difference of the two sides of (2.7) to be bounded on each fiber  $X_y$ . Such a bound can be obtained from Equation (2.7) for  $e_{[1,m-1]}$ -sedation, provided that we fix a lower bound on  $\|\text{pr}_{> e_m}(x)\|$ .

Thus, before applying the above strategy, we will treat points  $x$  with small  $\|\text{pr}_{>e_m}(x)\|$  separately. The idea for this is that for each fixed (small)  $d \geq 0$ , we can apply the same strategy as before to the subset  $\{x \in X : \|\text{pr}_{>e_m}(x)\| = d\}$ . Different  $d$  yield different sets  $Z_d$  to be removed from  $Y$  for (2.7) to hold. Instead of removing all of them from  $Y$  (which would be too much), we remove the limit (in a suitable sense) of  $Z_d$  for  $d \rightarrow 0$ ; this does not imply (2.7) b itself, but it does allow us to bound by how much it fails, and that is enough for applying the above strategy to the remainder of  $X$ .

Here are the details.

*Proof of Proposition 2.3.6.* During the proof, we will construct a set  $Z$  of dimension less than  $e_m$  which we will successively enlarge until the proposition is satisfied. More precisely, we will obtain something slightly stronger: We will find a  $Z \subseteq \mathcal{R}^{e_m}$  of dimension less than  $e_m$  such that

$$\text{val}(\partial_i f(x)) \geq \text{val}(u_{(v)}(x)) - \text{valdist}(\text{pr}_{\leq e_m}(x), Z) + \sum_{\ell=2}^m \text{val}(\sigma_\ell(x)) \quad (2.10)$$

holds for every  $x \in \hat{X} := X \setminus \text{pr}_{\leq e_m}^{-1}(Z)$ . This then implies that  $f \upharpoonright \hat{X}$  is  $(v)$ -sedated (using Remark 2.3.4 concerning the  $\sigma_\ell$ ).

In a very first step, we ensure that  $f$  is  $C^1$ : In Version (b), this is an assumption; in the other versions, if  $m \geq 2$ , it follows from the assumption that  $f$  is (say)  $e_{[1,1]}(v)$ -sedated, and if  $m = 1$ , this can be achieved by removing a suitable subset from  $X = Y$ .

Fix  $i \leq e_m$ . Equation (2.10) can be rewritten as

$$\text{val}(g_i(x)) \geq -\text{valdist}(\text{pr}_{\leq e_m}(x), Z), \quad (2.11)$$

where

$$g_i(x) := \partial_i f(x) \cdot u_{(v)}(x)^{-1} \cdot \prod_{\ell=2}^m \sigma_\ell^{-1}. \quad (2.12)$$

As in the above sketch of proof, given  $y \in Y$ , we write  $X_y$  for the fiber over  $X$  above  $y$ , and similarly, if  $X' \subseteq X$  is a subset, we set  $X'_y := X' \cap X_y$ .

We will prove the following.

**Claim 1:** Suppose that  $Z \subseteq \mathcal{R}^{e_m}$  is an  $\mathcal{L}$ -definable set of dimension less than  $e_m$  and that  $X' \subseteq X$  is an  $\mathcal{L}$ -definable subset such that for every  $y \in Y \setminus Z$  and every  $i \leq e_m$ ,  $|g_i|$  is bounded on the fiber  $X'_y$ . Then there exists an  $\mathcal{L}$ -definable set  $\hat{Z} \supseteq Z$  of dimension less than  $e_m$  such that we have

$$\text{val}(g_i(x)) \geq -\text{valdist}(\text{pr}_{\leq e_m}(x), \hat{Z}) \quad \text{for every } x \in X'. \quad (2.13)$$

Before proving Claim 1, we show how it implies the proposition. It suffices to prove that the set  $Z$  of  $y \in Y$  such that  $|g_i(x)|$  is unbounded on the fiber  $X_y$  has dimension less than  $m$ . Indeed, then we obtain (2.11) by applying the claim to  $X' := X$ .

If  $m = 1$ , then  $|g_i(x)|$  is bounded on each  $X_y$  for the trivial reason that  $X_y$  is a singleton; thus assume  $m \geq 2$ .

To bound  $|g_i(x)|$ , we first check that for every  $x \in X$ , we have

$$\text{val}(g_i(x)) \geq -\text{val}(\text{pr}_{>e_m}(x)). \quad (2.14)$$

We have

$$\text{val}(g_i(x)) = \underbrace{\text{val}(\partial_i f(x)) - \text{val}(u_{(v)}(x)) - \sum_{\ell=2}^{m-1} \text{val}(\sigma_\ell)}_{(*)} - \text{val}(\sigma_m).$$

In Version (b), if  $m = 2$  then the last term in  $(*)$  disappears, and hence (2.14) follows from the following three items: the assumption (2.9),  $\text{val}(u_{(b)}(x)) \leq \text{val}(\text{pr}_{>e_2}(x))$ , and  $\text{val}(\sigma_2) \leq 0$ . In all other cases, the assumption that  $f$  is  $e_{[1, m-1]}(v)$ -sedated implies  $(*) \geq -\text{val}(\zeta_{m-1})$ , which, together with  $\text{val}(\sigma_m) \leq \text{val}(\text{pr}_{>e_m}(x)) - \text{val}(\zeta_{m-1})$  (by the definition of  $\sigma_m$ ), implies (2.14).

For  $d \in \mathcal{R}_{\geq 0}$ , set

$$X'_d := \{x \in X : \|\text{pr}_{>e_m}(x)\| = d\};$$

by (2.14),  $|g_i|$  is bounded on  $X'_d$  for each fixed  $d$ , so Claim 1 (used in the language  $\mathcal{L}(d)$ ) yields an  $\mathcal{L}(d)$ -definable set  $Z_d \subseteq \mathcal{R}^{e_m}$  of dimension less than  $m$  and such that we have

$$\text{val}(g_i(x)) \geq -\text{val}(\text{dist}(y, Z_d)) \quad (2.15)$$

for  $x \in X'_d$  and  $y := \text{pr}_{\leq e_m}(x)$ . By the Compactness Theorem, we may assume that the sets  $Z_d$  are defined uniformly in  $d$ , so that the following sets are  $\mathcal{L}$ -definable:

$$Z_\bullet := \bigcup_{d \geq 0} (Z_d \times \{d\}) \subseteq \mathcal{R}^{e_m} \times \mathcal{R}_{\geq 0} \quad \text{and} \\ Z := \{y \in \mathcal{R}^{e_m} : (y, 0) \in \text{cl}(Z_\bullet)\},$$

(where  $\text{cl}(Z_\bullet)$  denotes the topological closure of  $Z_\bullet$ ). Since  $Z_d$  has dimension less than  $e_m$  for every  $d$ , we have  $\dim \partial Z_\bullet < \dim Z_\bullet \leq e_m$  and hence  $\dim Z < e_m$  (since  $Z \subseteq \partial Z_\bullet \cup Z_0$ ). We claim that  $|g_i|$  is bounded on each fiber  $X_y$  with  $y \neq Z$ .

Fix  $y \in Y \setminus Z$  and consider  $x \in X_y$  with  $y \in Y \setminus Z$  and set  $d := \|\text{pr}_{>e_m}(x)\|$  (so that  $x \in X'_d$ ). Inequality (2.14) provides a bound on  $|g_i(x)|$  for big  $d$ , and for small  $d$ , we will obtain a bound from (2.15). More precisely, set

$$d_0 := \text{dist}((y, 0), Z_\bullet)$$

(which is strictly positive, by definition of  $Z$ ). By (2.14) it suffices to bound  $|g_i(x)|$  for those  $x$  satisfying  $\text{val}(\text{pr}_{>e_m}(x)) > \text{val}(d_0)$ . This implies  $\text{valdist}((y, d), Z_\bullet) = \text{valdist}((y, 0), Z_\bullet)$ , and hence we must have

$$\text{valdist}(y, Z_d) = \text{valdist}((y, d), Z_d \times \{d\}) \leq \text{valdist}((y, d), Z_\bullet) = \text{val}(d_0).$$

So for such a  $d$ , we obtain

$$\text{val}(g_i(x)) \stackrel{(2.15)}{\geq} -\text{valdist}(y, Z_d) \geq -\text{val}(d_0).$$

Thus  $|g_i|$  is bounded on all of  $X_y$ , which finishes the proof that Claim 1 implies the proposition.

**Proof of Claim 1:** Even though the case  $m = 1$  (Versions (a), (c)) does not need to be treated separately, we do note that for  $m = 1$ , Claim 1 follows directly by applying Corollary 2.2.2 to  $f$  (and using (2.9) in Version (c)).

Fix  $i \leq e_m$  for the entire proof of the claim. (We can treat each  $g_i$  separately.)

Set  $Y' := \text{pr}_{\leq e_m}(X') \setminus Z$ . For  $y \in Y'$ ,  $|g_i|$  is bounded on  $X'_y$ , so we can apply Lemma 2.2.4 to the restriction  $g_i \upharpoonright X'_y$ , using the language  $\mathcal{L}(y)$ . Doing this for all  $y \in Y'$  (and applying the Compactness Theorem) yields an  $\mathcal{L}$ -definable function  $\tau: Y' \rightarrow X$  with  $\tau(y) \in X'_y$  such that

$$\text{val}(g_i(x)) \geq \text{val}(g_i(\tau(y))) \quad \text{for all } x \in X'_y \text{ and } y \in Y'. \quad (2.16)$$

We will prove that after a suitable enlargement of  $Z$ , we obtain

$$\text{val}(g_i(\tau(y))) \geq -\text{valdist}(y, Z) \quad \text{for every } y \in Y'; \quad (2.17)$$

together with (2.16), this implies (2.13).

In the remainder of the proof,  $\zeta_\ell$  and  $\sigma_\ell$  always refer to the point  $x := \tau(y)$ . Plugging (2.12) (the definition of  $g_i$ ) into (2.17) yields a condition on  $\partial_i f$ :

$$\text{val}(\partial_i f(\tau(y))) \geq \text{val}(u_{(v)}(\tau(y))) - \text{valdist}(y, Z) + \sum_{\ell=2}^m \text{val}(\sigma_\ell). \quad (2.18)$$

Consider the derivative of the function  $h(y) := f(\tau(y))$  with respect to the  $i$ th coordinate. Using the notation

$$\tau(y) = (y, \tau_{e_m+1}(y), \dots, \tau_{e_1}(y)), \quad (2.19)$$

we can write it as

$$\partial_i h(y) = \partial_i f(\tau(y)) + \sum_{k=e_m+1}^{e_1} \partial_k f(\tau(y)) \cdot \partial_i \tau_k(y), \quad (2.20)$$

so to obtain (2.18), it suffices to prove that in (2.20), (i) the left hand side and (ii) all summands of the sum over  $k$  have valuation at least that of the right hand side of (2.18).

For (i), apply Corollary 2.2.2 to  $h$  (extended trivially outside of  $Y'$ ). This yields that, by enlarging  $Z$ , we can achieve

$$\text{val}(\partial_i h(y)) \geq \text{val}(h(y)) - \text{val}\text{dist}(y, Z). \quad (2.21)$$

Since the sum in (2.18) is at most 0 (by definition of  $\sigma_\ell$ ), it remains to check that  $\text{val}(f(\tau(y))) \geq \text{val}(u_{(v)}(\tau(y)))$ ; this follows from the definition of  $u_{(v)}$ , and, in Version (c), (2.9).

For (ii), fix  $k$  (with  $e_m < k \leq e_1$ ) and choose  $m'$  such that  $e_{m'+1} < k \leq e_{m'}$ ; note that  $m' < m$ . Our goal is to prove

$$\text{val}(\partial_k f(x)) + \text{val}(\partial_i \tau_k(y)) \geq \text{val}(u_{(v)}(x)) - \text{val}\text{dist}(y, Z) + \sum_{\ell=2}^m \text{val}(\sigma_\ell) \quad (2.22)$$

(where  $x = \tau(y)$ ). Applying Corollary 2.2.2 to  $\tau_k$  (again, extended trivially outside of  $Y'$ ) yields, after further enlarging  $Z$ ,

$$\text{val}(\partial_i \tau_k(y)) \geq \text{val}(\tau_k(y)) - \text{val}\text{dist}(y, Z) \geq \text{val}(\text{pr}_{>e_{m'+1}}(x)) - \text{val}\text{dist}(y, Z). \quad (2.23)$$

In the case  $m' = 1$  of Version (b), (2.22) now follows from these three items: (2.9) (which implies  $\text{val}(\partial_k f(x)) \geq 0$ ),  $\text{val}(u_{(v)}(x)) \leq \text{val}(\text{pr}_{>e_2}(x))$ , and  $\text{val}(\sigma_\ell) \leq 0$ . Thus we may now suppose that either  $m' \geq 2$  or that we are not in Version (b). Then the assumption that  $f$  is  $e_{[1, m']}$ - $(v)$ -sedated implies

$$\text{val}(\partial_k f(x)) \geq \text{val}(u_{(v)}(x)) - \text{val}(\zeta_{m'}) + \sum_{\ell=2}^{m'} \text{val}(\sigma_\ell), \quad (2.24)$$

and (2.22) follows by taking the sum of (2.23) and (2.24) and then noting that  $\text{val}(\text{pr}_{>e_{m'+1}}(x)) - \text{val}(\zeta_{m'}) \geq \text{val}(\sigma_{m'+1})$  and  $\text{val}(\sigma_\ell) \leq 0$ .

This finishes the proof of (ii), and hence of (2.18), and hence of Claim 1, and hence of Proposition 2.3.6.  $\square$

The notion of (c)-sedation will be applied to the first derivatives of a function, to control its second derivatives. We introduce a corresponding notion. (Note that similar kinds of bounds also appear in [NV].)

**Definition 2.3.9** ( $(c_2)$ -sedated functions). Suppose that  $m$ ,  $e_\ell$  and  $X$  are given as in Assumption 2.3.1. We call an  $\mathcal{L}$ -definable function  $f: X \rightarrow \mathcal{R}$   $e_{[1, m]}$ - $(c_2)$ -sedated if it is  $C^2$ ,  $\text{val}(\text{Jac}_x f) \geq 0$  for every  $x \in X$ , and for  $1 \leq i \leq e_m$ ,  $1 \leq j \leq e_1$ , we have

$$\text{val}(\partial_{ij} f(x)) \geq -\text{val}(\zeta_m(x)) + \sum_{\ell=2}^m \text{val}(\sigma_\ell(x)), \quad (2.25)$$

where  $\zeta_m$  and  $\sigma_\ell$  are as in Notation 2.3.2. We call an  $\mathcal{L}$ -definable function  $X \rightarrow \mathcal{R}^n$   $e_{\leq m}$ - $(c_2)$ -sedated if each of its coordinate functions is  $e_{[1, m]}$ - $(c_2)$ -sedated.

**Corollary 2.3.10** (( $c_2$ )-sedating functions). *Let  $m, e_\ell, X, Y$  be as above, and suppose that  $f: X \rightarrow \mathcal{R}$  is an  $\mathcal{L}$ -definable function which is  $e_{[1,m']}$ -( $c_2$ )-sedated for all  $m' < m$ . Suppose moreover that  $\text{val}(\nabla f) \geq 0$  (this follows anyway if  $m \geq 2$ ). Then there exists an  $\mathcal{L}$ -definable set  $Z \subseteq Y$  of dimension less than  $e_m$  such that the restriction of  $f$  to  $X \setminus \text{pr}_{\leq e_m}^{-1}(Z)$  is  $e_{[1,m]}$ -( $c_2$ )-sedated.*

*Proof.* If  $m = 1$ , we start by removing a lower-dimensional subset from  $X = Y$  to ensure that  $f$  is  $C^2$ . (If  $m \geq 2$ ,  $f$  is already  $C^2$ .) Then we apply Proposition 2.3.6 (c) to each of the derivatives  $\partial_j f$  ( $1 \leq j \leq e_1$ ).  $\square$

We finish this subsection by proving that being sedated is preserved under certain kinds of transformations, which will be the building blocks of the rectilinearization maps mentioned in Subsection 1.9.

**Lemma 2.3.11** (Sedation and rectilinearization). *Fix  $(v) \in \{(a), (b), (c_2)\}$ , and let the following be given:*

- integers  $1 \leq m' < m$  ( $2 \leq m' < m$  in Version (b)),
- integers  $e_1 > \dots > e_m > 0$ ,
- $\mathcal{L}$ -definable sets  $X, \hat{X} \subseteq \mathcal{R}^{e_1}$ ,
- $\mathcal{L}$ -definable functions  $f: X \rightarrow \mathcal{R}$  and  $\hat{f}: \hat{X} \rightarrow \mathcal{R}$ .

*Suppose that there exists an  $\mathcal{L}$ -definable bijection  $\psi: \hat{X} \rightarrow X$  such that  $\hat{f} = f \circ \psi$  and which sends*

$$\begin{aligned} \hat{x} &= (x_m, \hat{x}_{m-1}, x_\star) \in \mathcal{R}^{e_m} \times \mathcal{R}^{e_{m-1}-e_m} \times \mathcal{R}^{e_1-e_{m-1}} & \text{to} \\ x &= (x_m, x_{m-1}, x_\star) = (x_m, \hat{x}_{m-1} + g(x_m), x_\star), \end{aligned}$$

*where  $g: \text{pr}_{\leq e_m}(X) \rightarrow \mathcal{R}^{e_{m-1}-e_m}$  is  $e_{[m,m]}$ -( $c_2$ )-sedated. Then  $f$  is  $e_{[1,m']}$ -( $v$ )-sedated iff  $\hat{f}$  is  $e_{[1,m']}$ -( $v$ )-sedated.*

*Proof.* The lemma is symmetric with respect to swapping  $\hat{X}$  and  $X$ ; we will carry out various arguments only in one direction without further notice.

We start by verifying that for  $\ell \leq m-1$ , the valuations  $\text{val}(\zeta_\ell)$  and  $\text{val}(\sigma_\ell)$  from Notation 2.3.2 are preserved by  $\psi$ . More precisely, we show that, for  $\hat{x} = (x_m, \hat{x}_{m-1}, x_\star) \in \hat{X}$  and  $x = (x_m, x_{m-1}, x_\star) = \psi(\hat{x})$ , we have

$$\text{valdist}(\text{pr}_{\leq e_\ell}(\hat{x}), \mathcal{R}^{e_\ell} \setminus \text{pr}_{\leq e_\ell}(\hat{X})) = \text{valdist}(\text{pr}_{\leq e_\ell}(x), \mathcal{R}^{e_\ell} \setminus \text{pr}_{\leq e_\ell}(X)) \quad (2.26)$$

for  $1 \leq \ell \leq m-1$ . Since  $\text{pr}_{> e_{m-1}} \circ \psi = \text{pr}_{> e_{m-1}}$ , this then also implies

$$\text{val}(\sigma_\ell(\hat{x})) = \text{val}(\sigma_\ell(x)), \quad (2.27)$$

where  $\sigma_\ell(\hat{x})$  is computed with respect to  $\hat{X}$  and  $\sigma_\ell(x)$  is computed with respect to  $X$ .

To prove (2.26), we assume  $\ell = 1$ ; for other  $\ell$ , the same proof applies, after replacing  $x, \hat{x}, X, \hat{X}$  by their projections to  $\mathcal{R}^{e_\ell}$ .

Both sides of (2.26) are no less than  $\mu := \text{val}(\text{dist}(x_m, \mathcal{R}^{e_m} \setminus \text{pr}_{\leq e_m}(X)))$ , so it suffices to verify that given an element  $\hat{y} \in \mathcal{R}^{e_1} \setminus \hat{X}$  satisfying  $\text{val}(\hat{y} - \hat{x}) > \mu$ , we can find an element  $y \in \mathcal{R}^{e_1} \setminus X$  satisfying  $\text{val}(y - x) = \text{val}(\hat{y} - \hat{x})$ .

We write  $\hat{y} = (y_m, \hat{y}_{m-1}, y_\star) \in \mathcal{R}^{e_m} \times \mathcal{R}^{e_{m-1}-e_m} \times \mathcal{R}^{e_1-e_{m-1}}$ . By definition of  $\mu$ , the function  $g$  is defined on the entire ball  $B := B_{>\mu}(x_m)$ . This means that, first of all,  $y := (y_m, \hat{y}_{m-1} + g(y_m), y_\star)$  is well-defined, and secondly, the Mean Value Theorem argument from Remark 2.1.13 applies, yielding

$$\text{val}(g(y_m) - g(x_m)) \geq \text{val}(y_m - x_m);$$

now an easy computation yields  $\text{val}(y - x) = \text{val}(\hat{y} - \hat{x})$ , as desired.

From (2.26) and (2.27), we obtain, for  $x = \psi(\hat{x})$ :

$$-\text{val}(\zeta_{m'}(x)) + \sum_{\ell=2}^{m'} \text{val}(\sigma_\ell(x)) = -\text{val}(\zeta_{m'}(\hat{x})) + \sum_{\ell=2}^{m'} \text{val}(\sigma_\ell(\hat{x})) =: \lambda(x).$$

That the function  $g$  is  $(c_2)$ -sedated in particular means that  $\text{val}(\text{Jac } g) \geq 0$ . This yields the following equations concerning the partial derivatives of  $\hat{f}(x_m, \hat{x}_{m-1}, x_\star) = f(x_m, \hat{x}_{m-1} + g(x_m), x_\star)$ :

$$\begin{aligned} \min_{1 \leq i \leq e_{m-1}} \text{val}(\partial_i \hat{f}(\hat{x})) &= \min_{1 \leq i \leq e_{m-1}} \text{val}(\partial_i f(x)) && \text{and} \\ \text{val}(\partial_i \hat{f}(\hat{x})) &= \text{val}(\partial_i f(x)) && \text{for } e_{m-1} < i \leq e_1. \end{aligned} \quad (2.28)$$

Now we are ready to prove the claims of the lemma. For (v) = (a), (b),  $f$  is  $e_{[1, m']}$ - $(v)$ -sedated iff

$$\min_{1 \leq i \leq e_{m'}} \text{val}(\partial_i f(x)) \geq \text{val}(u_{(v)}(x)) + \lambda(x) \quad \text{for every } x \in X, \quad (2.29)$$

and similarly for  $\hat{f}$ . The left hand sides of (2.29) are equal for  $f$  and  $\hat{f}$  by (2.28), and the right hand sides are equal since  $\hat{f}(\hat{x}) = f(x)$  and, in Version (b) (which implies  $m \geq 3$ ),  $\text{pr}_{>e_2}(x) = \text{pr}_{>e_2}(\hat{x})$ .

Finally, suppose that  $f$  is  $e_{[1, m']}$ - $(c_2)$ -sedated, i.e.,

$$\text{val}(\text{Jac } f) \geq 0 \quad \text{and} \quad \text{val}(\partial_{ij} f) \geq \lambda(x) \quad \text{for } 1 \leq i \leq e_{m'}, 1 \leq j \leq e_1.$$

(To simplify notation, from now on, we omit the points at which the derivatives are taken.) Using (2.28), we obtain  $\text{val}(\text{Jac } \hat{f}) \geq 0$ , and it remains to verify that  $\text{val}(\partial_{ij} \hat{f}) \geq \lambda$ . Direct

computation of this second derivative yields the following, where  $g = (g_{e_m+1}, \dots, g_{e_m-1})$  and where we set  $\partial_k g := 0$  for  $k > e_m$ :

$$\begin{aligned} \partial_{ij}(f \circ \psi) &= \partial_{ij}f + \sum_{e_m < \ell \leq e_{m-1}} \partial_{j\ell}f \cdot \partial_i g_\ell + \sum_{e_m < \ell \leq e_{m-1}} \partial_{i\ell}f \cdot \partial_j g_\ell \\ &+ \sum_{e_m < \ell, \ell' \leq e_{m-1}} \partial_{\ell\ell'}f \cdot \partial_i g_\ell \cdot \partial_j g_{\ell'} + \sum_{e_m < \ell \leq e_{m-1}} \partial_\ell f \cdot \partial_{ij}g_\ell. \end{aligned}$$

All the second derivatives of  $f$  appearing on the right hand side have valuation at least  $\lambda$  (note that  $\ell, \ell' \leq e_{m'}$ ). Together with  $\text{val}(\text{Jac } g) \geq 0$ , we get the desired bound for everything except the last sum. In that one, we have  $\text{val}(\partial_\ell f) \geq 0$  and  $\text{val}(\partial_{ij}g_\ell) \geq -\text{val}(\zeta_m)$  (since  $g$  is  $e_{[m,m]}(c_2)$ -sedated). Now  $-\text{val}(\zeta_m) \geq -\text{val}(\zeta_{m'}) \geq \lambda$  since  $\text{val}(\sigma_\ell) \leq 0$  for all  $\ell$ , so also here, we get the desired bound.  $\square$

### 3 The main proof

This entire section constitutes the proof of Theorem 1.6.7. We fix, once and for all, a closed,  $\mathcal{L}$ -definable set  $X \subseteq \mathcal{R}^n$ .

#### 3.1 Some notation

We fix some notation which will be useful at various places in the proof. Suppose that we have already fixed a stratification  $\mathcal{X}$  of  $X$  (in the sense of Definition 1.2.1); in particular, we assume that each  $X^i$  is closed. We moreover assume that the strata, i.e., the definably connected components of the skeletons  $\overset{\circ}{X}^i$ , form a bradycell decomposition in the sense of Definition 2.1.4. (Recall that definably connectedness always refers to the language  $\mathcal{L}$ .)

**Notation 3.1.1** (Aligning and groups of coordinates). Suppose that  $\mathcal{S} = (S^\ell)_{0 \leq \ell \leq m}$  is a sequence of strata with  $S^\ell \subseteq \overset{\circ}{X}^{e_\ell}$  for some  $e_0 \geq e_1 > e_2 > \dots > e_m$ . These inequalities, together with  $e_0 \leq n$ , imply  $m \leq n + 1$ , so Proposition 2.1.5 provides an aligner  $\kappa \in \mathcal{C}_n$  such that each transformed set  $\kappa(S^\ell)$  is aligned in the sense of Definition 2.1.1. In such a situation, i.e., when  $\mathcal{S}$  and an appropriate  $\kappa$  are given, we will assume that the strata  $S^\ell$  are already aligned by transforming our coordinate system using  $\kappa$ . (Why this assumption is harmless will be explained at the appropriate places.) We will moreover use the following notation, where  $0 \leq \ell \leq m$ :

- We write  $\bar{S}^\ell := \text{pr}_{\leq e_\ell}(S^\ell)$  for the base of  $S^\ell$  and  $\rho^\ell: \bar{S}^\ell \longrightarrow \mathcal{R}^{n-e_\ell}$  for the map whose graph is  $S^\ell$ .
- We introduce a notation for “groups of coordinates” of points  $x = (x_1, \dots, x_n) \in$

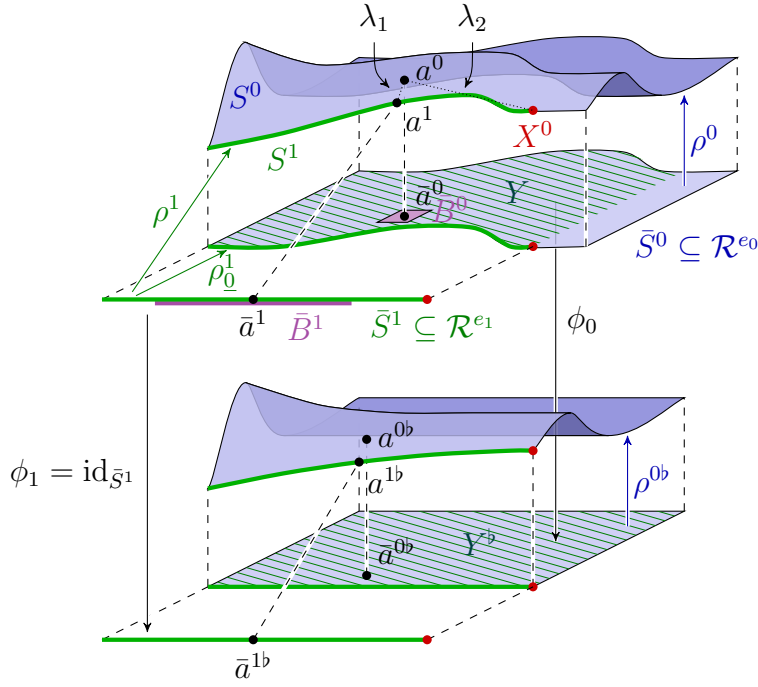


Figure 3.1: An overview of some of the notation from Subsections 3.1 and 3.2 (for a plain val-chain  $a^0, a^1$ ) and a bit of additional notation used later.

$\mathcal{R}^n$ :

$$\begin{aligned} x_{\underline{m}} &:= (x_1, x_2, \dots, x_{e_m}), \\ x_{\underline{\ell}} &:= (x_{e_{\ell+1}+1}, \dots, x_{e_{\ell}}) \quad \text{for } 0 \leq \ell < m \text{ and} \\ x_{\underline{\star}} &:= (x_{e_0+1}, \dots, x_n). \end{aligned}$$

In other words,

$$x = (x_m, x_{m-1}, \dots, x_0, x_{\underline{\star}}).$$

(Note that  $x_0$  might be the empty tuple since possibly  $e_0 = e_1$ .) We use a similar notation for points in  $\mathcal{R}^{e_{\ell}}$  and for the functions  $\rho^{\ell}$  ( $0 \leq \ell \leq m$ ):

$$\begin{aligned} x &= (x_m, x_{m-1}, \dots, x_{\ell+1}, x_{\underline{\ell}}) && \text{for } x \in \mathcal{R}^{e_{\ell}}; \\ \rho^{\ell} &= (\rho_{\underline{\ell-1}}^{\ell}, \dots, \rho_0^{\ell}, \rho_{\underline{\star}}^{\ell}). \end{aligned}$$

Now suppose that we additionally have a val-chain  $a^0, \dots, a^m$  with  $a^{\ell} \in S^{\ell}$  and with distances  $\lambda_1 > \dots > \lambda_{m+1}$  (and dimensions  $e_0 \geq e_1 > \dots > e_m$ ). There are natural balls  $\bar{B}^{\ell} \subseteq \bar{S}^{\ell}$  associated with such a val-chain, though it requires an argument to see that the balls, as defined below, are really subsets of  $\bar{S}^{\ell}$ .

**Notation 3.1.2.** Given a val-chain  $a^0, \dots, a^m$  with  $a^{\ell} \in S^{\ell}$  and with distances  $\lambda_1 > \dots > \lambda_{m+1}$ , we set

$$\bar{B}^{\ell} := B_{>\lambda_{\ell+1}}(\text{pr}_{\leq e_{\ell}}(a^0)) \subseteq \mathcal{R}^{e_{\ell}} \quad \text{for } 0 \leq \ell \leq m.$$

(If  $\lambda_{m+1} = -\infty$ , we set  $\bar{B}^m = \mathcal{R}^{e_m}$ .)

Note that  $\bar{B}^\ell$  also contains the projections  $\text{pr}_{\leq e_\ell}(a^1), \dots, \text{pr}_{\leq e_\ell}(a^\ell)$  and that it is the projection of the largest ball around  $a^0$  which is disjoint from  $X^{e_\ell-1}$ .

**Lemma 3.1.3.** *In the situation of Notations 3.1.1 and 3.1.2, the following hold for  $0 \leq \ell \leq m$ .*

1. *The ball  $\bar{B}^\ell$  is contained in  $\bar{S}^\ell$ . In particular,  $\text{pr}_{\leq e_\ell}(a^0) \in \bar{S}^\ell$ , so the function  $\rho^\ell$  is defined at the point  $\text{pr}_{\leq e_\ell}(a^0)$ .*
2. *The function  $\rho^\ell$  satisfies*

$$\text{val}(\rho^\ell(x^1) - \rho^\ell(x^2)) \geq \text{val}(x^1 - x^2) \quad \text{for } x^1, x^2 \in \bar{B}^\ell.$$

3. *For  $\bar{a}^0 := \text{pr}_{\leq e_\ell}(a^0)$  and  $a^{[\ell]} := (\bar{a}^0, \rho^\ell(\bar{a}^0)) \in S^\ell$ , we have  $\text{val}(a^0 - a^{[\ell]}) = \lambda_\ell$ . In particular, the sequences  $a^0, a^{[1]}, a^2, \dots, a^m$  and  $a^{[1]}, a^1, a^2, \dots, a^m$  are val-chains.*

Note that in the last statement, we might have  $a^{[1]} = a^0$ , namely when  $S^0 = S^1$ .

*Proof of Lemma 3.1.3.* (1) We have  $B := B_{>\lambda_{\ell+1}}(a^0) \cap S^\ell \neq \emptyset$  but  $B \cap \partial S^\ell = \emptyset$  (since  $\partial S^\ell \subseteq X^{e_\ell-1}$  and  $\text{val}(\text{dist}(a^0, X^{e_\ell-1})) = \lambda_{\ell+1}$ ), so Lemma 2.1.12 implies  $\bar{B}^\ell = \text{pr}_{\leq e_\ell}(B) \subseteq \bar{S}^\ell$ .

(2) This is just the Mean Value Theorem argument from Remark 2.1.13.

(3) The inequality  $\text{val}(a^0 - a^{[\ell]}) \leq \lambda_\ell$  follows from the definition of val-chain, since  $a^{[\ell]} \in X^{e_\ell}$ . For the other inequality, set  $\bar{a}^\ell := \text{pr}_{\leq e_\ell}(a^\ell)$ . Then  $\text{val}(a^0 - a^\ell) = \lambda_\ell$  implies  $\text{val}(\bar{a}^0 - \bar{a}^\ell) \geq \lambda_\ell$ , and then (2) yields  $\text{val}(a^{[\ell]} - a^\ell) \geq \lambda_\ell$ . This together with  $\text{val}(a^0 - a^\ell) = \lambda_\ell$  implies  $\text{val}(a^0 - a^{[\ell]}) \geq \lambda_\ell$ .

The ‘‘in particular’’ part is clear from the definition of val-chains. (Note that the second one is an augmented val-chain).  $\square$

## 3.2 Rectilinearization

In the setting of Notation 3.1.1, we will sometimes need to ‘‘rectilinearize’’ along the lower-dimensional strata: We will apply a map that translates the coordinates  $x_\ell$  by  $\rho_\ell^{\ell+1}$ . (Note that the maps  $\rho_j^{\ell+1}$  for  $j \leq \ell - 1$  are not used for rectilinearization.) Here is our notation for this:

**Notation 3.2.1** (Rectilinearization). For  $0 \leq \ell \leq m$ , and suitable  $x = (x_m, \dots, x_\ell) \in \mathcal{R}^{e_\ell}$ , we define

$$\begin{aligned} \phi_\ell(x) &:= x^b && \text{where } x^b \text{ is given by} \\ x_m^b &:= x_m, \\ x_{m-1}^b &:= x_{m-1} - \rho_{m-1}^m(x_m), \\ x_{m-2}^b &:= x_{m-2} - \rho_{m-2}^{m-1}(x_m, x_{m-1}), \\ &\vdots \\ x_\ell^b &:= x_\ell - \rho_\ell^{\ell+1}(x_m, \dots, x_{\ell+1}). \end{aligned}$$

Here, “suitable  $x$ ” means that all the involved maps  $\rho_j^{j+1}$  are defined, i.e.,  $\phi_\ell(x)$  is defined if  $\text{pr}_{\leq e_j}(x) \in \bar{S}^j$  for  $\ell < j \leq m$ .

*Remark 3.2.2.* The definition of  $\phi_\ell$  can also be written inductively:

$$\begin{aligned} \phi_m(x) &= x && \text{for } x \in \mathcal{R}^{e_m} \text{ and} \\ \phi_\ell((\bar{x}, x_\ell)) &= (\phi_{\ell+1}(\bar{x}), x_\ell - \rho_\ell^{\ell+1}(\bar{x})) && \text{for } (\bar{x}, x_\ell) \in \mathcal{R}^{e_{\ell+1}} \times \mathcal{R}^{e_\ell - e_{\ell+1}}, 0 \leq \ell < m. \end{aligned}$$

Note that if  $e_0 = e_1$ , then  $\phi_0 = \phi_1$ .

We fix some more notation:

**Notation 3.2.3.** We set

$$Y := \{x \in \mathcal{R}^{e_0} : \text{pr}_{\leq e_\ell}(x) \in \bar{S}^\ell \text{ for } 0 \leq \ell \leq m\},$$

which is a subset of the domain of  $\phi_0$ . We write  $Y^b := \phi_0(Y)$  for the rectilinearization of  $Y$  (note that  $\phi_0$  induces a bijection  $Y \rightarrow Y^b$ ) and

$$\rho^{\ell b} := \rho^\ell \circ \phi_\ell^{-1}$$

for the rectilinearization of  $\rho^\ell$ , where  $0 \leq \ell \leq m$ .

Note that the domain of  $\rho^{0b}$  is  $Y^b$ .

*Remark 3.2.4.* From  $\text{val}(\text{Jac } \rho^\ell) \geq 0$ , one easily deduces  $\text{Jac } \phi_\ell \in \text{GL}_{e_\ell}(\mathcal{O}_{\mathcal{R}})$  (at every point of the domain of  $\phi_\ell$ ), using Remark 3.2.2. (Intuitively, this follows because the expression of  $\text{Jac } \phi_\ell$  in terms of the partial derivatives of  $\rho^i$ ,  $\ell < i \leq m$ , is a “lower triangular matrix with identities on the diagonal”).

To transfer arguments between the rectilinearized and the unrectilinearized setting, we need the maps  $\phi_\ell$  to be isometries with respect to the valuation. This is not true everywhere, but it is true on the balls  $\bar{B}^\ell = B_{>\lambda_{\ell+1}}(\text{pr}_{\leq e_\ell}(a^0))$  introduced in Notation 3.1.2, which is what we really need. Here is the precise statement.

**Lemma 3.2.5.** *Suppose that  $a^0, \dots, a^m$  is a val-chain with  $a^\ell \in S^\ell$  and with distances  $\lambda_\ell$ . For  $0 \leq \ell \leq m$ ,  $\phi_\ell$  is defined on  $\bar{B}^\ell$  and the restriction  $\phi_\ell \upharpoonright \bar{B}^\ell$  is a valuative isometry (i.e.,  $\text{val}(\phi_\ell(x^1) - \phi_\ell(x^2)) = \text{val}(x^1 - x^2)$ ) with image  $B_{>\lambda_{\ell+1}}(\phi_\ell(\text{pr}_{\leq e_\ell}(a^0)))$ .*

*Proof.* Use induction and Remark 3.2.2. That  $\phi_\ell \upharpoonright \bar{B}^\ell$  is defined follows from Lemma 3.1.3 (1), that it is an isometry follows from Lemma 3.1.3 (2), and to obtain that the image is all of  $B_{>\lambda_{\ell+1}}(\phi_\ell(\text{pr}_{\leq e_\ell}(a^0)))$ , consider its inverse (which is easy to specify explicitly).  $\square$

### 3.3 Defining the stratification

In this section, we construct a stratification of the given set  $X \subseteq \mathcal{R}^n$ . (Afterwards, we will prove that this stratification has the desired properties.) The stratification is obtained by constructing the skeletons  $\mathring{X}^s$  one after another, starting with  $\mathring{X}^{\dim X}$ . More precisely, suppose that  $\mathring{X}^{s+1}, \dots, \mathring{X}^{\dim X}$  have already been constructed. We obtain  $\mathring{X}^s$  by starting with  $\mathring{X}^s := X \setminus \bigcup_{i>s} \mathring{X}^i$  and by removing closed subsets of dimension less than  $s$  in four steps.

**Step R1:** We start by partitioning  $\mathring{X}^s$  into bradycells (using Proposition 2.1.5) and remove all bradycells of dimension less than  $s$ . Moreover, for each bradycell  $S \subseteq X^s$  of dimension  $s$ , we remove its frontier  $\partial S$  from  $\mathring{X}^s$ . This ensures that afterwards, each definably connected component of  $\mathring{X}^s$  is a bradycell. (Recall that “definably connected” refers to the language  $\mathcal{L}$ .) Even though  $\mathring{X}^s$  is not yet final, let us already call those connected components strata.

By removing an additional closed subset of lower dimension from  $\mathring{X}^s$ , we ensure that the “border condition” holds, i.e., that for any strata  $S \subseteq \mathring{X}^s$ ,  $S' \subseteq \mathring{X}^{s'}$ , where  $s' > s$ , we have either  $S \subseteq \text{cl}(S')$  or  $S \cap \text{cl}(S') = \emptyset$ . (In the end, this will imply that  $\text{cl}(S')$  is a union of strata.)

Note that none of the properties achieved in this step can be destroyed by removing further closed, lower-dimensional subsets from  $\mathring{X}^s$ .

**Step R2:** Next, we choose a stratum  $S \subseteq \mathring{X}^s$  (i.e., a bradycell of dimension  $s$ ) and an aligner  $\kappa \in \mathcal{C}_n$  of  $S$  (see Definition 2.1.1). For each of these (finitely many) choices, we remove an  $\mathcal{L}$ -definable subset from  $\mathring{X}^s$  as follows.

As explained in Notation 3.1.1, we assume that  $S$  itself is a aligned. This assumption does not cause definability issues of the sets we remove, since  $\kappa$  is (by definition of  $\mathcal{C}_n$ )  $\mathcal{L}$ -definable. Set  $\bar{S} := \text{pr}_{\leq s}(S)$  and denote by  $\rho: \bar{S} \rightarrow \mathcal{R}^{n-s}$  the function whose graph is  $S$ . Moreover, set  $e_1 := s$ . By Corollary 2.3.10, there is a subset  $Z \subseteq \bar{S}$  of lower dimension such that  $\rho$  is  $e_{[1,1]}(c_2)$ -sedated on  $\bar{S} \setminus Z$ . The preimage  $S \cap \text{pr}_{\leq s}^{-1}(Z)$  is a subset of  $S$  of dimension less than  $s$ ; we remove its closure  $\text{cl}(S \cap \text{pr}_{\leq s}^{-1}(Z))$  from  $\mathring{X}^s$ .

**Step R3:** The next shrinking of  $\mathring{X}^s$  is similar, but instead of considering a single stratum in  $\mathring{X}^s$ , we consider a whole sequence  $\mathcal{S} = (S^\ell)_{0 \leq \ell \leq m}$ , with  $S^\ell \subseteq \mathring{X}^{e_\ell}$  for some

$e_0 \geq e_1 > e_2 > \cdots > e_m = s$ ,  $m \geq 0$ . (In fact, Step R2 is a special case of Step R3, but for R3 to work, we will need that this special case has been carried out before.) Similarly to Step R2, for any such sequence  $\mathcal{S}$  and any aligner  $\kappa \in \mathcal{C}_n$  of  $\mathcal{S}$ , we will obtain a subset  $Z \subseteq \bar{S}^m$  of dimension less than  $s$  (where we use Notation 3.1.1), and for each  $\mathcal{S}$  and  $\kappa$  as above, we remove the corresponding set  $\text{cl}(S^m \cap \text{pr}_{\leq s}^{-1}(Z))$  from  $\hat{X}^s$ .

The goal of Step R3 is to ensure that certain functions on the set  $Y^b \subseteq \mathcal{R}^{e_0}$  from Notation 3.2.3 are  $e_{[j,m]}(\text{v})$ -sedated. This will be achieved using Proposition 2.3.6 and Corollary 2.3.10, so we need to ensure that the functions are already  $e_{[j,m']}(\text{v})$ -sedated for  $m' < m$ . We use Notation 3.2.3 and set

$$\delta^b := \rho^{0b} - \rho_{\star}^{1b} \circ \text{pr}_{\leq e_1} : Y^b \longrightarrow \mathcal{R}^{n-e_0}. \quad (3.1)$$

(If  $e_0 = e_1$ , then  $\delta^b = \rho^{0b} - \rho^{1b}$ .) The precise goal of Step R3 is to ensure the following:

$$\begin{aligned} \text{If } m = 0 \text{ or } e_0 > e_1: \quad & \rho^{0b} \text{ is } e_{[0,m']}(\text{c}_2)\text{-sedated on } Y^b \text{ for } 0 \leq m' \leq m; \\ \text{If } m \geq 1 \text{ and } e_0 > e_1: \quad & \delta^b \text{ is } e_{[0,m']}(\text{b})\text{-sedated on } Y^b \text{ for } 1 \leq m' \leq m; \\ \text{If } m \geq 1 \text{ and } e_0 = e_1: \quad & \delta^b \text{ is } e_{[1,m']}(\text{a})\text{-sedated on } Y^b \text{ for } 1 \leq m' \leq m. \end{aligned} \quad (3.2)$$

(Note that in Subsection 2.3, the numbering starts with  $e_1$ , whereas for  $(\text{c}_2)$ - and  $(\text{b})$ -sedation, we now start with  $e_0$ .)

To obtain (3.2) for  $m' < m$ , nothing needs to be removed from  $S^m$ ; instead, we deduce this inductively from the corresponding result obtained in the construction of  $\hat{X}^{e_{m-1}}$  (using Lemma 2.3.11 and Step R2); then we can  $e_{[j,m]}(\text{v})$ -sedate the functions using Proposition 2.3.6 and Corollary 2.3.10. This is straightforward; here are the details.

*Proof of (3.2) for  $m' < m$ .* Fix  $m' < m$ . For any statement related to  $\delta^b$ , we shall implicitly assume  $m' \geq 1$ . We keep Notation 3.2.3 with respect to  $\mathcal{S}$ , but we now additionally consider the shortened sequence  $\hat{\mathcal{S}} = (S^\ell)_{0 \leq \ell \leq m-1}$  and put a hat on various objects relative to  $\hat{\mathcal{S}}$  introduced in Notations 3.2.1 and 3.2.3 and in (3.1):  $\hat{\phi}_\ell, \hat{Y}, \hat{Y}^b, \hat{\rho}^{\ell b}, \hat{\delta}^b$ . Note that we have  $Y \subseteq \hat{Y}$  and  $\phi_\ell = \psi_\ell \circ \hat{\phi}_\ell$  (for  $0 \leq \ell < m$ ), where  $\psi_\ell = \phi_{m-1} \times \text{id}_{\mathcal{R}^{e_\ell - e_{m-1}}}$  is the map that rectilinearizes only with respect to  $\rho_{m-1}^m$ . In particular,

$$\hat{\rho}^{0b} = \rho^{0b} \circ \psi_0, \quad \hat{\rho}^{1b} = \rho^{1b} \circ \psi_1, \quad \text{and} \quad \hat{\delta}^b = \delta^b \circ \psi_0.$$

By Step R3 for  $\hat{\mathcal{S}}$  (which has already been carried out when constructing  $\hat{X}^{e_{m-1}}$ ),  $\hat{\rho}^{0b}$  is  $e_{[0,m']}(\text{c}_2)$ -sedated if  $e_0 > e_1$  and  $\hat{\delta}^b$  is  $e_{[1,m']}(\text{a})$ -sedated or  $e_{[0,m']}(\text{b})$ -sedated (depending on whether  $e_0 > e_1$ ). The map  $\psi_0$  is of the form required by Lemma 2.3.11, since  $\rho_{m-1}^m$  is  $e_{[m,m]}(\text{c}_2)$ -sedated by Step R2, so that lemma implies (3.2) for  $m' < m$ .  $\square$

*Obtaining (3.2) for  $m' = m$ .* Suppose first that  $m = 0$  or  $e_0 > e_1$ . Using  $\text{val}(\text{Jac}_x \rho^0) \geq 0$  (for  $x \in Y$ ) and  $\text{val}(\text{Jac}_x \phi_0) = 0$  (by Remark 3.2.4), we obtain  $\text{val}(\text{Jac}_{x^b} \rho^{0b}) \geq 0$ , so

we can apply Corollary 2.3.10 to  $\rho^{0b}$  using  $e_{[0,m]}$ . This yields a subset  $Z \subseteq \text{pr}_{\leq e_m}(Y^b) = \text{pr}_{\leq e_m}(Y) \subseteq \bar{S}^m$  of dimension less than  $e_m = s$  such that  $\rho^{0b}$  is  $e_{[0,m]}(c_2)$ -sedated on  $Y^b \setminus Z'$ , where  $Z'$  is the preimage of  $Z$  in  $\mathcal{R}^{e_0}$  under the projection. We shrink  $Y^b$  to  $Y^b \setminus Z'$  by removing  $\text{cl}(S^m \cap \text{pr}_{\leq e_m}^{-1}(Z))$  from  $S^m$ .

In a similar way (but using Proposition 2.3.6 (b)), we ensure that  $\delta^b$  is  $e_{[0,m]}(b)$ -sedated if  $m \geq 1$ . For this, we have to check that  $\text{val}(\text{Jac}_{x^b} \delta^b) \geq 0$ ; this follows from the corresponding statements for  $\rho^{0b}$  and  $\rho^{1b}$ .

Finally, if  $m \geq 1$  and  $e_0 = e_1$ , then without checking any additional condition, we can apply Proposition 2.3.6 (a) to shrink  $S^m$  in such a way that  $\delta^b$  becomes  $e_{[1,m]}(a)$ -sedated.  $\square$

**Step R4:** We keep the notation from Step R3 and remove one more set from  $S^m$  (again, for each choice of  $\mathcal{S}$  and  $\kappa$ ), namely  $S^m \cap \text{pr}_{\leq e_m}^{-1}(\partial(\text{pr}_{\leq e_m}(Y)))$ . This ensures that if we choose a sequence  $(S^\ell)_\ell$  of strata after this step has been carried out and write  $Y$  for the set corresponding to this new sequence, then  $\partial(\text{pr}_{\leq e_m}(Y)) \cap \bar{S}^m = \emptyset$  and hence, since  $S^m$  is connected, we have either  $\bar{S}^m \subseteq \text{pr}_{\leq e_m}(Y)$  or  $\bar{S}^m \cap \text{pr}_{\leq e_m}(Y) = \emptyset$ . (Later, only sequences for which the first of these cases occurs will be relevant.)

This finishes the construction of  $\hat{X}^s$  and hence of the stratification of  $X$ . We will now prove that this stratification is indeed a valuative Lipschitz stratification.

### 3.4 Relating the stratification to val-chains

We fix a val-chain  $a^0, \dots, a^m$  with  $a^\ell \in S^\ell \subseteq \hat{X}^{e_\ell}$ , dimensions  $e_0 \geq e_1 > \dots > e_m$ , and distances  $\lambda_1 > \dots > \lambda_{m+1}$ . We use Notations 3.1.1, 3.1.2, 3.2.1 and 3.2.3. The main goal of this subsection is to prove Lemma 3.4.4, which can be considered as a bound on some kind of distance between the tangent spaces  $T_{a^0}(\hat{X}^{e_0})$  and  $T_{a^1}(\hat{X}^{e_1})$ . The three different properties obtained in (3.2) will roughly correspond to the following three different kinds of val-chains (in this order): augmented val-chains with  $S^0 = S^1$ , plain val-chains, and augmented val-chains with  $S^0 \neq S^1$ .

**Notation 3.4.1.** We set  $a := a^0$  and  $\bar{a} := \text{pr}_{\leq e_0}(a)$ . By Lemma 3.1.3 (1), we have  $\text{pr}_{\leq e_\ell}(a) \in \bar{S}^\ell$  for  $0 \leq \ell \leq m$ , so  $\bar{a} \in Y$  and we can define  $\bar{a}^b := \phi_0(\bar{a}) \in Y^b$ .

*Remark 3.4.2.* Since  $\text{pr}_{\leq e_m}(a) \in \text{pr}_{\leq e_m}(Y) \cap \bar{S}^m$ , this intersection is non-empty, so Step R4 implies  $\bar{S}^m \subseteq \text{pr}_{\leq e_m}(Y)$  and hence  $\bar{S}^m = \text{pr}_{\leq e_m}(Y)$ .

We apply Notation 2.3.2 to  $\bar{a}^b$ , relative to the set  $Y^b$ , starting with  $e_0$  instead of  $e_1$ , and we allow ourselves to use that notation even if  $e_0 = e_1$ :

$$\zeta_\ell(\bar{a}^b) = \text{dist}(\text{pr}_{\leq e_\ell}(\bar{a}^b), \mathcal{R}^{e_\ell} \setminus \text{pr}_{\leq e_\ell}(Y^b)) \quad \text{for } 0 \leq \ell \leq m \quad (3.3)$$

$$\sigma_\ell(\bar{a}^b) = \max\{1, \|\text{pr}_{> e_\ell}(\bar{a}^b)\| \cdot \zeta_{\ell-1}(\bar{a}^b)^{-1}\} \quad \text{for } 1 \leq \ell \leq m. \quad (3.4)$$

(Concerning the case  $e_0 = e_1$ , we consider the norm of the empty tuple as being 0 and its valuation as being  $\infty$ .)

**Lemma 3.4.3.** *We have*

$$\text{val}(\text{pr}_{>e_{\ell+1}}(\bar{a}^b)) \stackrel{(1)}{\geq} \lambda_{\ell+1} \stackrel{(2)}{\geq} \text{val}(\zeta_\ell(\bar{a}^b)) \quad (3.5)$$

for  $0 \leq \ell \leq m-1$  at (1) and  $0 \leq \ell \leq m$  at (2). In particular,

$$\text{val}(\sigma_\ell(\bar{a}^b)) = 0 \quad \text{for } 1 \leq \ell \leq m. \quad (3.6)$$

*Proof.* The “in particular” part follows directly from (3.5) and (3.4).

(1) We have  $\text{pr}_{>e_{\ell+1}}(\bar{a}^b) = (a_\ell^b, \dots, a_0^b)$ , so it suffices to check that  $\text{val}(a_j^b) \geq \lambda_{j+1}$  ( $\geq \lambda_{\ell+1}$ ) for  $0 \leq j \leq \ell$ . This follows from Lemma 3.1.3 (3); indeed,  $a_j^b = a_j - \rho_j^{\bar{j}+1}(\text{pr}_{\leq e_{j+1}}(a))$  is just one of the coordinates of  $a - a^{[j+1]}$ , where the notation  $a^{[j+1]}$  is the one from the Lemma 3.1.3.

(2) It is enough to check that we have an inclusion

$$B_{>\lambda_{\ell+1}}(\text{pr}_{\leq e_\ell}(\bar{a}^b)) \subseteq \text{pr}_{\leq e_\ell}(Y^b) = \phi_\ell(\text{pr}_{\leq e_\ell}(Y)). \quad (3.7)$$

By Lemma 3.2.5, we have  $B_{>\lambda_{\ell+1}}(\text{pr}_{\leq e_\ell}(\bar{a}^b)) = \phi_\ell(\bar{B}^\ell)$  (where  $\bar{B}^\ell$  was defined as  $B_{>\lambda_{\ell+1}}(\text{pr}_{\leq e_\ell}(\bar{a}))$ ); see Notation 3.1.2), so (3.7) is equivalent to

$$\bar{B}^\ell \subseteq \text{pr}_{\leq e_\ell}(Y). \quad (3.8)$$

The definition of  $Y$  yields

$$\text{pr}_{\leq e_\ell}(Y) = \text{pr}_{\leq e_\ell}(Y') \cap Y_{\ell+1} \cap \dots \cap Y_m, \quad (3.9)$$

where

$$Y' = \{x \in \mathcal{R}^{e_0} : \text{pr}_{\leq e_j}(x) \in \bar{S}^j \text{ for } 0 \leq j \leq \ell\}. \quad (3.10)$$

and where  $Y_j$  is the preimage of  $\bar{S}^j$  under the projection  $\mathcal{R}^{e_\ell} \rightarrow \mathcal{R}^{e_j}$  (for  $\ell+1 \leq j \leq m$ ). By Lemma 3.1.3 (1), for  $j \geq \ell+1$  we have  $\text{pr}_{\leq e_j}(\bar{B}^\ell) \subseteq \bar{B}^j \subseteq \bar{S}^j$  and hence  $\bar{B}^\ell \subseteq Y_j$ . By Remark 3.4.2 applied to the val-chain  $a^0, \dots, a^\ell$ , we have  $\bar{S}^\ell \subseteq \text{pr}_{\leq e_\ell}(Y')$ . Together with  $\bar{B}^\ell \subseteq \bar{S}^\ell$  this implies (3.8).  $\square$

Suppose now that  $m \geq 1$ . We keep Notation 3.4.1 and additionally set  $b := a^1$ ,  $\bar{b} := \text{pr}_{\leq e_1}(b)$  and  $\bar{b}^b := \phi_1(\bar{b})$ . (This is well-defined by the same argument as for  $\bar{a}^b$ , applied to the val-chain  $a^1, \dots, a^m$ ). Recall that in Notation 3.2.3, we introduced the rectilinearized maps  $\rho^{\ell b} := \rho^\ell \circ \phi_\ell^{-1}$ . The following is a key intermediate result.

**Lemma 3.4.4** (Bounding the difference of derivatives). *Suppose that  $m \geq 1$ . Then for  $1 \leq i \leq e_m$ , we have*

$$\text{val}(\partial_i \rho^{0b}(\bar{a}^b) - \partial_i \rho_{\bar{x}}^{1b}(\bar{b}^b)) \geq \lambda_1 - \lambda_{m+1}.$$

*Proof.* Set  $\bar{c} := \text{pr}_{\leq e_1}(a)$ ,  $c := (\bar{c}, \rho^1(\bar{c}))$  and  $\bar{c}^b := \phi_1(\bar{c})$ . Note that  $c = a^{[1]}$  in the notation of Lemma 3.1.3, so  $c, a^2, \dots, a^m$  is a val-chain and hence well-definedness of  $\bar{c}^b$  follows as for  $\bar{a}^b$  and  $\bar{b}^b$ .

To prove the lemma, we “use  $c$  as an intermediate step”, i.e., it suffices to prove

$$\text{val}(\partial_i \rho^{0b}(\bar{a}^b) - \partial_i \rho_{\star}^{1b}(\bar{c}^b)) \geq \lambda_1 - \lambda_{m+1} \quad \text{and} \quad (3.11)$$

$$\text{val}(\partial_i \rho_{\star}^{1b}(\bar{c}^b) - \partial_i \rho_{\star}^{1b}(\bar{b}^b)) \geq \lambda_1 - \lambda_{m+1}. \quad (3.12)$$

Since  $a, c, a^2, \dots, a^m$  and  $c, b, a^2, \dots, a^m$  are val-chains (by Lemma 3.1.3), these two inequalities follow from two special cases of the lemma itself: (3.11) is just the special case  $b = a^{[1]}$ , and (3.12) follows from the special case where  $S^0 = S^1$ . (The special case yields (3.12) with  $\rho_{\star}^{1b}$  replaced by  $\rho^{1b}$ .) Thus we will now prove the lemma in these two cases.

Case  $b = a^{[1]}$ : In this case,  $\text{pr}_{\leq e_1}(\bar{a}) = \bar{b}$  and hence also  $\text{pr}_{\leq e_1}(\bar{a}^b) = \bar{b}^b$ . Recall the definition of  $\delta^b$  from Step R3; we have

$$\delta^b(x^b) = \rho^{0b}(x^b) - \rho_{\star}^{1b}(\text{pr}_{\leq e_1}(x^b)) \quad \text{for } x^b \in Y^b$$

and hence

$$\partial_i \rho^{0b}(\bar{a}^b) - \partial_i \rho_{\star}^{1b}(\bar{b}^b) = \partial_i \delta^b(\bar{a}^b).$$

We now distinguish two sub-cases. If  $e_0 > e_1$ , then since  $\delta^b$  is  $e_{[0,m]}(b)$ -sedated on  $Y^b$  (by (3.2)), we get (for  $1 \leq i \leq e_m$ )

$$\begin{aligned} \text{val}(\partial_i \delta^b(\bar{a}^b)) &\stackrel{(2.7)}{\geq} \min\{\text{val}(\delta^b(\bar{a}^b)), \text{val}(\text{pr}_{> e_1}(\bar{a}^b))\} - \text{val}(\zeta_m(\bar{a}^b)) + \sum_{\ell=1}^m \text{val}(\sigma_{\ell}(\bar{a}^b)) \\ &\stackrel{(3.5),(3.6)}{\geq} \min\{\text{val}(\delta^b(\bar{a}^b)), \lambda_1\} - \lambda_{m+1}. \end{aligned}$$

If, on the other hand,  $e_0 = e_1$ , then  $\delta^b$  is  $e_{[1,m]}(a)$ -sedated on  $Y^b$  and we get

$$\begin{aligned} \text{val}(\partial_i \delta^b(\bar{a}^b)) &\stackrel{(2.7)}{\geq} \text{val}(\delta^b(\bar{a}^b)) - \text{val}(\zeta_m(\bar{a}^b)) + \sum_{\ell=2}^m \text{val}(\sigma_{\ell}(\bar{a}^b)) \\ &\stackrel{(3.5),(3.6)}{\geq} \text{val}(\delta^b(\bar{a}^b)) - \lambda_{m+1}. \end{aligned}$$

In both cases,  $\delta^b(\bar{a}^b) = \rho^0(\bar{a}) - \rho_{\star}^1(\bar{b}) = (a - b)_{\star}$ , so the valuation of this is at least  $\lambda_1$  (since  $a = a^0$  and  $b = a^1$ ) and we get  $\text{val}(\partial_i \delta^b(\bar{a}^b)) \geq \lambda_1 - \lambda_{m+1}$ , as desired.

Case  $S_0 = S_1$ : In that case, we have  $\rho_{\star}^{1b} = \rho^{0b}$ , so the claim of the lemma is

$$\text{val}(\partial_i \rho^{0b}(\bar{a}^b) - \partial_i \rho^{0b}(\bar{b}^b)) \geq \lambda_1 - \lambda_{m+1}; \quad (3.13)$$

we will prove this using the Mean Value Theorem argument from Remark 2.1.13.

Set  $B := B_{\geq \lambda_1}(\bar{a})$ . By Lemma 3.2.5,  $B^b := \phi_0(B)$  is also a ball (note that  $\phi_0 = \phi_1$  and that  $\bar{B}^0 \subseteq B \subseteq \bar{B}^1$ ) and, since  $B$  contains  $\bar{a}$  and  $\bar{b}$  and  $\phi_0$  is a valuative isometry on  $B$ , we have  $\text{val}(\bar{a}^b - \bar{b}^b) = \lambda_1$ . Thus for Remark 2.1.13 to yield (3.13), it remains to verify that on the entire ball  $B^b$ , we have

$$\text{val}(\text{Jac } \partial_i \rho^{0b}) \geq -\lambda_{m+1}.$$

Given any  $\bar{c}^b \in B^b$ , let  $\bar{c}$  be its preimage in  $B$  and  $c := (\bar{c}, \rho^0(\bar{c})) \in S^0$ . Applying (3.2) to the strata  $S^1, \dots, S^m$  yields that  $\rho^{0b} = \rho^{1b}$  is  $e_{[1,m]}(c_2)$ -sedated on  $Y^b$ . (Note that the set  $Y^b$  corresponding to  $S^1, \dots, S^m$  is the same as the one corresponding to  $S^0, \dots, S^m$ .) Together with Lemma 3.4.3, this yields

$$\begin{aligned} \text{val}(\text{Jac } \partial_i \rho^{0b}(\bar{c}^b)) &\stackrel{(2.25)}{\geq} -\text{val}(\zeta_m(\bar{c}^b)) + \sum_{\ell=2}^m \text{val}(\sigma_\ell(\bar{c}^b)) \\ &\stackrel{(3.5),(3.6)}{\geq} -\lambda_{m+1}. \end{aligned}$$

which is what we had to prove.  $\square$

### 3.5 Proving that we have a valuative Lipschitz stratification

We will use the characterization of valuative Lipschitz Stratifications given by Proposition 1.8.3. Thus suppose that  $a^0, \dots, a^m$  is a val-chain with  $a^\ell \in S^\ell \subseteq \overset{\circ}{X}^{e_\ell}$ , with dimensions  $e_0 \geq e_1 > \dots > e_m$ , and with distances  $\lambda_1 > \dots > \lambda_{m+1}$ . We need to find vector spaces

$$V_{k,m} \subseteq V_{k,m-1} \subseteq \dots \subseteq V_{k,k+1} \subseteq V_{k,k} = \mathbf{T}_{a^k} S^k \quad \text{for } 0 \leq k \leq m \quad (3.14)$$

with  $\dim V_{k,\ell} = e_\ell$  satisfying

$$\Delta(V_{k,\ell}, V_{k+1,\ell}) \geq \lambda_{k+1} - \lambda_{\ell+1} \quad \text{for } 0 \leq k < \ell \leq m, \quad (3.15)$$

(where  $\Delta(W_1, W_2)$  is the valuative metric on the Grassmannian; see Definition 1.8.1). The strategy is as follows. Given any val-chain as above and any aligner  $\kappa \in \mathcal{C}_n$  of  $(S^\ell)_{0 \leq \ell \leq m}$ , we will define an  $e_m$ -dimensional space denoted by  $V_{0,m}$  depending only on the val-chain and on  $\kappa$ . Let  $V_{k,\ell}$  be the space obtained by applying the same definition to the sub-val-chain  $a^k, a^{k+1}, \dots, a^\ell$  (and the same aligner  $\kappa$ ). Once the spaces are defined, we will prove:

$$V_{0,0} = \mathbf{T}_{a^0} S^0 \quad (\text{in the case } m = 0); \quad (3.16)$$

$$V_{0,m} \subseteq V_{0,m-1} \quad \text{if } m \geq 1; \quad (3.17)$$

$$\Delta(V_{0,m}, V_{1,m}) \geq \lambda_1 - \lambda_{m+1} \quad \text{if } m \geq 1. \quad (3.18)$$

By applying these results to various sub-val-chains of  $a^0, \dots, a^m$  one then obtains (3.14) and (3.15), i.e., we then are done with the proof of the theorem.

We start by defining  $V_{0,m}$ . As usual, we use Notation 3.1.1, 3.1.2, 3.2.1 and 3.2.3. In particular, we assume that the coordinate system has been transformed using  $\kappa$ . This is harmless, since such a transformation preserves the notion of val-chains on the one hand, and the properties we are about to prove on the other hand.

**Notation 3.5.1.** For  $0 \leq \ell \leq m$  and suitable  $x = (\bar{x}, x') \in \mathcal{R}^{e_\ell} \times \mathcal{R}^{n-e_\ell}$ , we define a variant of the rectilinearization maps, where “all coordinates of  $\mathcal{R}^{n-e_\ell}$  are rectilinearized along  $S^\ell$ ”:

$$\tilde{\phi}_\ell(x) := (\phi_\ell(\bar{x}), x' - \rho^\ell(\bar{x})). \quad (3.19)$$

(Note that if  $S^0 = S^1$ , then  $\tilde{\phi}_0 = \tilde{\phi}_1$ .) We moreover set

$$\begin{aligned} a &:= a^0, \\ W &:= \mathcal{R}^{e_m} \times \{0\}^{n-e_m} \quad \text{and} \\ V_{0,m} &:= (\text{Jac}_a \tilde{\phi}_0)^{-1}(W). \end{aligned}$$

That  $\tilde{\phi}_0$  is defined at  $a$  follows from Remark 3.4.2. As required, we have  $\dim V_{0,m} = e_m$ , so to finish the proof of the theorem, it remains to prove (3.16), (3.17) and (3.18).

*Proof of (3.16).* In the case  $m = 0$ ,  $\tilde{\phi}_0^{-1}$  sends  $\bar{S}^0 \times \{0\}^{n-e_0}$  to  $S^0$ , and we have  $W = \mathcal{R}^{e_0} \times \{0\}^{n-e_0}$ . Thus  $(\text{Jac}_a \tilde{\phi}_0)^{-1}(W)$  is the tangent space to  $S^0$  at  $a$ , as required.  $\square$

*Proof of (3.17).* Suppose that  $m \geq 1$ . We have

$$V_{0,m-1} = (\text{Jac}_a \hat{\phi}_0)^{-1}(\hat{W}) \quad (3.20)$$

where

$$\hat{W} = \mathcal{R}^{e_{m-1}} \times \{0\}^{n-e_{m-1}} \quad \text{and} \quad (3.21)$$

$$\hat{\phi}_0: (x_m, x_{m-1}, x_{m-2}, \dots, x_0, x_\star) \mapsto (x_m, x_{m-1}, x_{m-2}, \dots, x_0, x_\star - \rho^0(\text{pr}_{\leq e_0}(x))) \quad (3.22)$$

An easy computation shows that  $(\text{Jac}_a \hat{\phi}_0)^{-1}(\hat{W}) = (\text{Jac}_a \tilde{\phi}_0)^{-1}(W)$ ; indeed, we have  $\tilde{\phi}_0 = \psi \circ \hat{\phi}_0$ , where  $\psi = \phi_{m-1} \times \text{id}_{\mathcal{R}^{n-e_{m-1}}}$ , and  $(\text{Jac}_x \psi)^{-1}(\hat{W}) = \hat{W}$  for any  $x$ .

Together with  $\hat{W} \supseteq W$ , this implies  $V_{0,m-1} \supseteq V_{0,m}$ , as required.  $\square$

*Proof of (3.18).* We have  $V_{1,m} = (\text{Jac}_b \tilde{\phi}_1)^{-1}(W)$  where  $b := a^1$  (and  $\tilde{\phi}_1$  has been defined in (3.19)). To obtain  $\Delta(V_{0,m}, V_{1,m}) \geq \lambda_1 - \lambda_{m+1}$ , it suffices to prove that

$$\text{val} \left( (\text{Jac}_a \tilde{\phi}_0)^{-1} \upharpoonright W - (\text{Jac}_b \tilde{\phi}_1)^{-1} \upharpoonright W \right) \geq \lambda_1 - \lambda_{m+1} \quad (3.23)$$

(by Lemma 1.8.2).

From the definition of  $\tilde{\phi}_0$ , we get

$$\text{Jac}_a \tilde{\phi}_0 = \begin{pmatrix} \text{Jac}_{\bar{a}} \phi_0 & 0 \\ -\text{Jac}_{\bar{a}} \rho^0 & 1 \end{pmatrix} \quad \text{and hence} \quad (3.24)$$

$$(\text{Jac}_a \tilde{\phi}_0)^{-1} = \begin{pmatrix} (\text{Jac}_{\bar{a}} \phi_0)^{-1} & 0 \\ (\text{Jac}_{\bar{a}} \rho^0) \circ (\text{Jac}_{\bar{a}} \phi_0)^{-1} & 1 \end{pmatrix} = \begin{pmatrix} (\text{Jac}_{\bar{a}} \phi_0)^{-1} & 0 \\ \text{Jac}_{\bar{a}^b}(\rho^0 \circ \phi_0^{-1}) & 1 \end{pmatrix}, \quad (3.25)$$

where  $\bar{a}^b = \phi_0(\bar{a})$ . If we moreover set  $\bar{a} = \text{pr}_{\leq e_1}(\bar{a})$  and  $\bar{a}^b = \phi_1(\bar{a})$ , then we have  $\phi_0(\bar{a}) = (\phi_1(\bar{a}), \bar{a}_0 - \rho_0^1(\bar{a}))$ , and exactly the same computation as in (3.24) and (3.25) yields

$$(\text{Jac}_{\bar{a}} \phi_0)^{-1} = \begin{pmatrix} (\text{Jac}_{\bar{a}} \phi_1)^{-1} & 0 \\ \text{Jac}_{\bar{a}^b}(\rho_0^1 \circ \phi_1^{-1}) & 1 \end{pmatrix}. \quad (3.26)$$

Combining (3.25) with (3.26) yields

$$(\text{Jac}_a \tilde{\phi}_0)^{-1} = \begin{pmatrix} (\text{Jac}_{\bar{a}} \phi_1)^{-1} & 0 & 0 \\ \text{Jac}_{\bar{a}^b}(\rho_0^1 \circ \phi_1^{-1}) & 1 & 0 \\ \text{Jac}_{\bar{a}^b}(\rho^0 \circ \phi_0^{-1}) & & 1 \end{pmatrix} = \begin{pmatrix} (\text{Jac}_{\bar{a}} \phi_1)^{-1} & 0 & 0 \\ \text{Jac}_{\bar{a}^b} \rho_0^{1b} & 1 & 0 \\ \text{Jac}_{\bar{a}^b} \rho^{0b} & & 1 \end{pmatrix}, \quad (3.27)$$

where the coordinates are grouped according to  $\mathcal{R}^{e_1} \times \mathcal{R}^{e_0 - e_1} \times \mathcal{R}^{n - e_0}$ .

We also do the computation from (3.24) and (3.25) for  $\tilde{\phi}_1(b) = (\phi_1(\bar{b}), (b_0, b_x) - \rho^1(\bar{b}))$ , where  $\bar{b} = \text{pr}_{\leq e_1}(b)$  and  $\bar{b}^b = \phi_1(\bar{b})$ , and obtain (with the same grouping of coordinates as before)

$$(\text{Jac}_b \tilde{\phi}_1)^{-1} = \begin{pmatrix} (\text{Jac}_{\bar{b}} \phi_1)^{-1} & 0 & 0 \\ \text{Jac}_{\bar{b}^b}(\rho_0^1 \circ \phi_1^{-1}) & 1 & 0 \\ \text{Jac}_{\bar{b}^b}(\rho_x^1 \circ \phi_1^{-1}) & 0 & 1 \end{pmatrix} = \begin{pmatrix} (\text{Jac}_{\bar{b}} \phi_1)^{-1} & 0 & 0 \\ \text{Jac}_{\bar{b}^b} \rho_0^{1b} & 1 & 0 \\ \text{Jac}_{\bar{b}^b} \rho_x^{1b} & 0 & 1 \end{pmatrix}. \quad (3.28)$$

To prove (3.23), we have to prove the corresponding statements for the three sub-matrices where (3.27) and (3.28) differ.

For the lower most, this is exactly the statement of Lemma 3.4.4. For the middle sub-matrix, the result is obtained by applying Lemma 3.4.4 to the augmented val-chain  $a^{[1]}, a^1, a^2, \dots, a^m$ , where  $a^{[1]} = (\bar{a}, \rho^1(\bar{a})) \in S^1$  and  $\text{val}(a^{[1]} - a^1) \geq \lambda_1$  by Lemma 3.1.3 (3).

Finally, for the upper-most sub-matrix, we use an inductive argument. If  $m = 1$ , then  $\phi_1$  is the identity, so suppose now  $m \geq 2$ . Set  $\bar{\bar{a}} := \text{pr}_{\leq e_2}(a)$ ,  $a^{[2]} := (\bar{\bar{a}}, \rho^2(\bar{\bar{a}})) \in S^2$  and similarly  $\bar{\bar{b}} := \text{pr}_{\leq e_2}(b)$ ,  $b^{[2]} := (\bar{\bar{b}}, \rho^2(\bar{\bar{b}})) \in S^2$ . Using Lemma 3.1.3, we obtain that  $a^{[2]}, b^{[2]}, a^3, \dots, a^m$  is a val-chain with  $\text{val}(a^{[2]} - b^{[2]}) \geq \lambda_1$ . By induction, we may assume that (3.23) holds for this shorter val-chain, i.e.,

$$\text{val} \left( (\text{Jac}_{a^{[2]}} \tilde{\phi}_2)^{-1} \upharpoonright W - (\text{Jac}_{b^{[2]}} \tilde{\phi}_2)^{-1} \upharpoonright W \right) \geq \lambda_1 - \lambda_{m+1}. \quad (3.29)$$

This implies the desired inequality

$$\text{val} \left( (\text{Jac}_{\bar{a}} \phi_1)^{-1} \upharpoonright W - (\text{Jac}_{\bar{b}} \phi_1)^{-1} \upharpoonright W \right) \geq \lambda_1 - \lambda_{m+1}, \quad (3.30)$$

using that  $\phi_1$  is obtained from  $\tilde{\phi}_2$  by omitting some coordinates and that the derivatives of these functions only depend on the first  $e_2$  coordinates. More precisely, a computation as in (3.25) and (3.26) (applied to  $\tilde{\phi}_2$  and  $\phi_1$ ) yields that  $(\text{Jac}_{\bar{x}} \phi_1)^{-1}$  is a sub-matrix of  $(\text{Jac}_x \tilde{\phi}_2)^{-1}$  (for suitable  $x \in \mathcal{R}^n$  and  $\bar{x} = \text{pr}_{\leq e_1}(x)$ ) and that this sub-matrix only depends on  $\text{pr}_{\leq e_2}(x)$ .  $\square$

This finishes the proof of Theorem 1.6.7, and hence also of Theorem 1.

## References

- [Dri1] Lou van den Dries:  $T$ -convexity and tame extensions II. *Journal of Symbolic Logic* **62** (1997), no. 1, 14–34.
- [Dri2] Lou van den Dries: *Tame topology and o-minimal structures*. LMS Lecture Note Series, vol. 248, Cambridge University Press, Cambridge, UK, 1998.
- [DL] Lou van den Dries and Adam H. Lewenberg:  $T$ -convexity and tame extensions. *Journal of Symbolic Logic* **60** (1995), no. 1, 74–102.
- [DMM] Lou van den Dries, Angus Macintyre, and David Marker: The elementary theory of restricted analytic fields with exponentiation. *Annals of Mathematics* **140** (1994), no. 1, 183–205.
- [Fis] Andreas Fischer: o-minimal  $\Lambda^m$ -regular stratification. *Ann. Pure Appl. Logic* **147** (2007), no. 1-2, 101–112.
- [Hal] Immanuel Halupczok: Non-archimedean Whitney stratifications. *Proc. Lond. Math. Soc.* **109** (2014), no. 5, 1304–1362.
- [KP] Krzysztof Kurdyka and Adam Parusiński: Quasi-convex decomposition in o-minimal structures. Application to the gradient conjecture. *Singularity theory and its applications*, vol. 43, Mathematical Society of Japan, 2006, Proceedings of 12th MSJ-IRI symposium, pp. 137–177.
- [Mil1] Chris Miller: A growth dichotomy for o-minimal expansions of ordered fields. In *Logic: from foundations to applications (Staffordshire, 1993)*. Oxford Univ. Press, New York (1996), 385–399.
- [Mil2] Chris Miller: Exponentiation is hard to avoid. *Proc. Amer. Math. Soc.* **122** (1994), no. 1, 257–259.
- [Mos] Tadeusz Mostowski: Lipschitz equisingularity. *Dissertationes Math. (Rozprawy Mat.)* **243** (1985), 46.

- [Ngu] Nhan Nguyen: Structure métrique et géométrie des ensembles définissables dans des structures  $o$ -minimales. (2015), doctoral thesis.
- [NV] Nhan Nguyen and Guillaume Valette: Lipschitz stratifications in  $o$ -minimal structures. (2014), preprint.
- [Par1] Adam Parusiński: Lipschitz properties of semi-analytic sets. *Annales de l'institut Fourier* **38** (1988), no. 4, 189–213.
- [Par2] Adam Parusiński: Lipschitz stratification of subanalytic sets. *Annales scientifiques de l'École Normale Supérieure* **27** (1994), no. 6, 661–696.
- [Paw] Wiesław Pawłucki: Lipschitz cell decomposition in  $o$ -minimal structures. I. *Ill. J. Math.* **52** (2009), no. 3, 1045–1063.
- [Rol] Jean-Philippe Rolin: Construction of  $o$ -minimal structures from quasianalytic classes. (English summary) *Lecture notes on  $O$ -minimal structures and real analytic geometry*, 71–109, Fields Inst. Commun., **62**, Springer, New York, 2012.
- [Val] Guillaume Valette: Lipschitz triangulations. *Illinois J. Math.*, **49**, (2005), no. 3, 953–979 (electronic).
- [Ver] Jean-Louis Verdier: Stratifications de Whitney et théorème de Bertini-Sard. *Invent. Math.* **36** (1976), 295–312.
- [Yin] Yimu Yin: On polynomial-bounded  $T$ -convex fields. (2015), preprint.