Generalised Recombination Interpolation Method (GRIM)

Terry Lyons and Andrew D. McLeod

January 27, 2023

Abstract

In this paper we develop the *Generalised Recombination Interpolation Method* (GRIM) for finding sparse approximations of functions initially given as linear combinations of some (large) number of simpler functions. GRIM is a hybrid of dynamic growth-based interpolation techniques and thinning-based reduction techniques. We establish that the number of non-zero coefficients in the approximation returned by GRIM is controlled by the concentration of the data. In the case that the functions involved are $\mathrm{Lip}(\gamma)$ for some $\gamma > 0$ in the sense of Stein, we obtain improved convergence properties for GRIM. In particular, we prove that the level of data concentration required to guarantee that GRIM finds a good sparse approximation is decreasing with respect to the regularity parameter $\gamma > 0$.

Contents

1. Introduction

This article considers finding sparse approximations of functions with the aim of reducing computational complexity. Applications of sparse representations are wide ranging and include compressed sensing [DE03, CT05, Don06, CRT06], image processing [EMS08, BMPSZ08, BMPSZ09], facial recognition [GMSWY09], data assimilation [MM13], explainability [FMMT14, DF15], sensor placement in nuclear reactors [ABGMM16, ABCGMM18], reinforcement learning [BS18, HS19], DNA denoising [KK20], and inference acceleration within machine learning [ABDHP21, NPS22].

Whilst the particular formulation of the sparse approximation problem is situation dependent, the overall goal remains the same. Loosely speaking, the commonly shared aim is to approximate a complex system using only a few elementary features. In this generality, the problem is commonly tackled via *Least Absolute Shrinkage and Selection Operator (LASSO)* regression, which involves solving a minimisation problem under l^1-norm constraints. The terminology "LASSO" originates in [Tib96], though imposing l^1-norm constraints is considered in the earlier works [CM73, SS86]. More recent LASSO-type techniques may be found in, for example, [LY07, DGOY08, GO09, XZ16, TW19].

In this paper we consider the setting that the system of interest is known to be a linear combination of some (large) number of features. Within this setting the goal is to identify a linear combination of a strict sub-collection of the features (i.e. not *all* the features) that gives, in some sense, a good approximation of the system (i.e. of the original given linear combination of all the features). Depending on the particular context considered, techniques for finding such sparse approximations include the pioneering *Empirical Interpolation Method* [BMNP04, GMNP07, MNPP09], its subsequent generalisation the *Generalised Empirical Interpolation Method* [MM13, MMT14], *Pruning* [Ree93, AK13, CHXZ20, GLSWZ21], *Kernel Herding* [Wel09a, Wel09b, CSW10, BLL15, BCGMO18, TT21, PTT22], *Convex Kernel Quadrature* [HLO21], and *Kernel Thinning* [DM21a, DM21b, DMS21]. Of course one can still utilise the LASSO approach based on l^1-regularisation within this framework.

It is convenient for our purposes to consider the following loose categorisation of techniques for finding sparse approximations in this setting; those that are *growth-based*, and those that are *thinning-based*. Growth-based methods seek to inductively increase the size of a sub-collection of features, until the sub-collection is rich enough to well-approximate the entire collection of features. Thinning-based methods seek to inductively identify features that may be discarded without significantly affecting how well the remaining features can approximate the original entire collection. Of the techniques mentioned above, EIM, GEIM and Kernel Herding are growth-based, whilst LASSO, Convex Kernel Quadrature and Kernel Thinning are thinning-based.

We develop the *Generalised Recombination Interpolation Method* (GRIM) for finding sparse approximations of a given linear combination of functions. GRIM is a hybrid of growth-based and thinning-based techniques. The growth-based aspect takes the form of dynamic greedy selection in a similar spirit to that employed in GEIM [MM13, MMT14]. An important distinction from GEIM is that our greedy selection is *data-driven* rather than *feature-driven*. The thinning-based aspect arises through the use of *recombination* [LL12, LL16], which is a central technique within the Convex Kernel Quadrature approach proposed in [HLO21].

To allow a more detailed discussion of GRIM we must first fix notation. Let X be a Banach space and $\mathcal{N} \in \mathbb{Z}_{>0}$ be a (large) positive integer. Suppose we are interested in a system corresponding to an element $\varphi \in X$. Further suppose that, for given non-zero real numbers $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ and a given collection $\mathcal{F} = \{f_1, \ldots, f_{\mathcal{N}}\} \subset X$ of non-zero elements in X, we have

$$\varphi = \sum_{i=1}^{\mathcal{N}} a_i f_i. \tag{1.1}$$

Let X^* denote the dual of X and suppose that $\Sigma \subset X^*$ is a compact subset. We will often adopt the terminology that the set $\mathcal{F}$ consists of *features* whilst the set Σ consists of *data*. In this setting we consider the following sparse approximation problem. Given $\varepsilon > 0$, find an element $u = \sum_{i=1}^{\mathcal{N}} b_i f_i \in \mathrm{Span}(\mathcal{F})$ such that the cardinality of the set $\{i \in \{1, \ldots, \mathcal{N}\} : b_i \neq 0\}$ is *less* than $\mathcal{N}$ and that u is close to φ throughout Σ in the sense that, for every $\sigma \in \Sigma$, we have $|\sigma(\varphi - u)| \leq \varepsilon$.

We assume that the data over which we want to approximate φ is a compact subset of the dual space X^*. Consider the choice $X := C^0(\Omega)$ for a compact domain $\Omega \subset \mathbb{R}^d$. The identification of a point $p \in \Omega$ with the point mass $\delta_p \in C^0(\Omega)^*$ supported at p enables us to take $\Sigma := \Omega$. Hence seeking a sparse approximation of a sum of continuous functions $\Omega \to \mathbb{R}$ is within this framework. If we let $\mathcal{M}[\Omega]$ denote the collection of finite signed measures on Ω, then we may also consider the choice that $X := \mathcal{M}[\Omega]$. In this case we have that $C^0(\Omega) \subset X^*$ and hence, by choosing Σ to be the set of monomials with order no larger than some $k \in \mathbb{Z}_{\geq 0}$, the *cubature problem* [Str71] for empirical measures is within this framework. Combined with sampling, the *cubature problem* for empirical measures offers an approach to the *cubature problem* for general finite signed measures.

Suppose that $\mathcal{X}$ is a set and $\mathcal{H}_k$ is a *Reproducing Kernel Hilbert Space* (RKHS) associated to a positive semi-definite symmetric kernel function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ (the appropriate definitions can be found, for example, in [BT11]). In this context one can consider the *Kernel Quadrature* problem, for which the Kernel Herding [Wel09a, Wel09b, CSW10, BLL15, BCGMO18, TT21, PTT22], Convex Kernel Quadrature [HLO21] and Kernel Thinning [DM21a, DM21b, DMS21] methods have been developed. Given a probability measure $\mu \in \mathbb{P}[\mathcal{X}]$ the *Kernel Quadrature* problem involves finding, for some $n \in \mathbb{Z}_{\geq 1}$, points $z_1, \ldots, z_n \in \mathcal{X}$ and weights $w_1, \ldots, w_n \in \mathbb{R}$ so that the measure $\mu_n := \sum_{j=1}^{n} w_j \delta_{z_j}$ approximates μ in the sense that, for every $f \in \mathcal{H}_k$, we have $\mu_n(f) \approx \mu(f)$. Linearity ensures that this approximate equality will be valid for all $f \in \mathcal{H}_k$ provided it is true for every f in the closed unit ball $\overline{\mathbb{B}}_{\mathcal{H}_k}(0,1) \subset \mathcal{H}_k$. A consequence of the inclusion $\mathcal{H}_k \subset C^0(\mathcal{X})$ is that the closed unit ball $\overline{\mathbb{B}}_{\mathcal{H}_k}(0,1) \subset \mathcal{H}_k$ can be viewed as a subset of $\mathcal{M}[\mathcal{X}]^*$. Provided this closed unit ball is compact, which will be the case if, for example, the set $\mathcal{X}$ is finite, the choice of $X := \mathcal{M}[\mathcal{X}]$ and $\Sigma := \overline{\mathbb{B}}_{\mathcal{H}_k}(0,1)$ illustrates that the *Kernel Quadrature* problem for empirical probability distributions $\mu \in \mathbb{P}[\mathcal{X}]$ is within the framework we consider. Combined with sampling, the *Kernel Quadrature* problem for empirical probability distributions offers an approach to the *Kernel Quadrature* problem for general probability distributions.

Returning our attention to the general setting, GRIM uses a greedy selection algorithm to inductively grow a collection $L \subset \Sigma$ of linear functionals at which we require our approximation to coincide with φ. The following extension step determines how we select new linear functionals in Σ to be added to an existing collection $L' \subset \Sigma$.

Banach Extension Step
Assume $L' \subset \Sigma$ and $u \in \operatorname{Span}(\mathcal{F})$ are both given. Let $m \in \mathbb{Z}_{\geq 1}$ and take $\sigma_1 := \operatorname{argmax}\{|\sigma(\varphi - u)| : \sigma \in \Sigma\}$. Inductively for $j = 2, 3, \ldots, m$ take $\sigma_j := \operatorname{argmax}\{|\sigma(\varphi - u)| : \sigma \in \Sigma \setminus \{\sigma_1, \ldots, \sigma_{j-1}\}\}$. Once $\sigma_1, \ldots, \sigma_m \in \Sigma$ have been defined, we extend L' to $L := L' \cup \{\sigma_1, \ldots, \sigma_m\}$.

After each extension of the subset $L \subset \Sigma$ of linear functionals, we use recombination [LL12, LL16] to find an element $u \in \operatorname{Span}(\mathcal{F})$ satisfying that $u \equiv \varphi$ throughout L. Originating in [LL12] as a technique for reducing the support of a probability measure whilst preserving a specified list of moments, at its core recombination is a method of reducing the number of non-zero components in a solution of a system of linear equations. We apply recombination to the linear system determined by the set $\{\sigma(\varphi) : \sigma \in L\}$ via the expansion (1.1) (see Lemma 2.1).

We propose the following **Banach GRIM** algorithm to find a sparse approximation of φ.

Banach GRIM

(A) Fix $\varepsilon > 0$ as the target accuracy threshold.

(B) Choose $k_1 \in \mathbb{Z}_{\geq 1}$ and apply the **Banach Extension Step**, with $L' = \varnothing$, $u \equiv 0$ and $m := k_1$, to obtain a subset $\Sigma_1 = \{\sigma_{1,1}, \ldots, \sigma_{1,k_1}\} \subset \Sigma$. Apply recombination (cf. Lemma 2.1) to φ to find an element $u_1 \in \operatorname{Span}(\mathcal{F})$ satisfying, for every $\sigma \in \Sigma_1$, that $\sigma(\varphi - u_1) = 0$.

(C) For $s \in \mathbb{Z}_{\geq 2}$ we proceed inductively. If $|\sigma(\varphi - u_{s-1})| \leq \varepsilon$ for every $\sigma \in \Sigma$ then we stop since u_{s-1} is an approximation of φ possessing the desired level of accuracy. If this is not the case, then choose $k_s \in \mathbb{Z}_{\geq 1}$ and apply the **Banach Extension Step**, with $L' = \Sigma_{s-1}$, $u := u_{s-1}$ and $m := k_s$, to obtain a subset $\Sigma_s = \Sigma_{s-1} \cup \{\sigma_{s,1}, \ldots, \sigma_{s,k_s}\} \subset \Sigma$. Apply recombination (cf. Lemma 2.1) to φ to find an element $u_s \in \operatorname{Span}(\mathcal{F})$ satisfying, for every $\sigma \in \Sigma_s$, that $\sigma(\varphi - u_s) = 0$.

Similarly to GEIM [MM13, MMT14], GRIM involves dynamically growing a subset $L \subset \Sigma$ specifying the functionals in Σ at which an approximation is required to agree with φ. In GEIM this growth is primarily *feature-driven*. GEIM additionally dynamically grows a subset $S \subset \mathcal{F}$ determining the elements from $\mathcal{F}$ to be used to construct the approximation. The growth is driven at each step by first determining the new element from $\mathcal{F}$ to be added to S, before subsequently using this new elements to determine the new functional from Σ to be added to L. That is the new feature to be used by the approximation is selected first, and the new information to be matched by the next approximation is determined using the new feature.

GRIM, however, uses *data-driven* growth. For each $m \in \mathbb{Z}_{\geq 2}$, GRIM first determines the new linear functionals in Σ to be added to Σ_{m-1} to form $\Sigma_m \subset \Sigma$ before using recombination to find an approximation coinciding with φ on Σ_m. That is, we first choose the new information that we want our approximation to match *before* using recombination to both select the elements from $\mathcal{F}$ and use them to construct our approximation.

Evidently we have the nesting property that $\Sigma_{t_1} \subset \Sigma_{t_2}$ for integers $t_1 \leq t_2$, ensuring that at each step we are increasing the amount of information that we require our approximation to match. For each integer $m \in \mathbb{Z}_{\geq 1}$ let $S_m \subset \mathcal{F}$ denote the sub-collection of elements from $\mathcal{F}$ used to form the approximation u_m. Recombination is applied to a system of $1 + k_1 + \ldots + k_m$ linear equations when finding u_m, hence we may conclude that $\#(S_m) \leq \min\{1 + k_1 + \ldots + k_m, \mathcal{N}\}$ (cf. Lemma 2.1). Besides this upper bound for $\#(S_m)$, we have *no* control on the sets S_m. We impose only that the linear functionals are greedily chosen; the selection of the elements from $\mathcal{F}$ to form the approximation u_m is left up to recombination and determined by the data. In contrast to GEIM, there is *no* requirement that elements from $\mathcal{F}$ used to form u_m must also be used for u_l for $l > m$.

For the case that, for each $s \in \mathbb{Z}_{\geq 1}$, we choose $k_s := 1$ we are able to establish the following worst-case upper bound on the number of steps the **Banach GRIM** algorithm requires to find an approximation $u \in \operatorname{Span}(\mathcal{F})$ satisfying that $|\sigma(\varphi - u)| \leq \varepsilon$ for every $\sigma \in \Sigma$.

Theorem 1.1 (**Banach GRIM Convergence**)**.** *Let X be a Banach space and X^* denote its dual space. Assume $\varepsilon > 0$, that $\mathcal{N} \in \mathbb{Z}_{\geq 1}$, that $\mathcal{F} = \{f_1, \ldots, f_{\mathcal{N}}\} \subset X$ is a collection of non-zero elements in X, and that $\Sigma \subset X^*$ is compact. Let $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ and define $\varphi \in \operatorname{Span}(\mathcal{F})$ and a constant $C > 0$ by*

$$\varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i \qquad \text{and} \qquad C := \sum_{i=1}^{\mathcal{N}} |a_i| \, ||f_i||_X > 0. \tag{1.2}$$

Let $N \in \mathbb{Z}_{\geq 0}$ denote the $\varepsilon/2C$-packing number for Σ. That is,

$$N := \max\left\{ d \in \mathbb{Z} \;:\; \exists \sigma_1, \ldots, \sigma_d \in \Sigma \text{ for which } ||\sigma_a - \sigma_c||_{X^*} > \frac{\varepsilon}{2C} \text{ if } a \neq c \right\}. \tag{1.3}$$

*Then if we apply the **Banach GRIM** algorithm to approximate φ, with the target accuracy threshold in **Banach GRIM** (A) as ε and the choice that $k_s = 1$ for every $s \in \mathbb{Z}_{\geq 1}$, then after at most N steps the algorithm terminates. That is, if we let $M \in \{1, \ldots, N\}$ be the integer for which the algorithm terminates after step M and $Q_M := \min\{\mathcal{N}, M+1\}$, then there are coefficients $c_1, \ldots, c_{Q_M} \in \mathbb{R}$ and indices $e_1, \ldots, e_{Q_M} \in \{1, \ldots, \mathcal{N}\}$ for which $\sum_{s=1}^{Q_M} |c_s| \, ||f_{e_s}||_X = C$ and the element $u \in \operatorname{Span}(\mathcal{F})$ defined by*

$$u := \sum_{s=1}^{Q_M} c_s f_{e_s} \qquad \text{satisfies, for every } \sigma \in \Sigma, \text{ that} \qquad |\sigma(\varphi - u)| \leq \varepsilon. \tag{1.4}$$

Moreover, if the coefficients $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \backslash \{0\}$ corresponding to φ (cf. (1.2)) are all positive (i.e. $a_1, \ldots, a_{\mathcal{N}} > 0$) then the coefficients $c_1, \ldots, c_{Q_M} \in \mathbb{R}$ corresponding to u (cf. (1.4)) are all non-negative (i.e. $c_1, \ldots, c_{Q_M} \geq 0$).

This theorem will be a special case of the more elaborate Theorem 2.2 which will additionally establish a robustness estimate for the approximation u and account for numerical errors arising in the use of recombination.

The number of steps required for the **Banach GRIM** algorithm to converge is no greater than the $\varepsilon/2C$-packing number of Σ defined in (1.3). The packing number of a subset of a Banach space is closely related to the covering number of the subset. Covering and packing numbers, first studied by Kolmogorov [Kol56], arise in a variety of contexts including eigenvalue estimation [Car81, CS90, ET96], Gaussian Processes [LL99, LP04], and machine learning [EPP00, SSW01, Zho02, Ste03, SS07, Kuh11, MRT12, FS21].

If we let Z be a Banach space and $\mathcal{U} \subset Z$, then for any $r > 0$ we can define the r-packing number of $\mathcal{U}$, denoted by $N_{\text{pack}}(\mathcal{U}, Z, r)$, and the r-covering number of $\mathcal{U}$, denoted by $N_{\text{cov}}(\mathcal{U}, Z, r)$, by

$$\begin{aligned} N_{\text{pack}}(\mathcal{U}, Z, r) &:= \max\{d \in \mathbb{Z} : \exists z_1, \ldots, z_d \in \mathcal{U} \text{ such that } ||z_a - z_b||_Z > r \text{ whenever } a \neq b\} \quad \text{and} \\ N_{\text{cov}}(\mathcal{U}, Z, r) &:= \min\left\{ d \in \mathbb{Z} : \exists z_1, \ldots, z_d \in \mathcal{U} \text{ such that we have the inclusion } \mathcal{U} \subset \bigcup_{j=1}^{d} \overline{\mathbb{B}}_Z(z_j, r) \right\} \end{aligned} \tag{1.5}$$

respectively. Our convention throughout this article is that balls denoted by $\mathbb{B}$ are taken to be open, whilst those denoted by $\overline{\mathbb{B}}$ are taken to be closed. The the quantities defined in (1.5) satisfy, for any $r > 0$, that $N_{\text{pack}}(\mathcal{U}, Z, 2r) \leq N_{\text{cov}}(\mathcal{U}, Z, r) \leq N_{\text{pack}}(\mathcal{U}, Z, r)$. Consequently, the number of steps required for the **Banach GRIM** algorithm to converge is no greater than the $\varepsilon/4C$-covering number of Σ.

Theorem 1.1 establishes that the number of elements from $\mathcal{F}$ required to yield the desired approximation of φ is bounded above by $1 + N_{\text{pack}}(\Sigma, X^*, \varepsilon/2C)$. Thus when $N_{\text{pack}}(\Sigma, X^*, \varepsilon/2C) < \mathcal{N} - 1$, the conclusion (1.4) guarantees that the algorithm will find a linear combination of *fewer* than $\mathcal{N}$ of the elements $f_1, \ldots, f_{\mathcal{N}}$ that is within ε of φ throughout Σ. The $\varepsilon/2C$-packing number of Σ depends on both the features $\mathcal{F}$ through the constant C defined in (1.2) and the data Σ. The constant C itself depends only on the weights $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R}$ and the values $||f_1||_X, \ldots, ||f_{\mathcal{N}}||_X \in \mathbb{R}_{>0}$. No additional constraints are imposed on the collection of features $\mathcal{F}$; in particular, we do not assume the existence of a linear combination of fewer than $\mathcal{N}$ of the features in $\mathcal{F}$ giving a good approximation of φ throughout Σ.

As stated, Theorem 1.1 fixes $\varepsilon > 0$ and provides an upper bound on the number of features from the collection $\mathcal{F}$ that are used to form the approximation u resulting from an application of the **Banach GRIM** algorithm to approximate the element φ defined in (1.2). However, Theorem 1.1 can also be used to determine an implication in, loosely speaking, the opposite direction. To be more specific, assume the setting of Theorem 1.1 and consider a fixed $n_0 \in \{2, \ldots, \mathcal{N}\}$. Then let $\beta_0 = \beta_0(n_0, C, \Sigma) > 0$ be defined by

$$\beta_0 := \min\left\{ \lambda > 0 \;:\; N_{\text{pack}}\left(\Sigma, X^*, \frac{\lambda}{2C}\right) \leq n_0 - 1 \right\}. \tag{1.6}$$

Consider applying the **Banach GRIM** algorithm to approximate φ, with the target accuracy threshold in **Banach GRIM** (A) as β_0 and the choice that $k_s = 1$ for every $s \in \mathbb{Z}_{\geq 1}$. The definition of β_0 in (1.6) ensures that Theorem

1.1 tells us that the **Banach GRIM** algorithm terminates after no more than $n_0 - 1$ steps. Consequently, the algorithm returns an approximation u of φ that is a linear combination of at most n_0 of the features in $\mathcal{F}$, and that is within β_0 of φ on Σ in the sense that for every $\sigma \in \Sigma$ we have $|\sigma(\varphi - u)| \leq \beta_0$.

In this way, given any $n_0 \in \{2, \ldots, \mathcal{N}\}$, the relation given in (1.6) provides a guaranteed accuracy $\beta_0 = \beta_0(n_0, C, \Sigma) > 0$ for how well the **Banach GRIM** algorithm can approximate φ with the additional constraint that the approximation is a linear combination of no greater than n_0 of the features in $\mathcal{F}$. This guarantee ensures both that there is a linear combination of at most n_0 of the features in $\mathcal{F}$ that is within β_0 of φ throughout Σ and that the **Banach GRIM** algorithm will find such a linear combination.

The claim in Theorem 1.1 that the coefficients $a_1, \ldots, a_{\mathcal{N}}$ corresponding to φ (cf. (1.2)) all being positive (i.e. $a_1, \ldots, a_{\mathcal{N}} > 0$) results in the coefficients $c_1, \ldots, c_{Q_M} \in \mathbb{R}$ corresponding to u (cf. (1.4)) being non-negative (i.e. $c_1, \ldots, c_{Q_M} \geq 0$) is a direct consequence of the use of recombination in the **Banach GRIM** algorithm. Every application of recombination is made to the vector of weights $\mathbf{a} := (a_1, \ldots, a_{\mathcal{N}}) \in \mathbb{R}^{\mathcal{N}}$ with respect to a particular linear system of equations to which $\mathbf{a}$ is a solution (cf. Lemma 2.1). It is known that recombination reduces such collections of positive weights to a smaller collection of non-negative weights that remain a solution to the linear system [LL12, LL16]. Hence the positivity preservation claimed in Theorem 1.1 follows.

A particular consequence of this positivity preservation property is that if the target φ is a probability measure, then the approximation found by the **Banach GRIM** algorithm will also be a probability measure whose support is contained within the support of φ. Consequently, if the **Banach GRIM** algorithm is considered for the task of *Kernel Quadrature*, it will enjoy the benefits associated with *convex* weights as detailed in [HLO21].

Momentarily turning our attention to the *Kernel Quadrature* setting, assume that $\mathcal{X}$ is a set admitting a symmetric positive semi-definite kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. Let $\mathcal{H}_k$ denote the RKHS associated to k and fix $\varepsilon > 0$. Suppose that $x_1, \ldots, x_{\mathcal{N}} \in \mathcal{X}$ and $a_1, \ldots, a_{\mathcal{N}} > 0$ so that $\varphi := \sum_{i=1}^{\mathcal{N}} a_i \delta_{x_i} \in \mathbb{P}[\mathcal{X}]$. Under the choice that $X := \mathcal{M}[\mathcal{X}]$, we observe that the constant C corresponding to the definition in (1.2) satisfies that $C = 1$. Assuming that $\overline{\mathbb{B}}_{\mathcal{H}_k}(0,1) \subset C^0(\mathcal{X})$ is compact, Theorem 1.1 for the choice $\Sigma := \overline{\mathbb{B}}_{\mathcal{H}_k}(0,1)$ tells us that the performance of the **Banach GRIM** algorithm is controlled by the pointwise $\varepsilon/2$-packing number of $\overline{\mathbb{B}}_{\mathcal{H}_k}(0,1)$, i.e. by $N_{\text{pack}}(\overline{\mathbb{B}}_{\mathcal{H}_k}(0,1), C^0(\mathcal{X}), \varepsilon/2)$. To be precise, there is an $M \in \{1, \ldots, N_{\text{pack}}(\overline{\mathbb{B}}_{\mathcal{H}_k}(0,1), C^0(\mathcal{X}), \varepsilon/2)\}$ such that, for $Q_M := \min\{\mathcal{N}, M+1\}$, the **Banach GRIM** algorithm finds weights $c_1, \ldots, c_{Q(M)} \geq 0$ and indices $e(1), \ldots, e(Q_M) \in \{1, \ldots, \mathcal{N}\}$ such that $u := \sum_{s=1}^{Q_M} c_s \delta_{x_{e(s)}}$ is a probability measure satisfying, for every $f \in \overline{\mathbb{B}}_{\mathcal{H}_k}(0,1)$, that $|\varphi(f) - u(f)| \leq \varepsilon$.

Recall that the pointwise $\varepsilon/2$-packing number of $\overline{\mathbb{B}}_{\mathcal{H}_k}(0,1)$ is bounded above by the pointwise $\varepsilon/4$-covering number of $\overline{\mathbb{B}}_{\mathcal{H}_k}(0,1)$, i.e. $N_{\text{pack}}(\overline{\mathbb{B}}_{\mathcal{H}_k}(0,1), C^0(\mathcal{X}), \varepsilon/2) \leq N_{\text{cov}}(\overline{\mathbb{B}}_{\mathcal{H}_k}(0,1), C^0(\mathcal{X}), \varepsilon/4)$. Hence estimates for $N_{\text{cov}}(\overline{\mathbb{B}}_{\mathcal{H}_k}(0,1), C^0(\mathcal{X}), \varepsilon/4)$ lead to guarantees for the performance of the **Banach GRIM** algorithm for kernel quadrature. Many authors have considered estimating the covering number of the unit ball of a RKHS, see the works [Zho02, CSW11, Kuh11, SS13, HLLL18, Suz18, JJWWY20, FS21] for example. We now illustrate the performance guarantees available for the **Banach GRIM** algorithm considered for the task of kernel quadrature in a setting covered in [JJWWY20].

For this purpose, we now suppose that $\mathcal{X} \subset \mathbb{R}^d$ is compact, and that the kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is additionally continuous and bounded above by 1. Then the kernel has a pointwise convergent Mercer decomposition $k(x,y) = \sum_{m=1}^{\infty} \lambda_m e_m(x) e_m(y)$ with $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ and $\{e_m\}_{m=1}^{\infty} \subset L^2(\mathcal{X})$ being orthonormal [CS08, SS12]. The pairs $\{(\lambda_m, e_m)\}_{m=1}^{\infty}$ are the eigenpairs for the operator $T_k : L^2(\mathcal{X}) \to L^2(\mathcal{X})$ defined for $f \in L^2(\mathcal{X})$ by $T_k[f] := \int_{\mathcal{X}} k(\cdot, y) f(y) dy$. We assume that the eigenfunctions $\{e_m\}_{m=1}^{\infty}$ are uniformly bounded in the sense that for every $m \in \mathbb{Z}_{\geq 1}$ and every $x \in \mathcal{X}$ we have $|e_m(x)| \leq C_0$ for some constant $C_0 > 0$. Finally, we assume that the eigenvalues $\{\lambda_m\}_{m=1}^{\infty}$ decay exponentially as m increases in the sense that for every $m \in \mathbb{Z}_{\geq 1}$ we have $\lambda_m \leq C_1 e^{-C_2 j}$ for constants $C_1, C_2 > 0$. These assumptions are satisfied, for example, by the *squared exponential* (*Radial Basis Function*) kernel $k(s,t) := e^{-(s-t)^2}$ [JJWWY20]; more explicit estimates for this particular choice of kernel may be found in [FS21].

Given any $r \in (0,1)$, it is established in [JJWWY20] (cf. Lemma D.2 of [JJWWY20]) that under these assumptions we have

$$\log N_{\text{cov}}(\overline{\mathbb{B}}_{\mathcal{H}_k}(0,1), C^0(\mathcal{X}), r) \leq C_3 \left(\log\left(\frac{1}{r}\right) + C_4 \right)^2 \tag{1.7}$$

for constants $C_3 = C_3(C_0, C_1, C_2) > 0$ and $C_4 = C_4(C_0, C_1, C_2) > 0$. Assuming that $\varepsilon < 4$, by appealing to (1.7) for the choice $r := \varepsilon/4$ we may conclude that if $C_3 \left(\log\left(\frac{4}{\varepsilon}\right) + C_4\right)^2 < \log(\mathcal{N} - 1)$ then the **Banach GRIM** algorithm will return a probability measure $u \in \mathbb{P}[\mathcal{X}]$ given by a linear combination of fewer than $\mathcal{N}$ of the point masses $\delta_{x_1}, \ldots, \delta_{x_{\mathcal{N}}}$ satisfying, for every $f \in \overline{\mathbb{B}}_{\mathcal{H}_k}(0,1)$, that $|\varphi(f) - u(f)| \leq \varepsilon$.

Alternatively, given $n_0 \in \{2, \ldots, \mathcal{N}\}$ define $\beta_0 = \beta_0(C_0, C_1, C_2, n_0) > 0$ by

$$\beta_0 := 4e^{C_4} e^{-\left(\frac{\log(n_0-1)}{C_3}\right)^{\frac{1}{2}}} = 4e^{C_4}(n_0-1)^{-(C_3\log(n_0-1))^{-\frac{1}{2}}} > 0. \tag{1.8}$$

For $n_0 > 1 + e^{C_3C_4^2}$ we have that $\beta < 4$. Thus, by appealing to the **Banach GRIM** algorithm for the choice of β_0 as the target accuracy threshold and $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, we may combine (1.7) and (1.8) to deduce from Theorem 1.1 that the algorithm finds a probability measure $u \in \mathbb{P}[\mathcal{X}]$ given by a linear combination of no more than n_0 of the point masses $\delta_{x_1}, \ldots, \delta_{x_{\mathcal{N}}}$ satisfying, for every $f \in \overline{\mathbb{B}}_{\mathcal{H}_k}(0,1)$, that $|\varphi(f) - u(f)| \leq \beta_0$.

As n_0 increases, β_0 defined in (1.8) eventually decays slower than n_0^{-a} for any $a > 0$. This poor asymptotic behaviour is not unexpected for an estimate that is itself a combination of two worst-case scenario estimates. However, we may still observe that for any integer $A \in \mathbb{Z}_{\geq 1}$ large enough to ensure that $1 + e^{C_3C_4^2} < n_0 \leq 1 + e^{A^2/C_3}$ (which in particular requires $A > C_3C_4$), that $\beta_0 \leq 4e^{C_4}(n_0-1)^{-1/A}$.

The theoretical guarantees of Theorem 1.1 are of limited use in the common setting that $X = C^0(\Omega)$ and Σ is a collection of point masses supported at points in Ω. To illustrate the issue let $p, q \in \Omega$ with $p \neq q$ and consider the point masses $\delta_p, \delta_q \in C^0(\Omega)^*$. The existence of a function $f \in C^0(\Omega)$ with $||f||_{C^0(\Omega)} = 1$ and $f(p) = 1$ and $f(q) = -1$ means that $||\delta_p - \delta_q||_{C^0(\Omega)^*} = \sup\left\{|(\delta_p - \delta_q)(f)| : f \in C^0(\Omega) \text{ with } ||f||_{C^0(\Omega)} = 1\right\} = 2$. Consequently, if $\varepsilon/2C < 2$ then the positive integer N defined in (1.3) is simply given by the cardinality of Σ, i.e. $N = \#(\Sigma)$. In this case, the upper bound provided by Theorem 1.1 is the obvious (and essentially useless) fact that once we require our approximation u to coincide with φ at every $\sigma \in \Sigma$, we are guaranteed that u is within ε of φ at every $\sigma \in \Sigma$.

The underlying problem is that, without imposing further assumptions, knowing the value of a continuous function $\phi \in C^0(\Omega)$ at a point $p \in \Omega$ tells us nothing better than the bound $|\phi(q)| \leq ||\phi||_{C^0(\Omega)}$ at points $q \in \Omega \setminus \{p\}$. In order for the value $\phi(p)$ to provide information about the value of ϕ at other points $q \in \Omega$, we must necessarily impose greater regularity than continuity on ϕ. Recalling that in this setting we can view Σ as a subset of Ω, the case that Σ is finite influences our choice of greater regularity to impose. The class of Lip(γ) functions, for some $\gamma > 0$, in the sense of Stein [Ste70] exhibit an appropriate level of regularity.

To introduce Lip(γ) functions we suppose that V and W are both Banach spaces, and that $\Sigma \subset V$ is a closed subset. We additionally assume that all the tensor products of V are equipped with *admissible* norms; see Definition A.1 in this paper for the precise details. Given any $\gamma \in (0,1]$, the space $\mathrm{Lip}(\gamma, \Sigma, W)$ is the familiar space of bounded γ-Hölder continuous functions $\Sigma \to W$. That is, $\psi \in \mathrm{Lip}(\gamma, \Sigma, W)$ means that $\psi : \Sigma \to W$ and there exists a constant $C > 0$ such that whenever $x, y \in \Sigma$ we have both $||\psi(x)||_W \leq C$ and $||\psi(y) - \psi(x)||_W \leq C||y - x||_V^\gamma$. For $\gamma > 1$ the space $\mathrm{Lip}(\gamma, \Sigma, W)$ gives a sensible higher-order generalisation of Hölder continuity.

If we let $k \in \mathbb{Z}_{\geq 0}$ be such that $\gamma \in (k, k+1]$, then an element in $\mathrm{Lip}(\gamma, \Sigma, W)$ is a collection $\psi = \left(\psi^{(0)}, \ldots, \psi^{(k)}\right)$, where for each $l \in \{0, \ldots, k\}$ the function $\psi^{(l)}$ is defined on Σ and takes its values in the space of symmetric l-linear forms from V to W (denoted $\mathcal{L}(V^{\otimes l}; W)$), satisfying the following conditions. Firstly, for each $l \in \{0, \ldots, k\}$ the function $\psi^{(l)}$ is bounded on Σ. Secondly the collection ψ satisfy Taylor-like expansions with appropriately regular remainder terms. To be more precise, given $l \in \{0, \ldots, k\}$, $x, y \in \Sigma$ and $v \in V^{\otimes l}$, it is required that the difference between $\psi^{(l)}(y)[v]$ and $\psi_x^{(l)}(y)[v] := \sum_{s=0}^{k-l} \frac{1}{s!}\psi^{l+s}(x)\left[v \otimes (y-x)^{\otimes s}\right]$ is bounded above by $C||y - x||_V^{\gamma - l}$ for some constant $C \geq 0$. The precise definition of a $\mathrm{Lip}(\gamma, \Sigma, W)$ function may be found in Definition A.2 of this paper.

A good way to understand a $\mathrm{Lip}(\gamma, \Sigma, W)$ function is as a function that "locally looks like a polynomial function". Given any point $x \in \Sigma$, consider the function $\Psi_x : V \to W$ defined for $y \in V$ by

$$\Psi_x(y) := \sum_{s=0}^{k} \frac{1}{s!}\psi^{(s)}(x)\left[(y-x)^{\otimes s}\right]. \tag{1.9}$$

The function Ψ_x defined in (1.9) gives a proposal, based at the point $x \in \Sigma$, for how the function $\psi^{(0)} : \Sigma \to W$ could be extended to the entirety of V. The collection of functions $\psi^{(0)}, \ldots, \psi^{(k)}$ are related to Ψ_x in the following sense. For each $l \in \{0, \ldots, k\}$ the element $\psi^{(l)}(x) \in \mathcal{L}(V^{\otimes l}; W)$ is the l^{th} derivative of $\Psi_x(\cdot)$ at x. The $\mathrm{Lip}(\gamma, \Sigma, W)$ requirement ensures that the proposal $\Psi_x(y)$ varies in a Lipschitz manner as both the basepoint x and the evaluation point y vary across Σ. Loosely, if the points $x_1, x_2, y_1, y_2 \in \Sigma$ are all mututally close in the $||\cdot||_V$

norm sense, then the values $\Psi_{x_1}(y_1)$ and $\Psi_{x_2}(y_2)$ must be close in the $||\cdot||_W$ norm sense. The proposal functions Ψ_x for points $x \in \Sigma$ enable one to view a $\mathrm{Lip}(\gamma, \Sigma, W)$ function in a more traditional manner as the single function $\Sigma \times \Sigma \to W$ defined by the mapping $(x, y) \mapsto \Psi_x(y)$.

Returning to considering the collection $\psi = \left(\psi^{(0)}, \ldots, \psi^{(k)}\right) \in \mathrm{Lip}(\gamma, \Sigma, W)$, we observe that on the interior of Σ the functions $\psi^{(1)}, \ldots, \psi^{(k)}$ are determined by $\psi^{(0)}$. The bounds satisfied by $\psi^{(l)}(y)[v] - \psi_x^{(l)}(y)[v]$ for each $l \in \{0, \ldots, k\}$ and $v \in V^{\otimes l}$ ensure, for each $j \in \{1, \ldots, k\}$, that $\psi^{(j)}$ is the classical j^{th} derivative of $\psi^{(0)}$ on the interior of Σ. Thus, on the interior of Σ, the function $\psi^{(0)}$ is k times continuously differentiable, and its k^{th} derivative is $(\gamma - k)$-Hölder continuous. The original work of Stein [Ste70] provides further evidence for the sensibility of this notion of a $\mathrm{Lip}(\gamma, \Sigma, W)$ function. In a similar spirit to the original Whitney extension theorems (cf. Theorem I in [Whi34] and Theorem I in [Whi44]), Stein proves that when $V = \mathbb{R}^d$, $W = \mathbb{R}$ and $\Sigma \subset \mathbb{R}^d$ is closed, a function $\psi \in \mathrm{Lip}(\gamma, \Sigma, \mathbb{R})$ can be extended to an element $\tilde{\psi} \in \mathrm{Lip}(\gamma, \mathbb{R}^d, \mathbb{R})$ (cf. Chapter VI, Theorem 4 in [Ste70]).

A key benefit of the class $\mathrm{Lip}(\gamma, \Sigma, W)$ is that it well-defined for *any* closed subset Σ, even those with empty interiors such as finite subsets. The condition that $\psi = \left(\psi^{(0)}, \ldots, \psi^{(k)}\right) \in \mathrm{Lip}(\gamma, \Sigma, W)$ imposes a higher order notion of regularity using only the values of the functions $\psi^{(0)}, \ldots, \psi^{(k)}$ at the points in Σ. No knowledge of their behaviour outside Σ is required; we do not need to even define their value at *any* point in $V \setminus \Sigma$. Moreover, the value of ψ a point $p \in \Sigma$ determines the pointwise behaviour of ψ, up to arbitrarily small errors, within a ball centred at p whose radius increases as γ increases (see the *pointwise Lipschitz sandwich theorem* B.8 for precise details). These properties make the class of $\mathrm{Lip}(\gamma, \Sigma, W)$ functions attractive from a learning perspective.

We consider our sparse approximation problem in the $\mathrm{Lip}(\gamma, \Sigma, W)$ setting. That is, for each $i \in \{1, \ldots, \mathcal{N}\}$ we suppose that $f_i = \left(f_i^{(0)}, \ldots, f_i^{(k)}\right) \in \mathrm{Lip}(\gamma, \Sigma, W)$. A consequence is that φ defined in (1.1) is an element in $\mathrm{Lip}(\gamma, \Sigma, W)$. To ease notation, given $\phi = \left(\phi^{(0)}, \ldots, \phi^{(k)}\right) \in \mathrm{Lip}(\gamma, \Sigma, W)$ and $x \in \Sigma$ we define

$$\Lambda_\phi^k(x) := \max_{j \in \{0, \ldots, k\}} \left\{ \left|\left|\phi^{(j)}(x)\right|\right|_{\mathcal{L}(V^{\otimes j}; W)} \right\}. \tag{1.10}$$

Our strategy is as follows. We use a greedy selection algorithm to inductively grow a collection of points $L \subset \Sigma$ on which we will require an approximation u to agree with φ in the sense that $\Lambda_{\varphi-u}^k(x) = 0$ for every $x \in L$. At each step the thinning of recombination is used to find the desired approximation. We propose the following **Lip(γ) Extension Step** for the purpose of determining new points in Σ to be added to an existing collection $L' \subset \Sigma$.

Lip(γ) Extension Step
Assume that $L' \subset \Sigma$ and $u = \left(u^{(0)}, \ldots, u^{(k)}\right) \in \mathrm{Span}(\mathcal{F})$ are both given. Let $m \in \mathbb{Z}_{\geq 1}$ and first take $z_1 := \mathrm{argmax}\left\{\Lambda_{\varphi-u}^k(z) : z \in \Sigma\right\}$. Inductively for $j = 2, 3, \ldots, m$ we choose $z_j \in \Sigma$ by taking $z_j := \mathrm{argmax}\left\{\Lambda_{\varphi-u}^k(z) : z \in \Sigma \setminus \{z_1, \ldots, z_{j-1}\}\right\}$. Once $z_1, \ldots, z_m \in \Sigma$ have been defined, we extend L' to $L := L' \cup \{z_1, \ldots, z_m\}$.

For a given set $L \subset \Sigma$, recombination can be used to find an element $u \in \mathrm{Span}(\mathcal{F})$ such that φ and u coincide in the sense that $\Lambda_{\varphi-u}^k(z) = 0$ at every $z \in L$ (cf. Lemma 3.1). We propose the following GRIM algorithm to approximate φ.

Lip(γ) GRIM

(A) Fix $\varepsilon > 0$ as the target accuracy threshold.

(B) Choose $k_1 \in \mathbb{Z}_{\geq 1}$ and apply the **Lip(γ) Extension Step**, with $L' = \emptyset$, $u \equiv 0$ and $m := k_1$, to obtain a subset $\Sigma_1 = \{z_{1,1}, \ldots, z_{1,k_1}\} \subset \Sigma$. Apply recombination (cf. Lemma 3.1) to φ to find an element $u_1 \in \mathrm{Span}(\mathcal{F})$ satisfying, for every $z \in \Sigma_1$, that $\Lambda_{\varphi-u_1}^k(z) = 0$.

(C) For $s \in \mathbb{Z}_{\geq 2}$ we proceed inductively. If $|\varphi(z) - u_{s-1}(z)| \leq \varepsilon$ for every $z \in \Sigma$ then we stop since u_{s-1} is an approximation of φ possessing the desired level of accuracy. If this is not the case, then choose $k_s \in \mathbb{Z}_{\geq 1}$ and apply the **Lip(γ) Extension Step**, with $L' = \Sigma_{s-1}$, $u := u_{s-1}$ and $m := k_s$, to obtain a subset $\Sigma_s = \Sigma_{s-1} \cup \{z_{s,1}, \ldots, z_{s,k_s}\} \subset \Sigma$. Apply recombination (cf. Lemma 3.1) to φ to find an element $u_s \in \mathrm{Span}(\mathcal{F})$ satisfying, for every $z \in \Sigma_s$, that $\Lambda_{\varphi-u_s}^k(z) = 0$.

The dynamic growth is *data-driven*; we first choose the information that our approximation is required to match *before* using recombination to both select the elements from $\mathcal{F}$ and use them to construct the desired approximation.

The greedy selection of the points from Σ results in the nesting property that $\Sigma_{t_1} \subset \Sigma_{t_2}$ for integers $t_1 \leq t_2$, ensuring that at each step we increase the information our approximation will be required to match.

For each integer $m \in \mathbb{Z}_{\geq 1}$ let $S_m \subset \mathcal{F}$ denote the sub-collection of elements from $\mathcal{F}$ that are used to form the approximation u_m. If we let $D := \dim(W) \sum_{s=0}^{k} \binom{\dim(V)+s-1}{s}$ then recombination is applied to a system of $1 + (k_1 + \ldots + k_m)D$ linear equations when finding u_m. Hence we may conclude that $\#(S_m) \leq \min\{1 + (k_1 + \ldots + k_m)D, \mathcal{N}\}$ (cf. Lemma 3.1). Besides this upper bound for $\#(S_m)$, we have *no* control on the sets S_m. There is no requirement that elements from $\mathcal{F}$ used to form u_m must also be used for u_l for $l > m$.

For the case that, for each $s \in \mathbb{Z}_{\geq 1}$, we choose $k_s := 1$ we are able to establish the following worst-case upper bound on the number of steps the **Lip(γ) GRIM** algorithm requires to find approximation $u \in \mathrm{Span}(\mathcal{F})$ satisfying, for every $z \in \Sigma$, that $\Lambda^k_{\varphi-u}(z) \leq \varepsilon$.

Theorem 1.2 (**Lip(γ) GRIM Convergence**)**.** *Let V and W be finite dimensional Banach spaces with $\Sigma \subset V$ compact. Assume that the tensor products of V are all equipped with admissible norms (cf. Definition A.1). Let $\mathcal{N} \in \mathbb{Z}_{>0}$ be a positive integer and $\gamma > 0$ with $k \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k, k+1]$. . Assume that for every $i \in \{1, \ldots, \mathcal{N}\}$ the element $f_i = \left(f_i^{(0)}, \ldots, f_i^{(k)}\right) \in \mathrm{Lip}(\gamma, \Sigma, W)$ is non-zero and define $\mathcal{F} := \{f_i : i \in \{1, \ldots, \mathcal{N}\}\} \subset \mathrm{Lip}(\gamma, \Sigma, W)$. Given $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$, define $\varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(k)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\gamma, \Sigma, W)$ and constants $C, D > 0$ by*

$$\textbf{(I)} \quad \varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i, \quad \left(\begin{array}{c} \varphi^{(l)} := \sum_{i=1}^{\mathcal{N}} a_i f_i^{(l)} \\ \textit{for every } l \in \{0, \ldots, k\} \end{array} \right) \quad \textit{and} \quad \textbf{(II)} \quad C := \sum_{i=1}^{\mathcal{N}} |a_i| ||f_i||_{\mathrm{Lip}(\gamma,\Sigma,W)} > 0 \tag{1.11}$$

and $D := \dim(W) \sum_{s=0}^{k} \binom{\dim(V)+s-1}{s}$. Let $N = N(\Sigma, C, \gamma, \varepsilon) \in \mathbb{Z}_{\geq 0}$ denote the $(\varepsilon/2C)^{1/(\gamma-k)}$ packing number for Σ. That is

$$N := \max\left\{ d \in \mathbb{Z} \ : \ \exists x_1, \ldots, x_d \in \Sigma \textit{ for which } ||x_a - x_c||_V > \left(\frac{\varepsilon}{2C}\right)^{\frac{1}{\gamma-k}} \textit{ if } a \neq c \right\}. \tag{1.12}$$

*Then if we apply the **Lip(γ) GRIM** algorithm, with the choice that $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, to approximate φ on Σ, with the target accuracy threshold in **Lip(γ) GRIM** (A) as ε, then after at most N steps, the algorithm terminates. That is, if we let $M \in \{1, \ldots, N\}$ be the integer for which the algorithm terminates after step M and define $Q_M := \min\{\mathcal{N}, 1 + MD\}$, there are coefficients $c_{M,1}, \ldots, c_{M,Q_M} \in \mathbb{R}$ and indices $e_M(1), \ldots, e_M(Q_M) \in \{1, \ldots, \mathcal{N}\}$ such that $u_M = \left(u_M^{(0)}, \ldots u_M^{(k)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\gamma, \Sigma, W)$ defined by*

$$u_M := \sum_{s=1}^{Q_M} c_{M,s} f_{e_M(s)} \quad \left(\begin{array}{c} u_M^{(l)} := \sum_{s=1}^{Q_M} c_{M,s} f_{e_M(s)}^{(l)} \\ \textit{for every } l \in \{0, \ldots, k\} \end{array} \right) \quad \textit{satisfies} \quad \sup_{x \in \Sigma} \Lambda^k_{\varphi-u_M}(x) \leq \varepsilon. \tag{1.13}$$

*Moreover, if the coefficients $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ corresponding to φ (cf. (**I**) of (1.11)) are all positive (i.e. $a_1, \ldots, a_{\mathcal{N}} > 0$) then the coefficients $c_{M,1}, \ldots, c_{M,Q_M} \in \mathbb{R}$ corresponding to u_M (cf. (1.13)) are all non-negative (i.e. $c_{M,1}, \ldots, c_{M,Q_M} \geq 0$).*

This theorem will be a special case of the more elaborate Theorem 3.2 which will additionally establish a robustness estimate for the approximation u and account for numerical errors arising in the use of recombination. The claim that the coefficients $a_1, \ldots, a_{\mathcal{N}}$ corresponding to φ (cf. (**I**) of (1.11)) all being positive (i.e. $a_1, \ldots, a_{\mathcal{N}} > 0$) results in the coefficients $c_{M,1}, \ldots, c_{M,Q_M} \in \mathbb{R}$ corresponding to u (cf. (1.13)) being non-negative (i.e. $c_{M,1}, \ldots, c_{M,Q_M} \geq 0$) is a direct consequence of the use of recombination in the **Lip(γ) GRIM** algorithm. Every application of recombination is made to the vector of weights $\mathbf{a} := (a_1, \ldots, a_{\mathcal{N}}) \in \mathbb{R}^{\mathcal{N}}$ with respect to a particular linear system of equations to which $\mathbf{a}$ is a solution (cf. Lemma 3.1). It is known that recombination reduces such collections of positive weights to a smaller collection of non-negative weights that remain a solution to the linear system [LL12, LL16]. Hence the positivity preservation claimed in Theorem 1.2 follows.

Theorem 1.2 establishes that the number of elements from $\mathcal{F}$ required to yield the desired approximation of

φ is bounded above by $1 + DN_{\text{pack}}\left(\Sigma, V, (\varepsilon/2C)^{1/(\gamma-k)}\right)$. Thus when $N_{\text{pack}}\left(\Sigma, V, (\varepsilon/2C)^{1/(\gamma-k)}\right) < \frac{\mathcal{N}-1}{D}$, the conclusion (1.13) guarantees that the algorithm will find a linear combination of *fewer* than $\mathcal{N}$ of the elements $f_1, \ldots, f_{\mathcal{N}}$ that is within ε of φ throughout Σ pointwise sense that $\Lambda^k_{\varphi-u}(z) \leq \varepsilon$ for every $z \in \Sigma$. The $(\varepsilon/2C)^{1/(\gamma-k)}$-packing number of Σ depends on both the features $\mathcal{F}$ through the constant C defined in (1.2) and the data Σ. The constant C itself depends only on the original weights $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R}$ and the values $||f_1||_{\text{Lip}(\gamma,\Sigma,W)}, \ldots, ||f_{\mathcal{N}}||_{\text{Lip}(\gamma,\Sigma,W)} \in \mathbb{R}_{>0}$. No additional constraints are imposed on the collection of functions $\mathcal{F}$; in particular, we do not assume the existence of a linear combination of fewer than $\mathcal{N}$ of the features in $\mathcal{F}$ giving a good approximation of φ throughout Σ.

As stated, Theorem 1.2 fixes $\varepsilon > 0$ and provides an upper bound on the number of features from the collection $\mathcal{F}$ that are used to form the approximation u resulting from an application of the **Lip(γ) GRIM** algorithm to approximate the element φ defined in (**I**) of (1.11). However, in exactly the same way that the implication direction of Theorem 1.1 could be reversed, Theorem 1.2 can be used to determine an implication in, loosely speaking, the opposite direction. To be more specific, assume the setting of Theorem 1.2 and consider a fixed $n_0 \in \{2, \ldots, \mathcal{N}\}$. Then define $\beta_0 = \beta_0(n_0, C, \Sigma, \gamma) > 0$ by

$$\beta_0 := \min\left\{\lambda > 0 \ : \ N_{\text{pack}}\left(\Sigma, V, \left(\frac{\lambda}{2C}\right)^{\frac{1}{\gamma-k}}\right) \leq \frac{n_0 - 1}{D}\right\} > 0. \tag{1.14}$$

Consider applying the **Lip(γ) GRIM** algorithm to approximate φ, with the target accuracy threshold in **Banach GRIM** (A) as β_0 and the choice that $k_s = 1$ for every $s \in \mathbb{Z}_{\geq 1}$. The validity of (1.14) ensures that Theorem 1.2 tells us that the **Lip(γ) GRIM** algorithm terminates after no more than $\frac{n_0-1}{D}$ steps. Consequently, the algorithm returns an approximation u of φ that is a linear combination of at most n_0 of the features in $\mathcal{F}$, and that is within β_0 of φ on Σ in the sense that for every $x \in \Sigma$ we have $\Lambda^k_{\varphi-u}(x) \leq \beta_0$.

In this way, given any $n_0 \in \{2, \ldots, \mathcal{N}\}$, the relation given in (1.14) provides a guaranteed accuracy $\beta_0 = \beta_0(n_0, C, \Sigma, \gamma) > 0$ for how well the **Lip(γ) GRIM** algorithm can approximate φ with the additional constraint that the approximation is a linear combination of no greater than n_0 of the features in $\mathcal{F}$. This guarantee ensures both that there is a linear combination of at most n_0 of the features in $\mathcal{F}$ that is within β_0 of φ throughout Σ and that the **Lip(γ) GRIM** algorithm will find such a linear combination.

For illustration, suppose we are in the setting of Theorem 1.2 for the choices of $V := \mathbb{R}^d$ equipped with the Euclidean norm, $\Sigma := \overline{\mathbb{B}}^d(0,1) \subset \mathbb{R}^d$, $W := \mathbb{R}$, and $\gamma \in (0,1]$. Thus we have a collection $\mathcal{F} = \{f_1, \ldots, f_{\mathcal{N}}\} \subset \text{Lip}(\gamma, \overline{\mathbb{B}}^d(0,1), \mathbb{R})$ of non-zero elements and non-zero weights $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R}\backslash\{0\}$, and our goal is to approximate the element $\varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i$. In particular, for a given $n_0 \in \{1, \ldots, \mathcal{N}\}$ we are interested in how well we are able to approximate φ by a linear combination of at most n_0 elements from the collection $\mathcal{F}$.

Suppose that $r > 0$, $m \in \mathbb{Z}_{\geq 1}$ and that $y_1, \ldots, y_m \in \overline{\mathbb{B}}^d(0,1)$ are such that whenever $i, j \in \{1, \ldots, m\}$ with $i \neq j$ then $|y_i - y_j| > r$. Then the balls $\left\{\overline{\mathbb{B}}^d(y_i, r/2)\right\}_{i=1}^m$ are pairwise disjoint, and their union is contained within the ball $\overline{\mathbb{B}}^d(0, 1 + r/2)$. A volume comparison argument leads to the inequality that $m \leq (1 + 2/r)^d$. Consequently, we may conclude that $N_{\text{pack}}\left(\overline{\mathbb{B}}^d(0,1), \mathbb{R}^d, r\right) \leq (1 + 2/r)^d$.

Fix $n_0 \in \{3, \ldots, \mathcal{N}\}$ and define

$$\beta_0 := 2C\left(\frac{2}{(n_0-1)^{\frac{1}{d}} - 1}\right)^{\gamma} > 0. \tag{1.15}$$

Observe that (1.14) is valid since in this setting the constant D arising in Theorem 1.2 is equal to 1. As a result, the **Lip(γ) GRIM** algorithm, with this β_0 as the target accuracy threshold and the choice of $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, is guaranteed to find an approximation $u \in \text{Span}(\mathcal{F})$ of φ that is a linear combination of at most n_0 elements from $\mathcal{F}$ and is within β_0 of φ throughout $\overline{\mathbb{B}}^d(0,1)$ in the sense that, for every $x \in \overline{\mathbb{B}}^d(0,1)$, we have

$$\Lambda^0_{\varphi-u}(x) = |\varphi(x) - u(x)| \leq \beta_0 \overset{(1.15)}{=} 2C\left(\frac{2}{(n_0-1)^{\frac{1}{d}} - 1}\right)^{\gamma}. \tag{1.16}$$

Given $\varepsilon > 0$, it follows from (1.16) that if

$$n_0 \geq 1 + \left(1 + 2\left(\frac{2C}{\varepsilon}\right)^{\frac{1}{\gamma}}\right)^d \qquad \text{then} \qquad \sup_{x \in \overline{\mathbb{B}}^d(0,1)} |\varphi(x) - u(x)| \leq \varepsilon. \tag{1.17}$$

Of course (1.17) is only useful provided $\mathcal{N} > 1 + \left(1 + 2\left(\frac{2C}{\varepsilon}\right)^{\frac{1}{\gamma}}\right)^d$. It is important to keep in mind that the estimates (1.16) and (1.17) are worst-case scenario bounds. Beyond the assumption that the set $\mathcal{F} \subset \mathrm{Lip}(\gamma, \overline{\mathbb{B}}^d(0,1), \mathbb{R})$ is bounded, there are no additional conditions imposed on the functions in $\mathcal{F}$. In particular, there is no assumption made that the set $\mathcal{F}$ can be well-approximated by any strict subset of $\mathcal{F}$.

A key step in our proof of Theorem 1.2 is to establish that if two $\mathrm{Lip}(\gamma, \Sigma, W)$ functions coincide at a point $p \in \Sigma$, then they must remain close in a pointwise sense on a ball of a definite radius centred at p. A variant of this phenomenon is recorded in the following result.

Theorem 1.3 (**Pointwise Lipschitz Sandwich Theorem**). *Let V and W be Banach spaces, and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume that $\Sigma \subset V$ is closed. Let $K_0, \gamma, \varepsilon > 0$ with $k \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k, k+1]$. Then there exists a constant $\delta_0 = \delta_0(\varepsilon, K_0, \gamma) > 0$, defined by $\delta_0 := \min\left\{1\ ,\ \left(\frac{\varepsilon}{2K_0}\right)^{\frac{1}{\gamma - k}}\right\} > 0$, for which the following is true.*

Suppose $B \subset \Sigma$ is a δ_0-cover of Σ in the sense that $\Sigma \subset B_{\delta_0} := \{v \in V\ :\ \exists z \in B \text{ such that } ||v - z||_V \leq \delta_0\}$. Suppose $\psi = \left(\psi^{(0)}, \ldots, \psi^{(k)}\right), \varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(k)}\right) \in \mathrm{Lip}(\gamma, \Sigma, W)$ with $||\psi||_{\mathrm{Lip}(\gamma,\Sigma,W)}, ||\varphi||_{\mathrm{Lip}(\gamma,\Sigma,W)} \leq K_0$. Further suppose that $\psi \equiv \varphi$ throughout B. That is, for every $j \in \{0, \ldots, k\}$ and every $x \in B$ we have $\psi^{(j)}(x) = \varphi^{(j)}(x)$ in $\mathcal{L}(V^{\otimes j}; W)$. Then we may conclude that for every $s \in \{0, \ldots, k\}$ and every $x \in \Sigma$ that

$$\left|\left|\psi^{(s)}(x) - \varphi^{(s)}(x)\right|\right|_{\mathcal{L}(V^{\otimes s};w)} \leq \varepsilon. \tag{1.18}$$

In the case that $\gamma \in (0,1]$ and $W := \mathbb{R}$ this result is a well-known consequence of the combination of the maximal and minimal extensions of Whitney [Whi34] and McShane [McS34] respectively. The main content of this result is the setting that $\gamma > 1$, for which case we have been unable to locate a formal statement of this phenomenon within the existing literature.

Theorem 1.3 will be a special case of the more elaborate *pointwise Lipschitz sandwich theorem* B.8 which will establish a similar conclusion under the weaker assumption that the elements $\psi, \varphi \in \mathrm{Lip}(\gamma, \Sigma, W)$ are merely close throughout B in the sense that $\Lambda^k_{\psi - \varphi}(z)$ is small for every $z \in B$. Moreover, given any $\eta \in (0, \gamma)$, we additionally prove the *Lipschitz sandwich theorem* B.1 in which the pointwise conclusion (1.18) is replaced by the estimate that the $\mathrm{Lip}(\eta, \Sigma, W)$ norm of the difference $\psi - \varphi$ is no greater than ε. The restriction to $\eta < \gamma$ is essential and the result is *not* true for $\eta := \gamma$ (see Appendix B for full details).

The remainder of the paper is structured as follows. In Section 2 we present and analyse a more detailed version of the **Banach GRIM** algorithm. In particular, the convergence properties claimed in Theorem 1.1 are a particular consequence of the more detailed Theorem 2.2. In Section 3 we present a more detailed version of the **Lip(γ) GRIM** algorithm, and analyse it using the results of Section B. The convergence properties claimed in Theorem 1.2 are a particular consequence of the more elaborate $\mathrm{Lip}(\gamma)$ *GRIM Convergence* theorem 3.2. Additionally, given $\eta \in (0, \gamma)$, Theorem 3.3 establishes that the **Lip(γ) GRIM** algorithm can be applied to find an approximation of $\varphi \in \mathrm{Lip}(\gamma, \Sigma, W)$ that is close to φ in the space $\mathrm{Lip}(\eta, \Sigma, W)$.

The elements of the theory of $\mathrm{Lip}(\gamma, \Sigma, W)$ functions required in Section 3 are included in the appendices. In Appendix A we both rigorously define the notion, for $\gamma > 0$, of a $\mathrm{Lip}(\gamma, \Sigma, W)$ function, and collect together several well-known properties of $\mathrm{Lip}(\gamma, \Sigma, W)$ functions. In Appendix B we state several results detailing how the $\mathrm{Lip}(\gamma)$-behaviour of a function on a subset $B \subset \Sigma$ impacts its $\mathrm{Lip}(\eta)$, for $\eta < \gamma$, behaviour on the entirety of Σ. In particular, Theorem 1.3 is a consequence of the more involved *pointwise Lipschitz sandwich theorem* B.8, and we establish the full *Lipschitz sandwich theorem* B.1. All the results presented in Appendix B articulate properties of $\mathrm{Lip}(\gamma)$ functions in a form that is convenient for our purposes. When $\gamma \in (0,1]$ and $W := \mathbb{R}$, they are well-understood properties that can be deduced via the consideration of the maximal and minimal extensions of Whitney [Whi34] and McShane [McS34] respectively. The main content of Appendix B is that the phenomenon remain valid for the setting that $\gamma > 1$, for which case we have been unable to locate formal statements of these properties within the existing literature. In Appendix C we prove all the results stated in Appendix B along with several supplementary lemmata.

Acknowledgements: Terry Lyons was supported in part by the DataSıg programme funded by the EPSRC [grant number EP/S026347/1], in part by The Alan Turing Institute under the EPSRC grant EP/N510129/1, the Data Centric Engineering Programme (under the Lloyd's Register Foundation grant G0095), the Defence and Security Programme (funded by the UK Government) and the Office for National Statistics & The Alan Turing Institute (strategic partnership) and in part by the Hong Kong Innovation and Technology Commission (InnoHK Project CIMDA). Andrew McLeod was supported by the DataSıg programme funded by the EPSRC [grant number EP/S026347/1].

2. Banach GRIM

Let X be a Banach space and denote its dual space by X^*. Throughout, when referring to metric balls we use the convention that those denoted by $\mathbb{B}$ are taken to be open, whilst those denoted by $\overline{\mathbb{B}}$ are taken to be closed. In this section we consider searching for a sparse approximation of a given linear combination of elements in X.

To make this more precise, assume that $\Sigma \subset X^*$ is non-empty and compact, and that $\mathcal{N} \in \mathbb{Z}_{\geq 1}$ is a (large) positive integer. For every $i \in \{1, \ldots, \mathcal{N}\}$ we assume that $f_i \in X \setminus \{0\}$ and define $\mathcal{F} := \{f_1, \ldots, f_{\mathcal{N}}\} \subset X$. Given $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$, define $\varphi \in X$ by

$$\varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i \tag{2.1}$$

Given $\varepsilon > 0$ we try to find an element $u \in \mathrm{Span}(\mathcal{F}) \subset X$ that is both a sparse linear combination of the elements $f_1, \ldots, f_{\mathcal{N}}$ and gives a good approximation of φ in the following sense. For every linear functional $\sigma \in \Sigma$, we have that $|\sigma(\varphi - u)| \leq \varepsilon$.

To find our desired approximation u of φ, we use a combination of dynamic growth driven interpolation techniques and thinning techniques. A loose summary of our approach is as follows. We use a greedy selection algorithm to inductively grow a collection of linear functionals $L \subset \Sigma$ at which we want our approximation to agree with φ. At each step, recombination is used to thin the coefficients $a_1, \ldots, a_{\mathcal{N}}$ to find an element $u \in \mathrm{Span}(\mathcal{F}) \subset X$ that coincides with φ throughout the current choice of $L \subset \Sigma$.

The following **Banach Extension Step** is used to dynamically grow a collection of linear functionals from Σ at which we require our next approximation of φ to agree with φ.

Banach Extension Step
Assume that $L' \subset \Sigma$ and $u \in \mathrm{Span}(\mathcal{F})$ are both given. Let $m \in \mathbb{Z}_{\geq 1}$ and take

$$\sigma_1 := \operatorname{argmax}\left\{ |\sigma(\varphi - u)| \ : \ \sigma \in \Sigma \right\}. \tag{2.2}$$

Inductively for $j = 2, 3, \ldots, m$ take

$$\sigma_j := \operatorname{argmax}\left\{ |\sigma(\varphi - u)| \ : \ \sigma \in \Sigma \setminus \{\sigma_1, \ldots, \sigma_{j-1}\} \right\}. \tag{2.3}$$

Once $\sigma_1, \ldots, \sigma_m \in \Sigma$ have been defined, we extend L' to $L := L' \cup \{\sigma_1, \ldots, \sigma_m\}$.

For each choice of subset $L \subset \Sigma$, we use recombination to find $u \in \mathrm{Span}(\mathcal{F})$ agreeing with φ throughout L. That this is possible is the content of the following result.

Lemma 2.1 (Banach Space Recombination). *Assume X is a Banach space with dual space X^*. Let $\mathcal{N} \in \mathbb{Z}_{\geq 1}$ and $\mathcal{F} := \{f_1, \ldots, f_{\mathcal{N}}\} \subset X$ be a collection of non-zero elements. Suppose that $L \subset X^*$ is finite with cardinality $m \in \mathbb{Z}_{\geq 0}$, i.e. $\#(L) = m$. Let $a_1, \ldots, a_{\mathcal{N}} > 0$ and consider the element $\varphi \in \mathrm{Span}(\mathcal{F}) \subset X$ defined by $\varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i$. Set $\mathcal{M} := \min\{\mathcal{N}, m+1\}$. Then recombination can be applied to find non-negative coefficients $b_1, \ldots, b_{\mathcal{M}} \geq 0$ and indices $e(1), \ldots, e(\mathcal{M}) \in \{1, \ldots, \mathcal{N}\}$ satisfying that*

$$\sum_{j=1}^{\mathcal{M}} b_j = \sum_{i=1}^{\mathcal{N}} a_i, \tag{2.4}$$

and such that the element $u \in \mathrm{Span}(\mathcal{F}) \subset X$ defined by

$$u := \sum_{j=1}^{\mathcal{M}} b_j f_{e(j)} \quad \textit{satisfies, for every } \sigma \in L, \textit{ that} \quad \sigma(\varphi - u) = 0. \tag{2.5}$$

Proof of Lemma 2.1. Let $L = \{\sigma_1, \ldots, \sigma_m\} \subset X^*$. Since $\varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i$, the values $\sigma_1(\varphi), \ldots, \sigma_m(\varphi)$ and the sum of the coefficients $\sum_{i=1}^{\mathcal{N}} a_i$ give rise to the linear system of equations

$$\begin{pmatrix} 1 & 1 & \ldots & 1 \\ \sigma_1(f_1) & \sigma_1(f_2) & \ldots & \sigma_1(f_{\mathcal{N}}) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_m(f_1) & \sigma_m(f_2) & \ldots & \sigma_m(f_{\mathcal{N}}) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_{\mathcal{N}} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{\mathcal{N}} a_i \\ \sigma_1(\varphi) \\ \vdots \\ \sigma_m(\varphi) \end{pmatrix} \tag{2.6}$$

which we denote more compactly by $\mathbf{A}\mathbf{x} = \mathbf{y}$.

The positivity of the coefficients $a_1, \ldots, a_{\mathcal{N}}$ allows us to apply recombination [LL12] to this linear system. Recombination uses non-zero elements in the kernel $\ker(\mathbf{A})$ to kill (i.e. make 0) as many of the coefficients of $\mathbf{x}$ as possible. Altering $\mathbf{x}$ by elements $\mathbf{e} \in \ker(\mathbf{A})$ leaves the equality $\mathbf{A}\mathbf{x} = \mathbf{y}$ unchanged since $\mathbf{A}(\mathbf{x}+\theta\mathbf{e}) = \mathbf{A}\mathbf{x}$ for any $\theta \in \mathbb{R}$. Observe that $\dim(\ker(\mathbf{A})) \geq \mathcal{N} - \mathcal{M}$ for $\mathcal{M} := \min\{\mathcal{N}, m+1\}$. Therefore there are at least $\mathcal{N} - \mathcal{M}$ linearly independent elements in $\ker(\mathbf{A})$ which may be used to kill coefficients of $\mathbf{x}$; that is, recombination transforms $\mathbf{x}$ to a solution with at most $\mathcal{M}$ non-zero coefficients. Thus once recombination has been applied, examining (2.6) reveals that we have non-negative coefficients $b_1, \ldots, b_{\mathcal{M}} \geq 0$ and indices $e(1), \ldots, e(\mathcal{M}) \in \{1, \ldots, \mathcal{N}\}$ that satisfy (2.4), and such that the element u defined (2.5) satisfies that $\sigma_j(\varphi - u) = 0$ for every $j \in \{1, \ldots, m\}$. That is, the estimates claimed in (2.5) are satisfied. This completes the proof of Lemma 2.1. ■

Theoretically, Lemma 2.1 verifies that recombination can be used to find an approximation u of φ for which $\sigma(\varphi - u) = 0$ for all linear functionals $\sigma \in L$ for a given finite subset $L \subset \Sigma \subset X^*$. However, practical implementations of recombination will inevitably result in numerical errors entering into the equations. That is, the returned coefficients will only solve the equations modulo some (hopefully) small error term. To account for this in our analysis, whenever we apply Lemma 2.1 we will only assume that the resulting approximation $u \in \mathrm{Span}(\mathcal{F})$ is *close* to φ at each functional $\sigma \in L$. That is, for each $\sigma \in L$, we have that $|\sigma(\varphi - u)| \leq \varepsilon_0$ for some (small) constant $\varepsilon_0 \geq 0$.

We now rigorously define our proposed GRIM algorithm to find an approximation $u \in \mathrm{Span}(\mathcal{F})$ of $\varphi \in \mathrm{Span}(\mathcal{F})$ that is close to φ at every linear functional in Σ.

Banach GRIM

(A) Fix $\varepsilon > 0$ as the target accuracy threshold and fix $\varepsilon_0 \in [0, \varepsilon)$ as the acceptable recombination error bound.

(B) For each $i \in \{1, \ldots, \mathcal{N}\}$, if $a_i < 0$ then replace a_i and f_i by $-a_i$ and $-f_i$ respectively. This ensures that $a_1, \ldots, a_{\mathcal{N}} > 0$ whilst leaving the expansion $\varphi = \sum_{i=1}^{\mathcal{N}} a_i f_i$ unaltered. Additionally, rescale each f_i to have unit X norm. That is, for each $i \in \mathbb{N}$ we replace f_i by $h_i := \frac{f_i}{||f_i||_X}$. Then $\varphi = \sum_{i=1}^{\mathcal{N}} \alpha_i h_i$ where, for each $i \in \{1, \ldots, \mathcal{N}\}$, we have $\alpha_i := a_i ||f_i||_X$.

(C) Choose $k_1 \in \mathbb{Z}_{\geq 1}$ and apply the **Banach Extension Step**, with $L' = \varnothing$, $u \equiv 0$ and $m := k_1$, to obtain a subset $\Sigma_1 = \{\sigma_{1,1}, \ldots, \sigma_{1,k_1}\} \subset \Sigma$. Apply recombination (cf. Lemma 2.1) to find an element $u_1 \in \mathrm{Span}(\mathcal{F})$ satisfying, for every $\sigma \in \Sigma_1$, that $|\sigma(\varphi - u)| \leq \varepsilon_0$.

(D) For $s \geq 2$ we proceed inductively. If $|\sigma(\varphi - u_{s-1})| \leq \varepsilon$ for every $\sigma \in \Sigma$ then we stop and u_{s-1} is an approximation of φ possessing the desired level of accuracy. Otherwise, we choose $k_s \in \mathbb{Z}_{\geq 1}$ and apply the **Banach Extension Step**, with $L' = \Sigma_s$, $u := u_{s-1}$ and $m := k_s$, to obtain a subset $\Sigma_s := \Sigma_{s-1} \cup \{\sigma_{s,1}, \ldots, \sigma_{s,k_s}\} \subset \Sigma$. Apply recombination (cf. Lemma 2.1) to find an element $u \in \mathrm{Span}(\mathcal{F})$ satisfying, for every $\sigma \in \Sigma_s$, that $|\sigma(\varphi - u)| \leq \varepsilon_0$.

We can establish the following convergence theorem for the **Banach GRIM** algorithm with the choice that $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$.

Theorem 2.2 (**Banach GRIM Convergence**). *Assume X is a Banach space with dual-space X^*. Let $\varepsilon > \varepsilon_0 \geq 0$. Let $\mathcal{N} \in \mathbb{Z}_{\geq 1}$ and assume that $\mathcal{F} := \{f_1, \ldots, f_{\mathcal{N}}\} \subset X$ is a collection of non-zero elements in X. Further assume that $\Sigma \subset X^*$ is compact. Let $a_1, \ldots, a_{\mathcal{N}} > 0$ and define $\varphi \in \mathrm{Span}(\mathcal{F})$ and a constant $C > 0$ by*

$$(\mathbf{I}) \quad \varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i \qquad \text{and} \qquad (\mathbf{II}) \quad C := \sum_{i=1}^{\mathcal{N}} |a_i| ||f_i||_X > 0. \tag{2.7}$$

Then, given any $\lambda > 1$, there are positive constants $r_1 = r_1(C, \varepsilon, \varepsilon_0) > 0$ and $r_2 = r_2(C, \varepsilon, \lambda) > 0$, given by

$$(\mathbf{I}) \quad r_1 := \frac{\varepsilon - \varepsilon_0}{2C} > 0 \qquad \text{and} \qquad (\mathbf{II}) \quad r_2 := \frac{(\lambda - 1)\varepsilon}{2C} > 0 \tag{2.8}$$

and a non-negative integer $N = N(\Sigma, C, \varepsilon, \varepsilon_0) \in \mathbb{Z}_{\geq 0}$, given by the r_1-packing number of Σ

$$N := N_{\text{pack}}(\Sigma, X^*, r_1) = \max\left\{d \in \mathbb{Z} \ : \ \exists \sigma_1, \ldots, \sigma_d \in \Sigma \text{ for which } ||\sigma_a - \sigma_c||_{X^*} > r_1 \text{ if } a \neq c\right\}, \tag{2.9}$$

for which the following is true.

*If we apply the **Banach GRIM** algorithm, with the choice that $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, to approximate φ on Σ, with the target accuracy threshold and the acceptable recombination error bound in **Banach GRIM** (A) as ε and ε_0 respectively, then after at most N steps, the algorithm terminates. That is, if we let $M \in \{1, \ldots, N\}$ be the integer for which the algorithm terminates after step M and $Q_M := \min\{\mathcal{N}, 1 + M\}$, there are coefficients $c_{M,1}, \ldots, c_{M,Q_M} \in \mathbb{R}$ and indices $e_M(1), \ldots, e_M(Q_M) \in \{1, \ldots, \mathcal{N}\}$ with*

$$\sum_{s=1}^{Q_M} |c_{M,s}| \left|\left|f_{e_M(s)}\right|\right|_X = C, \tag{2.10}$$

and such that the element $u_M \in \mathrm{Span}(\mathcal{F})$ defined by

$$u_M := \sum_{s=1}^{Q_M} c_{M,s} f_{e_M(s)} \qquad \text{satisfies, for every } \sigma \in \Sigma, \text{ that} \qquad |\sigma(\varphi - u_M)| \leq \varepsilon. \tag{2.11}$$

*Moreover, if the coefficients $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ corresponding to φ (cf. (**I**) of (2.7)) are all positive (i.e. $a_1, \ldots, a_{\mathcal{N}} > 0$) then the coefficients $c_{M,1}, \ldots, c_{M,Q_M} \in \mathbb{R}$ corresponding to u_M (cf. (2.11)) are all non-negative (i.e. $c_1, \ldots, c_{Q_M} \geq 0$). Further, suppose $\Omega \subset \Sigma_{r_2} \subset X^*$ where Σ_{r_2} denotes the r_2-fattening of Σ in X^*, i.e. Σ_{r_2} is defined by $\Sigma_{r_2} := \{\omega \in X^* \ : \ \exists \sigma \in \Sigma \text{ such that } ||\sigma - \omega||_{X^*} \leq r_2\}$. Then φ remains* well-approximated *by u_M throughout Ω in the sense that for every $\sigma \in \Omega$ we have that*

$$|\sigma(\varphi - u_M)| \leq \lambda\varepsilon. \tag{2.12}$$

The remainder of this section is dedicated to proving the *Banach GRIM Convergence* theorem 2.2. The following lemma records that the linear functionals selected during the **Banach GRIM** algorithm, for the choice $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, are a definite X^*-distance apart.

Lemma 2.3 (**Banach GRIM Functional Separation**). *Assume X is a Banach space with dual-space X^*. Suppose that $\theta > \theta_0 \geq 0$. Let $\mathcal{N} \in \mathbb{Z}_{\geq 1}$ and assume that $\mathcal{F} := \{f_1, \ldots, f_{\mathcal{N}}\} \subset X$ is a collection of non-zero elements of X. Further suppose that $\Gamma \subset X^*$ is compact and that $D > 0$ is such that $\Gamma \subset \overline{\mathbb{B}}_{X^*}(0, D)$. Let $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ and define an element $\varphi \in \mathrm{Span}(\mathcal{F})$ and a constant $C > 0$ by*

$$(\mathbf{I}) \quad \varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i \qquad \text{and} \qquad (\mathbf{II}) \quad C := \sum_{i=1}^{\mathcal{N}} |a_i| ||f_i||_X > 0. \tag{2.13}$$

Then there exists a positive constant $r = r(C, \theta, \theta_0) > 0$, given by

$$r := \frac{\theta - \theta_0}{2C} > 0, \tag{2.14}$$

for which the following is true.

Consider applying the ***Banach GRIM*** *algorithm, with $k_s := 1$ for every $s \in \mathbb{Z}_{\geq 1}$, to approximate φ on Γ with θ as the accuracy threshold and θ_0 as the acceptable recombination error bound (cf.* ***Banach GRIM*** *(A)). Suppose that $m \in \mathbb{Z}_{\geq 2}$ and the algorithm reaches and carries out the m^{th} step without terminating. Then the linear functionals $\sigma_1, \ldots, \sigma_m \in \Gamma$ at which recombination ensures that, for each $s \in \{1, \ldots, m\}$, we have $|\sigma_s(\varphi - u_m)| \leq \theta_0$ (cf.* ***Banach GRIM*** *(C) and (D)) satisfy that for every $i, j \in \{1, \ldots, m\}$ with $i \neq j$ we have that*

$$||\sigma_i - \sigma_j||_{X^*} > r. \tag{2.15}$$

Moreover, for every $\sigma \in \Gamma$,

$$|\sigma(\varphi - u_m)| \leq \min\{2CD\ ,\ 2C\,\mathrm{dist}(\sigma, \Gamma_m) + \theta_0\} \tag{2.16}$$

where $\Gamma_m := \{\sigma_1, \ldots, \sigma_m\}$.

Proof of Lemma 2.3. Assume X is a Banach space with dual-space X^*. Let $\theta > \theta_0 \geq 0$ and $\mathcal{N} \in \mathbb{Z}_{\geq 1}$. Assume that $\mathcal{F} := \{f_1, \ldots, f_{\mathcal{N}}\} \subset X$ is a collection of non-zero element in X, and that $\Gamma \subset X^*$ is compact with $D > 0$ such that $\Gamma \subset \overline{\mathbb{B}}_{X^*}(0, D)$. Let $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ and define $\varphi \in \mathrm{Span}(\mathcal{F}) \subset X$ and $C > 0$ by **(I)** and **(II)** of (2.13) respectively.

With a view to later applying **Banach GRIM** to approximate φ on Γ, for each $i \in \{1, \ldots, \mathcal{N}\}$ let $\tilde{a}_i := |a_i|$ and $\tilde{f}_i$ be given by f_i if $a_i > 0$ and $-f_i$ if $a_i < 0$. For every $i \in \{1, \ldots, \mathcal{N}\}$ we evidently have $\left|\left|\tilde{f}_i\right|\right|_X = ||f_i||_X$. Moreover, we also have that $\tilde{a}_1, \ldots, \tilde{a}_{\mathcal{N}} > 0$ and $\varphi = \sum_{i=1}^{\mathcal{N}} \tilde{a}_i \tilde{f}_i$. Further, we rescale $\tilde{f}_i$ for each $i \in \{1, \ldots, \mathcal{N}\}$ to have unit X norm. That is (cf. **Banach GRIM** (B)), for each $i \in \{1, \ldots, \mathcal{N}\}$ set $h_i := \frac{\tilde{f}_i}{||f_i||_X}$ and $\alpha_i := \tilde{a}_i ||f_i||_X$. Then observe both that C satisfies

$$C := \sum_{i=1}^{\mathcal{N}} |a_i| ||f_i||_X = \sum_{i=1}^{\mathcal{N}} \tilde{a}_i ||f_i||_X = \sum_{i=1}^{\mathcal{N}} \alpha_i, \tag{2.17}$$

and, for every $i \in \{1, \ldots, \mathcal{N}\}$, that $\alpha_i h_i = \tilde{a}_i \tilde{f}_i = a_i f_i$. Therefore the expansion for φ in **(I)** of (2.13) is equivalent to

$$\varphi = \sum_{i=1}^{\mathcal{N}} \alpha_i h_i, \qquad \text{and hence} \qquad ||\varphi||_X \leq \sum_{i=1}^{\mathcal{N}} \alpha_i ||h_i||_X = \sum_{i=1}^{\mathcal{N}} \alpha_i \overset{(2.17)}{=} C. \tag{2.18}$$

Define a positive constant $r = r(C, \theta, \theta_0) > 0$ as in (2.14). That is

$$r := \frac{\theta - \theta_0}{2C} > 0. \tag{2.19}$$

With the constant r defined, we turn our attention to verifying the claims of the lemma.

Consider applying the **Banach GRIM** algorithm, with $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, to approximate φ on Γ, with the target accuracy threshold and the acceptable recombination error bound as θ and θ_0 here respectively (cf. **Banach GRIM** (A)). Suppose that $m \in \mathbb{Z}_{\geq 2}$ and that the algorithm reaches and carries out the m^{th} step without terminating. Let $\Gamma_m = \{\sigma_1, \ldots, \sigma_m\} \subset \Gamma$ denote the m linear functionals chosen after the m^{th} step is completed.

Then for every $l \in \{1, \ldots, m\}$, if we let $\Gamma_l := \{\sigma_1, \ldots, \sigma_l\} \subset \Gamma$, we have, recalling **Banach GRIM** (C) and (D), that recombination has been applied via Lemma 2.1 to find an approximation $u_l \in \mathrm{Span}(\mathcal{F}) \subset X$ of φ satisfying, for every $s \in \{1, \ldots, l\}$, that $|\sigma_s(\varphi - u_l)| \leq \theta_0$.

Let $Q_l := \min\{\mathcal{N}, l+1\}$. Then Lemma 2.1 additionally tells us that there are non-negative coefficients $b_{l,1}, \ldots, b_{l,Q_l} \geq 0$ and indices $e_l(1), \ldots, e_l(Q_l) \in \{1, \ldots, \mathcal{N}\}$ for which

$$u_l = \sum_{s=1}^{Q_l} b_{l,s} h_{e_l(s)} \qquad \text{and} \qquad \sum_{s=1}^{Q_l} b_{l,s} = \sum_{i=1}^{\mathcal{N}} \alpha_i. \tag{2.20}$$

A consequence of (2.20) is that

$$||u_l||_X \leq \sum_{s=1}^{Q_l} b_{l,s} \left|\left|h_{e_l(s)}\right|\right|_X = \sum_{s=1}^{Q_l} b_{l,s} \stackrel{(2.20)}{=} \sum_{i=1}^{\mathcal{N}} \alpha_i \stackrel{(2.17)}{=} C. \tag{2.21}$$

Suppose $D > 0$ is such that $\Gamma \subset \overline{\mathbb{B}}_{X^*}(0, D)$. Then given any $\sigma \in \Gamma$ we may compute

$$|\sigma(\varphi - u_l)| \leq ||\sigma||_{X^*} ||\varphi - u_l||_X \stackrel{(2.18)\ \&\ (2.21)}{\leq} 2CD. \tag{2.22}$$

Alternatively, we may consider any $n \in \{1, \ldots, l\}$ and use that $|\sigma_n(\varphi - u_l)| \leq \theta_0$ to compute that

$$|\sigma(\varphi - u_l)| \leq ||\sigma - \sigma_n||_{X^*} ||\varphi - u_l||_X + \theta_0 \stackrel{(2.18)\ \&\ (2.21)}{\leq} 2C||\sigma - \sigma_n||_{X^*} + \theta_0. \tag{2.23}$$

Taking the minimum over $n \in \{1, \ldots, l\}$ in (2.23) yields

$$|\sigma(\varphi - u_l)| \leq 2C \operatorname{dist}(\sigma, \Gamma_l) + \theta_0. \tag{2.24}$$

Together, (2.22) and (2.24) yield that for any $\sigma \in \Gamma$ we have

$$|\sigma(\varphi - u_l)| \leq \min \left\{ 2CD \ , \ 2C \operatorname{dist}(\sigma, \Gamma_l) + \theta_0 \right\}. \tag{2.25}$$

The choice of $l := m$ in (2.25) yields (2.16) as claimed.

To establish the separation of the linear functionals $\sigma_1, \ldots, \sigma_m$ making up Γ_m claimed in (2.15), we take $i, j \in \{1, \ldots, m\}$ such that $i \neq j$. Without loss of generality we may assume $i < j$. Then by assumption $\sigma_j = \operatorname{argmax}_{\sigma \in \Gamma} |\sigma(\varphi - u_{j-1})|$ and, since $j \leq m$ means that the **Banach GRIM** algorithm does not terminate before the j^{th} step is carried out, we must have that

$$|\sigma_j(\varphi - u_{j-1})| > \theta. \tag{2.26}$$

Hence by considering $l := j - 1$ and $\sigma := \sigma_j$ in (2.24) we observe that

$$\theta \stackrel{(2.26)}{<} |\sigma_j(\varphi - u_{j-1})| \stackrel{(2.24)}{\leq} 2C \operatorname{dist}(\sigma_j, \Gamma_{j-1}) + \theta_0 \leq 2C||\sigma_j - \sigma_i||_{X^*} + \theta_0, \tag{2.27}$$

where the last inequality holds since $i \leq j - 1$ means that $\sigma_i \in \Gamma_{j-1}$. A direct consequence of (2.27) is that

$$||\sigma_j - \sigma_i||_{X^*} > \frac{\theta - \theta_0}{2C} \stackrel{(2.19)}{=} r, \tag{2.28}$$

which is precisely the separation claimed in (2.15) This completes the proof of Lemma 2.3. ■

We can use Lemma 2.3 to establish an upper bound on the number of steps the **Banach GRIM** algorithm, for the choice that $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, requires to find an approximation $u \in \operatorname{Span}(\mathcal{F})$ that is *close* to φ for every linear functional in Σ. The precise statement is the following lemma.

Lemma 2.4 (Banach GRIM Number of Steps Bound). *Assume that X is a Banach space with dual-space X^*. Let $\theta > \theta_0 \geq 0$. Let $\mathcal{N} \in \mathbb{Z}_{\geq 1}$ and assume that $\mathcal{F} = \{f_1, \ldots, f_{\mathcal{N}}\} \subset X$ is a collection of non-zero elements in X. Further suppose that $\Gamma \subset X^*$ is compact. Let $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ and define an element $\varphi \in \operatorname{Span}(\mathcal{F})$ and a constant $C > 0$ by*

$$\textbf{(I)} \quad \varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i \qquad \textit{and} \qquad \textbf{(II)} \quad C := \sum_{i=1}^{\mathcal{N}} |a_i| ||f_i||_X > 0. \tag{2.29}$$

Then there is a positive constant $r = r(C, \theta, \theta_0) > 0$, given by

$$r := \frac{\theta - \theta_0}{2C} > 0, \tag{2.30}$$

and a non-negative integer $N = N(\Gamma, C, \theta, \theta_0) \in \mathbb{Z}_{\geq 0}$, given by the r-packing number of Γ

$$N := N_{\text{pack}}(\Gamma, X^*, r) = \max\left\{d \in \mathbb{Z} \;:\; \exists \sigma_1, \ldots, \sigma_d \in \Gamma \text{ such that } ||\sigma_a - \sigma_c||_{X^*} > r \text{ if } a \neq c\right\}, \tag{2.31}$$

for which the following is true.

*If we apply the **Banach GRIM** algorithm, with the choice that $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, to approximate φ on Γ, with the target accuracy threshold and the acceptable recombination error bound in **Banach GRIM** (A) as θ and θ_0 respectively, then after at most N steps the algorithm terminates. That is, if we let $M \in \{1, \ldots, N\}$ be the integer for which the algorithm terminates after step M and $Q_M := \min\{\mathcal{N}, M+1\}$, there are coefficients $c_1, \ldots, c_{Q_M} \in \mathbb{R}$ and indices $e(1), \ldots, e(Q_M) \in \{1, \ldots, \mathcal{N}\}$ with*

$$\sum_{s=1}^{Q_M} |c_s|\, ||f_{e(s)}||_X = C, \tag{2.32}$$

and such that the element $u_M \in \operatorname{Span}(\mathcal{F})$ defined by

$$u_M := \sum_{s=1}^{Q_M} c_s f_{e(s)} \qquad \text{satisfies, for every } \sigma \in \Gamma, \text{ that} \qquad |\sigma(\varphi - u_M)| \leq \theta. \tag{2.33}$$

*Moreover, if the coefficients $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ corresponding to φ (cf. (**I**) of (2.29)) are all positive (i.e. $a_1, \ldots, a_{\mathcal{N}} > 0$) then the coefficients $c_{M,1}, \ldots, c_{M,Q_M} \in \mathbb{R}$ corresponding to u_M (cf. (2.33)) are all non-negative (i.e. $c_1, \ldots, c_{Q_M} \geq 0$).*

Remark 2.5. Lemma 2.4 guarantees that if N defined in (2.31) satisfies that $N < \mathcal{N} - 1$ then the **Banach GRIM** algorithm will find an approximation $u \in \operatorname{Span}(\mathcal{F})$ of φ that is a linear combination of *less* than $\mathcal{N}$ of the elements $f_1, \ldots, f_{\mathcal{N}}$ but is within θ of φ throughout Γ in the sense that $|\sigma(\varphi - u)| \leq \theta$ for every $\sigma \in \Gamma$.

Remark 2.6. By invoking Lemma 2.3, we can additionally conclude that, for every $m \in \{1, \ldots, M\}$ and any $D > 0$ for which $\Gamma \subset \overline{\mathbb{B}}_{X^*}(0, D)$, the approximation $u_m \in \operatorname{Span}(\mathcal{F})$ found at the m^{th} step satisfies that for every $\sigma \in \Gamma$ that $|\sigma(\varphi - u_m)| \leq \min\{2CD, \theta_0 + 2C \operatorname{dist}(\sigma, \Gamma_m)\}$ where Γ_m denotes the m-linear functionals in Γ that have been selected once the m^{th} step is complete.

Proof of Lemma 2.4. Assume that X is a Banach space with dual-space X^*. Let $\theta > \theta_0 \geq 0$. Let $\mathcal{N} \in \mathbb{Z}_{\geq 1}$ and assume that $\mathcal{F} = \{f_1, \ldots, f_{\mathcal{N}}\} \subset X$ is a collection of non-zero elements in X. Further suppose that $\Gamma \subset X^*$ is compact. Let $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ and define an element $\varphi \in \operatorname{Span}(\mathcal{F})$ and a constant $C > 0$ by (**I**) and (**II**) of (2.29) respectively.

With a view to later applying **Banach GRIM** to approximate φ on Γ, for each $i \in \{1, \ldots, \mathcal{N}\}$ let $\tilde{a}_i := |a_i|$ and $\tilde{f}_i$ be given by f_i if $a_i > 0$ and $-f_i$ if $a_i < 0$. For every $i \in \{1, \ldots, \mathcal{N}\}$ we evidently have $\left|\left|\tilde{f}_i\right|\right|_X = ||f_i||_X$. Moreover, we also have that $\tilde{a}_1, \ldots, \tilde{a}_{\mathcal{N}} > 0$ and $\varphi = \sum_{i=1}^{\mathcal{N}} \tilde{a}_i \tilde{f}_i$. Further, we rescale $\tilde{f}_i$ for each $i \in \{1, \ldots, \mathcal{N}\}$ to have unit X norm. That is (cf. **Banach GRIM** (B)), for each $i \in \{1, \ldots, \mathcal{N}\}$ set $h_i := \frac{\tilde{f}_i}{||f_i||_X}$ and $\alpha_i := \tilde{a}_i ||f_i||_X$. Observe both that C satisfies

$$C := \sum_{i=1}^{\mathcal{N}} |a_i| ||f_i||_X = \sum_{i=1}^{\mathcal{N}} \tilde{a}_i ||f_i||_X = \sum_{i=1}^{\mathcal{N}} \alpha_i, \tag{2.34}$$

and, for every $i \in \{1, \ldots, \mathcal{N}\}$, that $\alpha_i h_i = \tilde{a}_i \tilde{f}_i = a_i f_i$. Therefore the expansion for φ in (**I**) of (2.13) is equivalent to

$$\varphi = \sum_{i=1}^{\mathcal{N}} \alpha_i h_i, \qquad \text{and hence} \qquad ||\varphi||_X \leq \sum_{i=1}^{\mathcal{N}} \alpha_i ||h_i||_X = \sum_{i=1}^{\mathcal{N}} \alpha_i \overset{(2.34)}{=} C. \tag{2.35}$$

Define a positive constant $r = r(C, \theta, \theta_0) > 0$ by

$$r := \frac{\theta - \theta_0}{2C} > 0, \tag{2.36}$$

and a non-negative integer $N = N(\Gamma, C, \theta, \theta_0) \in \mathbb{Z}_{\geq 0}$ by the r-packing number of Γ

$$N := N_{\text{pack}}(\Gamma, X^*, r) = \max \left\{ d \in \mathbb{Z} \; : \; \exists \sigma_1, \ldots, \sigma_d \in \Gamma \text{ such that } ||\sigma_a - \sigma_c||_{X^*} > r \text{ if } a \neq c \right\}. \tag{2.37}$$

Observe that the constant $r > 0$ defined in (2.36) agrees with the constant r arising in Lemma 2.3. Now consider applying the **Banach GRIM** algorithm, with the choice that $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, to approximate φ on Γ with the target accuracy threshold and the acceptable recombination error bound in **Banach GRIM** (A) as θ and θ_0 respectively.

Suppose $m \in \mathbb{Z}_{\geq 1}$ and that the algorithm does not terminate before or at the m^{th} step. Therefore, recalling **Banach GRIM** (C) and (D), distinct linear functionals $\sigma_1, \ldots, \sigma_m \in \Gamma$ have been selected, and recombination has been used via Lemma 2.1 to find an approximation $u_m \in \text{Span}(\mathcal{F}) \subset X$ of φ satisfying, for every $s \in \{1, \ldots, m\}$, that $|\sigma_s(\varphi - u_m)| \leq \theta_0$. Moreover, we may appeal to Lemma 2.3 to deduce that whenever $s, t \in \{1, \ldots, m\}$ with $s \neq t$ we have

$$||\sigma_s - \sigma_t||_{X^*} > r. \tag{2.38}$$

A consequence of (2.38) is that N defined in (2.37) must satisfy that $N \geq m$. Therefore if the **Banach GRIM** algorithm, with the choice that $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, does not terminate at the m^{th} step we must have that $m \leq N$, and so the algorithm terminates after at most N steps as claimed.

Let $M \in \{1, \ldots, N\}$ be the integer for which the **Banach GRIM** algorithm terminates after step M. The termination criterion (cf. **Banach GRIM** (D)) means that the element $u_M \in \text{Span}(\mathcal{F}) \subset X$ found at the M^{th} step must satisfy, for every $\sigma \in \Gamma$, that $|\sigma(\varphi - u_M)| \leq \theta$ as claimed in the second part of (2.33). Moreover, if we let $Q_M := \min \{\mathcal{N}, M + 1\}$, then Lemma 2.1 tells us that there are non-negative coefficients $b_{M,1}, \ldots, b_{M,Q_M} \geq 0$, with

$$\sum_{s=1}^{Q_M} b_{M,s} = \sum_{i=1}^{\mathcal{N}} \alpha_i \overset{(2.34)}{=} C, \tag{2.39}$$

and indices $e_M(1), \ldots, e_M(Q_M) \in \{1, \ldots, \mathcal{N}\}$ for which the element $u_M \in \text{Span}(\mathcal{F})$ is given by

$$u_M = \sum_{s=1}^{Q_M} b_{M,s} h_{e_M(s)} = \sum_{s=1}^{Q_M} \frac{b_{M,s}}{||f_{e_M(s)}||_X} \tilde{f}_{e_M(s)}. \tag{2.40}$$

For each $s \in \{1, \ldots, Q_M\}$, we define $c_{M,s} := \frac{b_{M,s}}{||f_{e_M(s)}||_X}$ if $\tilde{f}_{e_M(s)} = f_{e_M(s)}$ (which we recall is the case if $a_{e_M(s)} > 0$) and $c_{M,s} := -\frac{b_{M,s}}{||f_{e_M(s)}||_X}$ if $\tilde{f}_{e_M(s)} = -f_{e_M(s)}$ (which we recall is the case if $a_{e_M(s)} < 0$). Then (2.40) gives the expansion for $u_M \in \text{Span}(\mathcal{F}) \subset X$ in terms of the elements $f_1, \ldots, f_{\mathcal{N}}$ claimed in the first part of (2.33). Moreover, from (2.39) we have that

$$\sum_{s=1}^{Q_M} |c_{M,s}| \, ||f_{e_M(s)}||_X = \sum_{s=1}^{Q_M} b_{M,s} \overset{(2.39)}{=} C \tag{2.41}$$

as claimed in (2.32).

It remains only to prove that if the coefficients $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ are all positive (i.e. $a_1, \ldots, a_{\mathcal{N}} > 0$), then the resulting coefficients $c_{M,1}, \ldots, c_{M,Q_M} \in \mathbb{R}$ are all non-negative (i.e. $c_{M,1}, \ldots, c_{M,Q_M} \geq 0$). To see this, observe that if $a_1, \ldots, a_{\mathcal{N}} > 0$ then, for every $i \in \{1, \ldots, \mathcal{N}\}$, we have that $\tilde{f}_i = f_i$. Consequently, for every $s \in \{1, \ldots, Q_M\}$ we have that $\tilde{f}_{e_M(s)} = f_{e_M(s)}$, and so by definition we have $c_{M,s} = \frac{b_{M,s}}{||f_{e_M(s)}||_X}$. Since $b_{M,s} \geq 0$, it follows that $c_{M,s} \geq 0$. This completes the proof of Lemma 2.4. ■

The following result records a quantified version of the statement that an approximation of φ on Γ, of the form found by the **Banach GRIM** algorithm, remains a good approximation of φ on an enlargement of the set Γ. The precise result is the following lemma.

Lemma 2.7 (Banach GRIM Robustness). *Assume that X is a Banach space with dual-space X^*. Let $\theta > \theta_0 \geq 0$. Let $\mathcal{N} \in \mathbb{Z}_{\geq 1}$ and assume that $\mathcal{F} = \{f_1, \ldots, f_{\mathcal{N}}\} \subset X$ is a collection of non-zero elements in X. Further suppose* *that* $\Gamma \subset X^*$ *is compact. Let* $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ *and define an element* $\varphi \in \operatorname{Span}(\mathcal{F})$ *and a constant* $C > 0$ *by*

$$\textbf{(I)} \quad \varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i \qquad \textit{and} \qquad \textbf{(II)} \quad C := \sum_{i=1}^{\mathcal{N}} |a_i| ||f_i||_X > 0. \tag{2.42}$$

Then there exists a positive constant $r = r(C, \theta, \theta_0) > 0$*, given by*

$$r := \frac{\theta - \theta_0}{2C} > 0, \tag{2.43}$$

for which the following is true.

Suppose that $u \in \operatorname{Span}(\mathcal{F})$ *satisfies* $||u||_X \leq C$ *and, for every* $\sigma \in \Gamma$ *we have* $|\sigma(\varphi - u)| \leq \theta_0$*. Then for every* $\sigma \in \Gamma_r$*, where* $\Gamma_r := \{\sigma \in X^* \ : \ \exists \omega \in \Gamma$ *such that* $||\sigma - \omega||_{X^*} \leq r\}$ *is the* r*-fattening of* Γ*, we have that*

$$|\sigma(\varphi - u)| \leq \theta. \tag{2.44}$$

Proof of Lemma 2.7. Assume that X is a Banach space with dual-space X^*. Let $\theta > \theta_0 \geq 0$. Let $\mathcal{N} \in \mathbb{Z}_{\geq 1}$ and assume that $\mathcal{F} = \{f_1, \ldots, f_{\mathcal{N}}\} \subset X$ is a collection of non-zero elements in X. Further suppose that $\Gamma \subset X^*$. Let $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ and define an element $\varphi \in \operatorname{Span}(\mathcal{F})$ and a constant $C > 0$ by **(I)** and **(II)** in (2.42) respectively. Now define $r = r(C, \theta, \theta_0) > 0$ by

$$r := \frac{\theta - \theta_0}{2C} > 0. \tag{2.45}$$

With the constant r defined we turn to verifying that the claims of the lemma are valid.

Suppose that $u \in \operatorname{Span}(\mathcal{F})$ with $||u||_X \leq C$ and such that for every $\sigma \in \Gamma$ we have $|\sigma(\varphi - u)| \leq \theta_0$. Observe that the definition of φ and the constant C (cf. **(I)** and **(II)** of (2.42)) allow us to deduce that $||\varphi||_X \leq C$. Now consider $\sigma \in \Gamma_r$ for Γ_r denoting the r-fattening of Γ defined by $\Gamma_r := \{\sigma \in X^* \ : \ \exists \omega \in \Gamma$ such that $||\sigma - \omega||_{X^*} \leq r\}$. Then for any choice of $\omega \in \Gamma$ we have

$$|\sigma(\varphi - u)| \leq |(\sigma - \omega)(\varphi - u)| + |\omega(\varphi - u)| \leq ||\sigma - \omega||_{X^*} ||\varphi - u||_X + \theta_0 \leq 2C ||\sigma - \omega||_{X^*} + \theta_0. \tag{2.46}$$

Since $\sigma \in \Gamma_r$ we know that we may choose $\omega \in \Gamma$ such that $||\sigma - \omega||_{X^*} \leq r$. By making such a choice for ω in (2.46), we see that $|\sigma(\varphi - u)| \leq 2Cr + \theta_0 \overset{(2.45)}{=} \theta$ as claimed in (2.44). This completes the proof of Lemma 2.7. ■

We now prove the *Banach GRIM Convergence* theorem by combining Lemmas 2.3, 2.4, and 2.7.

Proof of Theorem 2.2. Assume X is a Banach space with dual-space X^*. Let $\varepsilon > \varepsilon_0 \geq 0$. Let $\mathcal{N} \in \mathbb{Z}_{\geq 1}$ and assume that $\mathcal{F} := \{f_1, \ldots, f_{\mathcal{N}}\} \subset X$ is a collection of non-zero elements in X. Further assume that $\Sigma \subset X^*$ is compact. Let $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ and define $\varphi \in \operatorname{Span}(\mathcal{F})$ and a constant $C > 0$ by

$$\textbf{(I)} \quad \varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i \qquad \text{and} \qquad \textbf{(II)} \quad C := \sum_{i=1}^{\mathcal{N}} |a_i| ||f_i||_X > 0. \tag{2.47}$$

Let $\lambda > 1$ and define positive constants $r_1 = r_1(C, \varepsilon, \varepsilon_0) > 0$ and $r_2 = r_2(C, \varepsilon, \lambda) > 0$ by

$$\textbf{(I)} \quad r_1 := \frac{\varepsilon - \varepsilon_0}{2C} > 0 \qquad \text{and} \qquad \textbf{(II)} \quad r_2 := \frac{(\lambda - 1)\varepsilon}{2C} > 0. \tag{2.48}$$

Further define a non-negative integer $N = N(\Sigma, C, \varepsilon, \varepsilon_0) \in \mathbb{Z}_{\geq 0}$ to be the r_1-packing number of Σ

$$N := \max\{d \in \mathbb{Z} \ : \ \exists \sigma_1, \ldots, \sigma_d \in \Sigma \text{ for which } ||\sigma_a - \sigma_c||_{X^*} > r_1 \text{ if } a \neq c\}. \tag{2.49}$$

Now consider applying the **Banach GRIM** algorithm, with the choice that $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, to approximate φ on Σ, with the target accuracy threshold and the acceptable recombination error bound in **Banach GRIM** (A) as ε and ε_0 respectively. Since the constant r_1 defined in **(I)** of (2.48) and the integer N defined in (2.49) agree with the definition of the constant r and the integer N arising in Lemma 2.4, with $\Gamma := \Sigma$, $C := C$, $\theta := \varepsilon$ and $\theta_0 := \varepsilon_0$, we may apply Lemma 2.4 to conclude that the algorithm terminates after M steps for some integer $M \in \mathbb{Z}_{\geq 1}$ satisfying $M \leq N$ as claimed. Additionally, if $Q_M := \min\{\mathcal{N}, M+1\}$, then Lemma 2.4 tells us there are coefficients $c_{M,1}, \ldots, c_{M,Q_M} \geq 0$ and indices $e_M(1), \ldots, e_M(Q_M) \in \{1, \ldots, \mathcal{N}\}$ with

$$\sum_{s=1}^{Q_M} |c_{M,s}| \, ||f_{e_M(s)}||_X = C, \tag{2.50}$$

and such that the element $u_M \in \mathrm{Span}(\mathcal{F}) \subset X$ defined by

$$u_M := \sum_{s=1}^{Q_M} c_{M,s} f_{e_M(s)} \qquad \text{satisfies, for every } \sigma \in \Sigma, \text{ that} \qquad |\sigma(\varphi - u_M)| \leq \varepsilon. \tag{2.51}$$

Together (2.50) and (2.51) establish the claims (2.10) and (2.11) respectively.

Lemma 2.4 additionally establishes that if the coefficients $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ associated to φ (cf. (**I**) of (2.47)) are all positive (i.e. $a_1, \ldots, a_{\mathcal{N}} > 0$) then the coefficients $c_{M,1}, \ldots, c_{M,Q_M} \in \mathbb{R}$ associated with u_M (cf. (2.51)) are all non-negative (i.e. $c_{M,1}, \ldots, c_{M,Q_M} \geq 0$, which is precisely as claimed in Theorem 2.2.

Now suppose $\Omega \subset \Sigma_{r_2} \subset X^*$ where $\Sigma_{r_2} := \{\omega \in X^* : \exists \sigma \in \Sigma \text{ such that } ||\sigma - \omega||_{X^*} \leq r_2\}$ is the r_2-fattening of Σ in X^*. Since the constant r_2 defined in (**II**) of (2.48) coincides with the constant r arising in Lemma 2.7 for the choices $C := C$, $\theta_0 := \varepsilon$ and $\theta := \lambda\varepsilon$, we may apply Lemma 2.7 to conclude, for every $\sigma \in \Sigma_{r_2}$, that $|\sigma(\varphi - u_M)| \leq \lambda\varepsilon$. Since $\Omega \subset \Sigma_{r_2}$ this establishes (2.12) and completes the proof of Theorem 2.2. ■

3. Lipschitz GRIM

Consider Banach spaces V and W and a compact subset $\Sigma \subset V$. Assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1 in Appendix A). Let $\gamma > 0$ with $k \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k, k+1]$. In this section we consider trying to find a sparse approximation of a given linear combination of $\mathrm{Lip}(\gamma, \Sigma, W)$ functions (cf. Definition A.2 in Appendix A).

To make this more precise, assume that $\mathcal{N} \in \mathbb{Z}$ is a (large) integer. For every $i \in \{1, \ldots, \mathcal{N}\}$ we assume that $f_i = \left(f_i^{(0)}, \ldots, f_i^{(k)}\right) \in \mathrm{Lip}(\gamma, \Sigma, W)$ is non-zero, and set $\mathcal{F} := \{f_i : i \in \{1, \ldots, \mathcal{N}\}\} \subset \mathrm{Lip}(\gamma, \Sigma, W)$. Given $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$, define $\varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(k)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\gamma, \Sigma, W)$ by

$$\varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i \quad \text{so that for every } l \in \{0, \ldots, k\} \text{ we have} \quad \varphi^{(l)} = \sum_{i=1}^{\mathcal{N}} a_i f_i^{(l)}. \tag{3.1}$$

We aim to find an element $u = \left(u^{(0)}, \ldots, u^{(k)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\gamma, \Sigma, W)$ that is both a sparse linear combination of the elements $f_1, \ldots, f_{\mathcal{N}}$, and gives a good approximation of φ in the following sense. Given a fixed $l \in \{0, \ldots, k\}$, we require that for every $s \in \{0, \ldots, l\}$ the functions $u^{(s)}, \varphi^{(s)} : \Sigma \to \mathcal{L}(V^{\otimes s}; W)$ are close in $\mathcal{L}(V^{\otimes s}; W)$ throughout Σ. That is, for any given $x \in \Sigma$ the norm $\left|\left|\varphi^{(s)}(x) - u^{(s)}(x)\right|\right|_{\mathcal{L}(V^{\otimes s}; W)}$ is small.

To ease notation, given $\phi = \left(\phi^{(0)}, \ldots, \phi^{(k)}\right) \in \mathrm{Lip}(\gamma, \Sigma, W)$, $l \in \{0, \ldots, k\}$, and $x \in \Sigma$ we define

$$\Lambda^l_\phi(x) := \max_{j \in \{0, \ldots, l\}} \left\{ \left|\left|\phi^{(j)}(x)\right|\right|_{\mathcal{L}(V^{\otimes j}; W)} \right\} \tag{3.2}$$

A consequence of (3.2) is that for any $l \in \{0, \ldots, k\}$ and any $x \in \Sigma$ we have that $\Lambda^l_\phi(x) \leq ||\phi||_{\mathrm{Lip}(\gamma, \mathcal{A}, W)}$ for any subset $\mathcal{A} \subset \Sigma$ with $x \in \mathcal{A}$. Our problem of interest can then be stated as follows. Given $\varepsilon > 0$ and $q \in \{0, \ldots, k\}$, find an element $u = \left(u^{(0)}, \ldots, u^{(k)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\gamma, \Sigma, W)$ that is a sparse linear combination of the elements $f_1, \ldots, f_{\mathcal{N}}$ and satisfies, for every $x \in \Sigma$, that $\Lambda^q_{\varphi - u}(x) \leq \varepsilon$.

The choice of $q \in \{0, \ldots, k\}$ provides a gauge for the strength of approximation we seek. By examination of (3.2), the choice $q := 0$ means we aim only to approximate φ by u at the "order zero level" by requiring $||\varphi^{(0)} - u^{(0)}||_{C^0(\Sigma; W)} \leq \varepsilon$. By examination of (3.2), the choice $q := k$ corresponds to u approximating φ at the "order k level" by requiring, for every $s \in \{0, \ldots, k\}$ and every $x \in \Sigma$, that $\left|\left|\varphi^{(s)}(x) - u^{(s)}(x)\right|\right|_{\mathcal{L}(V^{\otimes s};W)} \leq \varepsilon$.

To find our desired approximation u of φ, we use a combination of dynamic growth driven interpolation techniques and thinning techniques. A loose summary of our approach is as follows. We use a greedy selection algorithm to inductively grow a collection of points $L \subset \Sigma$ at which we want our approximation to agree with φ. At each step the thinning of recombination is used to find an element $u = \left(u^{(0)}, \ldots, u^{(k)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\gamma, \Sigma, W)$ that coincides with φ throughout the current choice of subset $L \subset \Sigma$.

We now provide the rigorous details of our algorithm. The following **Lip(γ) Extension Step** is used to dynamically grow the collection of points $L \subset \Sigma$ at which we require our next approximation of φ to agree with φ.

Lip(γ) Extension Step

Assume that $L' \subset \Sigma$, $u = \left(u^{(0)}, \ldots, u^{(k)}\right) \in \mathrm{Span}(\mathcal{F})$ and $q \in \{0, \ldots, k\}$ are given. Let $m \in \mathbb{Z}_{\geq 1}$ and take

$$z_1 := \mathrm{argmax}\left\{\Lambda^q_{\varphi - u}(z) \;:\; z \in \Sigma\right\}. \tag{3.3}$$

Inductively for $j = 2, 3, \ldots, m$ take

$$z_j := \mathrm{argmax}\left\{\Lambda^q_{\varphi - u}(z) \;:\; z \in \Sigma \setminus \{z_1, \ldots, z_{j-1}\}\right\}. \tag{3.4}$$

Once $z_1, \ldots, z_m \in \Sigma$ have been defined, we extend L' to $L := L' \cup \{z_1, \ldots, z_m\}$.

For each choice of subset $L \subset \Sigma$, we use recombination to find $u = \left(u^{(0)}, \ldots, u^{(k)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\gamma, \Sigma, W)$ agreeing with φ throughout L. That this is possible is the content of the following result.

Lemma 3.1 ($\mathrm{Lip}(\gamma, \Sigma, W)$ Recombination). *Assume that V and W are finite dimensional Banach spaces. Suppose that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1), and that $\Sigma \subset V$ is closed. Let $\mathcal{N} \in \mathbb{Z}_{>0}$ be a (large) positive integer and $\gamma > 0$ with $k \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k, k+1]$. Assume that for every $i \in \{1, \ldots, \mathcal{N}\}$ we have a non-zero element $f_i = \left(f_i^{(0)}, \ldots, f_i^{(k)}\right) \in \mathrm{Lip}(\gamma, \Sigma, W)$ and define $\mathcal{F} := \{f_i \,:\, i \in \{1, \ldots, \mathcal{N}\}\} \subset \mathrm{Lip}(\gamma, \Sigma, W)$. Given $a_1, \ldots, a_{\mathcal{N}} > 0$, define $\varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(k)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\gamma, \Sigma, W)$ by*

$$\varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i \quad \text{so that for every } j \in \{0, \ldots, k\} \text{ we have} \quad \varphi^{(j)} = \sum_{i=1}^{\mathcal{N}} a_i f_i^{(j)}. \tag{3.5}$$

Further define both

$$Q := 1 + m \dim(W) \sum_{s=0}^{k} \binom{\dim(V) + s - 1}{s} \qquad \text{and} \qquad \mathcal{M} := \min\{Q, \mathcal{N}\}. \tag{3.6}$$

Then given points $p_1, \ldots, p_m \in \Sigma$, recombination can find $u = \left(u^{(0)}, \ldots, u^{(k)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\gamma, \Sigma, W)$ satisfying the following conditions. Firstly

$$u = \sum_{s=1}^{\mathcal{M}} b_s f_{e(s)}, \quad \text{so that for every } j \in \{0, \ldots, k\} \text{ we have} \quad u^{(j)} = \sum_{s=1}^{\mathcal{M}} b_s f_{e(s)}^{(j)}, \tag{3.7}$$

for coefficients $b_1, \ldots, b_{\mathcal{M}} \geq 0$ and indices $e(1), \ldots, e(\mathcal{M}) \in \{1, \ldots, \mathcal{N}\}$. Secondly, the sum of the coefficients is preserved in the sense that

$$\sum_{s=1}^{\mathcal{M}} b_s = \sum_{i=1}^{\mathcal{N}} a_i. \tag{3.8}$$

Thirdly, $u \equiv \varphi$ throughout $\{p_1, \ldots, p_m\}$. That is, for every $a \in \{1, \ldots, m\}$ and every $j \in \{0, \ldots, k\}$ we have that $\varphi^{(j)}(p_a) \equiv u^{(j)}(p_a)$ in $\mathcal{L}(V^{\otimes j}; W)$.

Proof of Lemma 3.1. Assume that V and W are Banach spaces with $d, c \in \mathbb{Z}_{\geq 1}$ for which $\dim(V) = d$ and $\dim(W) = c$. Suppose that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1), and that $\Sigma \subset V$ is closed. Let $\mathcal{N} \in \mathbb{Z}_{>0}$ be a (large) positive integer and $\gamma > 0$ with $k \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k, k+1]$. Assume that for every $i \in \{1, \ldots, \mathcal{N}\}$ we have a non-zero $f_i = \left(f_i^{(0)}, \ldots, f_i^{(k)}\right) \in \mathrm{Lip}(\gamma, \Sigma, W)$ such that $\mathcal{F} := \{f_i : i \in \{1, \ldots, \mathcal{N}\}\}$ is a linearly independent subset of $\mathrm{Lip}(\gamma, \Sigma, W)$. Let $a_1, \ldots, a_{\mathcal{N}} > 0$ and define $\varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(k)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\gamma, \Sigma, W)$ by

$$\varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i \qquad \text{so that for every } j \in \{0, \ldots, k\} \text{ we have} \qquad \varphi^{(j)} = \sum_{i=1}^{\mathcal{N}} a_i f_i^{(j)}. \tag{3.9}$$

Choose bases $v_1, \ldots, v_d$ of V and $w_1, \ldots, w_c$ of W. Then for each $n \in \mathbb{N}$ the set

$$\mathcal{V}_n := \{v_{l_1} \otimes \ldots \otimes v_{l_n} : (l_1, \ldots, l_n) \in \{1, \ldots, d\}^n\} \tag{3.10}$$

is a basis for $V^{\otimes n}$. For each $j \in \{1, \ldots, k\}$ a symmetric j-linear form $h^{(j)} \in \mathcal{L}(V^{\otimes j}; W)$ is determined by its action on the subset $\mathcal{V}_j^{\mathrm{ord}} \subset \mathcal{V}_j$ defined by

$$\mathcal{V}_j^{\mathrm{ord}} := \left\{v_{l_1} \otimes \ldots \otimes v_{l_j} : (l_1, \ldots, l_j) \in \{1, \ldots, d\}^j \text{ such that } l_1 \leq \ldots \leq l_j\right\} \subset \mathcal{V}_j. \tag{3.11}$$

That is, the collection of values $h^{(j)}(p)\left[v_{l_1} \otimes \ldots \otimes v_{l_j}\right]$ for every $(l_1, \ldots, l_j) \in \{1, \ldots, d\}^j$ with $l_1 \leq \ldots \leq l_j$ determine the value $h^{(j)}(p)[v]$ for every $v \in V^{\otimes j}$. For future use we record that the cardinality of $\mathcal{V}_j^{\mathrm{ord}}$ is given by $\beta(d, j) := \begin{pmatrix} d+j-1 \\ j \end{pmatrix} = \frac{(d+j-1)!}{(d-1)!j!}$.

The basis $w_1, \ldots w_c$ of W allows us to define an inner product $\langle \cdot, \cdot \rangle_W$ on W. This is done as follows. Given $a, b \in \{1, \ldots, c\}$ we set $\langle w_a, w_b \rangle_W := 1$ if $a = b$ and $\langle w_a, w_b \rangle_W := 0$ if $a \neq b$, before subsequently extending $\langle \cdot, \cdot \rangle_W$ to the entirety of W via linearity. This inner product allows us to decompose any $w \in W$ as $w = \langle w, w_1 \rangle_W w_1 + \ldots + \langle w, w_c \rangle_W w_c$. Therefore for any $p \in \Sigma$ the value $\varphi^{(0)}(p) \in W$ is determined by the c real-numbers

$$\{\langle \varphi(p), w_r \rangle_W : r \in \{1, \ldots, c\}\}, \tag{3.12}$$

whilst for each $j \in \{1, \ldots, k\}$ the symmetric j-linear form $\varphi^{(j)}(p)$ is determined by the $\beta(d, j)c$ real numbers

$$\left\{\left\langle \varphi^{(j)}(p)[v_{l_1} \otimes \ldots \otimes v_{l_j}], w_r \right\rangle_W : (l_1, \ldots, l_j) \in \{1, \ldots, d\}^j,\ l_1 \leq \ldots \leq l_j,\ r \in \{1, \ldots, c\}\right\}. \tag{3.13}$$

Together, (3.12) and (3.13) illustrate that matching the value of $\varphi^{(0)}(p)$ and the maps $\varphi^{(1)}(p), \ldots, \varphi^{(k)}(p)$ is equivalent to solving $c(1 + \beta(d, 1) + \ldots + \beta(d, k))$ real-valued equations. Hence for points $p_1, \ldots, p_m \in \Sigma$, the values $\varphi(p_1), \ldots, \varphi(p_m) \in W$ and the maps $\varphi^{(j)}(p_1), \ldots, \varphi^{(j)}(p_m) \in \mathcal{L}(V^{\otimes j}; W)$ for each $j \in \{1, \ldots, k\}$ determine a linear system of $mc(1 + \beta(d, 1) + \ldots + \beta(d, k)) \overset{(3.6)}{=} Q - 1$ real-valued equations involving the functions f_i and the coefficients a_i for $i \in \{1, \ldots, \mathcal{N}\}$. To be more precise, for each choice of $s \in \{1, \ldots, m\}$, $j \in \{1, \ldots, k\}$, $(l_1, \ldots, l_j) \in \{1, \ldots, d\}^j$ with $l_1 \leq \ldots \leq l_j$ and $r \in \{1, \ldots, c\}$ we have the equation

$$\left\langle \varphi^{(j)}(p_s)\left[v_{l_1} \otimes \ldots v_{l_j}\right], w_r \right\rangle_W \overset{(3.9)}{=} \sum_{i=1}^{\mathcal{N}} a_i \left\langle f_i^{(j)}(p_s)\left[v_{l_1} \otimes \ldots v_{l_j}\right], w_r \right\rangle_W. \tag{3.14}$$

Considering (3.14) for all possible choices of $s \in \{1, \ldots, m\}$, $j \in \{1, \ldots, k\}$, $(l_1, \ldots, l_j) \in \{1, \ldots, d\}^j$ with $l_1 \leq \ldots \leq l_j$ and $r \in \{1, \ldots, c\}$ results in a collection of $Q - 1$ equations. By taking the sum of the coefficients a_i to be an additional equation we obtain a total of Q equations that can be expressed as the linear system $\mathbf{A}\mathbf{x} = \mathbf{y}$ where $\mathbf{x} \in \mathbb{R}^{\mathcal{N}}$ is given by

$$\mathbf{x} := (a_1, \ldots, a_{\mathcal{N}})^T, \tag{3.15}$$

$\mathbf{y} \in \mathbb{R}^Q$ is given by

$$\mathbf{y} = \left(\sum_{i=1}^{\mathcal{N}} a_i,\ \left\langle \varphi^{(0)}(p_1), w_1 \right\rangle_W,\ \ldots, \left\langle \varphi^{(k)}(p_m)\left[v_d \otimes \ldots \otimes v_d\right],\ w_c \right\rangle_W \right)^T, \tag{3.16}$$

and $\mathbf{A} \in \mathbb{R}^{\mathcal{N} \times Q}$ is the $\mathcal{N} \times Q$ matrix given by

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ \left\langle f_1^{(0)}(p_1), w_1 \right\rangle_W & \left\langle f_2^{(0)}(p_1), w_1 \right\rangle_W & \cdots & \left\langle f_{\mathcal{N}}^{(0)}(p_1), w_1 \right\rangle_W \\ \vdots & \vdots & \ddots & \vdots \\ \left\langle f_1^{(k)}(p_m)[v_d \otimes \ldots \otimes v_d], w_c \right\rangle_W & \left\langle f_2^{(k)}(p_m)[v_d \otimes \ldots \otimes v_d], w_c \right\rangle_W & \cdots & \left\langle f_{\mathcal{N}}^{(k)}(p_m)[v_d \otimes \ldots \otimes v_d], w_c \right\rangle_W \end{pmatrix}. \tag{3.17}$$

The positivity of the coefficients $a_1, \ldots, a_{\mathcal{N}} > 0$ allow us to apply recombination [LL12] to this system. Recombination uses non-zero elements in the kernel $\ker(\mathbf{A})$ to kill (i.e. make 0) as many of the coefficients of $\mathbf{x}$ as possible. Altering $\mathbf{x}$ by elements $\mathbf{e} \in \ker(\mathbf{A})$ leaves the equality $\mathbf{A}\mathbf{x} = \mathbf{y}$ unchanged since $\mathbf{A}(\mathbf{x} + \theta\mathbf{e}) = \mathbf{A}\mathbf{x}$ for any $\theta \in \mathbb{R}$. Observe that $\dim(\ker(\mathbf{A})) = \mathcal{N} - \dim(\mathrm{Im}(\mathbf{A})) \geq \mathcal{N} - \mathcal{M}$ for $\mathcal{M} := \min\{Q, \mathcal{N}\}$. Therefore there are at least $\mathcal{N} - \mathcal{M}$ linearly independent elements in $\ker(\mathbf{A})$ which may be used to kill coefficients of $\mathbf{x}$; that is, recombination transforms $\mathbf{x}$ to a solution with at most $\mathcal{M}$ non-zero coefficients. That is, recombination will return non-negative coefficients $b_1, \ldots, b_{\mathcal{M}} \geq 0$ and indices $e(1), \ldots, e(\mathcal{M}) \in \{1, \ldots, \mathcal{N}\}$ such that the vector $\mathbf{x}' \in \mathbb{R}^{\mathcal{N}}$ with entries $\mathbf{x}'_\alpha = b_\alpha$ if $\alpha \in \{e(1), \ldots, e(\mathcal{M})\}$ and 0 otherwise still solves the equation $\mathbf{A}\mathbf{x}' = \mathbf{y}$. Define $u \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\gamma, \Sigma, W)$ by (cf. (3.7))

$$u := \sum_{j=1}^{\mathcal{M}} b_j f_{e(j)} \qquad \text{so that for each } s \in \{0, \ldots, k\} \qquad u^{(s)} := \sum_{j=1}^{\mathcal{M}} b_j f_{e(j)}^{(s)}. \tag{3.18}$$

Examining (3.17) and (3.16) reveals that $\mathbf{A}\mathbf{x}' = \mathbf{y}$ means that $\sum_{j=1}^{\mathcal{M}} b_j = \sum_{i=1}^{\mathcal{N}} a_i$, as claimed in (3.8), and that for each $a \in \{1, \ldots, m\}$ we have

$$\varphi(p_a) = u(p_a), \quad \varphi^{(1)}(p_a) = u^{(1)}(p_a), \quad \ldots \quad \varphi^{(k-1)}(p_a) = u^{(k-1)}(p_a) \quad \text{and} \quad \varphi^{(k)}(p_a) = u^{(k)}(p_a). \tag{3.19}$$

A consequence of (3.19) is that u and φ to coincide throughout $\{p_1, \ldots, p_m\}$ as claimed. This completes the proof of Lemma 3.1. ■

Theoretically, Lemma 3.1 verifies that recombination can be used to find an approximation $u = \left(u^{(0)}, \ldots, u^{(k)}\right) \in \mathrm{Lip}(\gamma, \Sigma, W)$ coinciding with φ at a specified finite collection of points $\{p_1, \ldots, p_m\} \subset \Sigma$, say. However, practical implementations of recombination will inevitably result in numerical errors entering into the equations. That is, the returned coefficients will only solve the equations modulo some (hopefully) small error term. To account for this in our analysis, whenever we apply Lemma 3.1 we will only assume that the resulting approximation u is *close* to φ at each of the points $p_1, \ldots, p_m \in \Sigma$ in the following sense. For each $j \in \{0, \ldots, k\}$ and each $c \in \{1, \ldots, m\}$, we assume that $\left|\left|\varphi^{(j)}(p_c) - u^{(j)}(p_c)\right|\right|_{\mathcal{L}(V^{\otimes l}; W)} \leq \varepsilon_0$ for some (small) constant $\varepsilon_0 \geq 0$. This may alternatively be denoted by assuming, for every $c \in \{1, \ldots, m\}$, that $\Lambda^k_{\varphi - u}(p_c) \leq \varepsilon_0$.

We now rigorously define our proposed GRIM algorithm designed to find an approximation $u \in \mathrm{Span}(\mathcal{F})$ of $\varphi \in \mathrm{Span}(\mathcal{F})$ that is close to φ throughout Σ in the sense, for a fixed $q \in \{0, \ldots, k\}$, that $\Lambda^q_{\varphi - u}(x) \leq \varepsilon$ for every $x \in \Sigma$.

Lip(γ) GRIM

(A) Fix $\varepsilon > 0$ as the target accuracy threshold, fix $\varepsilon_0 \in [0, \varepsilon)$ as the acceptable recombination error bound, and fix a choice of $q \in \{0, \ldots, k\}$ as the order level.

(B) For each $i \in \{1, \ldots, \mathcal{N}\}$, if $a_i < 0$ then replace a_i and f_i by $-a_i$ and $-f_i$ respectively. This ensures that $a_1, \ldots, a_{\mathcal{N}} > 0$ whilst leaving the expansion $\varphi = \sum_{i=1}^{\mathcal{N}} a_i f_i$ unaltered. Additionally, rescale each f_i to have unit $\mathrm{Lip}(\gamma, \Sigma, W)$ norm; that is, for each $i \in \{1, \ldots, \mathcal{N}\}$ replace f_i by $h_i := \frac{f_i}{||f_i||_{\mathrm{Lip}(\gamma, \Sigma, W)}}$ so that $h_i \in \mathrm{Lip}(\gamma, \Sigma, W)$ with $||h_i||_{\mathrm{Lip}(\gamma, \Sigma, W)} = 1$, and that $\varphi = \sum_{i=1}^{\mathcal{N}} \alpha_i h_i$ where $\alpha_i := a_i ||f_i||_{\mathrm{Lip}(\gamma, \Sigma, W)}$.

(C) Choose $k_1 \in \mathbb{Z}_{\geq 1}$ and apply the **Lip(γ) Extension Step**, with $L' := \emptyset$, $u \equiv 0$ and $m := k_1$, to obtain a subset $\Sigma_1 = \{z_{1,1}, \ldots, z_{1,k_1}\} \subset \Sigma$. Apply recombination via Lemma 3.1 to find an element $u_1 \in \mathrm{Span}(\mathcal{F})$ satisfying, for every $z \in \Sigma_1$, that $\Lambda^k_{\varphi - u_1}(z) \leq \varepsilon_0$.

(D) For $s \in \mathbb{Z}_{\geq 2}$ we proceed inductively. If $\Lambda^q_{\varphi - u_{s-1}}(x) \leq \varepsilon$ for every $x \in \Sigma$ then we stop and u_{s-1} is an approximation of φ possessing the desired level of accuracy. If this is not the case, then choose $k_s \in \mathbb{Z}_{\geq 1}$

and apply the **Lip(γ) Extension Step**, with $L' := \Sigma_{s-1}$, $u := u_{s-1}$ and $m := k_s$, to obtain a subset $\Sigma_s = \Sigma_{s-1} \cup \{z_{s,1}, \ldots, z_{s,k_s}\} \subset \Sigma$. Apply recombination via Lemma 3.1 to find an element $u_s \in \mathrm{Span}(\mathcal{F})$ satisfying, for every $z \in \Sigma_s$, that $\Lambda^k_{\varphi - u_s}(z) \leq \varepsilon_0$.

We can establish the following convergence theorem for the **Lip(γ) GRIM** algorithm with the choice of $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$.

Theorem 3.2 (Lip(γ) GRIM Convergence). *Let V and W be finite dimensional Banach spaces with $\Sigma \subset V$ compact. Assume that the tensor products of V are all equipped with admissible norms (cf. Definition A.1). Let $\mathcal{N} \in \mathbb{Z}_{>0}$ be a positive integer and $\gamma > 0$ with $k \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k, k+1]$. Assume that for every $i \in \{1, \ldots, \mathcal{N}\}$ the element $f_i = \left(f_i^{(0)}, \ldots, f_i^{(k)}\right) \in \mathrm{Lip}(\gamma, \Sigma, W)$ is non-zero and define $\mathcal{F} := \{f_i \ : \ i \in \{1, \ldots, \mathcal{N}\}\} \subset \mathrm{Lip}(\gamma, \Sigma, W)$. Given $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$, define $\varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(k)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\gamma, \Sigma, W)$ and constants $C, D > 0$ by*

$$\textbf{(I)} \quad \varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i, \quad \begin{pmatrix} \varphi^{(l)} := \sum_{i=1}^{\mathcal{N}} a_i f_i^{(l)} \\ \text{for every } l \in \{0, \ldots, k\} \end{pmatrix} \quad \text{and} \quad \textbf{(II)} \quad C := \sum_{i=1}^{\mathcal{N}} |a_i| ||f_i||_{\mathrm{Lip}(\gamma, \Sigma, W)} > 0 \tag{3.20}$$

and

$$D := \dim(W) \sum_{s=0}^{k} \begin{pmatrix} \dim(V) + s - 1 \\ s \end{pmatrix}. \tag{3.21}$$

Let $0 \leq \varepsilon_0 < \varepsilon < 2C$, $q \in \{0, \ldots, k\}$ and $K \geq 1$, $T > 1$ with $T\varepsilon \leq 2KC$. Then there are positive constants $r_1 = r_1(C, \gamma, \varepsilon, \varepsilon_0, q) > 0$, and $r_2 = r_2(C, \gamma, \varepsilon, R, T, q) > 0$, given by

$$\begin{cases} \textbf{(I)} \quad r_1 := \sup\left\{\lambda > 0 \ : \ 2C\lambda^{\gamma - q} + \varepsilon_0 e^{\lambda} \leq \varepsilon\right\} \\ \textbf{(II)} \quad r_2 := \sup\left\{\lambda > 0 \ : \ \varepsilon e^{\lambda} + 2KC\left(\lambda^{\gamma-q} + e^{\lambda} - 1\right) \leq T\varepsilon\right\}, \end{cases} \tag{3.22}$$

and a non-negative integer $N = N(\Sigma, C, \gamma, \varepsilon, \varepsilon_0, q) \in \mathbb{Z}_{\geq 0}$, given by the r_1-packing number of Σ

$$N = N_{\mathrm{pack}}(\Sigma, V, r_1) := \max\left\{d \in \mathbb{Z} \ : \ \exists x_1, \ldots, x_d \in \Sigma \text{ for which } ||x_a - x_c||_V > r_1 \text{ if } a \neq c\right\}, \tag{3.23}$$

for which the following is true.

*If we apply the **Lip(γ) GRIM** algorithm, with the choice that $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, to approximate φ on Σ, with the target accuracy threshold, the acceptable recombination error bound, and the order level in **Lip(γ) GRIM** (A) as ε, ε_0, and q respectively, then after at most N steps, the algorithm terminates. That is, if we let $M \in \{1, \ldots, N\}$ be the integer for which the algorithm terminates after step M and define $Q_M := \min\{\mathcal{N}, 1 + MD\}$, there are coefficients $c_{M,1}, \ldots, c_{M,Q_M} \in \mathbb{R}$ and indices $e_M(1), \ldots, e_M(Q_M) \in \{1, \ldots, \mathcal{N}\}$ with*

$$\sum_{s=1}^{Q_M} |c_{M,s}| \, ||f_{e_M(s)}||_{\mathrm{Lip}(\gamma, \Sigma, W)} = C, \tag{3.24}$$

and such that $u_M = \left(u_M^{(0)}, \ldots u_M^{(k)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\gamma, \Sigma, W)$ defined by

$$u_M := \sum_{s=1}^{Q_M} c_{M,s} f_{e_M(s)} \begin{pmatrix} u_M^{(l)} := \sum_{s=1}^{Q_M} c_{M,s} f_{e_M(s)}^{(l)} \\ \text{for every } l \in \{0, \ldots, k\} \end{pmatrix} \quad \text{satisfies} \quad \sup_{x \in \Sigma} \Lambda^q_{\varphi - u_M}(x) \leq \varepsilon. \tag{3.25}$$

*Moreover, if the coefficients $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ corresponding to φ (cf. **(I)** of (3.20)) are all positive (i.e. $a_1, \ldots, a_{\mathcal{N}} > 0$) then the coefficients $c_{M,1}, \ldots, c_{M,Q_M} \in \mathbb{R}$ corresponding to u_M (cf. (3.25)) are all non-negative (i.e. $c_{M,1}, \ldots, c_{M,Q_M} \geq 0$).*

Further, suppose $\Omega \subset V$ is closed and $\Omega \subset \Sigma_{r_2}$ where Σ_{r_2} denotes the r_2-fattening of Σ, i.e. Σ_{r_2} is defined by $\Sigma_{r_2} := \{u \in V \ : \ \exists z \in \Sigma \text{ such that } ||z - u||_V \leq r_2\}$. Further suppose that for every $i \in \{1, \ldots, \mathcal{N}\}$ the element $f_i \in \mathrm{Lip}(\gamma, \Sigma, W)$ can be extended to an element in $F_i \in \mathrm{Lip}(\gamma, \Omega, W)$ with $||F_i||_{\mathrm{Lip}(\gamma, \Omega, W)} \leq K ||f_i||_{\mathrm{Lip}(\gamma, \Sigma, W)}$.

Then the extension $\hat{\varphi} \in \mathrm{Lip}(\gamma,\Omega,W)$ *of* φ*, defined by*

$$\hat{\varphi} := \sum_{i=1}^{\mathcal{N}} a_i F_i \quad \textit{so that for every } l \in \{0,\ldots,k\} \textit{ we have} \quad \hat{\varphi}^{(l)} = \sum_{i=1}^{\mathcal{N}} a_i F_i^{(l)}, \tag{3.26}$$

remains well-approximated *by the extension* $\hat{u}_M \in \mathrm{Lip}(\gamma,\Omega,W)$ *of* u_M*, defined by*

$$\hat{u}_M := \sum_{s=1}^{Q_M} c_{M,s} F_{e_M(s)} \quad \textit{so that for every } l \in \{0,\ldots,k\} \textit{ we have} \quad \hat{u}_M^{(l)} = \sum_{s=1}^{Q_M} c_{M,s} F_{e_M(s)}^{(l)}, \tag{3.27}$$

in the sense that for every $z \in \Omega$ *we have*

$$\Lambda^q_{\hat{\varphi}-\hat{u}_M}(z) = \max_{j\in\{0,\ldots,q\}} \left|\left|\hat{\varphi}^{(j)}(z) - \hat{u}_M^{(j)}(z)\right|\right|_{\mathcal{L}(V^{\otimes j};W)} \leq T\varepsilon. \tag{3.28}$$

The choice of q in the $\mathrm{Lip}(\gamma)$ *GRIM Convergence* Theorem 3.2 is determined by the order to which we wish to approximate the element $\varphi = \left(\varphi^{(0)},\ldots,\varphi^{(k)}\right) \in \mathrm{Lip}(\gamma,\Sigma,W)$. For example, if we are only interested in approximating the function $\varphi^{(0)} : \Sigma \to W$, then it may be appropriate to choose $q := 0$. The choice of $q := 0$ ensures, via (3.25), that the algorithm finds an approximation $u_M = \left(u_M^{(0)},\ldots,u_M^{(k)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\gamma,\Sigma,W)$ for which the function $u_M^{(0)} : \Sigma \to W$ is close to $\varphi^{(0)}$ throughout Σ in the sense that $\left|\left|\varphi^{(0)} - u_M^{(0)}\right|\right|_{C^0(\Sigma;W)} \leq \varepsilon$. Moreover, since **(I)** in (3.22) means that the constant $r_1 \leq 1$, the choice of $q := 0$ results in the largest value of r_1. Since the maximum number of steps before termination N is the r_1-packing number of Σ defined in (3.23), the choice $q := 0$ additionally has the benefit of minimising the upper bound for the maximum number of steps the algorithm runs for before terminating.

Larger values of q result in the algorithm finding an approximation of φ in a stronger sense throughout Σ, at the cost of a worse upper bound for the maximum number of steps the algorithm runs for before terminating. The choice $q := k$ provides the estimates (via (3.25)) that, for every $j \in \{0,\ldots,k\}$, the functions $u_M^{(j)} : \Sigma \to \mathcal{L}(V^{\otimes j};W)$ are close to $\varphi^{(j)} : \Sigma \to \mathcal{L}(V^{\otimes j};W)$ throughout Σ in the sense that, for every $x \in \Sigma$, we have $\left|\left|\varphi^{(j)}(x) - u_M^{(j)}(x)\right|\right|_{\mathcal{L}(V^{\otimes j};W)} \leq \varepsilon$. However, given any $\eta \in (0,\gamma)$, the choice $q := k$ can be used to additionally ensure that the algorithm returns an approximation $u_M = \left(u_M^{(0)},\ldots,u_M^{(k)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\gamma,\Sigma,W)$ that is close to φ in the $\mathrm{Lip}(\eta,\Sigma,W)$ sense that $\left|\left|\varphi_{[b]} - (u_M)_{[b]}\right|\right|_{\mathrm{Lip}(\eta,\Sigma,W)} \leq \varepsilon$ for $b \in \{0,\ldots,k\}$ such that $\eta \in (b,b+1]$. Here we use the notation that $\varphi_{[b]} = \left(\varphi^{(0)},\ldots,\varphi^{(b)}\right)$ and $(u_M)_{[b]} = \left(u_M^{(0)},\ldots,u_M^{(b)}\right)$. That this is possible forms the content of the following result.

Theorem 3.3 (**$\mathrm{Lip}(\eta,\Sigma,W)$ Approximation via $\mathrm{Lip}(\gamma)$ GRIM**). *Let* V *and* W *be finite dimensional Banach spaces,* $\Sigma \subset V$ *a non-empty compact subset, and assume that all the tensor powers of* V *are equipped with admissible norms (cf. Definition A.1). Let* $\mathcal{N} \in \mathbb{Z}_{>0}$ *be a positive integer, and suppose that* $\gamma > \eta > 0$ *with* $k,b \in \mathbb{Z}_{\geq 0}$ *such that* $\gamma \in (k,k+1]$ *and* $\eta \in (b,b+1]$*. Assume, for every* $i \in \{1,\ldots,\mathcal{N}\}$*, that* $f_i = \left(f_i^{(0)},\ldots,f_i^{(k)}\right) \in \mathrm{Lip}(\gamma,\Sigma,W)$ *is non-zero and define* $\mathcal{F} := \{f_i : i \in \{1,\ldots,\mathcal{N}\}\} \subset \mathrm{Lip}(\gamma,\Sigma,W)$*. Given* $a_1,\ldots,a_{\mathcal{N}} \in \mathbb{R}\setminus\{0\}$*, define* $\varphi := \left(\varphi^{(0)},\ldots,\varphi^{(k)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\gamma,\Sigma,W)$ *and constants* $C,D > 0$ *by*

$$\textbf{(I)} \quad \varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i, \quad \left(\begin{array}{c} \varphi^{(l)} := \sum_{i=1}^{\mathcal{N}} a_i f_i^{(l)} \\ \textit{for every } l \in \{0,\ldots,k\} \end{array}\right) \quad \textit{and} \quad \textbf{(II)} \quad C := \sum_{i=1}^{\mathcal{N}} |a_i| ||f_i||_{\mathrm{Lip}(\gamma,\Sigma,W)} > 0 \tag{3.29}$$

and

$$D := \dim(W) \sum_{s=0}^{k} \left(\begin{array}{c} \dim(V)+s-1 \\ s \end{array}\right). \tag{3.30}$$

Let $\varepsilon > 0$ *be given. Then there are positive constants* $\varepsilon_0 = \varepsilon_0(\varepsilon,C,\gamma,\eta) > 0$*,* $\delta_0 = \delta_0(\varepsilon,C,\gamma,\eta) > 0$*, a* *non-negative integer* $J = J(\Sigma, \varepsilon, C, \gamma, \eta) \in \mathbb{Z}_{\geq 0}$ *given by*

$$J := N_{\text{cov}}(\Sigma, V, \delta_0) = \min \left\{ d \in \mathbb{Z}_{\geq 0} \ : \ \exists x_1, \ldots, x_d \in \Sigma \text{ such that } \Sigma \subset \bigcup_{l=1}^{d} \overline{\mathbb{B}}_V(x_l, \delta_0) \right\}, \tag{3.31}$$

and a finite subset $B \subset \Sigma$ *with cardinality* $\#(B) = J$ *for which the following is true.*

Retrieve both the positive constant r_1 *and the non-negative integer* N *arising in the* $\mathrm{Lip}(\gamma)$ GRIM Convergence *Theorem 3.2 for the choices of the subset* Σ *and the constants* C, ε, ε_0, γ *and* q *there as the subset* B *and the constants* C, ε_0, $\varepsilon_0/2$, γ *and* k *here respectively. Examination of the* $\mathrm{Lip}(\gamma)$ GRIM Convergence *Theorem 3.2 reveals that this amounts to taking* $r_1 = r_1(C, \varepsilon, \varepsilon_0, \gamma) > 0$ *as (cf.* **(I)** *of (3.22))*

$$r_1 := \sup \left\{ \lambda > 0 \ : \ 2C\lambda^{\gamma - k} + \frac{\varepsilon_0}{2} e^{\lambda} \leq \varepsilon_0 \right\} \tag{3.32}$$

and $N = N(B, C, \varepsilon, \varepsilon_0, \gamma) \in \mathbb{Z}_{\geq 0}$ *to be the* r_1*-packing number of* B *defined by (cf. (3.23))*

$$N := N_{\text{pack}}(\Sigma, V, r_1) = \max \{ d \in \mathbb{Z} \ : \ \exists x_1, \ldots, x_d \in B \text{ for which } ||x_a - x_c||_V > r_1 \text{ whenever } a \neq c \} \tag{3.33}$$

If we apply the $\mathbf{Lip(\gamma)}$ ***GRIM*** *algorithm, with the choice that* $k_s := 1$ *for every integer* $s \in \mathbb{Z}_{\geq 1}$*, to approximate* φ *on* B*, with the target accuracy threshold, the acceptable recombination error bound, and the order level in* $\mathbf{Lip(\gamma)}$ ***GRIM*** *(A) as* ε_0*,* $\varepsilon_0/2$*, and* k *respectively, then after at most* N *steps, the algorithm terminates. Moreover, if* $M \in \{1, \ldots, N\}$ *is the integer for which the algorithm terminates after step* M *and* $Q_M := \min\{\mathcal{N}, 1 + MD\}$*, there are coefficients* $c_{M,1}, \ldots, c_{M,Q_M} \in \mathbb{R}$ *and indices* $e_M(1), \ldots, e_M(Q_M) \in \{1, \ldots, \mathcal{N}\}$ *with*

$$\sum_{s=1}^{Q_M} |c_{M,s}| \, \left|\left| f_{e_M(s)} \right|\right|_{\mathrm{Lip}(\gamma, \Sigma, W)} = C, \tag{3.34}$$

and such that $u_M = \left(u_M^{(0)}, \ldots u_M^{(k)} \right) \in \mathrm{Span}(F) \subset \mathrm{Lip}(\gamma, \Sigma, W)$ *defined by*

$$u_M := \sum_{s=1}^{Q_M} c_{M,s} f_{e_M(s)} \quad \left(\begin{array}{c} u_M^{(l)} := \sum_{s=1}^{Q_M} c_{M,s} f_{e_M(s)}^{(l)} \\ \text{for every } l \in \{0, \ldots, k\} \end{array} \right) \quad \text{satisfies} \quad \left|\left| \varphi_{[b]} - (u_M)_{[b]} \right|\right|_{\mathrm{Lip}(\eta, \Sigma, W)} \leq \varepsilon \tag{3.35}$$

where $\varphi_{[b]} = \left(\varphi^{(0)}, \ldots, \varphi^{(b)} \right)$ *and* $(u_M)_{[b]} = \left(u_M^{(0)}, \ldots, u_M^{(b)} \right)$*. Moreover, if the coefficients* $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ *corresponding to* φ *(cf.* **(I)** *of (3.29)) are all positive (i.e.* $a_1, \ldots, a_{\mathcal{N}} > 0$*) then the coefficients* $c_{M,1}, \ldots, c_{M,Q_M} \in \mathbb{R}$ *corresponding to* u_M *(cf. (3.35)) are all non-negative (i.e.* $c_{M,1}, \ldots, c_{M,Q_M} \geq 0$*).*

Remark 3.4. The integer N defined in (3.33) satisfies that $N \leq \#(B) = J$. Consequently we may conclude that the $\mathbf{Lip(\gamma)}$ **GRIM** algorithm terminates after, at most, J steps. Moreover, if $N < \frac{\mathcal{N}-1}{D}$ then $Q_M < \mathcal{N}$ and we are guaranteed that the returned element u_M is a linear combination of *less* than $\mathcal{N}$ of the elements in $\mathcal{F}$ that is nevertheless within ε of φ throughout Σ in the $\mathrm{Lip}(\eta, \Sigma, W)$ sense specified in (3.35).

Proof of Theorem 3.3. Let V and W be finite dimensional Banach spaces, $\Sigma \subset V$ a non-empty compact subset, and assume that all the tensor powers of V are equipped with admissible norms (cf. Definition A.1). Let $\mathcal{N} \in \mathbb{Z}_{>0}$ be a positive integer, and suppose that $\gamma > \eta > 0$ with $k, b \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k, k+1]$ and $\eta \in (b, b+1]$. Assume, for every $i \in \{1, \ldots, \mathcal{N}\}$ that $f_i = \left(f_i^{(0)}, \ldots, f_i^{(k)} \right) \in \mathrm{Lip}(\gamma, \Sigma, W)$ is non-zero and define $\mathcal{F} := \{ f_i \ : \ i \in \{1, \ldots, \mathcal{N}\} \} \subset \mathrm{Lip}(\gamma, \Sigma, W)$. Given $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$, define $\varphi := \left(\varphi^{(0)}, \ldots, \varphi^{(k)} \right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\gamma, \Sigma, W)$ and constants $C, D > 0$ by

$$\textbf{(I)} \quad \varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i, \quad \left(\begin{array}{c} \varphi^{(l)} := \sum_{i=1}^{\mathcal{N}} a_i f_i^{(l)} \\ \text{for every } l \in \{0, \ldots, k\} \end{array} \right) \quad \text{and} \quad \textbf{(II)} \quad C := \sum_{i=1}^{\mathcal{N}} |a_i| ||f_i||_{\mathrm{Lip}(\gamma, \Sigma, W)} > 0 \tag{3.36}$$

and

$$D := \dim(W) \sum_{s=0}^{k} \binom{\dim(V) + s - 1}{s}. \tag{3.37}$$

Let $\varepsilon > 0$ be given. Retrieve the constants $\varepsilon_0, \delta_0 > 0$ resulting from the *Lipschitz Sandwich Theorem* B.1 for the choices of ε, K_0, γ and η there as ε, C, γ and η here respectively. Note we are not actually applying the *Lipschitz Sandwich Theorem* B.1 here, but merely retrieving constants in preparation for its future application. An examination of the *Lipschitz Sandwich Theorem* B.1 reveals that the resulting δ_0 and ε_0 depend only on ε, C, γ and η. We now fix the values of both δ_0 and ε_0 for the remainder of the proof.

Having specified both δ_0 and ε_0, we now let $J = J(\Sigma, \varepsilon, C, \gamma, \eta) \in \mathbb{Z}_{\geq 0}$ be defined as in (3.31). That is

$$J := N_{\text{cov}}(\Sigma, V, \delta_0) = \min \left\{ d \in \mathbb{Z}_{\geq 0} \ : \ \exists x_1, \ldots, x_d \in \Sigma \text{ such that } \Sigma \subset \bigcup_{l=1}^{d} \overline{\mathbb{B}}_V(x_l, \delta_0) \right\}. \tag{3.38}$$

Since Σ is non-empty and compact we may conclude both that $J \geq 1$ and that J is finite. Let $B := \{z_1, \ldots, z_J\} \subset \Sigma$ be any choice of J points in Σ for which

$$\Sigma \subset \bigcup_{j=1}^{J} \overline{\mathbb{B}}_V(z_j, \delta_0) = B_{\delta_0} := \{ v \in V \ : \ \exists z \in B \text{ such that } ||z - v||_V \leq \delta_0 \}. \tag{3.39}$$

Evidently $\#(B) = J$. We now fix the value of the integer J and the choice of the finite subset $B \subset \Sigma$ with $\#(B) = J$ for the remainder of the proof. With the constants ε_0, δ_0, the integer J and the subset B specified, we set about verifying the assertions of the theorem.

Retrieve both the positive constant r_1 and the non-negative integer N arising in the Lip(γ) *GRIM Convergence* Theorem 3.2 for the choices of the subset Σ and the constants $C, \varepsilon, \varepsilon_0, \gamma$ and q there as the subset B and the constants $C, \varepsilon_0, \varepsilon_0/2, \gamma$ and k here respectively. Examination of the Lip(γ) *GRIM Convergence* Theorem 3.2 reveals that this amounts to taking $r_1 = r_1(C, \varepsilon, \varepsilon_0, \gamma) > 0$ as (cf. (**I**) of (3.22))

$$r_1 := \sup \left\{ \lambda > 0 \ : \ 2C\lambda^{\gamma - k} + \frac{\varepsilon_0}{2} e^{\lambda} \leq \varepsilon_0 \right\} \tag{3.40}$$

and $N = N(B, C, \varepsilon, \varepsilon_0, \gamma) \in \mathbb{Z}_{\geq 0}$ to be the r_1-packing number of B defined by (cf. (3.23))

$$N := N_{\text{pack}}(\Sigma, V, r_1) = \max \{ d \in \mathbb{Z} \ : \ \exists x_1, \ldots, x_d \in B \text{ for which } ||x_a - x_c||_V > r_1 \text{ whenever } a \neq c \} \tag{3.41}$$

as claimed in (3.32) and (3.33) respectively. Consider applying **Lip(γ) GRIM** algorithm, with the choice that $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, to approximate φ on B, with the target accuracy threshold, the acceptable recombination error bound, and the order level in **Lip(γ) GRIM** (A) as ε_0, $\varepsilon_0/2$, and k respectively.

Our choice of r_1 and N in (3.32) and (3.41) respectively allow us to apply the Lip(γ) *GRIM Convergence* Theorem 3.2 and conclude that after at most N steps, the algorithm terminates. Theorem 3.2 additionally tells us that if $M \in \{1, \ldots, N\}$ is the integer for which the algorithm terminates after step M and $Q_M := \min\{\mathcal{N}, 1 + MD\}$, there are coefficients $c_{M,1}, \ldots, c_{M,Q_M} \in \mathbb{R}$ and indices $e_M(1), \ldots, e_M(Q_M) \in \{1, \ldots, \mathcal{N}\}$ with

$$\sum_{s=1}^{Q_M} |c_{M,s}| \, ||f_{e_M(s)}||_{\text{Lip}(\gamma, \Sigma, W)} = C, \tag{3.42}$$

and such that $u_M = \left(u_M^{(0)}, \ldots u_M^{(k)} \right) \in \text{Span}(\mathcal{F}) \subset \text{Lip}(\gamma, \Sigma, W)$ defined by

$$u_M := \sum_{s=1}^{Q_M} c_{M,s} f_{e_M(s)} \quad \left(\begin{array}{c} u_M^{(l)} := \sum_{s=1}^{Q_M} c_{M,s} f_{e_M(s)}^{(l)} \\ \text{for every } l \in \{0, \ldots, k\} \end{array} \right) \quad \text{satisfies} \quad \sup_{x \in B} \Lambda^k_{\varphi - u_M}(x) \leq \varepsilon_0. \tag{3.43}$$

We note that (3.42) is precisely the equality claimed in (3.34), and that (3.43) establishes the expansion claimed for u_M in the first part of (3.35). The $\mathrm{Lip}(\gamma)$ *GRIM Convergence* Theorem 3.2 additionally establishes that if the coefficients $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ corresponding to φ (cf. **(I)** of (3.36)) are all positive (i.e. $a_1, \ldots, a_{\mathcal{N}} > 0$) then the coefficients $c_{M,1}, \ldots, c_{M,Q_M} \in \mathbb{R}$ corresponding to u_M (cf. (3.43)) are all non-negative (i.e. $c_{M,1}, \ldots, c_{M,Q_M} \geq 0$), which is precisely as claimed in Theorem 3.3. It remains only to establish the $\mathrm{Lip}(\eta, \Sigma, W)$ estimate claimed in the second part of (3.35).

For this purpose we first note that **(I)** and **(II)** in (3.36) ensure that $||\varphi||_{\mathrm{Lip}(\gamma,\Sigma,W)} \leq C$. Further, (3.42) and (3.43) ensure that $||u_M||_{\mathrm{Lip}(\gamma,\Sigma,W)} \leq C$. Moreover, (3.43) additionally tells us that for every $x \in B$ we have

$$\Lambda^k_{\varphi - u_M}(x) := \max_{l \in \{0,\ldots,k\}} \left|\left| \varphi^{(l)}(x) - u_M^{(l)}(x) \right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \varepsilon_0. \tag{3.44}$$

Recalling how we specified the constants δ_0 and ε_0, we have the hypotheses required in order to apply the *Lipschitz Sandwich Theorem* B.8 for the choices of the subset B and the constants K_0, ε, γ and η of that result as the subset B and the constants C, ε, γ and η here respectively. That is, we have verified that we have the required $\mathrm{Lip}(\gamma, \Sigma, W)$ bounds $||\varphi||_{\mathrm{Lip}(\gamma,\Sigma,W)}, ||u_M||_{\mathrm{Lip}(\gamma,\Sigma,W)} \leq C$, the δ_0-cover hypothesis required in (B.2) is given by (3.39), and pointwise estimates required, for every $l \in \{0, \ldots, k\}$, for the difference $\varphi^{(l)} - u_M^{(l)}$ in (B.3) are provided by (3.44). Therefore we are able to apply the *Lipschitz Sandwich Theorem* B.8 and conclude that (cf. (B.4) for q there as b here)

$$\left|\left| \varphi_{[b]} - (u_M)_{[b]} \right|\right|_{\mathrm{Lip}(\eta,\Sigma,W)} \leq \varepsilon \tag{3.45}$$

where $\varphi_{[b]} = \left(\varphi^{(0)}, \ldots, \varphi^{(b)}\right)$ and $(u_M)_{[b]} = \left(u_M^{(0)}, \ldots, u_M^{(b)}\right)$. The estimate (3.45) is precisely the estimate claimed in the second part of (3.35). This completes the proof of Theorem 3.3. ■

The remainder of this section is dedicated to the proof of the $\mathrm{Lip}(\gamma)$ *GRIM Convergence* Theorem 3.2. We first prove that the points selected during the **Lip(γ) GRIM** algorithm, with the choice of $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, have to be a definite distance apart. This is the content of the following result.

Lemma 3.5 (Lip(γ) GRIM Point Separation). *Assume V and W are finite dimensional Banach spaces with $\Gamma \subset V$ compact. Assume that the tensor products of V are all equipped with admissible norms (cf. Definition A.1). Let $\theta, \mu > 0$ with $n \in \mathbb{Z}_{\geq 0}$ such that $\mu \in (n, n+1]$, $b \in \{0, \ldots, n\}$, and let $\mathcal{N} \in \mathbb{Z}_{>0}$ be a positive integer. Assume that for every $i \in \{1, \ldots, \mathcal{N}\}$ the element $f_i = \left(f_i^{(0)}, \ldots, f_i^{(n)}\right) \in \mathrm{Lip}(\mu, \Gamma, W)$ is non-zero and define $\mathcal{F} := \{f_i : i \in \{1, \ldots, \mathcal{N}\}\} \subset \mathrm{Lip}(\mu, \Gamma, W)$. Given $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$, define $\varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(n)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\mu, \Gamma, W)$ and $C > 0$ by*

$$\textbf{(I)} \quad \varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i \quad \left(\begin{array}{c} \varphi^{(l)} := \sum_{i=1}^{\mathcal{N}} a_i f_i^{(l)} \\ \text{for every } l \in \{0,\ldots,n\} \end{array} \right) \quad \text{and} \quad \textbf{(II)} \quad C := \sum_{i=1}^{\mathcal{N}} |a_i| \, ||f_i||_{\mathrm{Lip}(\mu,\Gamma,W)} > 0. \tag{3.46}$$

Let $0 \leq \theta_0 < \min\{2C, \theta\}$. Then there exists a positive constant $r = r(C, \mu, \theta, \theta_0, b) > 0$, given by

$$r := \sup\left\{\lambda > 0 \;:\; 2C\lambda^{\mu-b} + \theta_0 e^{\lambda} \leq \min\{2C, \theta\}\right\}, \tag{3.47}$$

for which the following is true.

Consider applying the **Lip(μ) GRIM** *algorithm, with the choice of $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, to approximate φ on Γ with the target accuracy threshold, the acceptable recombination error bound, and the order level in* **Lip(μ) GRIM** *(A) as θ, θ_0, and b here respectively. Given $m \in \mathbb{Z}_{\geq 2}$, if the algorithm reaches and carries out the m^{th} step without terminating, let $u_m = \left(u_m^{(0)}, \ldots, u_m^{(n)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\mu, \Gamma, W)$ denote the approximation found at the m^{th} step and $\Gamma_m := \{z_1, \ldots, z_m\} \subset \Gamma$ denote the points selected such that at every $z \in \Gamma_m$ we have $\Lambda^n_{\varphi - u_m}(z) \leq \theta_0$ (cf.* **Lip(μ) GRIM** *(C) and (D)). Then whenever $a, c \in \{1, \ldots, m\}$ with $a \neq c$ we have that*

$$||z_a - z_c||_V > r. \tag{3.48}$$

Moreover, for every $z \in \Gamma$*, the quantity* $\Lambda^b_{\varphi-u_m}(z) := \max_{h\in\{0,\dots,b\}} \left|\left|\varphi^{(h)}(z) - u_m^{(h)}(z)\right|\right|_{\mathcal{L}(V^{\otimes j};W)}$ *satisfies*

$$\Lambda^b_{\varphi-u_m}(z) \leq \min\left\{2C\ ,\ 2C \max_{h\in\{0,\dots,b\}}\left\{\mathrm{dist}(z,\Gamma_m)^{\mu-h}\right\} + \theta_0 e^{\mathrm{dist}(z,\Gamma_m)}\right\}. \tag{3.49}$$

Proof of Lemma 3.5. Let V and W are finite dimensional Banach spaces with $\Gamma \subset V$ compact. Assume that the tensor products of V are all equipped with admissible norms (cf. Definition A.1). Let $\theta, \mu > 0$ with $n \in \mathbb{Z}_{\geq 0}$ such that $\mu \in (n, n+1]$, $b \in \{0,\dots,n\}$, and let $\mathcal{N} \in \mathbb{Z}_{>0}$ be a positive integer. Assume $a_1,\dots,a_{\mathcal{N}} \in \mathbb{R}\setminus\{0\}$, and that for every $i \in \{1,\dots,\mathcal{N}\}$ the function $f_i = \left(f_i^{(0)},\dots,f_i^{(n)}\right) \in \mathrm{Lip}(\mu,\Gamma,W)$ is non-zero and define $\mathcal{F} := \{f_i : i \in \{1,\dots,\mathcal{N}\}\} \subset \mathrm{Lip}(\mu,\Gamma,W)$. Define $\varphi = \left(\varphi^{(0)},\dots,\varphi^{(n)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\mu,\Gamma,W)$ as in (**I**) of (3.46) and the constant C as in (**II**) of (3.46).

With a view to applying **Lip(μ) GRIM** to approximate φ on Γ, for each $i \in \{1,\dots,\mathcal{N}\}$ let $\tilde{a}_i := |a_i|$ and $\tilde{f}_i$ be given by f_i if $a_i > 0$ and $-f_i$ if $a_i < 0$. Observe that for every $i \in \{1,\dots,\mathcal{N}\}$ we have that $\left|\left|\tilde{f}_i\right|\right|_{\mathrm{Lip}(\mu,\Gamma,W)} = ||f_i||_{\mathrm{Lip}(\mu,\Gamma,W)}$. Moreover, we also have that $\tilde{a}_1,\dots,\tilde{a}_{\mathcal{N}} > 0$ and that $\varphi = \sum_{i=1}^{\mathcal{N}} \tilde{a}_i \tilde{f}_i$. Further, we rescale $\tilde{f}_i$ for each $i \in \{1,\dots,\mathcal{N}\}$ to have unit $\mathrm{Lip}(\mu,\Gamma,W)$ norm. That is (cf. **Lip(μ) GRIM** (B)), for each $i \in \{1,\dots,\mathcal{N}\}$ set $h_i := \frac{\tilde{f}_i}{||f_i||_{\mathrm{Lip}(\mu,\Gamma,W)}}$ and $\alpha_i := \tilde{a}_i ||f_i||_{\mathrm{Lip}(\mu,\Gamma,W)}$. Then observe both that C satisfies

$$C = \sum_{i=1}^{\mathcal{N}} |a_i|\, ||f_i||_{\mathrm{Lip}(\mu,\Gamma)} = \sum_{i=1}^{\mathcal{N}} \tilde{a}_i ||f_i||_{\mathrm{Lip}(\mu,\Gamma)} = \sum_{i=1}^{\mathcal{N}} \alpha_i, \tag{3.50}$$

and, for every $i \in \{1,\dots,\mathcal{N}\}$, that $\alpha_i h_i = \tilde{a}_i \tilde{f}_i = a_i f_i$. Thus the expansion for φ in (**I**) of (3.46) is equivalent to

$$\varphi = \sum_{i=1}^{\mathcal{N}} \alpha_i h_i, \qquad \text{and hence} \qquad ||\varphi||_{\mathrm{Lip}(\mu,\Gamma,W)} \leq \sum_{i=1}^{\mathcal{N}} \alpha_i ||h_i||_{\mathrm{Lip}(\mu,\Gamma,W)} = \sum_{i=1}^{\mathcal{N}} \alpha_i \overset{(3.50)}{=} C. \tag{3.51}$$

Let $0 \leq \theta_0 < \min\{2C,\theta\}$. Define $D > 0$ by

$$D := \dim(W) \sum_{s=0}^{k} \binom{\dim(V)+s-1}{s} \tag{3.52}$$

Finally, define $r = r(C,\mu,\theta,\theta_0,b) > 0$ by

$$r := \sup\left\{\lambda > 0 : 2C\lambda^{\mu-b} + \theta_0 e^{\lambda} \leq \min\{2C,\theta\}\right\}. \tag{3.53}$$

Now consider applying the **Lip(μ) GRIM** algorithm to approximate φ on Γ with the target accuracy threshold, the acceptable recombination error bound, and the order level in **Lip(μ) GRIM** (A) as θ, θ_0, and b here respectively. Suppose that $m \in \mathbb{Z}_{\geq 2}$ and that the **Lip(μ) GRIM** algorithm reaches and carries out the m^{th} step without terminating. Let $\Gamma_m = \{z_1,\dots,z_m\} \subset \Gamma$ denote the points chosen after the m^{th} step is completed. Then for every $l \in \{1,\dots,m\}$, if we let $\Gamma_l := \{z_1,\dots,z_l\} \subset \Gamma$, we have, recalling **Lip(μ) GRIM** (C) and (D), that recombination via Lemma 3.1 has found an approximation $u_l = \left(u_l^{(0)},\dots,u_l^{(n)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\mu,\Gamma,W)$ of φ satisfying, for each $s \in \{1,\dots,l\}$, that $\Lambda^n_{\varphi-u_l}(z_s) \leq \theta_0$. That is, given any $j \in \{0,\dots,n\}$ and any $s \in \{1,\dots,l\}$, we have $\left|\left|\varphi^{(j)}(z_s) - u_l^{(j)}(z_s)\right|\right|_{\mathcal{L}(V^{\otimes j};W)} \leq \theta_0$.

Let $Q_l := \min\{\mathcal{N}, 1+lD\}$. Then Lemma 3.1 additionally tells us that there are non-negative coefficients $b_{l,1},\dots,b_{l,Q_l} \geq 0$ and indices $e_l(1),\dots,e_l(Q_l) \in \{1,\dots,\mathcal{N}\}$ for which

$$u_l = \sum_{s=1}^{Q_l} b_{l,s} h_{e_l(s)} \quad \left(\begin{array}{c} u_l^{(j)} := \sum_{s=1}^{Q_l} b_{l,s} h^{(j)}_{e_l(s)} \\ \text{for every } j \in \{0,\dots,n\}\end{array}\right) \quad \text{and} \quad \sum_{s=1}^{Q_l} b_{l,s} = \sum_{i=1}^{\mathcal{N}} \alpha_i. \tag{3.54}$$

A consequence of (3.54) is that

$$||u_l||_{\mathrm{Lip}(\mu,\Gamma,W)} \leq \sum_{s=1}^{Q_l} b_{l,s} \left|\left|h_{e_l(s)}\right|\right|_{\mathrm{Lip}(\mu,\Gamma,W)} = \sum_{s=1}^{Q_l} b_{l,s} \overset{(3.54)}{=} \sum_{i=1}^{\mathcal{N}} \alpha_i \overset{(3.50)}{=} C. \tag{3.55}$$

Consider a point $x \in \Gamma$ and let $j \in \{1, \ldots, l\}$ be such that $\mathrm{dist}(x, \Gamma_l) = ||x - z_j||_V = \mathrm{dist}(x, z_j)$. Consider $h \in \{0, \ldots, b\}$. By applying Lemma C.1 to the function $\varphi - u_l$, with the A, r_0, p, ρ, l, n and q of that result as $2C$, θ_0, z_j, μ, h, n and n here respectively, we deduce that (cf. (C.3))

$$\left|\left|\varphi^{(h)}(x) - u_l^{(h)}(x)\right|\right|_{\mathcal{L}(V^{\otimes h};W)} \leq \min\left\{2C \;,\; 2C||x - z_j||_V^{\mu-h} + \theta_0 \sum_{s=0}^{n-h} \frac{1}{s!}||x - z_j||_V^s\right\}. \tag{3.56}$$

From (3.56) we deduce that

$$\left|\left|\varphi^{(h)}(x) - u_l^{(h)}(x)\right|\right|_{\mathcal{L}(V^{\otimes h};W)} \leq \min\left\{2C \;,\; 2C\,\mathrm{dist}(x, \Gamma_l)^{\mu-h} + \theta_0 e^{\mathrm{dist}(x,\Gamma_l)}\right\}. \tag{3.57}$$

The arbitrariness of $h \in \{0, \ldots, b\}$ and $x \in \Gamma$ allows us to conclude that (3.57) is valid for every $h \in \{0, \ldots, b\}$ and every $x \in \Gamma$. As a result we have that

$$\sup_{x\in\Gamma}\left\{\Lambda^b_{\varphi-u_l}(x)\right\} \leq \min\left\{2C \;,\; 2C \max_{h\in\{0,\ldots,b\}}\left\{\mathrm{dist}(x, \Gamma_l)^{\mu-h}\right\} + \theta_0 e^{\mathrm{dist}(x,\Gamma_l)}\right\}. \tag{3.58}$$

The estimate claimed in (3.49) is obtained by taking $l := m$ in (3.58).

It remains only to establish the separation of the points $z_1, \ldots, z_m$ as claimed in (3.48). For this purpose, consider distinct $s, j \in \{1, \ldots, m\}$ and without loss of generality assume $s < j$. By assumption the **Lip(μ) GRIM** algorithm does not terminate on any of the first m steps. Therefore $z_j := \mathrm{argmax}_{x\in\Gamma} \Lambda^b_{\varphi-u_{j-1}}(x)$ must satisfy that

$$\Lambda^b_{\varphi-u_{j-1}}(z_j) = \max_{h\in\{0,\ldots,b\}} \left|\left|\varphi^{(h)}(z_j) - u_{j-1}^{(h)}(z_j)\right|\right|_{\mathcal{L}(V^{\otimes h};W)} > \theta. \tag{3.59}$$

Observe that $\varphi, u_{j-1} \in \mathrm{Lip}(\mu, \Gamma, W)$, with (cf. (3.51) and (3.55)) $||\varphi||_{\mathrm{Lip}(\mu,\Gamma,W)}, ||u_{j-1}||_{\mathrm{Lip}(\mu,\Gamma,W)} \leq C$. Further, for any $a \in \{0, \ldots, n\}$ and any $i \in \{1, \ldots, j-1\}$, we have $\left|\left|\varphi^{(a)}(z_i) - u_{j-1}^{(a)}(z_i)\right|\right|_{\mathcal{L}(V^{\otimes a};W)} \leq \Lambda^n_{\varphi-u_{j-1}}(z_i) \leq \theta_0$. Since the definition of r here in (3.53) matches the specification of the constant δ_0 in (B.13) of Theorem B.8 for l there as b here, we can apply TheoremB.8, with the choices $\Sigma := \overline{\mathbb{B}}_V(z_i, r) \cap \Gamma$, $B := \{z_i\}$, $\varepsilon := \theta$, $\varepsilon_0 := \theta_0$, $K_0 := C$ and $\gamma := \mu$, to conclude that for every $h \in \{0, \ldots, b\}$ and every $x \in \overline{\mathbb{B}}_V(z_i, r) \cap \Gamma$ we have

$$\left|\left|\varphi^{(h)}(x) - u_{j-1}^{(h)}(x)\right|\right|_{\mathcal{L}(V^{\otimes h};W)} \leq \theta. \tag{3.60}$$

The arbitrariness of $h \in \{0, \ldots, b\}$ allows us to take the maximum over $h \in \{0, \ldots, b\}$ in (3.60) and conclude that for every $x \in \overline{\mathbb{B}}_V(z_i, r) \cap \Gamma$ we have

$$\Lambda^b_{\varphi-u_{j-1}}(x) \leq \theta. \tag{3.61}$$

The arbitrariness of $i \in \{1, \ldots, j-1\}$ allows us to conclude that (3.61) is valid for every $i \in \{1, \ldots, j-1\}$. Together, (3.59) and (3.61) mean that for every $i \in \{1, \ldots, j-1\}$ we must have that $z_j \notin \overline{\mathbb{B}}_V(z_i, r) \cap \Gamma$. Since $s \in \{1, \ldots, j-1\}$, we conclude that $z_j \notin \overline{\mathbb{B}}_V(z_s, r) \cap \Gamma$, i.e. that $||z_j - z_s||_V > r$ as claimed in (3.48). This completes the proof of Lemma 3.5. ■

The point separation obtained in Lemma 3.5 allows us to establish an upper bound for the number of steps the **Lip(γ) GRIM** algorithm, with the choice of $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, carries out before terminating. This bound is the content of the following result.

Lemma 3.6 (Lip(μ) GRIM Number of Steps Bound). *Assume V and W are finite dimensional Banach spaces with $\Gamma \subset V$ compact. Assume that the tensor products of V are all equipped with admissible norms (cf. Definition A.1). Let $\theta, \mu > 0$ with $n \in \mathbb{Z}_{\geq 0}$ such that $\mu \in (n, n+1]$, $b \in \{0, \ldots, n\}$, and let $\mathcal{N} \in \mathbb{Z}_{>0}$ be a positive integer.*

Assume that for every $i \in \{1, \ldots, \mathcal{N}\}$ the function $f_i = \left(f_i^{(0)}, \ldots, f_i^{(n)}\right) \in \operatorname{Lip}(\mu, \Gamma, W)$ is non-zero and define $\mathcal{F} := \{f_i : i \in \{1, \ldots, \mathcal{N}\}\} \subset \operatorname{Lip}(\mu, \Gamma, W)$. Given $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$, define $\varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(n)}\right) \in \operatorname{Span}(\mathcal{F}) \subset \operatorname{Lip}(\mu, \Gamma, W)$ and $C > 0$ by

$$\textbf{(I)} \quad \varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i, \quad \left(\begin{array}{c} \varphi^{(l)} := \sum_{i=1}^{\mathcal{N}} a_i f_i^{(l)} \\ \textit{for every } l \in \{0, \ldots, n\} \end{array} \right) \quad \textit{and} \quad \textbf{(II)} \quad C := \sum_{i=1}^{\mathcal{N}} |a_i| \, ||f_i||_{\operatorname{Lip}(\mu,\Gamma,W)} > 0, \tag{3.62}$$

and $D > 0$ by $D := \dim(W) \sum_{s=0}^{n} \left(\begin{array}{c} \dim(V) + s - 1 \\ s \end{array} \right)$. Let $0 \leq \theta_0 < \min\{2C, \theta\}$. Then there is a positive constant $r = r(C, \mu, \theta, \theta_0, b) > 0$, given by

$$r := \sup\left\{\lambda > 0 \ : \ 2C\lambda^{\mu - b} + \theta_0 e^{\lambda} \leq \min\{2C, \theta\}\right\}, \tag{3.63}$$

and a non-negative integer $N = N(\Gamma, C, \mu, \theta, \theta_0, b) \in \mathbb{Z}_{\geq 0}$, given by the r-packing number of Γ

$$N := N_{\text{pack}}(\Gamma, V, r) \max\left\{d \in \mathbb{Z} : \exists x_1, \ldots, x_d \in \Gamma \textit{ for which } ||x_a - x_c||_V > r \textit{ if } a \neq c\right\}, \tag{3.64}$$

for which the following is true.

If we apply the **$\operatorname{Lip}(\boldsymbol{\mu})$ GRIM** *algorithm, with the choice of $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, to approximate φ on Γ, with the target accuracy threshold,the acceptable recombination error bound, and the order level in* **$\operatorname{Lip}(\boldsymbol{\mu})$ GRIM** *(A) as θ, θ_0, and b respectively, then after at most N steps the algorithm terminates. That is, if we let $M \in \{1, \ldots, N\}$ be the integer for which the algorithm terminates after step M and $Q_M := \min\{\mathcal{N}, 1 + MD\}$, there are coefficients $c_{M,1}, \ldots, c_{M,Q_M} \in \mathbb{R}$ and indices $e_M(1), \ldots, e_M(Q_M) \in \{1, \ldots, \mathcal{N}\}$ with*

$$\sum_{s=1}^{Q_M} |c_{M,s}| \, \left|\left|f_{e_M(s)}\right|\right|_{\operatorname{Lip}(\mu,\Gamma,W)} = C, \tag{3.65}$$

and such that $u_M = \left(u_M^{(0)}, \ldots, u_M^{(n)}\right) \in \operatorname{Span}(F) \subset \operatorname{Lip}(\mu, \Gamma, W)$ defined by

$$u_M := \sum_{s=1}^{Q_M} c_{M,s} f_{e_M(s)}, \quad \textit{so that for every } l \in \{0, \ldots, n\} \textit{ we have} \quad u_M^{(l)} := \sum_{s=1}^{Q_M} c_{M,s} f_{e_M(s)}^{(l)}, \tag{3.66}$$

satisfies, for every $z \in \Gamma$, that

$$\Lambda^b_{\varphi - u_M}(z) = \max_{s \in \{0, \ldots, b\}} \left|\left|\varphi^{(s)}(z) - u_M^{(s)}(z)\right|\right|_{\mathcal{L}(V^{\otimes s}; W)} \leq \theta. \tag{3.67}$$

Moreover, if the coefficients $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ corresponding to φ (cf. **(I)** *of (3.62)) are all positive (i.e. $a_1, \ldots, a_{\mathcal{N}} > 0$) then the coefficients $c_{M,1}, \ldots, c_{M,Q_M} \in \mathbb{R}$ corresponding to u_M (cf. (3.66)) are all non-negative (i.e. $c_{M,1}, \ldots, c_{M,Q_M} \geq 0$).*

Remark 3.7. Lemma 3.6 guarantees that if N defined in (3.64) satisfies $N < \frac{\mathcal{N}-1}{D}$ then the **$\operatorname{Lip}(\boldsymbol{\mu})$ GRIM** algorithm, with the choice that $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, will find an approximation $u = \left(u^{(0)}, \ldots, u^{(n)}\right) \in \operatorname{Span}(\mathcal{F}) \subset \operatorname{Lip}(\mu, \Gamma, W)$ of φ that is a linear combination of *less* than $\mathcal{N}$ of the elements $f_1, \ldots, f_{\mathcal{N}}$ but such that, for every $j \in \{0, \ldots, b\}$, the function $u^{(j)}$ is within θ of $\varphi^{(j)}$ throughout Γ in the sense that for every $p \in \Gamma$ we have $\left|\left|\varphi^{(j)}(p) - u^{(j)}(p)\right|\right|_{\mathcal{L}(V^{\otimes j}; W)} \leq \theta$.

Remark 3.8. By invoking Lemma 3.5 (cf. (3.49)) we can additionally conclude that, for every $m \in \{1, \ldots, M\}$, the approximation $u_m = \left(u_m^{(0)}, \ldots, u_m^{(n)}\right) \in \operatorname{Span}(\mathcal{F}) \subset \operatorname{Lip}(\mu, \Gamma, W)$ of φ found at the m^{th} step satisfies that for every $z \in \Gamma$ the quantity $\Lambda^b_{\varphi - u_m}(z) := \max_{h \in \{0, \ldots, b\}} \left|\left|\varphi^{(h)}(z) - u_m^{(h)}(z)\right|\right|_{\mathcal{L}(V^{\otimes j}; W)}$ satisfies

$$\Lambda^b_{\varphi - u_m}(z) \leq \min\left\{2C \ , \ 2C \max_{h \in \{0, \ldots, b\}} \left\{\operatorname{dist}(z, \Gamma_m)^{\mu - h}\right\} + \theta_0 e^{\operatorname{dist}(z, \Gamma_m)}\right\} \tag{3.68}$$

where Γ_m denotes the m-points in Γ that have been selected once the m^{th} step is complete.

Proof of Lemma 3.6. Let V and W are finite dimensional Banach spaces with $\Gamma \subset V$ compact. Assume that the tensor products of V are all equipped with admissible norms (cf. Definition A.1). Let $\theta, \mu > 0$ with $n \in \mathbb{Z}_{\geq 0}$ such that $\mu \in (n, n+1]$, $b \in \{0, \ldots, n\}$, and let $\mathcal{N} \in \mathbb{Z}_{>0}$ be a positive integer. Assume $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$, and that for every $i \in \{1, \ldots, \mathcal{N}\}$ the function $f_i = \left(f_i^{(0)}, \ldots, f_i^{(n)}\right) \in \mathrm{Lip}(\mu, \Gamma, W)$ is non-zero and define $\mathcal{F} := \{f_i \ : \ i \in \{1, \ldots, \mathcal{N}\}\} \subset \mathrm{Lip}(\mu, \Gamma, W)$. Define $\varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(n)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\mu, \Gamma, W)$ as in **(I)** of (3.62) and the constant C as in **(II)** of (3.62).

With a view to applying **Lip(μ) GRIM** to approximate φ on Γ, for each $i \in \{1, \ldots, \mathcal{N}\}$ let $\tilde{a}_i := |a_i|$ and $\tilde{f}_i$ be given by f_i if $a_i > 0$ and $-f_i$ if $a_i < 0$. Observe that for every $i \in \{1, \ldots, \mathcal{N}\}$ we have that $\left|\left|\tilde{f}_i\right|\right|_{\mathrm{Lip}(\mu,\Gamma,W)} = ||f_i||_{\mathrm{Lip}(\mu,\Gamma,W)}$. Moreover, we also have that $\tilde{a}_1, \ldots, \tilde{a}_{\mathcal{N}} > 0$ and that $\varphi = \sum_{i=1}^{\mathcal{N}} \tilde{a}_i \tilde{f}_i$. Further, we rescale $\tilde{f}_i$ for each $i \in \{1, \ldots, \mathcal{N}\}$ to have unit $\mathrm{Lip}(\mu, \Gamma, W)$ norm. That is (cf. **Lip(μ) GRIM** (B)), for each $i \in \{1, \ldots, \mathcal{N}\}$ set $h_i := \frac{\tilde{f}_i}{||f_i||_{\mathrm{Lip}(\mu,\Gamma,W)}}$ and $\alpha_i := \tilde{a}_i ||f_i||_{\mathrm{Lip}(\mu,\Gamma,W)}$. Then observe both that C satisfies

$$C = \sum_{i=1}^{\mathcal{N}} |a_i| \, ||f_i||_{\mathrm{Lip}(\mu,\Gamma)} = \sum_{i=1}^{\mathcal{N}} \tilde{a}_i ||f_i||_{\mathrm{Lip}(\mu,\Gamma)} = \sum_{i=1}^{\mathcal{N}} \alpha_i, \tag{3.69}$$

and, for every $i \in \{1, \ldots, \mathcal{N}\}$, that $\alpha_i h_i = \tilde{a}_i \tilde{f}_i = a_i f_i$. Thus the expansion for φ in **(I)** of (3.62) is equivalent to

$$\varphi = \sum_{i=1}^{\mathcal{N}} \alpha_i h_i, \qquad \text{and hence} \qquad ||\varphi||_{\mathrm{Lip}(\mu,\Gamma,W)} \leq \sum_{i=1}^{\mathcal{N}} \alpha_i ||h_i||_{\mathrm{Lip}(\mu,\Gamma,W)} = \sum_{i=1}^{\mathcal{N}} \alpha_i \overset{(3.69)}{=} C. \tag{3.70}$$

Let $0 \leq \theta_0 < \min\{2C, \theta\}$. Define $D > 0$ by $D := \dim(W) \sum_{s=0}^{k} \binom{\dim(V) + s - 1}{s}$. Retrieve the constant $r = r(C, \mu, \theta, \theta_0, b) > 0$ arising in Lemma 3.5. Examining (3.47) in Lemma 3.5 reveals this amounts to defining

$$r := \sup\left\{\lambda > 0 \ : \ 2C\lambda^{\mu-b} + \theta_0 e^{\lambda} \leq \min\{2C, \theta\}\right\} \tag{3.71}$$

as claimed in (3.63). Finally define $N = N(\Gamma, C, \mu, \theta, \theta_0, b) \in \mathbb{Z}_{\geq 1}$ to be the r-packing number of Γ as in (3.64). That is

$$N := N_{\mathrm{pack}}(\Gamma, V, r) = \max\{d \in \mathbb{Z} : \exists x_1, \ldots, x_d \in \Gamma \text{ for which } ||x_a - x_c||_V > r \text{ if } a \neq c\}. \tag{3.72}$$

Now consider applying the **Lip(μ) GRIM** algorithm, with the choice $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, to approximate φ on Γ with the target accuracy threshold, the acceptable recombination error bound, and the order level in **Lip(μ) GRIM** (A) as θ, θ_0, and b here respectively.

Suppose $m \in \mathbb{Z}_{\geq 1}$ and that the algorithm does not terminate before or at the m^{th} step. Therefore, recalling **Lip(μ) GRIM** (C) and (D), distinct points $z_1, \ldots, z_m \in \Gamma$ have been selected, and recombination has been used via Lemma 3.1 to find an approximation $u_m = \left(u_m^{(0)}, \ldots, u_m^{(n)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\mu, \Gamma, W)$ of φ satisfying, for each $s \in \{1, \ldots, m\}$, that $\Lambda^n_{\varphi - u_m}(z_s) \leq \theta_0$. That is, given any $j \in \{0, \ldots, n\}$ and any $s \in \{1, \ldots, m\}$ we have $\left|\left|\varphi^{(j)}(z_s) - u_m^{(j)}(z_s)\right|\right|_{\mathcal{L}(V^{\otimes j};W)} \leq \theta_0$. Lemma 3.5 may be applied to obtain, for every $s, t \in \{1, \ldots, m\}$ with $s \neq t$, that

$$||z_s - z_t||_V > r. \tag{3.73}$$

A consequence of (3.73) is that N defined in (3.72) must satisfy that $N \geq m$. Therefore if the **Lip(μ) GRIM** algorithm, with the choice that $k_s := 1$ for every integer $s \in \mathbb{Z}_{\geq 1}$, does not terminate at the m^{th} step we must have that $m \leq N$, and so the algorithm terminates after at most N steps as claimed.

Let $M \in \{1, \ldots, N\}$ be the integer for which the **Lip(μ) GRIM** algorithm must terminate after step M. The termination criterion (cf. **Lip(μ) GRIM** (D)) means that the $\mathrm{Lip}(\mu, \Gamma, W)$ function $u_M = \left(u_M^{(0)}, \ldots, u_M^{(n)}\right) \in$

$\mathrm{Span}(\mathcal{F})$ found at the M^{th} step must satisfy, for every $z \in \Gamma$, that

$$\Lambda^{b}_{\varphi - u_M}(z) := \max_{s \in \{0,\dots,b\}} \left|\left| \varphi^{(s)}(z) - u_M^{(s)}(z) \right|\right|_{\mathcal{L}(V^{\otimes s};W)} \leq \theta \tag{3.74}$$

as claimed in (3.67). Moreover, if we let $Q_M := \min\{\mathcal{N}, 1 + MD\}$, then Lemma 3.1 tells us that there are non-negative coefficients $b_{M,1},\dots,b_{M,Q_M} \geq 0$, with

$$\sum_{s=1}^{Q_M} b_{M,s} = \sum_{i=1}^{\mathcal{N}} \alpha_i \overset{(3.69)}{=} C, \tag{3.75}$$

and indices $e_M(1),\dots,e_M(Q_M) \in \{1,\dots,\mathcal{N}\}$, for which the element u_M is given by

$$u_M := \sum_{s=1}^{Q_M} b_{M,s} h_{e_M(s)} \quad \text{so that for every } j \in \{0,\dots,n\} \quad u_M^{(j)} = \sum_{s=1}^{Q_M} b_{M,s} h^{(j)}_{e_M(s)}. \tag{3.76}$$

A consequence of (3.76) is that

$$u_M := \sum_{s=1}^{Q_M} b_{M,s} h_{e_M(s)} = \sum_{s=1}^{Q_M} \frac{b_{M,s}}{||f_{e_M(s)}||_{\mathrm{Lip}(\mu,\Gamma,W)}} \tilde{f}_{e_M(s)}. \tag{3.77}$$

For each $s \in \{1,\dots,Q_M\}$, define $c_{M,s} := \frac{b_{M,s}}{||f_{e_M(s)}||_{\mathrm{Lip}(\mu,\Gamma,W)}}$ if $\tilde{f}_{e_M(s)} = f_{e_M(s)}$ (which, we recall, is the case if $a_{e_M(s)} > 0$) and $c_{M,s} := -\frac{b_{M,s}}{||f_{e_M(s)}||_{\mathrm{Lip}(\mu,\Gamma,W)}}$ if $\tilde{f}_{e_M(s)} = -f_{e_M(s)}$ (which, we recall, is the case if $a_{e_M(s)} < 0$). Then (3.77) gives the expansion for $u_M \in \mathrm{Lip}(\mu,\Gamma,W)$ in terms of the functions $f_1,\dots,f_{\mathcal{N}}$ claimed in (3.66). Moreover, from (3.75) we have that

$$\sum_{s=1}^{Q_M} |c_{M,s}| \, ||f_{e_M(s)}||_{\mathrm{Lip}(\mu,\Gamma,W)} = \sum_{s=1}^{Q_M} b_{M,s} \overset{(3.75)}{=} C \tag{3.78}$$

as claimed in (3.65).

It remains only to prove that if the coefficients $a_1,\dots,a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ corresponding to φ (cf. (**I**) of (3.62)) are all positive (i.e. $a_1,\dots,a_{\mathcal{N}} > 0$) then the coefficients $c_{M,1},\dots,c_{M,Q_M} \in \mathbb{R}$ corresponding to u_M (cf. (3.77)) are all non-negative (i.e. $c_{M,1},\dots,c_{M,Q_M} \geq 0$). First note that $a_1,\dots,a_{\mathcal{N}} > 0$ means, for every $i \in \{1,\dots,\mathcal{N}\}$, that $\tilde{f}_i = f_i$. Consequently, for every $s \in \{1,\dots,Q_M\}$, we have that $\tilde{f}_{e_M(s)} = f_{e_M(s)}$, and so be definition $c_{M,s} = \frac{b_{M,s}}{||f_{e_M(s)}||_{\mathrm{Lip}(\mu,\Gamma,W)}}$. Since $b_{M,s} \geq 0$, it follows that $c_{M,s} \geq 0$. This completes the proof of Lemma 3.6. ■

Suppose that $\Gamma \subset \Omega \subset V$, and that for every $i \in \{1,\dots,\mathcal{N}\}$ the element f_i admits a $\mathrm{Lip}(\mu)$ extension to Ω. Then the following result records a quantified version of the statement that an approximation of φ on Γ, of the form found by the **Lip(μ) GRIM** algorithm, may be extended to approximate φ on an enlargement of the set Γ. The precise result is the following lemma.

Lemma 3.9 (Lip(μ) GRIM Robustness). *Assume V and W are finite dimensional Banach spaces with $\Gamma \subset \Omega \subset V$ both closed subsets. Assume that the tensor products of V are all equipped with admissible norms (cf. Definition A.1). Let $\theta, \mu > 0$ with $n \in \mathbb{Z}_{\geq 0}$ such that $\mu \in (n, n+1]$, $b \in \{0,\dots,n\}$, $R \geq 1$, $T > 1$, and let $\mathcal{N} \in \mathbb{Z}_{>0}$ be a positive integer. Assume that for every $i \in \{1,\dots,\mathcal{N}\}$ the function $f_i = \left(f_i^{(0)},\dots,f_i^{(n)}\right) \in \mathrm{Lip}(\mu,\Gamma,W)$ is non-zero, and that $F_i = \left(F_i^{(0)},\dots,F_i^{(n)}\right) \in \mathrm{Lip}(\mu,\Omega,W)$ is an extension of f_i, in the sense that $F_i|_\Gamma \equiv f_i$, satisfying the norm estimate that $||F_i||_{\mathrm{Lip}(\mu,\Omega,W)} \leq R||f_i||_{\mathrm{Lip}(\mu,\Gamma,W)}$. Let $a_1,\dots,a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ and define*

$\varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(n)}\right) \in \mathrm{Lip}(\mu, \Gamma, W)$ and $C > 0$ by

$$\textbf{(I)} \quad \varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i, \quad \left(\begin{array}{c} \varphi^{(l)} := \sum_{i=1}^{\mathcal{N}} a_i f_i^{(l)} \\ \textit{for every } l \in \{0, \ldots, n\} \end{array} \right) \quad \textit{and} \quad \textbf{(II)} \quad C := \sum_{i=1}^{\mathcal{N}} |a_i| ||f_i||_{\mathrm{Lip}(\mu,\Gamma,W)} > 0. \tag{3.79}$$

Assume that $T\theta \leq 2RC$. Then there is a positive constant $r = r(C, \mu, \theta, R, T, b) > 0$, given by

$$r := \sup \left\{ \lambda > 0 \ : \ \theta e^{\lambda} + 2RC \left(\lambda^{\mu - b} + e^{\lambda} - 1 \right) \leq T\theta \right\}, \tag{3.80}$$

for which the following is true.

Suppose that $Q \in \mathbb{Z}_{\geq 1}$ and there are coefficients $c_1, \ldots, c_Q \in \mathbb{R}$ and indices $e(1), \ldots, e(Q) \in \{1, \ldots, \mathcal{N}\}$ such that

$$\sum_{s=1}^{Q} |c_s| \, \left|\left| f_{e(s)} \right|\right|_{\mathrm{Lip}(\mu,\Gamma,W)} \leq C \quad \textit{and} \quad u := \sum_{s=1}^{Q} c_s f_{e(s)} \quad \textit{satisfies} \quad \Lambda^b_{\varphi - u}(z) \leq \theta \tag{3.81}$$

for every $z \in \Gamma$. Then the extensions $\hat{\varphi}, \hat{u} \in \mathrm{Lip}(\mu, \Omega, W)$, of φ and u respectively, defined by

$$\hat{\varphi} := \sum_{i=1}^{\mathcal{N}} a_i F_i \qquad \textit{and} \qquad \hat{u} := \sum_{s=1}^{Q} c_s F_{e(s)} \tag{3.82}$$

satisfy, for every $z \in \Omega \cap \Gamma_r$, that

$$\Lambda^b_{\hat{\varphi} - \hat{u}}(z) = \max_{s \in \{0, \ldots, b\}} \left|\left| \hat{\varphi}^{(s)}(z) - \hat{u}^{(s)}(z) \right|\right|_{\mathcal{L}(V^{\otimes s}; W)} \leq T\theta \tag{3.83}$$

where Γ_r denotes the r-fattening of Γ, i.e. $\Gamma_r := \{u \in V \ : \ \exists z \in \Gamma \text{ such that } ||z - u||_V \leq r\}$.

Proof of Lemma 3.9. Let V and W are finite dimensional Banach spaces with $\Gamma \subset \Omega \subset V$ both closed subsets. Assume that the tensor products of V are all equipped with admissible norms (cf. Definition A.1). Let $\theta, \mu > 0$ with $n \in \mathbb{Z}_{\geq 0}$ such that $\mu \in (n, n+1]$, $b \in \{0, \ldots, n\}$, $R \geq 1$, $T > 1$, and let $\mathcal{N} \in \mathbb{Z}_{>0}$ be a positive integer. Assume that for every $i \in \{1, \ldots, \mathcal{N}\}$ the function $f_i = \left(f_i^{(0)}, \ldots, f_i^{(n)} \right) \in \mathrm{Lip}(\mu, \Gamma, W)$ is non-zero, and that $F_i = \left(F_i^{(0)}, \ldots, F_i^{(n)} \right) \in \mathrm{Lip}(\mu, \Omega, W)$ is an extension of f_i, in the sense that $F_i|_{\Gamma} \equiv f_i$, satisfying the norm estimate that $||F_i||_{\mathrm{Lip}(\mu,\Omega,W)} \leq R||f_i||_{\mathrm{Lip}(\mu,\Gamma,W)}$. Let $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ and define $\varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(n)}\right) \in \mathrm{Lip}(\mu, \Gamma, W)$ and $C > 0$ as in **(I)** and **(II)** of (3.79) respectively. Assume that $T\theta \leq 2RC$. Define $r = r(C, \mu, \theta, R, T, b) > 0$ as in (3.80), that is

$$r := \sup \left\{ \lambda \in (0, 1] \ : \ \theta e^{\lambda} + 2RC \left(\lambda^{\mu - b} + e^{\lambda} - 1 \right) \leq T\theta \right\}. \tag{3.84}$$

A consequence of (3.84) is that

$$\theta e^r + 2RC \left(r^{\mu - s} + e^r - 1 \right) \leq \theta e^r + 2RC \left(r^{\mu - b} + e^r - 1 \right) \leq T\theta \leq 2RC. \tag{3.85}$$

Fix this value of r for the remainder of the proof.

Now assume $Q \in \mathbb{Z}_{\geq 1}$ with coefficients $c_1, \ldots, c_Q \in \mathbb{R}$ and indices $e(1), \ldots, e(Q) \in \{1, \ldots, \mathcal{N}\}$ such that

$$\sum_{s=1}^{Q} |c_s| \, \left|\left| f_{e(s)} \right|\right|_{\mathrm{Lip}(\mu,\Gamma,W)} \leq C \quad \text{and} \quad u := \sum_{s=1}^{Q} c_s f_{e(s)} \quad \text{satisfies, for every } z \in \Gamma, \text{ that} \quad \Lambda^b_{\varphi - u}(z) \leq \theta. \tag{3.86}$$

The final part of (3.86) means that whenever $l \in \{0, \ldots, b\}$ and $z \in \Gamma$ we have $\left|\left| \varphi^{(l)}(z) - u^{(l)}(z) \right|\right|_{\mathcal{L}(V^{\otimes l}; W)} \leq \theta$. Consider the extensions $\hat{\varphi}, \hat{u} \in \mathrm{Lip}(\mu, \Omega, W)$, of φ and u respectively, defined by

$$\hat{\varphi} := \sum_{i=1}^{\mathcal{N}} a_i F_i \qquad \text{and} \qquad \hat{u} := \sum_{s=1}^{Q} c_s F_{e(s)}. \tag{3.87}$$

Then it follows from (3.87) that

$$||\hat{\varphi}||_{\mathrm{Lip}(\mu,\Omega,W)} \overset{(3.87)}{\leq} \sum_{i=1}^{\mathcal{N}} |a_i| ||F_i||_{\mathrm{Lip}(\mu,\Omega,W)} \leq R \sum_{i=1}^{\mathcal{N}} |a_i| ||f_i||_{\mathrm{Lip}(\mu,\Gamma,W)} \overset{\textbf{(II)} \text{ in } (3.79)}{=} RC, \tag{3.88}$$

and that

$$||\hat{u}||_{\mathrm{Lip}(\mu,\Omega,W)} \overset{(3.87)}{\leq} \sum_{s=1}^{Q} |c_s| \, ||F_{e(s)}||_{\mathrm{Lip}(\mu,\Omega,W)} \leq R \sum_{s=1}^{Q} |c_s| \, ||f_{e(s)}||_{\mathrm{Lip}(\mu,\Gamma,W)} \overset{(3.86)}{\leq} RC. \tag{3.89}$$

Let $z \in \Omega$ and $x \in \Gamma$. Then $\hat{\varphi} - \hat{u} \in \mathrm{Lip}(\mu,\Omega,W)$ satisfies that $||\hat{\varphi} - \hat{u}||_{\mathrm{Lip}(\mu,\Omega,W)} \leq 2RC$ and, for every $j \in \{0,\ldots,b\}$, that $\left|\left|\hat{\varphi}^{(j)}(x) - \hat{u}^{(j)}(x)\right|\right|_{\mathcal{L}(V^{\otimes j};W)} \leq \theta$. Let $s \in \{0,\ldots,b\}$. By applying Lemma C.1 to $\hat{\varphi} - \hat{u}$, with the A, r_0, p, ρ, n and q of that result as $2RC$, θ, x, μ, n and b here respectively, we deduce that (cf. (C.3))

$$\left|\left|\hat{\varphi}^{(s)}(z) - \hat{u}^{(s)}(z)\right|\right|_{\mathcal{L}(V^{\otimes s};W)} \leq \min\left\{2RC \ , \ \theta \sum_{j=0}^{b-s} \frac{1}{j!} \operatorname{dist}(z,x)^j + 2RC\left(\operatorname{dist}(z,x)^{\mu-s} + S(z,x)\right)\right\} \tag{3.90}$$

where (cf. (C.4))

$$S(z,x) := \begin{cases} \sum_{j=b-s+1}^{n-s} \frac{1}{j!} \operatorname{dist}(z,x)^j & \text{if } b < n \\ 0 & \text{if } b = n. \end{cases} \tag{3.91}$$

It follows from (3.91) that $S(z,x) \leq e^{\operatorname{dist}(z,x)} - 1$, and so (3.90) yields

$$\left|\left|\hat{\varphi}^{(s)}(z) - \hat{u}^{(s)}(z)\right|\right|_{\mathcal{L}(V^{\otimes s};W)} \leq \min\left\{2RC \ , \ \theta e^{\operatorname{dist}(z,x)} + 2RC\left(\operatorname{dist}(z,x)^{\mu-s} + e^{\operatorname{dist}(z,x)} - 1\right)\right\}. \tag{3.92}$$

The arbitrariness of $x \in \Gamma$ allows us to take the infimum over $x \in \Gamma$ in (3.92) to obtain

$$\left|\left|\hat{\varphi}^{(s)}(z) - \hat{u}^{(s)}(z)\right|\right|_{\mathcal{L}(V^{\otimes s};W)} \leq \min\left\{2RC \ , \ \theta e^{\operatorname{dist}(z,\Gamma)} + 2RC\left(\operatorname{dist}(z,\Gamma)^{\mu-s} + e^{\operatorname{dist}(z,\Gamma)} - 1\right)\right\}. \tag{3.93}$$

Moreover, the arbitrariness of $z \in \Omega$ allows us to conclude that (3.93) is valid for every $z \in \Omega$. Now define $\Phi := \Omega \cap \Gamma_r$, with $\Gamma_r := \{u \in V \ : \ \exists z \in \Gamma \text{ such that } ||z-u||_V \leq r\}$. Since $\Phi \subset \Omega$, given any $z \in \Phi$ we may conclude that (3.93) holds for z. But $z \in \Phi$ means $\operatorname{dist}(z,\Gamma) \leq r$, and so (3.93) becomes

$$\begin{aligned}\left|\left|\hat{\varphi}^{(s)}(z) - \hat{u}^{(s)}(z)\right|\right|_{\mathcal{L}(V^{\otimes s};W)} &\leq \min\left\{2RC \ , \ \theta e^r + 2RC\left(r^{\mu-s} + e^r - 1\right)\right\} \\ &\leq \min\left\{2RC \ , \ \theta e^r + 2RC\left(r^{\mu-b} + e^r - 1\right)\right\} \\ &\overset{(3.85)}{\leq} \min\left\{2RC, T\theta\right\} = T\theta\end{aligned}$$

where we have used that $r \leq 1$ means $r^{\mu-s} \leq r^{\mu-b}$ since $s \in \{0,\ldots,b\}$. The arbitrariness of $s \in \{0,\ldots,b\}$ means that we have established that

$$\Lambda^b_{\hat{\varphi}-\hat{u}}(z) = \max_{s \in \{0,\ldots,b\}} \left|\left|\hat{\varphi}^{(s)}(z) - \hat{u}^{(s)}(z)\right|\right|_{\mathcal{L}(V^{\otimes s};W)} \leq T\theta. \tag{3.94}$$

The arbitrariness of $z \in \Phi$ means that (3.94) is valid for any $z \in \Phi$ as claimed in (3.83). This completes the proof of Lemma 3.9 ∎

We combine Lemmas 3.5, 3.6 and 3.9 to prove Theorem 3.2.

Proof of Theorem 3.2. Let V and W be finite dimensional Banach spaces with $\Sigma \subset V$ compact. Assume that the tensor products of V are all equipped with admissible norms (cf. Definition A.1). Let $\mathcal{N} \in \mathbb{Z}_{>0}$ be a positive integer and $\gamma > 0$ with $k \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k, k+1]$. Assume that for every $i \in \{1,\ldots,\mathcal{N}\}$ the function $f_i = \left(f_i^{(0)},\ldots,f_i^{(k)}\right) \in \mathrm{Lip}(\gamma,\Sigma,W)$ is non-zero and define $\mathcal{F} := \{f_i \ : \ i \in \{1,\ldots,\mathcal{N}\}\} \subset \mathrm{Lip}(\gamma,\Sigma,W)$.

Given $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$, define $\varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(k)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\gamma, \Sigma, W)$ and constants $C, D > 0$ by

$$\textbf{(I)} \quad \varphi := \sum_{i=1}^{\mathcal{N}} a_i f_i, \quad \left(\begin{array}{c} \varphi^{(l)} := \sum_{i=1}^{\mathcal{N}} a_i f_i^{(l)} \\ \text{for every } l \in \{0, \ldots, k\} \end{array} \right) \quad \text{and} \quad \textbf{(II)} \quad C := \sum_{i=1}^{\mathcal{N}} |a_i| ||f_i||_{\mathrm{Lip}(\gamma,\Sigma,W)} > 0 \tag{3.95}$$

and $D := \dim(W) \sum_{s=0}^{k} \left(\begin{array}{c} \dim(V) + s - 1 \\ s \end{array} \right)$. Let $0 \leq \varepsilon_0 < \varepsilon < 2C$, $q \in \{0, \ldots, k\}$ and $K \geq 1$, $T > 1$ with $T\varepsilon \leq 2KC$. Define positive constant $r_1 = r_1(C, \gamma, \varepsilon, \varepsilon_0, q) > 0$ and $r_2 = r_2(C, \gamma, \varepsilon, R, T, q) > 0$ as in **(I)** and **(II)** of (3.22) respectively. That is,

$$\begin{cases} \textbf{(I)} \quad r_1 := \sup\left\{\lambda > 0 \ : \ 2C\lambda^{\gamma - q} + \varepsilon_0 e^{\lambda} \leq \varepsilon\right\} \\ \textbf{(II)} \quad r_2 := \sup\left\{\lambda \in (0,1] \ : \ \varepsilon e^{\lambda} + 2KC\left(\lambda^{\gamma-q} + e^{\lambda} - 1\right) \leq T\varepsilon\right\}, \end{cases} \tag{3.96}$$

Define $N = N(\Sigma, C, \gamma, \varepsilon, \varepsilon_0, q) \in \mathbb{Z}_{\geq 0}$ to be the r_1-packing number of Σ as in (3.23). That is,

$$N := N_{\mathrm{pack}}(\Sigma, V, r_1) = \max\left\{d \in \mathbb{Z} \ : \ \exists x_1, \ldots, x_d \in \Sigma \text{ for which } ||x_a - x_c||_V > r_1 \text{ if } a \neq c\right\}, \tag{3.97}$$

Now consider applying the **Lip(γ) GRIM** algorithm to approximate φ on Σ, with the target accuracy threshold, the acceptable recombination error bound, and the order level in **Lip(γ) GRIM** (A) as ε, ε_0, and q respectively. Since the constant r_1 and the integer N here agree with the definition of the constant r and the integer N arising in Lemma 3.6, with $\Gamma := \Sigma$, $C := C$, $\mu := \gamma$, $\theta := \varepsilon$, $\theta_0 := \varepsilon_0$ and $b := q$ (and recalling that since we are assuming $\varepsilon < 2C$, we have that $\min\{2C, \theta\} = \min\{2C, \varepsilon\} = \varepsilon$), we may apply Lemma 3.6 to conclude that the algorithm terminates after M steps for some integer M satisfying $M \leq N$ as required.

A further consequence of Lemma 3.6 is that, if we let $Q_M := \min\{\mathcal{N}, 1 + MD\}$, then there are coefficients $c_{M,1}, \ldots, c_{M,Q_M} \in \mathbb{R}$ and indices $e_M(1), \ldots, e_M(Q_M) \in \{1, \ldots, \mathcal{N}\}$ with

$$\sum_{s=1}^{Q_M} |c_{M,s}| \left|\left|f_{e_M(s)}\right|\right|_{\mathrm{Lip}(\gamma,\Sigma,W)} = C, \tag{3.98}$$

and such that $u_M = \left(u_M^{(0)}, \ldots u_M^{(k)}\right) \in \mathrm{Span}(\mathcal{F}) \subset \mathrm{Lip}(\gamma, \Sigma, W)$ defined by (cf. (3.66))

$$u_M := \sum_{s=1}^{Q_M} c_{M,s} f_{e_M(s)} \quad \text{so that for every} l \in \{0, \ldots, k\} \text{ we have} \quad u_M^{(l)} := \sum_{s=1}^{Q_M} c_{M,s} f_{e_M(s)}^{(l)} \tag{3.99}$$

satisfies, for every $x \in \Sigma$, that (cf. (3.67))

$$\Lambda^q_{\varphi - u_M}(x) = \max_{j \in \{0,\ldots,q\}} \left|\left|\varphi^{(j)}(x) - u_M^{(j)}(x)\right|\right|_{\mathcal{L}(V^{\otimes j}; W)} \leq \varepsilon. \tag{3.100}$$

The claim (3.24) is a consequence of (3.98), whilst the claims in (3.25) are consequences of (3.99) and (3.100).

Lemma 3.6 additionally establishes that if the coefficients $a_1, \ldots, a_{\mathcal{N}} \in \mathbb{R} \setminus \{0\}$ associated to φ (cf. **(I)** of (3.95)) are all positive (i.e. $a_1, \ldots, a_{\mathcal{N}} > 0$) then the coefficients $c_{M,1}, \ldots, c_{M,Q_M} \in \mathbb{R}$ associated to u_M (cf. (3.99)) are all non-negative (i.e. $c_{M,1}, \ldots, c_{M,Q_M} \geq 0$, which is precisely as claimed in Theorem 3.2.

Now suppose that $\Omega \subset V$ is closed with $\Omega \subset \Sigma_{r_2}$. Here Σ_{r_2} denotes the r_2-fattening of Σ, i.e. it is defined by $\Sigma_{r_2} := \{u \in V \ : \ \exists z \in \Sigma \text{ such that } ||z - u||_V \leq r_2\}$. Assume that, for each $i \in \{1, \ldots, \mathcal{N}\}$, the element $f_i = \left(f_i^{(0)}, \ldots, f_i^{(k)}\right)$ can be extended to an element $F_i = \left(F_i^{(0)}, \ldots, F_i^{(k)}\right) \in \mathrm{Lip}(\gamma, \Omega, W)$ satisfying the estimate that $||F_i||_{\mathrm{Lip}(\gamma,\Omega,W)} \leq K||f_i||_{\mathrm{Lip}(\gamma,\Sigma,W)}$. Define an extension $\hat{\varphi} = \left(\hat{\varphi}^{(0)}, \ldots, \hat{\varphi}^{(k)}\right) \in \mathrm{Lip}(\gamma, \Omega, W)$ of φ by (cf. (3.26)) $\hat{\varphi} := \sum_{i=1}^{\mathcal{N}} a_i F_i$ so that for every $l \in \{0, \ldots, k\}$ we have $\hat{\varphi}^{(l)} = \sum_{i=1}^{\mathcal{N}} a_i F_i^{(l)}$, and define an extension $\hat{u}_M = \left(\hat{u}_M^{(0)}, \ldots, \hat{u}_M^{(k)}\right) \in \mathrm{Lip}(\gamma, \Omega, W)$ of u_M by (cf. (3.27)) $\hat{u}_M := \sum_{s=1}^{Q_M} c_{M,s} F_{e_M(s)}$ so that for every $l \in \{0, \ldots, k\}$ we have $\hat{u}_M^{(l)} = \sum_{s=1}^{Q_M} c_{M,s} F_{e_M(s)}^{(l)}$.

Since the constant r_2 here agrees with the definition of the constant r arising in Lemma 3.9, with $C := C$,

$\mu := \gamma$, $\theta := \varepsilon$, $b := q$, $R := K$ and $T := T$, we may apply Lemma 3.9 to conclude, for every $z \in \Omega \cap \Gamma_r$, that (cf. (3.83))

$$\Lambda^q_{\hat{\varphi}-\hat{u}_M}(z) = \max_{s\in\{0,\dots,b\}} \left|\left|\hat{\varphi}^{(s)}(z) - \hat{u}^{(s)}_M(z)\right|\right|_{\mathcal{L}(V^{\otimes s};W)} \leq T\varepsilon. \tag{3.101}$$

The estimates claimed in (3.28) now follow from (3.101) and the fact that $\Omega \subset \Sigma_{r_2}$ means that $\Omega \cap \Sigma_{r_2} = \Omega$. This completes the proof of Theorem 3.2. ■

Appendices

A. Lipschitz Functions

Let V and W be Banach spaces and assume that $\Sigma \subset V$ is a closed subset. In this section we define the space $\mathrm{Lip}(\gamma, \Sigma, W)$ and record some related properties. To define $\mathrm{Lip}(\gamma, \Sigma, W)$ functions, we must first make a choice of norm on the tensor powers of V. We restrict to considering norms that are *admissible* in the following sense.

Definition A.1 (Admissible Norms on Tensor Powers)**.** Let V be a Banach space. We say that its tensor powers are endowed with admissible norms if for each $n \in \mathbb{Z}_{\geq 1}$ we have equipped the tensor power $V^{\otimes n}$ of V with a norm $||\cdot||_{V^{\otimes n}}$ such that the following conditions hold.

- For each $n \in \mathbb{Z}_{\geq 1}$ the symmetric group S_n acts by isometries on $V^{\otimes n}$, i.e. for any $\rho \in S_n$ and any $v \in V^{\otimes n}$ we have
$$||\rho(v)||_{V^{\otimes n}} = ||v||_{V^{\otimes n}}. \tag{A.1}$$
The action of S_n on $V^{\otimes n}$ is given by permuting the order of the letters, i.e. if $a_1 \otimes \ldots \otimes a_n \in V^{\otimes n}$ and $\rho \in S_n$ then $\rho(a_1 \otimes \ldots \otimes a_n) := a_{\rho(1)} \otimes \ldots \otimes a_{\rho(n)}$, and the action is extended to the entirety of $V^{\otimes n}$ by linearity.

- For any $n, m \in \mathbb{Z}_{\geq 1}$ and any $v \in V^{\otimes n}$ and $w \in V^{\otimes m}$ we have
$$||v \otimes w||_{V^{\otimes (n+m)}} \leq ||v||_{V^{\otimes n}} ||w||_{V^{\otimes m}}. \tag{A.2}$$

- For any $n, m \in \mathbb{Z}_{\geq 1}$ and any $\phi \in (V^{\otimes n})^*$ and $\sigma \in (V^{\otimes m})^*$ we have
$$||\phi \otimes \sigma||_{\left(V^{\otimes (n+m)}\right)^*} \leq ||\phi||_{(V^{\otimes n})^*} ||\sigma||_{(V^{\otimes m})^*}. \tag{A.3}$$
Here, given any $k \in \mathbb{Z}_{\geq 1}$, the norm $||\cdot||_{(V^{\otimes k})^*}$ denotes the dual-norm induced by $||\cdot||_{V^{\otimes k}}$.

It turns out (see [Rya13]) that having *both* the inequalities (A.2) and (A.3) ensures that we in fact have equality in both estimates. Hence if the tensor powers of V are equipped with admissible norms, we have equality in both (A.2) and (A.3).

We can now give a rigorous definition of a $\mathrm{Lip}(\gamma, \Sigma, W)$ function.

Definition A.2 ($\mathrm{Lip}(\gamma, \Sigma, W)$ functions)**.** Let V and W be Banach spaces, $\Sigma \subset V$ a closed subset, and assume that the tensor powers of V are all equipped with admissible norms. Let $\gamma > 0$ with $k \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k, k+1]$. Suppose that $\psi^{(0)} : \Sigma \to W$ is a function, and that for each $l \in \{1, \ldots, k\}$ we have a function $\psi^{(l)} : \Sigma \to \mathcal{L}(V^{\otimes l}; W)$ taking its values in the space of symmetric l-linear forms from V to W. Then the collection $\psi = (\psi^{(0)}, \psi^{(1)}, \ldots, \psi^{(k)})$ is an element of $\mathrm{Lip}(\gamma, \Sigma, W)$ if there exists a constant $M \geq 0$ for which the following conditions hold:

- For each $l \in \{0, \ldots, k\}$ and every $x \in \Sigma$ we have that
$$||\psi^{(l)}(x)||_{\mathcal{L}(V^{\otimes l};W)} \leq M \tag{A.4}$$

- For each $j \in \{0, \ldots, k\}$ define $R_j^{\psi} : \Sigma \times \Sigma \to \mathcal{L}(V^{\otimes j}; W)$ for $z, p \in \Sigma$ and $v \in V^{\otimes j}$ by

$$R_j^{\psi}(z,p)[v] := \psi^{(j)}(p)[v] - \sum_{s=0}^{k-j} \frac{1}{s!} \psi^{(j+s)}(z) \left[v \otimes (p-z)^{\otimes s} \right]. \tag{A.5}$$

Then whenever $l \in \{0, \ldots, k\}$ and $x, y \in \Sigma$ we have

$$\left|\left| R_l^{\psi}(x,y) \right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq M ||y-x||_V^{\gamma - l}. \tag{A.6}$$

We sometimes say that $\psi \in \mathrm{Lip}(\gamma, \Sigma, W)$ without explicitly mentioning the functions $\psi^{(0)}, \ldots, \psi^{(k)}$. Furthermore, given $l \in \{0, \ldots, k\}$, we introduce the notation that $\psi_{[l]} := (\psi^{(0)}, \ldots, \psi^{(l)})$. The $\mathrm{Lip}(\gamma, \Sigma, W)$ norm of ψ, denoted by $||\psi||_{\mathrm{Lip}(\gamma,\Sigma,W)}$, is the smallest $M \geq 0$ satisfying the requirements (A.4) and (A.6).

Suppose $\gamma > 0$ with $k \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k, k+1]$, and let $\psi = (\psi^{(0)}, \ldots, \psi^{(k)}) \in \mathrm{Lip}(\gamma, \Sigma, W)$. A good way to understand a $\mathrm{Lip}(\gamma, \Sigma, W)$ function is as a function that "locally looks like a polynomial function". That is, for each $x \in \Sigma$ we can define the function $\Psi_x : V \to W$ for $y \in V$ by

$$\Psi_x(y) := \sum_{s=0}^{k} \frac{1}{s!} \psi^{(s)}(x) \left[(y-x)^{\otimes s} \right]. \tag{A.7}$$

The function Ψ_x defined in (A.7) gives a proposal, based at the point $x \in \Sigma$, for how the function $\psi^{(0)} : \Sigma \to W$ could be extended to the entirety of V. It is related to the collection of functions $\psi^{(0)}, \ldots, \psi^{(k)}$ in the sense that, for each $l \in \{0, \ldots, k\}$, the element $\psi^{(l)}(x) \in \mathcal{L}(V^{\otimes l}; W)$ is the l^{th} derivative of $\Psi_x(\cdot)$ at x. The expansions (A.5) and the estimates (A.6) then ensure that the proposals $\Psi_x(y)$ must vary in a Lipschitz manner as both the basepoint x and the evaluation point y vary across Σ. Loosely, if the points $x_1, x_2, y_1, y_2 \in \Sigma$ are all mututally close in the $|| \cdot ||_V$ norm sense, then the values $\Psi_{x_1}(y_1)$ and $\Psi_{x_2}(y_2)$ must be close in the $|| \cdot ||_W$ norm sense.

On the interior of Σ the functions $\psi^{(1)}, \ldots, \psi^{(k)}$ are the classical derivatives of $\psi^{(0)}$, though this interior may be empty. In general, the functions $\psi^{(1)}, \ldots, \psi^{(k)}$ are not necessarily uniquely determined by $\psi^{(0)}$. A consequence of Definition A.2 is that $\psi^{(0)} \in C^0(\Sigma; W)$ with $\left|\left|\psi^{(0)}\right|\right|_{C^0(\Sigma;W)} \leq ||\psi||_{\mathrm{Lip}(\gamma,\Sigma,W)}$. Moreover, in [Ste70] Stein proves that the $\mathrm{Lip}(\gamma)$ spaces are nested in the sense that given any $\eta \in (0, \gamma)$, if $q \in \mathbb{Z}_{\geq 0}$ is such that $\eta \in (q, q+1]$, then $\psi_{[q]} = (\psi^{(0)}, \ldots, \psi^{(q)}) \in \mathrm{Lip}(\eta, \Sigma, W)$. But it is *not* necessarily true that $||\psi_{[q]}||_{\mathrm{Lip}(\eta,\Sigma,W)} \leq ||\psi_{[k]}||_{\mathrm{Lip}(\gamma,\Sigma,W)}$.

For example, consider $\Sigma := [-1, 1] \subset \mathbb{R}$ and define functions $\psi^{(0)} : \Sigma \to \mathbb{R}$ and $\psi^{(1)} : \Sigma \to \mathcal{L}(\mathbb{R}; \mathbb{R})$ by $\psi^{(0)}(x) := x^2$ and $\psi^{(1)}(x)[v] := 2xv$ respectively. Let $\psi = (\psi^{(0)}, \psi^{(1)})$. Then the associated remainder terms are $R_0^{\psi}(x,y) := \psi^{(0)}(y) - \psi^{(0)}(x) - \psi^{(1)}(x)[y-x] = (y-x)^2$ and $R_1^{\psi}(x,y)[v] := \psi^{(1)}(y)[v] - \psi^{(1)}(x)[v] = 2(y-x)v$. It follows that $\psi \in \mathrm{Lip}(2, \Sigma, \mathbb{R})$ with $||\psi||_{\mathrm{Lip}(2,\Sigma,\mathbb{R})} = 2$. However $R_1^{\psi}(-1,1)[v] = 4v = 2\sqrt{2}|1-(-1)|^{\frac{1}{2}}v$ and so $||\psi||_{\mathrm{Lip}(3/2,\Sigma,\mathbb{R})} = 2\sqrt{2} = \sqrt{2}||\psi||_{\mathrm{Lip}(2,\Sigma,\mathbb{R})}$.

Our next goal is to obtain an estimate of the form $|| \cdot ||_{\mathrm{Lip}(\eta,\Sigma,W)} \leq C || \cdot ||_{\mathrm{Lip}(\gamma,\Sigma,W)}$ for an explicit constant $C \geq 1$ in a form convenient for our purposes. The remainder term bounds recorded in the following result form our first step in this direction.

Lemma A.3 (Remainder Term Estimates)**.** *Let V and W be Banach spaces, and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume that $\Gamma \subset V$ is closed. Let $\rho > 0$ with $n \in \mathbb{Z}_{\geq 0}$ such that $\rho \in (n, n+1]$, let $\theta \in (0, \rho)$, and suppose $\psi = (\psi^{(0)}, \ldots, \psi^{(n)}) \in \mathrm{Lip}(\rho, \Gamma, W)$. For $l \in \{0, \ldots, n\}$ let $R_l^{\psi} : \Gamma \times \Gamma \to \mathcal{L}(V^{\otimes l}; W)$ denote the remainder term associated to $\psi^{(l)}$ (cf. (A.5) in Definition A.2). If $\theta \in (n, \rho)$ then for every $l \in \{0, \ldots, n\}$ we have that for every $x, y \in \Gamma$ with $x \neq y$ that*

$$\frac{\left|\left| R_l^{\psi}(x,y) \right|\right|_{\mathcal{L}(V^{\otimes l};W)}}{||y-x||_V^{\theta-l}} \leq \min \left\{ \mathrm{diam}(\Gamma)^{\rho-\theta} \,,\, G(\rho, \theta, l, \Gamma) \right\} ||\psi||_{\mathrm{Lip}(\rho,\Gamma,W)} \tag{A.8}$$

where $G(\rho,\theta,l,\Gamma)$ is defined by

$$G(\rho,\theta,l,\Gamma) := \inf_{r\in(0,\mathrm{diam}(\Gamma))} \left\{ \max\left\{ r^{\rho-\theta} \,,\, \frac{1}{r^{\theta-l}}\left(1+\sum_{s=0}^{n-l}\frac{r^s}{s!}\right)\right\}\right\}. \tag{A.9}$$

If $\theta \in (0,n]$ (which is only possible if $n \geq 1$) then let $q \in \{0,\dots,n-1\}$ be such that $\theta \in (q,q+1]$. For each $l \in \{0,\dots,q\}$ let $\tilde{R}^\psi_l : \Gamma\times\Gamma \to \mathcal{L}(V^{\otimes l};W)$ denote the alteration of the remainder term R^ψ_l defined for $x,y\in\Gamma$ and $v \in V^{\otimes l}$ by

$$\tilde{R}^\psi_l(x,y)[v] := R^\psi_l(x,y)[v] + \sum_{s=q-l+1}^{n-l}\frac{1}{s!}\psi^{l+s}(x)\left[v\otimes(y-x)^{\otimes s}\right]. \tag{A.10}$$

Then for every $l\in\{0,\dots,q\}$ and every $x,y\in\Gamma$ with $x\neq y$ we have that

$$\frac{\left|\left|\tilde{R}^\psi_l(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)}}{||y-x||_V^{\theta-l}} \leq \min\left\{\mathrm{diam}(\Gamma)^{\rho-\theta}+\sum_{i=q+1}^{n}\frac{\mathrm{diam}(\Gamma)^{i-\theta}}{(i-l)!} \,,\, H(\rho,\theta,l,\Gamma)\right\}||\psi||_{\mathrm{Lip}(\rho,\Gamma,W)} \tag{A.11}$$

where $H(\rho,\theta,l,\Gamma)$ is defined by

$$H(\rho,\theta,l,\Gamma) := \inf_{r\in(0,\mathrm{diam}(\Gamma))}\left\{\max\left\{r^{\rho-\theta}+\sum_{i=q+1}^{n}\frac{r^{i-\theta}}{(i-l)!} \,,\, \frac{1}{r^{\theta-l}}\left(1+\sum_{s=0}^{q-l}\frac{r^s}{s!}\right)\right\}\right\}. \tag{A.12}$$

Proof of Lemma A.3. Let V and W be Banach spaces, and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume that $\Gamma\subset V$ is closed and that $\rho>0$ with $n\in\mathbb{Z}_{\geq 0}$ such that $\rho\in(n,n+1]$. Suppose that, for $l\in\{0,\dots,n\}$, we have functions $\psi^{(l)}:\Gamma\to\mathcal{L}(V^{\otimes l};W)$ such that $\psi=(\psi^{(0)},\dots,\psi^{(n)})$ defines an element of $\mathrm{Lip}(\rho,\Gamma,W)$. For each $l\in\{0,\dots,n\}$ define $R^\psi_l:\Gamma\times\Gamma\to\mathcal{L}(V^{\otimes l};W)$ for $x,y\in\Gamma$ and $v\in V^{\otimes l}$ by

$$R^\psi_l(x,y)[v] := \psi^{(l)}(y)[v] - \sum_{s=0}^{n-l}\frac{1}{s!}\psi^{l+s}(x)\left[v\otimes(y-x)^{\otimes s}\right]. \tag{A.13}$$

We claim that the estimates (A.8) and (A.11) are immediate when $||\psi||_{\mathrm{Lip}(\rho,\Gamma,W)}=0$. To see this, note that if $||\psi||_{\mathrm{Lip}(\rho,\Gamma,W)}=0$ then for each $l\in\{0,\dots,n\}$ and any $x\in\Gamma$ we have that $\psi^{(l)}(x)\equiv 0$ in $\mathcal{L}(V^{\otimes l};W)$. Conseuqently, for each $l\in\{0,\dots,n\}$ and any $x,y\in\Gamma$, we have via (A.13) that $R^\psi_l(x,y)\equiv 0$ in $\mathcal{L}(V^{\otimes l};W)$. When $\theta\in(n,\rho)$, this tells us that the estimate (A.8) is true since both sides are zero.

When $\theta\in(0,n]$ (which is only possible if $n\geq 1$) then, if $q\in\{0,\dots,n-1\}$ is such that $\theta\in(q,q+1]$, for each $l\in\{0,\dots,q\}$ and any $x,y\in\Gamma$ we have via (A.10) that the alteration $\tilde{R}^\psi_l$ of R^ψ_l satisfies that $\tilde{R}^\psi_l(x,y)\equiv 0$ in $\mathcal{L}(V^{\otimes l};W)$. Hence the estimate (A.11) is true since both sides are again zero.

If $||\psi||_{\mathrm{Lip}(\rho,\Gamma,W)}\neq 0$ then by replacing ψ by $\psi/||\psi||_{\mathrm{Lip}(\rho,\Gamma,W)}$ it suffices to prove the estimates (A.8) and (A.11) under the additional assumption that $||\psi||_{\mathrm{Lip}(\rho,\Gamma,W)}=1$. As a consequence, whenever $l\in\{0,\dots,n\}$ and $x,y\in\Gamma$, we have the bounds (cf. (A.4) and cf. (A.6))

$$\textbf{(I)}\quad \left|\left|\psi^{(l)}(x)\right|\right|_{\mathcal{L}(V^{\otimes l};W)}\leq 1 \qquad\text{and}\qquad \textbf{(II)}\quad \left|\left|R^\psi_l(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)}\leq ||y-x||_V^{\rho-l}. \tag{A.14}$$

First suppose $\theta\in(n,\rho)$ and let $l\in\{0,\dots,n\}$. For any $x,y\in\Gamma$ we have that

$$\left|\left|R^\psi_l(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \overset{\textbf{(II)}\text{ of (A.14)}}{\leq} ||y-x||_V^{\rho-l} = ||y-x||_V^{\rho-\theta}||y-x||_V^{\theta-l}. \tag{A.15}$$

A first consequence of (A.15) is

$$\left|\left|R^\psi_l(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \mathrm{diam}(\Gamma)^{\rho-\theta}||y-x||_V^{\theta-l}. \tag{A.16}$$

A second consequence of (A.15) is that, for any fixed $r \in (0, \operatorname{diam}(\Gamma))$, if $||y-x||_V \leq r$ then

$$\left|\left|R_l^{\psi}(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq r^{\rho-\theta}||y-x||_V^{\theta-l}. \tag{A.17}$$

If $||y-x||_V > r$ then we use (A.13) and that the tensor powers of V are equipped with admissible norms (cf. Definition A.1) to compute for any $v \in V^{\otimes l}$ that

$$\begin{aligned}\left|\left|R_l^{\psi}(x,y)[v]\right|\right|_W &\overset{(A.13)}{\leq} \left|\left|\psi^{(l)}(y)[v]\right|\right|_W + \sum_{j=0}^{n-l}\frac{1}{j!}\left|\left|\psi^{(l+j)}(x)\left[v\otimes(y-x)^{\otimes j}\right]\right|\right|_W \\ &\overset{(\mathbf{I}) \text{ of } (A.14)}{\leq} ||v||_{V^{\otimes l}} + \sum_{j=0}^{n-l}\frac{1}{j!}||y-x||_V^j||v||_{V^{\otimes l}} \leq r^{-(\theta-l)}\left(1+\sum_{j=0}^{n-l}\frac{r^j}{j!}\right)||y-x||_V^{\theta-l}||v||_{V^{\otimes l}}.\end{aligned}$$

In the last line we have used that, for any $j \in \{0,\ldots,n-l\}$, that $||y-x||_V^{j-(\theta-l)} < r^{j-(\theta-l)}$. This is itself a consequence of the facts that for any $j \in \{0,\ldots,n-l\}$ that $j-(\theta-l) \leq n-\theta < 0$, and that $r < ||y-x||_V$. By taking the supremum over $v \in V^{\otimes l}$ with unit $V^{\otimes l}$ norm, we may conclude that when $||y-x||_V > r$ we have

$$\left|\left|R_l^{\psi}(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq r^{-(\theta-l)}\left(1+\sum_{s=0}^{n-l}\frac{r^s}{s!}\right)||y-x||_V^{\theta-l}. \tag{A.18}$$

By combining (A.17) and (A.18) we deduce that for every $x, y \in \Gamma$ we have

$$\left|\left|R_l^{\psi}(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \max\left\{r^{\rho-\theta}\, ,\, \frac{1}{r^{\theta-l}}\left(1+\sum_{s=0}^{n-l}\frac{r^s}{s!}\right)\right\}||y-x||_V^{\theta-l}. \tag{A.19}$$

Recall that the choice of $r \in (0, \operatorname{diam}(\Gamma))$ was arbitrary. Consequently we may take the infimum over the choice of $r \in (0, \operatorname{diam}(\Gamma))$ in (A.19) to obtain that whenever $x \neq y$ we have

$$\frac{\left|\left|R_l^{\psi}(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)}}{||y-x||_V^{\theta-l}} \leq \inf_{r\in(0,\operatorname{diam}(\Gamma))}\left\{\max\left\{r^{\rho-\theta}\, ,\, \frac{1}{r^{\theta-l}}\left(1+\sum_{s=0}^{n-l}\frac{r^s}{s!}\right)\right\}\right\}. \tag{A.20}$$

If we define

$$G(\rho,\theta,l,\Gamma) := \inf_{r\in(0,\operatorname{diam}(\Gamma))}\left\{\max\left\{r^{\rho-\theta}\, ,\, \frac{1}{r^{\theta-l}}\left(1+\sum_{s=0}^{n-l}\frac{r^s}{s!}\right)\right\}\right\}, \tag{A.21}$$

then (A.16) and (A.20) yield that

$$\frac{\left|\left|R_l^{\psi}(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)}}{||y-x||_V^{\theta-l}} \leq \min\left\{\operatorname{diam}(\Gamma)^{\rho-\theta}, G(\rho,\theta,l,\Gamma)\right\}. \tag{A.22}$$

The arbitrariness of $l \in \{0,\ldots,n\}$ and the points $x, y \in \Gamma$ with $x \neq y$ mean that (A.22) establishes the estimates claimed in (A.8) for the case that $||\psi||_{\operatorname{Lip}(\rho,\Gamma,W)} = 1$.

Now assume that $0 < \theta \leq n < \rho \leq n+1$ which requires $n \geq 1$. Let $q \in \{0,\ldots,n-1\}$ be such that $\theta \in (q, q+1]$. For each $l \in \{0,\ldots,q\}$ let $\tilde{R}_l^{\psi} : \Gamma\times\Gamma \to \mathcal{L}(V^{\otimes l};W)$ denote the alteration of the remainder term R_l^{ψ} defined for $x, y \in \Gamma$ and $v \in V^{\otimes l}$ by

$$\tilde{R}_l^{\psi}(x,y)[v] := R_l^{\psi}(x,y)[v] + \sum_{s=q-l+1}^{n-l}\frac{1}{s!}\psi^{l+s}(x)\left[v\otimes(y-x)^{\otimes s}\right]. \tag{A.23}$$

Let $l \in \{0,\ldots,q\}$, $x, y \in \Gamma$ and $v \in V^{\otimes l}$. Recalling that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1), we may compute that

$$\left|\left|\tilde{R}^{\psi}_l(x,y)[v]\right|\right|_W \overset{\text{(A.23)}}{\leq} \left|\left|R^{\psi}_l(x,y)[v]\right|\right|_W + \sum_{s=q-l+1}^{n-l} \frac{1}{s!} \left|\left|\psi^{l+s}(x)\left[v \otimes (y-x)^{\otimes s}\right]\right|\right|_W$$
$$\overset{\text{(A.14)}}{\leq} \left(||y-x||_V^{\rho-l} + \sum_{s=q-l+1}^{n-l} \frac{1}{s!} ||y-x||_V^{s} \right) ||v||_{V^{\otimes l}}$$
$$= \left(||y-x||_V^{\rho-\theta} + \sum_{i=q+1}^{n} \frac{1}{(i-l)!} ||y-x||_V^{i-\theta} \right) ||y-x||_V^{\theta-l} ||v||_{V^{\otimes l}}.$$

By taking the supremum over $v \in V^{\otimes l}$ with unit $V^{\otimes l}$ norm we may conclude that

$$\left|\left|\tilde{R}^{\psi}_l(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \left(||y-x||_V^{\rho-\theta} + \sum_{i=q+1}^{n} \frac{||y-x||_V^{i-\theta}}{(i-l)!} \right) ||y-x||_V^{\theta-l}. \tag{A.24}$$

A first consequence of (A.24) is that

$$\left|\left|\tilde{R}^{\psi}_l(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \left(\operatorname{diam}(\Gamma)^{\rho-\theta} + \sum_{i=q+1}^{n} \frac{\operatorname{diam}(\Gamma)^{i-\theta}}{(i-l)!} \right) ||y-x||_V^{\theta-l}. \tag{A.25}$$

Now consider a fixed choice of constant $r \in (0, \operatorname{diam}(\Gamma))$. If $||y-x||_V \leq r$ then a consequence of (A.24) is that

$$\left|\left|\tilde{R}^{\psi}_l(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \left(r^{\rho-\theta} + \sum_{i=q+1}^{n} \frac{r^{i-\theta}}{(i-l)!} \right) ||y-x||_V^{\theta-l}. \tag{A.26}$$

If $||y-x||_V > r$ then we may first observe via (A.13) and (A.23) that for any $v \in V^{\otimes l}$ we have

$$\tilde{R}^{\psi}_l(x,y)[v] = \psi^{(l)}(y)[v] - \sum_{s=0}^{q-l} \frac{1}{s!} \psi^{(l+s)}(x)\left[v \otimes (y-x)^{\otimes s}\right]. \tag{A.27}$$

We may use (A.27) and that the tensor powers of V are equipped with admissible norms (cf. Definition A.1) to compute for any $v \in V^{\otimes l}$ that

$$\left|\left|\tilde{R}^{\psi}_l(x,y)[v]\right|\right|_W \overset{\text{(A.27)}}{\leq} \left|\left|\psi^{(l)}(y)[v]\right|\right|_W + \sum_{j=0}^{q-l} \frac{1}{j!} \left|\left|\psi^{(l+j)}(x)\left[v \otimes (y-x)^{\otimes j}\right]\right|\right|_W$$
$$\overset{\text{(A.14)}}{\leq} ||v||_{V^{\otimes l}} + \sum_{j=0}^{q-l} \frac{1}{j!} ||y-x||_V^{j} ||v||_{V^{\otimes l}} \leq r^{-(\theta-l)} \left(1 + \sum_{j=0}^{q-l} \frac{r^j}{j!} \right) ||y-x||_V^{\theta-l} ||v||_{V^{\otimes l}}.$$

For the last inequality we have used that, for any $j \in \{0, \ldots, q-l\}$, that $||y-x||_V^{j-(\theta-l)} < r^{j-(\theta-l)}$. This is itself a consequence of the facts that for any $j \in \{0, \ldots, q-l\}$ that $j-(\theta-l) \leq q-\theta < 0$, and that $r < ||y-x||_V$. By taking the supremum over $v \in V^{\otimes l}$ with unit $V^{\otimes l}$ norm, we may conclude that when $||y-x||_V > r$ we have

$$\left|\left|\tilde{R}^{\psi}_l(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq r^{-(\theta-l)} \left(1 + \sum_{s=0}^{q-l} \frac{r^s}{s!} \right) ||y-x||_V^{\theta-l}. \tag{A.28}$$

Together (A.26) and (A.28) give that for every $x, y \in \Gamma$ with $x \neq y$ we have

$$\frac{\left|\left|\tilde{R}_l^{\psi}(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)}}{||y-x||_V^{\theta-l}} \leq \max\left\{ r^{\rho-\theta} + \sum_{i=q+1}^{n} \frac{r^{i-\theta}}{(i-l)!} \,,\, \frac{1}{r^{\theta-l}}\left(1+\sum_{s=0}^{q-l}\frac{r^s}{s!}\right)\right\}. \tag{A.29}$$

Recall that the choice of $r \in (0, \mathrm{diam}(\Gamma))$ was arbitrary. Consequently we may take the infimum over the choice of $r \in (0, \mathrm{diam}(\Gamma))$ in (A.29) to obtain that whenever $x \neq y$ we have

$$\frac{\left|\left|\tilde{R}_l^{\psi}(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)}}{||y-x||_V^{\theta-l}} \leq H(\rho,\theta,l,\Gamma) \tag{A.30}$$

for $H(\rho,\theta,l,\Gamma)$ defined by

$$H(\rho,\theta,l,\Gamma) := \inf_{r\in(0,\mathrm{diam}(\Gamma))}\left\{\max\left\{ r^{\rho-\theta} + \sum_{i=q+1}^{n} \frac{r^{i-\theta}}{(i-l)!} \,,\, \frac{1}{r^{\theta-l}}\left(1+\sum_{s=0}^{q-l}\frac{r^s}{s!}\right)\right\}\right\}. \tag{A.31}$$

Together (A.25) and (A.30) yield

$$\frac{\left|\left|\tilde{R}_l^{\psi}(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)}}{||y-x||_V^{\theta-l}} \leq \min\left\{ \mathrm{diam}(\Gamma)^{\rho-\theta} + \sum_{i=q+1}^{n} \frac{\mathrm{diam}(\Gamma)^{i-\theta}}{(i-l)!} \,,\, H(\rho,\theta,l,\Gamma)\right\}. \tag{A.32}$$

The arbitrariness of $l \in \{0,\ldots,q\}$ and the points $x, y \in \Gamma$ with $x \neq y$ mean that (A.32) establishes the estimates claimed in (A.11) for the case that $||\psi||_{\mathrm{Lip}(\rho,\Gamma,W)} = 1$. This completes the proof of Lemma A.3. ■

We use Lemma A.3 to obtain an estimate of the form $||\cdot||_{\mathrm{Lip}(\eta,\Sigma,W)} \leq C||\cdot||_{\mathrm{Lip}(\gamma,\Sigma,W)}$ for an explicit constant $C \geq 1$ in a form that is convenient for our purposes.

Lemma A.4 ($\mathrm{Lip}(\gamma,\Sigma,W)$ Nesting)**.** *Let V and W be Banach spaces, and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume that $\Gamma \subset V$ is closed. Let $\rho > 0$ with $n \in \mathbb{Z}_{\geq 0}$ such that $\rho \in (n, n+1]$, and $\theta \in (0,\rho)$ with $q \in \{0,\ldots,n\}$ such that $\theta \in (q, q+1]$. Suppose that $\psi = (\psi^{(0)},\ldots,\psi^{(n)}) \in \mathrm{Lip}(\rho,\Gamma,W)$. Then $\psi_{[q]} = (\psi^{(0)},\ldots,\psi^{(q)}) \in \mathrm{Lip}(\theta,\Gamma,W)$. Further, if $\theta \in (n,\rho)$ then we have the estimate that*

$$\left|\left|\psi_{[q]}\right|\right|_{\mathrm{Lip}(\theta,\Gamma,W)} \leq \max\left\{1\,,\ \min\left\{1+e\,,\ \mathrm{diam}(\Gamma)^{\rho-\theta}\right\}\right\}||\psi||_{\mathrm{Lip}(\rho,\Gamma,W)}. \tag{A.33}$$

If $\theta \in (0,n]$ then we have the estimate that

$$\left|\left|\psi_{[q]}\right|\right|_{\mathrm{Lip}(\theta,\Gamma,W)} \leq \min\left\{C_1\,,\ C_2\right\}||\psi||_{\mathrm{Lip}(\rho,\Gamma,W)} \tag{A.34}$$

where $C_1, C_2 > 0$ are constants, depending only on $\mathrm{diam}(\Gamma)$, ρ, and θ, defined by

$$C_1 := \max\left\{1\,,\ \min\left\{1+e\,,\ \mathrm{diam}(\Sigma)^{\rho-\theta} + \sum_{j=q+1}^{n}\frac{\mathrm{diam}(\Gamma)^{j-\theta}}{(j-q)!}\right\}\right\} \tag{A.35}$$

and

$$C_2 = \max\left\{1, \min\left\{1+e, \mathrm{diam}(\Gamma)^{q+1-\theta}\right\}\right\}\left(1+\min\left\{e, \mathrm{diam}(\Gamma)^{\rho-n}\right\}\right)\left(1+\min\left\{e, \mathrm{diam}(\Gamma)\right\}\right)^{n-(q+1)}. \tag{A.36}$$

Finally, as a simple consequence of (A.33) and (A.34), for any $\theta \in (0,\rho)$ we have the estimate

$$\left|\left|\psi_{[q]}\right|\right|_{\mathrm{Lip}(\theta,\Gamma,W)} \leq (1+e)||\psi||_{\mathrm{Lip}(\rho,\Gamma,W)}. \tag{A.37}$$

Remark A.5. The integers $n, q \in \mathbb{Z}_{\geq 0}$ are determined by ρ and θ respectively. Consequently, any apparent dependence on n and q in (A.35) and (A.36) is really dependence on ρ and θ respectively.

Remark A.6. We can have equality in (A.33). To see this, let $\Gamma := [-1, 1] \subset \mathbb{R}$ and define $\psi^{(0)} : \Gamma \to \mathbb{R}$ by $\psi^{(0)}(x) := x^2$, $\psi^{(1)} : \Gamma \to \mathcal{L}(\mathbb{R}; \mathbb{R})$ by $\psi^{(1)}(x)[v] := 2xv$, $R_0(x, y) := \psi^{(0)}(y) - \psi^{(0)}(x) - \psi^{(1)}[y - x] = (y - x)^2$ and $R_1(x, y)[v] := \psi^{(1)}(y)[v] - \psi^{(1)}(x)[v] = 2(y - x)v$. Then $\psi = \left(\psi^{(0)}, \psi^{(1)}\right) \in \mathrm{Lip}(2, \Gamma, \mathbb{R})$ with $||\psi||_{\mathrm{Lip}(2,\Gamma,\mathbb{R})} = 2$. However $R_1(-1, 1)[v] = 4v = 2\sqrt{2}|1 - (-1)|^{\frac{1}{2}} v$ and so $||\psi||_{\mathrm{Lip}(3/2,\Gamma,\mathbb{R})} = 2\sqrt{2} = \sqrt{2}||\psi||_{\mathrm{Lip}(2,\Gamma,\mathbb{R})}$. Here $\mathrm{diam}(\Gamma) = 2$, $\rho = 2$ and $\theta = 3/2$. Thus we observe that $1 < \mathrm{diam}(\Gamma)^{2-\frac{3}{2}} = \sqrt{2} < 1 + e$, which establishes equality in (A.33).

Remark A.7. We can have equality in (A.34). As an example, let $\Gamma := \{0, 1\} \subset \mathbb{R}$ and define $\psi^{(0)} : \Gamma \to \mathbb{R}$ by $\psi^{(0)}(0) := -A$ and $\psi^{(0)}(1) := A$ for some $A > 0$, and define $\psi^{(1)} : \Gamma \to \mathcal{L}(\mathbb{R}; \mathbb{R})$ by $\psi^{(1)}(x)[v] := Av$ for every $x \in \Gamma$. Then given $x, y \in \Gamma$

$$\psi^{(0)}(y) - \psi^{(0)}(x) - \psi^{(1)}[y - x] = \begin{cases} A & \text{if } x = 0, y = 1 \\ -A & \text{if } x = 1, y = 0 \\ 0 & \text{if } x = y. \end{cases} \tag{A.38}$$

It follows from (A.38) that $\psi = \left(\psi^{(0)}, \psi^{(1)}\right) \in \mathrm{Lip}(2, \Gamma, \mathbb{R})$ with $||\psi||_{\mathrm{Lip}(2,\Gamma,\mathbb{R})} = A$. Moreover, we also have that $\psi_{[0]} = \psi^{(0)} \in \mathrm{Lip}(1, \Gamma, \mathbb{R})$ with $\left|\left|\psi_{[0]}\right|\right|_{\mathrm{Lip}(1,\Gamma,\mathbb{R})} = 2A = 2||\psi||_{\mathrm{Lip}(2,\Gamma,\mathbb{R})}$. Here $\mathrm{diam}(\Gamma) = 1$, $n := 1$, $\rho := 2$, $\theta := 1$ and $q = 0$. Consequently, both C_1 defined in (A.35) and C_2 defined in (A.36) are equal to 2. Hence $\min\{C_1, C_2\} = 2$, and so we have equality in (A.34).

Proof of Lemma A.4. Let V and W be Banach spaces, and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume that $\Gamma \subset V$ is closed. Let $\rho > 0$ with $n \in \mathbb{Z}_{\geq 0}$ such that $\rho \in (n, n+1]$, and let $\theta \in (0, \rho)$ with $q \in \{0, \ldots, n\}$ such that $\theta \in (q, q+1]$. To deal with the case that $\theta \in (n, \rho)$, we first establish the following claim.

Claim A.8. *Suppose V and W be Banach spaces, and that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume that $\mathcal{D} \subset V$ is closed. Let $\lambda > 0$ with $m \in \mathbb{Z}_{\geq 0}$ such that $\lambda \in (m, m+1]$, and $\sigma \in (m, \lambda)$. If $\phi = (\phi^{(0)}, \ldots, \phi^{(m)}) \in \mathrm{Lip}(\lambda, \mathcal{D}, W)$ then $\phi \in \mathrm{Lip}(\sigma, \mathcal{D}, W)$, and we have the estimate that*

$$||\phi||_{\mathrm{Lip}(\sigma,\mathcal{D},W)} \leq \max\left\{1\ ,\ \min\left\{1 + e\ ,\ \mathrm{diam}(\mathcal{D})^{\lambda-\sigma}\right\}\right\} ||\phi||_{\mathrm{Lip}(\lambda,\mathcal{D},W)}. \tag{A.39}$$

Proof of Claim A.8. For each $l \in \{0, \ldots, m\}$ define $R_l^\phi : \mathcal{D} \times \mathcal{D} \to \mathcal{L}(V^{\otimes l}; W)$ for $x, y \in \mathcal{D}$ and $v \in V^{\otimes l}$ by

$$R_l^\phi(x, y)[v] := \phi^{(l)}(y)[v] - \sum_{s=0}^{n-l} \frac{1}{s!} \phi^{l+s}(x) \left[v \otimes (y - x)^{\otimes s}\right]. \tag{A.40}$$

Since the estimate (A.39) is trivial when $||\phi||_{\mathrm{Lip}(\lambda,\mathcal{D},W)} = 0$, we need only establish the validity of (A.39) when $||\phi||_{\mathrm{Lip}(\lambda,\mathcal{D},W)} \neq 0$. But in this case, by replacing ϕ by $\phi/||\phi||_{\mathrm{Lip}(\lambda,\mathcal{D},W)}$ it suffices to prove (A.39) under the additional assumption that $||\phi||_{\mathrm{Lip}(\lambda,\mathcal{D},W)} = 1$

A consequence of $\phi \in \mathrm{Lip}(\lambda, \mathcal{D}, W)$ with $||\phi||_{\mathrm{Lip}(\lambda,\mathcal{D},W)} = 1$ is that, whenever $l \in \{0, \ldots, m\}$ and $x, y \in \mathcal{D}$, we have the bounds (cf. (A.4) and (A.6))

$$\textbf{(I)} \quad \left|\left|\phi^{(l)}(x)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq 1 \quad \text{and} \quad \textbf{(II)} \quad \left|\left|R_l^\phi(x, y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq ||y - x||_V^{\lambda-l}. \tag{A.41}$$

Given any $l \in \{0, \ldots, m\}$ and any point $x \in \mathcal{D}$, we can conclude from (A.40) that $R_l^\psi(x, x) \equiv 0$ in $\mathcal{L}(V^{\otimes l}; W)$. Hence controlling the $\mathcal{L}(V^{\otimes l}; W)$ norm of the remainder term R_l^ψ is trivial on the diagonal of $\mathcal{D} \times \mathcal{D}$.

Given any $l \in \{0, \ldots, m\}$, we now estimate the $\mathcal{L}(V^{\otimes l}; W)$ norm of R_l^ψ off the diagonal of $\mathcal{D} \times \mathcal{D}$. For any points $x, y \in \mathcal{D}$ with $x \neq y$, we apply Lemma A.3, recalling that $||\phi||_{\mathrm{Lip}(\lambda,\mathcal{D},W)} = 1$, to conclude that (cf. (A.8))

$$\frac{\left|\left|R_l^\phi(x, y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)}}{||y - x||_V^{\sigma-l}} \leq \min\left\{\mathrm{diam}(\mathcal{D})^{\lambda-\sigma}\ ,\ G(\lambda, \sigma, l, \mathcal{D})\right\} \tag{A.42}$$

where $G(\lambda, \sigma, l, \mathcal{D})$ is defined by

$$G(\lambda, \sigma, l, \mathcal{D}) := \inf_{r \in (0, \operatorname{diam}(\mathcal{D}))} \left\{ \max \left\{ r^{\lambda - \sigma} \ , \ \frac{1}{r^{\sigma - l}} \left(1 + \sum_{s=0}^{m-l} \frac{r^s}{s!} \right) \right\} \right\}. \quad \text{(A.43)}$$

We now prove that

$$\min \left\{ \operatorname{diam}(\mathcal{D})^{\lambda - \sigma}, G(\lambda, \sigma, l, \mathcal{D}) \right\} \leq 1 + e. \quad \text{(A.44)}$$

If $\operatorname{diam}(\mathcal{D}) \leq 1$, then (A.44) is obtained by observing that

$$\min \left\{ \operatorname{diam}(\mathcal{D})^{\lambda - \sigma}, G(\lambda, \sigma, l, \mathcal{D}) \right\} \leq \operatorname{diam}(\mathcal{D})^{\lambda - \sigma} \leq 1 < 1 + e.$$

If $\operatorname{diam}(\mathcal{D}) > 1$ then (A.44) is obtained by observing that

$$\min \left\{ \operatorname{diam}(\mathcal{D})^{\lambda - \sigma}, G(\lambda, \sigma, l, \mathcal{D}) \right\} \leq G(\lambda, \sigma, l, \mathcal{D}) \leq \max \left\{ 1 \ , \ 1 + \sum_{s=0}^{m-l} \frac{1}{s!} \right\} \leq (1 + e).$$

Hence (A.44) is proven. Together (A.42) and (A.44) yield that

$$\frac{\left|\left| R_l^{\phi}(x, y) \right|\right|_{\mathcal{L}(V^{\otimes l}; W)}}{||y - x||_V^{\sigma - l}} \leq 1 + e \quad \text{(A.45)}$$

Thus we may combine (A.42) and (A.45) to conclude that

$$\frac{\left|\left| R_l^{\phi}(x, y) \right|\right|_{\mathcal{L}(V^{\otimes l}; W)}}{||y - x||_V^{\sigma - l}} \leq \min \left\{ \operatorname{diam}(\mathcal{D})^{\lambda - \sigma} \ , \ 1 + e \right\}. \quad \text{(A.46)}$$

Both the choice of $l \in \{0, \ldots, m\}$ and the choice of points $x, y \in \mathcal{D}$ with $x \neq y$ were arbitrary. Hence we may conclude that the estimate (A.46) is valid for every $l \in \{0, \ldots, m\}$ and all points $x, y \in \mathcal{D}$ with $x \neq y$. The pointwise bounds for the functions $\phi^{(0)}, \ldots, \phi^{(m)}$ given in (**I**) of (A.41) and the remainder term bounds (A.46) establish that

$$||\phi||_{\operatorname{Lip}(\sigma, \mathcal{D}, W)} \leq \max \left\{ 1 \ , \ \min \left\{ \operatorname{diam}(\mathcal{D})^{\lambda - \sigma} \ , \ 1 + e \right\} \right\}. \quad \text{(A.47)}$$

The estimate (A.47) is precisely the estimate claimed in (A.39) for the case that $||\phi||_{\operatorname{Lip}(\lambda, \mathcal{D}, W)} = 1$. This completes the proof of Claim A.8. ■

The estimate claimed in the case that $\theta \in (n, \rho)$ is an immediate consequence of Claim A.8. Indeed, assuming that $\theta \in (n, \rho)$, we appeal to Claim A.8 with $\mathcal{D} := \Gamma$, $m := n$, $\lambda := \rho$ and $\sigma := \theta$ to conclude from (A.39) that

$$||\psi||_{\operatorname{Lip}(\theta, \Gamma, W)} \leq \max \left\{ 1 \ , \ \min \left\{ 1 + e \ , \ \operatorname{diam}(\Gamma)^{\rho - \theta} \right\} \right\} ||\psi||_{\operatorname{Lip}(\rho, \Gamma, W)}, \quad \text{(A.48)}$$

which is precisely the estimate claimed in (A.33).

It remains only to establish the estimate claimed in (A.34) for the case that $0 < \theta \leq n < \rho \leq n + 1$. Observe that this requires $n \geq 1$ and $q \in \{0, \ldots, n - 1\}$. We begin by establishing the following claim.

Claim A.9. *Suppose V and W are Banach spaces, and that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume that $\mathcal{D} \subset V$ is closed. Let $\lambda > 1$ with $m \in \mathbb{Z}_{\geq 1}$ such that $\lambda \in (m, m + 1]$, and $\sigma \in (0, m]$ with $p \in \{0, \ldots, m - 1\}$ such that $\sigma \in (p, p + 1]$. If $\phi = (\phi^{(0)}, \ldots, \phi^{(m)}) \in \operatorname{Lip}(\lambda, \mathcal{D}, W)$ then $\phi_{[p]} = (\phi^{(0)}, \ldots, \phi^{(p)}) \in \operatorname{Lip}(\sigma, \mathcal{D}, W)$, and we have the estimate that*

$$\left|\left| \phi_{[p]} \right|\right|_{\operatorname{Lip}(\sigma, \mathcal{D}, W)} \leq \max \left\{ 1 \ , \ \min \left\{ 1 + e \ , \ \operatorname{diam}(\mathcal{D})^{\lambda - \sigma} + \sum_{j=p+1}^{m} \frac{\operatorname{diam}(\mathcal{D})^{j - \sigma}}{(j - p)!} \right\} \right\} ||\phi||_{\operatorname{Lip}(\lambda, \mathcal{D}, W)}. \quad \text{(A.49)}$$

Proof of Claim A.9. For each $l \in \{0, \ldots, m\}$ define $R_l^\phi : \mathcal{D} \times \mathcal{D} \to \mathcal{L}(V^{\otimes l}; W)$ for $x, y \in \mathcal{D}$ and $v \in V^{\otimes l}$ by

$$R_l^\phi(x,y)[v] := \phi^{(l)}(y)[v] - \sum_{j=0}^{m-l} \frac{1}{j!} \phi^{(j+l)}(x) \left[v \otimes (y-x)^{\otimes j} \right]. \tag{A.50}$$

Since the estimate (A.49) is trivial when $||\phi||_{\mathrm{Lip}(\lambda,\mathcal{D},W)} = 0$, we need only establish the validity of (A.49) when $||\phi||_{\mathrm{Lip}(\lambda,\mathcal{D},W)} \neq 0$. But in this case, by replacing ϕ by $\phi / ||\phi||_{\mathrm{Lip}(\lambda,\mathcal{D},W)}$ it suffices to prove (A.49) under the additional assumption that $||\phi||_{\mathrm{Lip}(\lambda,\mathcal{D},W)} = 1$

A consequence of $\phi \in \mathrm{Lip}(\lambda, \mathcal{D}, W)$ with $||\phi||_{\mathrm{Lip}(\lambda,\mathcal{D},W)} = 1$ is that, whenever $l \in \{0, \ldots, m\}$ and $x, y \in \mathcal{D}$, we have the bounds (cf. (A.4) and (A.6))

$$\textbf{(I)} \quad \left|\left| \phi^{(l)}(x) \right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq 1 \qquad \text{and} \qquad \textbf{(II)} \quad \left|\left| R_l^\phi(x,y) \right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq ||y-x||_V^{\lambda - l}. \tag{A.51}$$

Our goal is to show that $\phi_{[p]} = (\phi^{(0)}, \ldots, \phi^{(p)})$ is in $\mathrm{Lip}(\sigma, \mathcal{D}, W)$. For this purpose, given $s \in \{0, \ldots, p\}$, let $\tilde{R}_s^\phi : \mathcal{D} \times \mathcal{D} \to \mathcal{L}(V^{\otimes s}; W)$ be defined for all $x, y \in \mathcal{D}$ and $v \in V^{\otimes s}$ by

$$\tilde{R}_s^\phi(x,y)[v] := \phi^{(s)}(y)[v] - \sum_{j=0}^{p-s} \frac{1}{j!} \phi^{(s+j)}(x) \left[v \otimes (y-x)^{\otimes j} \right]. \tag{A.52}$$

Together, (A.50) and (A.52) yield that

$$\tilde{R}_s^\phi(x,y)[v] = R_s^\phi(x,y)[v] + \sum_{j=p+1-s}^{m-s} \frac{1}{(k-j)!} \phi^{(s+j)}(x) \left[v \otimes (y-x)^{\otimes j} \right]. \tag{A.53}$$

Given any $l \in \{0, \ldots, p\}$ and any point $x \in \mathcal{D}$, we conclude from (A.53) that $\tilde{R}_l^\psi(x,x) \equiv 0$ in $\mathcal{L}(V^{\otimes l}; W)$. Hence controlling the $\mathcal{L}(V^{\otimes l}; W)$ norm of $\tilde{R}_l^\psi$ is trivial on the diagonal of $\mathcal{D} \times \mathcal{D}$.

Given any $l \in \{0, \ldots, p\}$ we now estimate the $\mathcal{L}(V^{\otimes l}; W)$ norm of $\tilde{R}_l^\psi$ off of the diagonal of $\mathcal{D} \times \mathcal{D}$. The alteration in (A.53) is exactly the same as the alteration defined in (A.10) of Lemma A.3. Consequently, recalling that $||\phi||_{\mathrm{Lip}(\lambda,\mathcal{D},W)} = 1$, we may apply that result (Lemma A.3) to deduce that for every $x, y \in \mathcal{D}$ with $x \neq y$ we have that (cf. (A.11))

$$\frac{\left|\left| \tilde{R}_l^\phi(x,y) \right|\right|_{\mathcal{L}(V^{\otimes l};W)}}{||y-x||_V^{\sigma - l}} \leq \min \left\{ \mathrm{diam}(\mathcal{D})^{\lambda - \sigma} + \sum_{i=p+1}^{m} \frac{\mathrm{diam}(\mathcal{D})^{i-\sigma}}{(i-l)!} \,,\, H(\lambda, \sigma, l, \mathcal{D}) \right\} \tag{A.54}$$

where $H(\lambda, \sigma, l, \mathcal{D})$ is defined by (cf. (A.12))

$$H(\lambda, \sigma, l, \mathcal{D}) := \inf_{r \in (0, \mathrm{diam}(\mathcal{D}))} \left\{ \max \left\{ r^{\lambda - \sigma} + \sum_{i=p+1}^{m} \frac{r^{i-\sigma}}{(i-l)!} \,,\, \frac{1}{r^{\sigma - l}} \left(1 + \sum_{s=0}^{p-l} \frac{r^s}{s!} \right) \right\} \right\}. \tag{A.55}$$

We now prove that

$$\mathcal{H} := \min \left\{ \mathrm{diam}(\mathcal{D})^{\lambda - \sigma} + \sum_{i=p+1}^{m} \frac{\mathrm{diam}(\mathcal{D})^{i-\sigma}}{(i-l)!} \,,\, H(\lambda, \sigma, l, \mathcal{D}) \right\} \leq 1 + e. \tag{A.56}$$

If $\mathrm{diam}(\mathcal{D}) \leq 1$ then we obtain (A.56) by observing, for every $i \in \{p+1, \ldots, m\}$, that $(i-l)! \geq (i-p)!$ and hence

$$\mathcal{H} \leq \mathrm{diam}(\mathcal{D})^{\lambda - \sigma} + \sum_{i=p+1}^{m} \frac{\mathrm{diam}(\mathcal{D})^{i-\sigma}}{(i-l)!} \leq 1 + \sum_{i=p+1}^{m} \frac{1}{(i-p)!} < 1 + e.$$

If $\operatorname{diam}(\mathcal{D}) > 1$ then we obtain (A.56) by observing that

$$\mathcal{H} \leq H(\lambda,\sigma,l,\mathcal{D}) \leq \max\left\{1+\sum_{i=p+1}^{m}\frac{1}{(i-l)!}, 1+\sum_{s=0}^{p-l}\frac{1}{s!}\right\} \leq \max\left\{1+\sum_{i=p+1}^{m}\frac{1}{(i-p)!}, 1+\sum_{s=0}^{p-l}\frac{1}{s!}\right\} < 1+e.$$

Hence (A.56) is proven. Together (A.54) and (A.56) establish that

$$\frac{\left|\left|\tilde{R}^{\phi}_l(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)}}{||y-x||_V^{\sigma-l}} \leq 1+e \tag{A.57}$$

Thus we may combine (A.54), (A.57), and the observation that for every $i \in \{p+1,\ldots,m\}$ we have $(i-l)! \geq (i-p)!$ to conclude that

$$\frac{\left|\left|\tilde{R}^{\phi}_l(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)}}{||y-x||_V^{\sigma-l}} \leq \min\left\{\operatorname{diam}(\mathcal{D})^{\lambda-\sigma}+\sum_{i=p+1}^{m}\frac{\operatorname{diam}(\mathcal{D})^{i-\sigma}}{(i-p)!}\ ,\ 1+e\right\}. \tag{A.58}$$

The arbitrariness of $l \in \{0,\ldots,p\}$ and the points $x,y \in \mathcal{D}$ with $x \neq y$ ensure that the estimate (A.58) is valid for every $l \in \{0,\ldots,p\}$ and every $x,y \in \mathcal{D}$ with $x \neq y$. Together, the definitions (A.52), the bounds in (**I**) of (A.51), and the estimates (A.58) allow us to conclude that $\phi_{[p]} = (\phi^{(0)},\ldots,\phi^{(p)}) \in \operatorname{Lip}(\sigma,\mathcal{D},W)$, and that

$$||\phi_{[p]}||_{\operatorname{Lip}(\sigma,\mathcal{D},W)} \leq \max\left\{1\ ,\ \min\left\{1+e\ ,\ \operatorname{diam}(\mathcal{D})^{\lambda-\sigma}+\sum_{j=p+1}^{m}\frac{\operatorname{diam}(\mathcal{D})^{j-\sigma}}{(j-p)!}\right\}\right\}. \tag{A.59}$$

The estimate (A.59) is precisely the estimate claimed in (A.49) for the case that $||\phi||_{\operatorname{Lip}(\lambda,\mathcal{D},W)} = 1$. This completes the proof of Claim A.9. ■

Returning to the proof of Lemma A.4 itself, suppose $\theta \in (0,n]$. A direct application of Claim A.9 with $\mathcal{D} := \Gamma$, $m := n$, $\lambda := \rho$ and $\sigma := \theta$ means that (A.49) yields that

$$||\psi_{[q]}||_{\operatorname{Lip}(\theta,\Gamma,W)} \leq \max\left\{1\ ,\ \min\left\{1+e\ ,\ \operatorname{diam}(\Gamma)^{\rho-\theta}+\sum_{j=q+1}^{n}\frac{\operatorname{diam}(\Gamma)^{j-\theta}}{(j-q)!}\right\}\right\}||\psi||_{\operatorname{Lip}(\rho,\Gamma,W)} \tag{A.60}$$

where $q \leq n-1$ since $\theta \in (0,n]$. By examining the definition of C_1 in (A.35), we see that (A.60) is the first part of the estimate claimed in (A.34). To derive the remaining estimate we note that $\theta \in (q,q+1]$. By appealing to Claim A.9, with $\mathcal{D} := \Gamma$, $m := n$, $\lambda := \rho$ and $\sigma := n$, we deduce via (A.49) that

$$||\psi_{[n-1]}||_{\operatorname{Lip}(n,\Gamma,W)} \leq \min\left\{1+e\ ,\ 1+\operatorname{diam}(\Gamma)^{\rho-n}\right\}||\psi||_{\operatorname{Lip}(\rho,\Gamma,W)}. \tag{A.61}$$

If we now appeal to Claim A.9 for $\mathcal{D} := \Gamma$, $m := n-1$, $\lambda := n$ and $\sigma := n-1$, then (A.49) and (A.61) give

$$\begin{aligned}
||\psi_{[n-2]}||_{\operatorname{Lip}(n-1,\Gamma,W)} &\overset{\text{(A.49)}}{\leq} \max\left\{1\ ,\ \min\left\{1+e\ ,\ 1+\operatorname{diam}(\Gamma)\right\}\right\}||\psi_{[n-1]}||_{\operatorname{Lip}(n,\Gamma,W)} \\
&\overset{\text{(A.61)}}{\leq} \min\left\{1+e\ ,\ 1+\operatorname{diam}(\Gamma)^{\rho-n}\right\}\min\left\{1+e\ ,\ 1+\operatorname{diam}(\Gamma)\right\}||\psi||_{\operatorname{Lip}(\rho,\Gamma,W)} \\
&= \left(1+\min\left\{e\ ,\ \operatorname{diam}(\Gamma)^{\rho-n}\right\}\right)\left(1+\min\left\{e\ ,\ \operatorname{diam}(\Gamma)\right\}\right)||\psi||_{\operatorname{Lip}(\rho,\Gamma,W)}.
\end{aligned}$$

We can now appeal to Claim A.9 for $\mathcal{D} := \Gamma$, $m := n-2$, $\lambda := n-1$, and $\sigma := n-2$. Proceeding inductively as $r = 0,1,\ldots,n-1$ increases, we establish via applying Claim A.9 for $\mathcal{D} := \Gamma$, $m := n-r$, $\lambda := n-(r-1)$, and $\sigma := n-r$ that for every $r \in \{0,1,\ldots,n-1\}$ we have that

$$||\psi_{[n-r-1]}||_{\operatorname{Lip}(n-r,\Gamma,W)} \leq \left(1+\min\left\{e\ ,\ \operatorname{diam}(\Gamma)^{\rho-n}\right\}\right)\left(1+\min\left\{e\ ,\ \operatorname{diam}(\Gamma)\right\}\right)^r||\psi||_{\operatorname{Lip}(\rho,\Gamma,W)}. \tag{A.62}$$

Taking $r := n - (q+1)$ in (A.62) yields that

$$||\psi_{[q]}||_{\mathrm{Lip}(q+1,\Gamma,W)} \leq \left(1 + \min\left\{e \ , \ \mathrm{diam}(\Gamma)^{\rho-n}\right\}\right)\left(1 + \min\left\{e \ , \ \mathrm{diam}(\Gamma)\right\}\right)^{n-(q+1)} ||\psi||_{\mathrm{Lip}(\rho,\Gamma,W)}. \quad (A.63)$$

As $\theta \in (q, q+1]$, we can appeal to Claim A.8 with $\mathcal{D} := \Gamma$, $m := q$, $\lambda := q+1$ and $\sigma := \theta$ to deduce via (A.39) that

$$||\psi_{[q]}||_{\mathrm{Lip}(\theta,\Gamma,W)} \leq \max\left\{1 \ , \ \min\left\{1+e \ , \ \mathrm{diam}(\Gamma)^{q+1-\theta}\right\}\right\} ||\psi_{[q]}||_{\mathrm{Lip}(q+1,\Gamma,W)}. \quad (A.64)$$

Together, (A.63) and (A.64) yield that

$$||\psi_{[q]}||_{\mathrm{Lip}(\theta,\Gamma,W)} \leq C_2 ||\psi||_{\mathrm{Lip}(\rho,\Gamma,W)} \quad (A.65)$$

for $C_2 > 0$ defined by

$$C_2 := \max\left\{1, \min\left\{1+e, \mathrm{diam}(\Gamma)^{q+1-\theta}\right\}\right\}\left(1 + \min\left\{e, \mathrm{diam}(\Gamma)^{\rho-n}\right\}\right)\left(1+\min\left\{e, \mathrm{diam}(\Gamma)\right\}\right)^{n-(q+1)}$$

as claimed in (A.36). The estimates (A.60) and (A.65) combine to give

$$||\psi_{[q]}||_{\mathrm{Lip}(\theta,\Gamma,W)} \leq \min\left\{C_1 \ , \ C_2\right\} ||\psi||_{\mathrm{Lip}(\rho,\Gamma,W)}$$

where

$$C_1 = \max\left\{1 \ , \ \min\left\{1+e \ , \ \mathrm{diam}(\Gamma)^{\rho-\eta} + \sum_{j=q+1}^{n} \frac{\mathrm{diam}(\Gamma)^{j-\theta}}{(j-q)!}\right\}\right\}$$

and

$$C_2 := \max\left\{1, \min\left\{1+e, \mathrm{diam}(\Gamma)^{q+1-\theta}\right\}\right\}\left(1 + \min\left\{e, \mathrm{diam}(\Gamma)^{\rho-n}\right\}\right)\left(1+\min\left\{e, \mathrm{diam}(\Gamma)\right\}\right)^{n-(q+1)}$$

as claimed in (A.34).

Finally, since $C_1 \leq 1+e$, (A.33) and (A.34) combine to yield that, for any $\theta \in (0, \rho)$, we have the estimate $||\psi_{[q]}||_{\mathrm{Lip}(\theta,\Gamma,W)} \leq (1+e)||\psi||_{\mathrm{Lip}(\rho,\Gamma,W)}$ as claimed in (A.37). This completes the proof of Lemma A.4. ■

B. Lipschitz Sandwich Theorems - Statements

Let V and W be Banach spaces and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Throughout, when referring to metric balls we use the convention that those denoted by $\mathbb{B}$ are taken to be open and those denoted by $\overline{\mathbb{B}}$ are taken to be closed.

Let $\Sigma \subset V$ be a closed subset. Suppose $\gamma > 0$ with $k \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k, k+1]$, and that $\psi = \left(\psi^{(0)}, \ldots, \psi^{(k)}\right) \in \mathrm{Lip}(\gamma, \Sigma, W)$. Given any fixed point $x \in \Sigma$ and any $l \in \{0, \ldots, k\}$ we may consider $\psi_x^{(l)} : V \to \mathcal{L}(V^{\otimes l}; W)$, defined for $y \in V$ and $v \in V^{\otimes l}$ by

$$\psi_x^{(l)}(y)[v] := \sum_{s=0}^{k-l} \frac{1}{s!}\psi^{(l+s)}(x)\left[v \otimes (y-x)^{\otimes s}\right], \quad (B.1)$$

as a proposal for how $\psi^{(l)}$ may be extended to the entirety of V. In [Ste70], for each $l \in \{0, \ldots, k\}$ and any point $y \in V \setminus \Sigma$, Stein uses an appropriately weighted average of the collection $\left\{\psi_x^{(l)}(y) \ : \ x \in \Sigma\right\}$ to define an extension of $\psi^{(l)}$ at the point $y \in V \setminus \Sigma$. Moreover, Stein proves that the collection of extensions of the functions $\psi^{(0)}, \ldots, \psi^{(k)}$ to the entirety of V gives an extension of ψ to the entirety of V as an element in $\mathrm{Lip}(\gamma, V, W)$; see Theorem 4 in Chapter VI of [Ste70]. It is unreasonable to expect this extension to be unique. For example, take $\Sigma := [-1, 1] \subset \mathbb{R}$ and $\psi^{(0)} : \Sigma \to \mathbb{R}$ defined for $t \in \Sigma$ by $\psi^{(0)}(t) := |t|$. Then $\psi^{(0)} \in \mathrm{Lip}(1, \Sigma, \mathbb{R})$ and there are clearly numerous distinct ways to extend ψ to an element in $\mathrm{Lip}(1, \mathbb{R}, \mathbb{R})$.

Suppose that $B \subset \Sigma$ is a non-empty closed subset. Then *Stein's extension theorem* (Theorem 4 in Chapter VI of [Ste70]) tells us that any element in $\mathrm{Lip}(\gamma, B, W)$ can be extended to an element in $\mathrm{Lip}(\gamma, \Sigma, W)$. We are interested in understanding when extensions of an element in $\mathrm{Lip}(\gamma, B, W)$ to $\mathrm{Lip}(\gamma, \Sigma, W)$ are forced to remain, in some sense, close throughout Σ. We consider the following problem. Given elements $\psi = \left(\psi^{(0)}, \ldots, \psi^{(k)}\right)$ and $\varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(k)}\right)$ in $\mathrm{Lip}(\gamma, \Sigma, W)$, when does knowing that ψ and φ are, in some sense, "close" on B ensures that ψ and φ remain "close", in some possibly different sense, throughout Σ.

The results presented in this section regarding this problem articulate properties of $\mathrm{Lip}(\gamma)$ functions in a form that is convenient for our purposes. In particular, when $\gamma \in (0,1]$ and $W := \mathbb{R}$, all the results of this section follow from consideration of the maximal and minimal extensions of Whitney [Whi34] and McShane [McS34] respectively. The main content of our results is the setting that $\gamma > 1$, for which case we have been unable to locate formal statements of these properties within the existing literature.

The following *Lipschitz Sandwich Theorem* gives a condition for the subset B and precise meanings for the notions of closeness to be considered between ψ and φ on B and Σ respectively under which this problem has an affirmative answer.

Theorem B.1 (Lipschitz Sandwich Theorem). *Let V and W be Banach spaces, and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume that $\Sigma \subset V$ is non-empty and closed. Let $\varepsilon, K_0 > 0$, and $\gamma > \eta > 0$ with $k, q \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k, k+1]$ and $\eta \in (q, q+1]$. Then there exist constants $\delta_0 = \delta_0(\varepsilon, K_0, \gamma, \eta) > 0$ and $\varepsilon_0 = \varepsilon_0(\varepsilon, K_0, \gamma, \eta) > 0$ for which the following is true.*

Suppose $B \subset \Sigma$ is a closed subset that is a δ_0-cover of Σ in the sense that

$$\Sigma \subset \bigcup_{x \in B} \overline{\mathbb{B}}_V(x, \delta_0) = B_{\delta_0} := \{v \in V \ : \ \exists z \in B \text{ such that } ||v - z||_V \leq \delta_0\}. \tag{B.2}$$

Suppose $\psi = \left(\psi^{(0)}, \ldots, \psi^{(k)}\right), \varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(k)}\right) \in \mathrm{Lip}(\gamma, \Sigma, W)$ with $||\psi||_{\mathrm{Lip}(\gamma,\Sigma,W)}, ||\varphi||_{\mathrm{Lip}(\gamma,\Sigma,W)} \leq K_0$. Further suppose that for every $l \in \{0, \ldots, k\}$ and every $x \in B$ the difference $\psi^{(l)}(x) - \varphi^{(l)}(x) \in \mathcal{L}(V^{\otimes l}; W)$ satisfies the bound

$$\left|\left|\psi^{(l)}(x) - \varphi^{(l)}(x)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \varepsilon_0. \tag{B.3}$$

Then we may conclude that

$$\left|\left|\psi_{[q]} - \varphi_{[q]}\right|\right|_{\mathrm{Lip}(\eta,\Sigma,W)} \leq \varepsilon \tag{B.4}$$

where $\psi_{[q]} := \left(\psi^{(0)}, \ldots, \psi^{(q)}\right)$ and $\varphi_{[q]} := \left(\varphi^{(0)}, \ldots, \varphi^{(q)}\right)$.

Remark B.2. Using the same notation as in Theorem B.1, the estimates (B.3) throughout B are a weaker condition than $||\psi - \varphi||_{\mathrm{Lip}(\gamma,B,W)} \leq \varepsilon_0$. The bound $||\psi - \varphi||_{\mathrm{Lip}(\gamma,B,W)} \leq \varepsilon_0$ implies that the pointwise estimates in (B.3) are valid. But the converse is *not* true since the pointwise estimates in (B.3) alone are insufficient to establish the required estimates for the remainder terms associated to the difference $\psi - \varphi$ (cf. Definition A.2).

Remark B.3. The restriction that $\eta \in (0, \gamma)$ in Theorem B.1 is necessary; the theorem is *false* for $\eta := \gamma$. As an example, fix $K_0, \varepsilon > 0$ with $\varepsilon < 2K_0$, let $\delta > 0$ and consider a fixed $N \in \mathbb{Z}_{\geq 1}$ for which $1/N < \delta$. Define $\Sigma := \{0, 1/N\} \subset \mathbb{R}$ and $B := \Sigma \setminus \{1/N\} = \{0\} \subset \mathbb{R}$. Then we have that $\Sigma \subset [-\delta, \delta]$ and so B is a δ-cover of Σ as required in (B.2). Define $\psi, \varphi : \Sigma \to \mathbb{R}$ by $\psi(0) := 0$, $\psi(1/N) := K_0/N$ and $\varphi(0) := 0$, $\varphi(1/N) := -K_0/N$. Then $\psi, \varphi \in \mathrm{Lip}(1, \Sigma, \mathbb{R})$ with $||\psi||_{\mathrm{Lip}(1,\Sigma,W)} = ||\varphi||_{\mathrm{Lip}(1,\Sigma,W)} = K_0$ and $\psi - \varphi \equiv 0$ throughout B, establishing the validity of the bounds (B.3) for any $\varepsilon_0 \geq 0$. However $|(\psi - \varphi)(1/N) - (\psi - \varphi)(0)| = 2K_0/N = 2K_0|1/N - 0|$, which means that $||\psi - \varphi||_{\mathrm{Lip}(1,\Sigma,\mathbb{R})} = 2K_0 > \varepsilon$.

Remark B.4. It may initially appear that the Theorem should be valid for any fixed ε_0 with $\varepsilon_0 < \varepsilon$ by suitably restricting δ_0, rather than having to allow ε_0 to depend on ε, K_0, γ and η. But this is *not* the case. If we only assume $\varepsilon_0 < \varepsilon$, then the estimates in (B.3) can even be insufficient to establish that $||\psi - \varphi||_{\mathrm{Lip}(\eta,B,W)} \leq \varepsilon$. For example, let $\gamma := 1$, $\eta := 1/2$, and fix $0 < \varepsilon_0 < \varepsilon < 1 < K_0$ such that $2\varepsilon_0 K_0 > \varepsilon^2$. Define $x_0 := 2\varepsilon_0/K_0 > 0$ and consider $\Sigma = B := \{0, x_0\}$. Define $\psi, \varphi : \Sigma \to \mathbb{R}$ by $\psi(0) := -\varepsilon_0$, $\psi(x_0) := \varepsilon_0$ and $\varphi(0) := 0 =: \varphi(x_0)$. Then $\psi, \varphi \in \mathrm{Lip}(1, \Sigma, \mathbb{R})$, with $||\varphi||_{\mathrm{Lip}(1,\Sigma,\mathbb{R})} = 0$ and $||\psi||_{\mathrm{Lip}(1,\Sigma,\mathbb{R})} = K_0$. Moreover, $|\psi - \varphi| = |\psi| \leq \varepsilon_0$ throughout $\Sigma = B$ so that the estimates (B.3) are valid. However we may also compute that $|(\psi - \varphi)(x_0) - (\psi - \varphi)(0)| = 2\varepsilon_0 = 2\varepsilon_0\sqrt{1/x_0}\sqrt{|x_0 - 0|} = \sqrt{2\varepsilon_0 K_0}\sqrt{|x_0 - 0|}$ so that $||\psi - \varphi||_{\mathrm{Lip}(1/2,B,\mathbb{R})} = \sqrt{2\varepsilon_0 K_0} > \varepsilon$.

The issue described in Remark B.4 is only present when the cardinality of the subset B is greater than 1, i.e. when B contains at least two distinct points. When B consists of a single point we can in fact allow for an arbitrary

$\varepsilon_0 \in [0, \varepsilon)$ in Theorem B.1 rather than having to allow ε_0 to depend on ε, K_0, γ and η. The precise statement is recorded in the following theorem.

Theorem B.5 (Single-Point Lipschitz Sandwich Theorem). *Let V and W be Banach spaces and assume that the tensor powers of V are all equipped with admissible tensor norms (cf. Definition A.1). Assume that $\Sigma \subset V$ is closed and non-empty. Let $\varepsilon, K_0 > 0$, $\gamma > \eta > 0$ with $k, q \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k, k+1]$ and $\eta \in (q, q+1]$, and $0 \leq \varepsilon_0 < \min\{2K_0, \varepsilon\}$. Then there exists a constant $\delta_0 = \delta_0(\varepsilon, \varepsilon_0, K_0, \gamma, \eta) > 0$ for which the following is true.*

Suppose $p \in \Sigma$ and that $\psi = \left(\psi^{(0)}, \ldots, \psi^{(k)}\right)$ and $\varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(k)}\right)$ are elements in $\mathrm{Lip}(\gamma, \Sigma, W)$ with $||\psi||_{\mathrm{Lip}(\gamma,\Sigma,W)}, ||\varphi||_{\mathrm{Lip}(\gamma,\Sigma,W)} \leq K_0$. Further suppose that for every $l \in \{0, \ldots, k\}$ the difference $\psi^{(l)}(p) - \varphi^{(l)}(p) \in \mathcal{L}(V^{\otimes l}; W)$ satisfies the bound

$$\left|\left|\psi^{(l)}(p) - \varphi^{(l)}(p)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \varepsilon_0. \tag{B.5}$$

Then we may conclude that

$$\left|\left|\psi_{[q]} - \varphi_{[q]}\right|\right|_{\mathrm{Lip}(\eta, \overline{\mathbb{B}}_V(p,\delta_0) \,\cap\, \Sigma, W)} \leq \varepsilon \tag{B.6}$$

where $\psi_{[q]} := \left(\psi^{(0)}, \ldots, \psi^{(q)}\right)$ and $\varphi_{[q]} := \left(\varphi^{(0)}, \ldots, \varphi^{(q)}\right)$.

Establishing Theorem B.5 will form the first step in our proof of Theorem B.1.

Continuing to discuss the consequences of Theorem B.1, when the subset $\Sigma \subset V$ is compact we may use Theorem B.1 to prove the following result.

Corollary B.6. *Let V and W be Banach spaces, and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume that $\Sigma \subset V$ is compact. Let $\varepsilon, K_0 > 0$, and $\gamma > \eta > 0$ with $k, q \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k, k+1]$ and $\eta \in (q, q+1]$. Let $\delta_0 = \delta_0(\varepsilon, K_0, \gamma, \eta) > 0$ and $\varepsilon_0 = \varepsilon_0(\varepsilon, K_0, \gamma, \eta) > 0$ denote the constants arising from Theorem B.1, and let $N = N(\Sigma, \varepsilon, K_0, \gamma, \eta) \in \mathbb{Z}_{\geq 0}$ denote the δ_0-covering number of Σ. That is,*

$$N := N_{\mathrm{cov}}(\Sigma, V, \delta_0) = \min\left\{ a \in \mathbb{Z} : \text{ There exists } x_1, \ldots, x_a \in \Sigma \text{ such that } \Sigma \subset \bigcup_{j=1}^{a} \overline{\mathbb{B}}_V(x_j, \delta_0) \right\}. \tag{B.7}$$

Then there is a finite subset $\Sigma_N = \{z_1, \ldots, z_N\} \subset \Sigma$ for which the following is true.

Suppose $\psi = \left(\psi^{(0)}, \ldots, \psi^{(k)}\right), \varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(k)}\right) \in \mathrm{Lip}(\gamma, \Sigma, W)$ with $||\psi||_{\mathrm{Lip}(\gamma,\Sigma,W)}, ||\varphi||_{\mathrm{Lip}(\gamma,\Sigma,W)} \leq K_0$. Further suppose that for every $l \in \{0, \ldots, k\}$ and every $j \in \{1, \ldots, N\}$ the difference $\psi^{(l)}(z_j) - \varphi^{(l)}(z_j) \in \mathcal{L}(V^{\otimes l}; W)$ satisfies the bound

$$\left|\left|\psi^{(l)}(z_j) - \varphi^{(l)}(z_j)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \varepsilon_0. \tag{B.8}$$

Then we may conclude that

$$\left|\left|\psi_{[q]} - \varphi_{[q]}\right|\right|_{\mathrm{Lip}(\eta,\Sigma)} \leq \varepsilon \tag{B.9}$$

where $\psi_{[q]} := \left(\psi^{(0)}, \ldots, \psi^{(q)}\right)$ and $\varphi_{[q]} := \left(\varphi^{(0)}, \ldots, \varphi^{(q)}\right)$.

Remark B.7. When $N \in \mathbb{Z}_{\geq 1}$ defined in (B.7) is less than the cardinality of Σ, Corollary B.6 guarantees that we are able to identify a strictly smaller collection of points at which the behaviour of a $\mathrm{Lip}(\gamma, \Sigma, W)$ function determines the functions $\mathrm{Lip}(\eta)$-behaviour up to an arbitrarily small error over the entire set Σ. That is, using the notation of Corollary B.6, if $F \in \mathrm{Lip}(\gamma, \Sigma_N, W)$ then any two extensions ψ and φ of F to elements in $\mathrm{Lip}(\gamma, \Sigma, W)$ with $\mathrm{Lip}(\gamma, \Sigma, W)$-norms bounded above by K_0 can differ, in the $\mathrm{Lip}(\eta)$-sense, by at most ε throughout Σ.

Proof of Corollary B.6. Let V and W be Banach spaces, and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume that $\Sigma \subset V$ is compact. Let $\varepsilon, K_0 > 0$, and $\gamma > \eta > 0$ with $k, q \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k, k+1]$ and $\eta \in (q, q+1]$. Retrieve the constants $\delta_0 = \delta_0(\varepsilon, K_0, \gamma, \eta) > 0$ and $\varepsilon_0 = \varepsilon_0(\varepsilon, K_0, \gamma, \eta) > 0$ arising from Theorem B.1 for these choices of ε, K_0, γ and η. Note that we are not actually applying Theorem B.1, but simply retrieving constants in preparation for its future application. Define

$N = N(\Sigma, \varepsilon, K_0, \gamma, \eta) \in \mathbb{Z}_{\geq 0}$ to be the δ_0-covering number for Σ. That is,

$$N := N_{\text{cov}}(\Sigma, V, \delta_0) = \min \left\{ a \in \mathbb{Z} : \text{ There exists } x_1, \ldots, x_a \in \Sigma \text{ such that } \Sigma \subset \bigcup_{j=1}^{a} \overline{\mathbb{B}}_V(x_j, \delta_0) \right\}. \quad \text{(B.10)}$$

Let $z_1, \ldots, z_N \in \Sigma$ be any collection of N-points in Σ for which

$$\Sigma \subset \bigcup_{j=1}^{N} \overline{\mathbb{B}}_V(z_j, \delta_0). \quad \text{(B.11)}$$

Set $\Sigma_N := \{z_1, \ldots, z_N\}$.

Suppose $\psi = \left(\psi^{(0)}, \ldots, \psi^{(k)}\right), \varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(k)}\right) \in \text{Lip}(\gamma, \Sigma, W)$ with $||\psi||_{\text{Lip}(\gamma,\Sigma,W)}, ||\varphi||_{\text{Lip}(\gamma,\Sigma,W)} \leq K_0$. Further suppose that for every $l \in \{0, \ldots, k\}$ and every $j \in \{1, \ldots, N\}$ the difference $\psi^{(l)}(z_j) - \varphi^{(l)}(z_j) \in \mathcal{L}(V^{\otimes l}; W)$ satisfies the bound

$$\left|\left|\psi^{(l)}(z_j) - \varphi^{(l)}(z_j)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \varepsilon_0. \quad \text{(B.12)}$$

Then (B.11) and (B.12) enable us to appeal to Theorem B.1, with $B := \Sigma_N$, to conclude that $||\psi_{[q]} - \varphi_{[q]}||_{\text{Lip}(\eta,\Sigma,W)} \leq \varepsilon$ where $\psi_{[q]} := \left(\psi^{(0)}, \ldots, \psi^{(q)}\right)$ and $\varphi_{[q]} := \left(\varphi^{(0)}, \ldots, \varphi^{(q)}\right)$. This is precisely the estimate claimed in (B.9). This completes the proof of Corollary B.6. ■

Returning our attention to Theorem B.1 itself, the Lipschitz estimates obtained in the conclusion (B.4) yield pointwise estimates for the difference $\psi^{(0)} - \varphi^{(0)} : \Sigma \to W$. In particular, we may conclude that $\left|\left|\psi^{(0)} - \varphi^{(0)}\right|\right|_{C^0(\Sigma;W)} \leq \varepsilon$. However such pointwise estimates can be established directly without needing to appeal to Theorem B.1. Moreover, this direct approach allows us to obtain estimates for the difference $\psi^{(l)} - \varphi^{(l)} : \Sigma \to \mathcal{L}(V^{\otimes l}; W)$ for every $l \in \{0, \ldots, k\}$. An additional benefit is that we are able to provide a more explicit constant δ_0 for which we require the subset $B \subset \Sigma$ to be a δ_0-cover of Σ. The precise result is recorded in the following theorem.

Theorem B.8 (Pointwise Lipschitz Sandwich Theorem). *Let V and W be Banach spaces, and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume that $\Sigma \subset V$ is closed. Let $K_0, \gamma, \varepsilon > 0$ with $k \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k, k+1]$ and $0 \leq \varepsilon_0 < \min\{2K_0, \varepsilon\}$. Then given any $l \in \{0, \ldots, k\}$, there exists a constant $\delta_0 = \delta_0(\varepsilon, \varepsilon_0, K_0, \gamma, l) > 0$, defined by*

$$\delta_0 := \sup\left\{\theta > 0 \ : \ 2K_0\theta^{\gamma - l} + \varepsilon_0 e^{\theta} \leq \min\{2K_0 \ , \ \varepsilon\}\right\} > 0, \quad \text{(B.13)}$$

for which the following is true.

Suppose $B \subset \Sigma$ is a δ_0-cover of Σ in the sense that

$$\Sigma \subset \bigcup_{x \in B} \overline{\mathbb{B}}_V(x, \delta_0) = B_{\delta_0} := \{v \in V \ : \ \exists z \in B \text{ such that } ||v - z||_V \leq \delta_0\}. \quad \text{(B.14)}$$

Suppose $\psi = \left(\psi^{(0)}, \ldots, \psi^{(k)}\right), \varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(k)}\right) \in \text{Lip}(\gamma, \Sigma, W)$ with $||\psi||_{\text{Lip}(\gamma,\Sigma,W)}, ||\varphi||_{\text{Lip}(\gamma,\Sigma,W)} \leq K_0$. Further suppose that for every $j \in \{0, \ldots, k\}$ and every $x \in B$ the difference $\psi^{(j)}(x) - \varphi^{(j)}(x) \in \mathcal{L}(V^{\otimes j}; W)$ satisfies the bound

$$\left|\left|\psi^{(j)}(x) - \varphi^{(j)}(x)\right|\right|_{\mathcal{L}(V^{\otimes j};W)} \leq \varepsilon_0. \quad \text{(B.15)}$$

Then we may conclude that for every $s \in \{0, \ldots, l\}$ and every $x \in \Sigma$ that

$$\left|\left|\psi^{(s)}(x) - \varphi^{(s)}(x)\right|\right|_{\mathcal{L}(V^{\otimes s};w)} \leq \varepsilon. \quad \text{(B.16)}$$

Remark B.9. In contrasts to Theorem B.1 we are able to deal with arbitrary $\varepsilon_0 < \varepsilon$ by suitably restricting δ_0. The issue outlined in Remark B.4 is no longer a problem in this setting since the same notion of closeness is used in both the assumption (B.15) and the conclusion (B.16).

Remark B.10. It follows from (B.13) that, for every $l \in \{0,\ldots,k\}$, the constant $\delta_0 = \delta_0(\varepsilon,\varepsilon_0,K_0,\gamma,l) > 0$ is bounded above by 1. Consequently, if the constants ε, ε_0, K_0, and γ remain fixed, we may conclude that the constant δ_0 is decreasing with respect to the argument $l \in \{0,\ldots,k\}$ in the sense that the mapping $l \mapsto \delta_0(\varepsilon,\varepsilon_0,K_0,\gamma,l)$ is a decreasing function on $\{0,\ldots,k\}$.

The following result is a consequence of Theorem B.8 when the subset $\Sigma \subset V$ is compact.

Corollary B.11. *Let V and W be Banach spaces, and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume that $\Sigma \subset V$ is compact. Let $K_0,\gamma,\varepsilon > 0$, with $k \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k,k+1]$, and $0 \leq \varepsilon_0 < \min\{2K_0,\varepsilon\}$. Given $l \in \{0,\ldots,k\}$ let $\delta_0 = \delta_0(\varepsilon,\varepsilon_0,K_0,\gamma,l) > 0$ denote the constant arising from Theorem B.8 (cf. (B.13)). Let $N = N(\Sigma,\varepsilon,\varepsilon_0,K_0,\gamma,l) \in \mathbb{Z}_{\geq 0}$ denote the δ_0-covering number of Σ. That is,*

$$N := N_{\mathrm{cov}}(\Sigma,V,\delta_0) = \min\left\{ a \in \mathbb{Z} : \textit{ There exists } x_1,\ldots,x_a \in \Sigma \textit{ such that } \Sigma \subset \bigcup_{j=1}^{a} \overline{\mathbb{B}}_V(x_j,\delta_0) \right\}. \tag{B.17}$$

Then there is a finite subset $\Sigma_N = \{z_1,\ldots,z_N\} \subset \Sigma$ for which the following is true.

Suppose $\psi = \left(\psi^{(0)},\ldots,\psi^{(k)}\right), \varphi = \left(\varphi^{(0)},\ldots,\varphi^{(k)}\right) \in \mathrm{Lip}(\gamma,\Sigma,W)$ with $||\psi||_{\mathrm{Lip}(\gamma,\Sigma,W)}, ||\varphi||_{\mathrm{Lip}(\gamma,\Sigma,W)} \leq K_0$. Further suppose that for every $i \in \{0,\ldots,k\}$ and every $j \in \{1,\ldots,N\}$ the difference $\psi^{(i)}(z_j) - \varphi^{(i)}(z_j) \in \mathcal{L}(V^{\otimes i};W)$ satisfies the bound

$$\left|\left|\psi^{(i)}(z_j) - \varphi^{(i)}(z_j)\right|\right|_{\mathcal{L}(V^{\otimes s};W)} \leq \varepsilon_0. \tag{B.18}$$

Then we may conclude that for every $s \in \{0,\ldots,l\}$ and every $x \in \Sigma$ that

$$\left|\left|\psi^{(s)}(x) - \varphi^{(s)}(x)\right|\right|_{\mathcal{L}(V^{\otimes s};w)} \leq \varepsilon. \tag{B.19}$$

Remark B.12. When $N \in \mathbb{Z}_{\geq 1}$ defined in (B.17) is less than the cardinality of Σ, Corollary B.11 guarantees that we are able to identify a strictly smaller collection of points Σ_N such that the behaviour of a $\mathrm{Lip}(\gamma,\Sigma,W)$ function $F = \left(F^{(0)},\ldots,F^{(k)}\right)$ on Σ_N determines the pointwise behaviour of $F_{[l]} = \left(F^{(0)},\ldots,F^{(l)}\right)$ over the entire set Σ up to an arbitrarily small error. That is, using the notation of Corollary B.11, if $F \in \mathrm{Lip}(\gamma,\Sigma_N,W)$ and $\psi = \left(\psi^{(0)},\ldots,\psi^{(k)}\right)$ and $\varphi = \left(\varphi^{(0)},\ldots,\varphi^{(k)}\right)$ are both extensions of F to $\mathrm{Lip}(\gamma,\Sigma,W)$ with $\mathrm{Lip}(\gamma,\Sigma,W)$-norms bounded above by K_0, then for every $s \in \{0,\ldots,l\}$ the functions $\psi^{(l)}$ and $\varphi^{(l)}$ may only differ, in the pointwise sense, by at most ε throughout Σ.

Proof of Corollary B.11. Let V and W be Banach spaces, and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume that $\Sigma \subset V$ is compact. Let $K_0,\gamma,\varepsilon > 0$, with $k \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k,k+1]$, and $0 \leq \varepsilon_0 < \min\{2K_0,\varepsilon\}$. Given $l \in \{0,\ldots,k\}$ retrieve the constant $\delta_0 = \delta_0(\varepsilon,\varepsilon_0,K_0,\gamma,l) > 0$ denote the constant arising from Theorem B.8 for these choices of ε, ε_0, K_0, γ, and l. Note that we are not actually applying Theorem B.1, but simply retrieving a constant in preparation for its future application. Define $N = N(\Sigma,\varepsilon,\varepsilon_0,K_0,\gamma,l) \in \mathbb{Z}_{\geq 0}$ to be the δ_0-covering number of Σ. That is,

$$N := N_{\mathrm{cov}}(\Sigma,V,\delta_0) = \min\left\{ a \in \mathbb{Z} : \text{ There exists } x_1,\ldots,x_a \in \Sigma \text{ such that } \Sigma \subset \bigcup_{j=1}^{a} \overline{\mathbb{B}}_V(x_j,\delta_0) \right\}. \tag{B.20}$$

Let $z_1,\ldots,z_N \in \Sigma$ be any collection of N-points in Σ for which

$$\Sigma \subset \bigcup_{j=1}^{N} \overline{\mathbb{B}}_V(z_j,\delta_0). \tag{B.21}$$

Set $\Sigma_N := \{z_1,\ldots,z_N\}$.

Now suppose that both $\psi = \left(\psi^{(0)},\ldots,\psi^{(k)}\right), \varphi = \left(\varphi^{(0)},\ldots,\varphi^{(k)}\right) \in \mathrm{Lip}(\gamma,\Sigma,W)$ have their norms bounded by K_0, i.e. $||\psi||_{\mathrm{Lip}(\gamma,\Sigma,W)}, ||\varphi||_{\mathrm{Lip}(\gamma,\Sigma,W)} \leq K_0$. Further suppose that for every $i \in \{0,\ldots,k\}$ and every $j \in \{1, \ldots, N\}$ the difference $\psi^{(i)}(z_j) - \varphi^{(i)}(z_j) \in \mathcal{L}(V^{\otimes i}; W)$ satisfies the bound

$$\left|\left|\psi^{(i)}(z_j) - \varphi^{(i)}(z_j)\right|\right|_{\mathcal{L}(V^{\otimes s};W)} \leq \varepsilon_0. \tag{B.22}$$

Together, (B.21) and (B.22) provide the hypotheses required to allow us to appeal to Theorem B.8 with the subset B of that result as the subset Σ_N here. A consequence of doing so is that, for every $s \in \{0, \ldots, l\}$ and $x \in \Sigma$, we have that $\left|\left|\psi^{(s)}(x) - \varphi^{(s)}(x)\right|\right|_{\mathcal{L}(V^{\otimes s};W)} \leq \varepsilon$ as claimed in (B.19). This completes the proof of Corollary B.11. ■

C. Lipschitz Sandwich Theorems - Proofs

In this section we provide rigorous proofs of the results stated in Appendix B. Let V and W be Banach spaces and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Throughout, when referring to metric balls we use the convention that those denoted by $\mathbb{B}$ are taken to be open and those denoted by $\overline{\mathbb{B}}$ are taken to be closed.

Let $\Sigma \subset V$ be closed and $\gamma > 0$ with $k \in \mathbb{Z}_{\geq 0}$ with $\gamma \in (k, k+1]$. We begin by establishing estimates for an element $F = \left(F^{(0)}, \ldots, F^{(k)}\right) \in \operatorname{Lip}(\gamma, \Sigma, W)$ at points $x \in \Sigma$ in terms of the value of F at some fixed point $p \in \Sigma$. The precise result is the following lemma.

Lemma C.1 (Pointwise Estimates). *Let V and W be Banach spaces, and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume $\Gamma \subset V$ is closed with $p \in \Gamma$. Let $A, \rho > 0$, $r_0 \geq 0$, and $n \in \mathbb{Z}_{\geq 0}$ such that $\rho \in (n, n+1]$. Let $F = \left(F^{(0)}, \ldots, F^{(n)}\right) \in \operatorname{Lip}(\rho, \Gamma, W)$ with $||F||_{\operatorname{Lip}(\rho,\Gamma,W)} \leq A$. For every $j \in \{0, \ldots, n\}$ let $R^F_j : \Gamma \times \Gamma \to \mathcal{L}(V^{\otimes j}; W)$ denote the remainder term associated to $F^{(j)}$, defined for $x, y \in \Gamma$ and $v \in V^{\otimes j}$ by*

$$R^F_j(x, y)[v] := F^{(j)}(y)[v] - \sum_{s=0}^{n-j} \frac{1}{s!} F^{(j+s)}(x) \left[v \otimes (y - x)^{\otimes s}\right]. \tag{C.1}$$

Then for for every $l \in \{0, \ldots, n\}$, any $x, y \in \Sigma$, and any $\theta \in (n, \rho)$, we have that

$$\left|\left|R^F_l(x, y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq A \left(\operatorname{dist}(x, p) + \operatorname{dist}(y, p)\right)^{\rho - \theta} ||y - x||^{\theta - l}_V. \tag{C.2}$$

Further, suppose $q \in \{0, \ldots, n\}$ and that for every $s \in \{0, \ldots, q\}$ we have $\left|\left|F^{(s)}(p)\right|\right|_{\mathcal{L}(V^{\otimes s};W)} \leq r_0$. Then for any $l \in \{0, \ldots, q\}$ and any $x \in \Gamma$ we have that

$$\left|\left|F^{(l)}(x)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \min\left\{ A\ ,\ A\left[\operatorname{dist}(x, p)^{\rho - l} + S_{l,q}(x, p)\right] + r_0 \sum_{j=0}^{q-l} \frac{1}{j!} \operatorname{dist}(x, p)^j \right\} \tag{C.3}$$

where

$$S_{l,q}(x, p) := \begin{cases} \sum_{j=q+1-l}^{n-l} \frac{1}{j!} \operatorname{dist}(x, p)^j & \text{if } q < n \\ 0 & \text{if } q = n. \end{cases} \tag{C.4}$$

Proof of Lemma C.1. Let V and W be Banach spaces, and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume that $\Gamma \subset V$ is closed and that $\rho > 0$ with $n \in \mathbb{Z}_{\geq 0}$ such that $\rho \in (n, n+1]$. Suppose that, for $l \in \{0, \ldots, n\}$, we have functions $F^{(l)} : \Gamma \to \mathcal{L}(V^{\otimes l}; W)$ such that $F = (F^{(0)}, \ldots, F^{(n)}) \in \operatorname{Lip}(\rho, \Gamma, W)$ with $||F||_{\operatorname{Lip}(\rho,\Gamma,W)} \leq A$. For each $j \in \{0, \ldots, n\}$ let $R^F_j : \Gamma \times \Gamma \to \mathcal{L}(V^{\otimes j}; W)$ denote the remainder term associated to $F^{(j)}$, defined for $x, y \in \Gamma$ and $v \in V^{\otimes j}$ by

$$R^F_j(x, y)[v] := F^{(j)}(y)[v] - \sum_{s=0}^{n-j} \frac{1}{s!} F^{(j+s)}(x) \left[v \otimes (y - x)^{\otimes s}\right]. \tag{C.5}$$

As a consequence of $F \in \operatorname{Lip}(\rho, \Gamma, W)$ with $||F||_{\operatorname{Lip}(\rho,\Gamma,W)} \leq A$, whenever $l \in \{0, \ldots, n\}$ and $x, y \in \Gamma$, we have the bounds (cf. (A.4) and (A.6))

$$\textbf{(I)} \quad \left|\left|F^{(l)}(x)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq A \qquad \text{and} \qquad \textbf{(II)} \quad \left|\left|R_l^F(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq A||y-x||_V^{\rho-l}. \tag{C.6}$$

If $l \in \{0,\ldots,n\}$, $x,y \in \Gamma$, and $\theta \in (n,\rho)$, we use **(II)** of (C.6) to compute that

$$\begin{aligned}\left|\left|R_l^F(x,y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \stackrel{\textbf{(II)} \text{ of (C.6)}}{\leq} A||y-x||_V^{\rho-l} &= A||y-x||_V^{\rho-\theta}||y-x||_V^{\theta-l} \\ &\leq A\left(||x-p||_V + ||y-p||_V\right)^{\rho-\theta}||y-x||_V^{\theta-l} \\ &= A\left(\mathrm{dist}(x,p)+\mathrm{dist}(y,p)\right)^{\rho-\theta}||y-x||_V^{\theta-l}\end{aligned}$$

as claimed in (C.2).

Now suppose that $q \in \{0,\ldots,n\}$ and that for every $s \in \{0,\ldots,q\}$ we have $\left|\left|F^{(s)}(p)\right|\right|_{\mathcal{L}(V^{\otimes s};W)} \leq r_0$. Given $l \in \{0,\ldots,q\}$, $x \in \Gamma$, and $v \in V^{\otimes l}$, recalling that the tensor powers of V are equipped with admissible norms (cf. Definition A.1), we may use (C.5) and **(II)** of (C.6) to obtain that

$$\left|\left|F^{(l)}(x)[v]\right|\right|_W \leq \sum_{s=0}^{n-l}\frac{1}{s!}\left|\left|F^{(s)}(p)\right|\right|_{\mathcal{L}(V^{\otimes s};W)}||x-p||_V^s||v||_{V^{\otimes l}} + A||x-p||_V^{\rho-l}||v||_{V^{\otimes l}}. \tag{C.7}$$

If $q=n$ then (C.7) tells us that

$$\left|\left|F^{(l)}(x)[v]\right|\right|_W \leq \left(r_0\sum_{s=0}^{q-l}\frac{1}{s!}||x-p||_V^s + A||x-p||_V^{\rho-l}\right)||v||_{V^{\otimes l}}. \tag{C.8}$$

Whilst if $q<n$, we deduce from (C.7) that

$$\left|\left|F^{(l)}(x)[v]\right|\right|_W \leq \left(r_0\sum_{s=0}^{q-l}\frac{1}{s!}||x-p||_V^s + A\sum_{s=q-l+1}^{n-l}\frac{1}{s!}||x-p||_V^s + A||x-p||_V^{\rho-l}\right)||v||_{V^{\otimes l}}. \tag{C.9}$$

If we let $S_{l,q}(x,p)$ be defined as in (C.4), then (C.8) and (C.9) combine to yield that

$$\left|\left|F^{(l)}(x)[v]\right|\right|_W \leq \left(r_0\sum_{s=0}^{q-l}\frac{1}{s!}\,\mathrm{dist}(x,p)^s + A\left[\mathrm{dist}(x,p)^\rho + S_{l,q}(x,p)\right]\right)||v||_{V^{\otimes l}}. \tag{C.10}$$

Taking the supremum over $v \in V^{\otimes l}$ with unit $V^{\otimes l}$ norm in (C.10) yields the second estimate claimed in (C.3). The first estimate claimed in (C.3) follows from **(I)** in (C.6). This completes the proof of Lemma C.1. ■

We can use the pointwise estimates in Lemma C.1 to prove the *Pointwise Lipschitz Sandwich Theorem* stated as Theorem B.8 in Section B.

Proof of Theorem B.8. Assume that V and W are Banach spaces and that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume $\Sigma \subset V$ is a closed subset. Let $K_0,\gamma,\varepsilon > 0$ with $k \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k,k+1]$, and $0 \leq \varepsilon_0 < \min\{2K_0,\varepsilon\}$. Let $l \in \{0,\ldots,k\}$ and define $\delta_0 = \delta_0(\varepsilon,\varepsilon_0,K_0,\gamma,l) > 0$ by

$$\delta_0 := \sup\left\{\theta > 0 \ : \ 2K_0\theta^{\gamma-l} + \varepsilon_0 e^\theta \leq \min\{2K_0,\varepsilon\}\right\}. \tag{C.11}$$

A first consequence of (C.11) is that $\delta_0 \leq 1$, and so for every $s \in \{0,\ldots,l\}$ we have that $\delta_0^{\gamma-s} \leq \delta_0^{\gamma-l}$. A second consequence of (C.11) is that

$$2K_0\delta_0^{\gamma-l} + \varepsilon_0 e^{\delta_0} \leq \min\{2K_0,\varepsilon\} \leq \varepsilon. \tag{C.12}$$

Now assume that $B \subset \Sigma$ is a δ_0-cover of Σ in the sense that

$$\Sigma \subset \bigcup_{x \in B} \overline{\mathbb{B}}_V(x,\delta_0) = B_{\delta_0} := \{v \in V \ : \ \exists z \in B \text{ such that } ||z-v||_V \leq \delta_0\}. \tag{C.13}$$

Suppose that $\psi = \left(\psi^{(0)},\ldots,\psi^{(k)}\right), \varphi = \left(\varphi^{(0)},\ldots,\varphi^{(k)}\right) \in \mathrm{Lip}(\gamma,\Sigma,W)$ with $||\psi||_{\mathrm{Lip}(\gamma,\Sigma,W)}, ||\varphi||_{\mathrm{Lip}(\gamma,\Sigma,W)} \leq K_0$. Further suppose that for every $j \in \{0,\ldots,k\}$ and every $x \in B$ the difference $\psi^{(j)}(x) - \varphi^{(j)}(x) \in \mathcal{L}(V^{\otimes j};W)$ satisfies the bound

$$\left|\left|\psi^{(j)}(x) - \varphi^{(j)}(x)\right|\right|_{\mathcal{L}(V^{\otimes j};W)} \leq \varepsilon_0. \tag{C.14}$$

Define $F \in \mathrm{Lip}(\gamma,\Sigma,W)$ by $F := \psi - \varphi$ so that for every $j \in \{0,\ldots,k\}$ we have $F^{(j)} := \psi^{(j)} - \varphi^{(j)}$. Then $||F||_{\mathrm{Lip}(\gamma,\Sigma,W)} \leq 2K_0$ and, for every $j \in \{0,\ldots,k\}$ and every $x \in B$, (C.14) gives that $\left|\left|F^{(j)}(x)\right|\right|_{\mathcal{L}(V^{\otimes j};W)} \leq \varepsilon_0$.

Now fix $x \in \Sigma$ and $s \in \{0,\ldots,l\}$. From (C.13) we conclude that there exists a point $z \in B$ with $||z-x||_V \leq \delta_0$. Then apply Lemma C.1, with $A := 2K_0$, $r_0 := \varepsilon_0$, $\rho := \gamma$, $p := z$, $n := k$ and $q := k$, to conclude that (cf. (C.3))

$$\left|\left|F^{(s)}(x)\right|\right|_{\mathcal{L}(V^{\otimes s};W)} \leq \min\left\{2K_0, 2K_0\delta_0^{\gamma-s} + \varepsilon_0\sum_{j=0}^{k-s}\frac{1}{j!}\delta_0^j\right\} \leq \min\left\{2K_0, 2K_0\delta_0^{\gamma-l} + \varepsilon_0 e^{\delta_0}\right\} \overset{(C.12)}{\leq} \varepsilon \tag{C.15}$$

where we have used that $\delta_0^{\gamma-s} \leq \delta_0^{\gamma-l}$. Since $F = \psi - \varphi$, the arbitrariness of $s \in \{0,\ldots,l\}$ and of $x \in \Sigma$ ensure that (C.15) gives the bounds claimed in (B.16) This completes the proof of Theorem B.8. ■

We now turn our attention to the *Single-Point Lipschitz Sandwich Theorem* B.5. Our strategy to prove this result will be to convert the pointwise estimates achieved in Lemma C.1 to full Lipschitz norm estimates. To be more precise, recall that $\Sigma \subset V$ is closed and $\gamma > 0$ with $k \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k,k+1]$. Let $\eta \in (0,\gamma)$, $K_0, \varepsilon > 0$, and $0 \leq \varepsilon_0 < \min\{2K_0,\varepsilon\}$. Retrieve the constant $\delta_0 = \delta_0(K_0,\varepsilon,\varepsilon_0,\gamma) > 0$ arising in Theorem B.8 for the choice $l := k$. Given a point $p \in \Sigma$, define $\Omega := \Sigma \cap \overline{\mathbb{B}}_V(p,\delta_0)$.

Suppose $\psi = \left(\psi^{(0)},\ldots,\psi^{(k)}\right) \in \mathrm{Lip}(\gamma,\Sigma,W), \varphi = \left(\varphi^{(0)},\ldots,\varphi^{(k)}\right) \in \mathrm{Lip}(\gamma,\Sigma,W)$ satisfy the norm bounds $||\psi||_{\mathrm{Lip}(\gamma,\Sigma,W)}, ||\varphi||_{\mathrm{Lip}(\gamma,\Sigma,W)} \leq K_0$. Further suppose that for every $j \in \{0,\ldots,k\}$ the difference $\psi^{(j)}(p) - \varphi^{(j)}(p) \in \mathcal{L}(V^{\otimes j};W)$ satisfies $\left|\left|\psi^{(j)}(p) - \varphi^{(j)}(p)\right|\right|_{\mathcal{L}(V^{\otimes j};W)} \leq \varepsilon_0$. Then by applying Theorem B.8, for the choices l, Σ and B there as k, Ω and $\{p\}$ here respectively, we may conclude that for every $s \in \{0,\ldots,k\}$ and every $x \in \Omega$ we have $\left|\left|\psi^{(s)}(x) - \varphi^{(s)}(x)\right|\right|_{\mathcal{L}(V^{\otimes s};W)} \leq \varepsilon$.

We will prove the *Single-Point Lipschitz Sandwich Theorem* B.5, by establishing that after reducing the constant δ_0, allowing it to additionally depend on η, we may strengthen these pointwise bounds into a bound on the full $\mathrm{Lip}(\eta,\Omega,W)$ norm of $\psi - \varphi$. There is a natural dichotomy within this strategy between the case that $\eta \in (k,\gamma)$ and the case that $\eta \in (0,k]$. We first deal with the simpler case that $\eta \in (k,\gamma)$.

In order to prove Theorem B.5, we first strengthen the pointwise estimates obtained in Lemma C.1 to full Lipschitz norm bounds on a local neighbourhood of the point p. The precise statement is the following lemma.

Lemma C.2 (Local Lipschitz Bounds I)**.** *Let V and W be Banach spaces, and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume that $\Gamma \subset V$ is non-empty and closed, and that $z \in \Gamma$. Let $A, \rho > 0$ with $n \in \mathbb{Z}_{\geq 0}$ such that $\rho \in (n,n+1]$, $r_0 \in [0,A]$, and $\theta \in (n,\rho)$. Suppose that $F = \left(F^{(0)},\ldots,F^{(n)}\right) \in \mathrm{Lip}(\rho,\Gamma,W)$ satisfies that $||F||_{\mathrm{Lip}(\rho,\Gamma,W)} \leq A$, and that for every $j \in \{0,\ldots,n\}$ we have the bound $\left|\left|F^{(j)}(z)\right|\right|_{\mathcal{L}(V^{\otimes j};W)} \leq r_0$. Then for any $\delta \in [0,1]$ we have that*

$$||F||_{\mathrm{Lip}(\theta,\Omega,W)} \leq \max\left\{(2\delta)^{\rho-\theta}A\ , \ \min\left\{A\ , \ A\delta^{\rho-n} + r_0 e^{\delta}\right\}\right\} \tag{C.16}$$

where $\Omega := \Gamma \cap \overline{\mathbb{B}}_V(z,\delta)$.

Proof of Lemma C.2. Let V and W be Banach spaces, and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume that $\Gamma \subset V$ is non-empty and closed, and that $z \in \Gamma$. Let $A, \rho > 0$ with $n \in \mathbb{Z}_{\geq 0}$ such that $\rho \in (n,n+1]$, $r_0 \in [0,A]$, and $\theta \in (n,\rho)$. Suppose that $F = \left(F^{(0)},\ldots,F^{(n)}\right) \in \mathrm{Lip}(\rho,\Gamma,W)$ satisfies that $||F||_{\mathrm{Lip}(\rho,\Gamma,W)} \leq A$, and that for every $j \in \{0,\ldots,n\}$ we have the bound $||F^{(j)}(z)||_{\mathcal{L}(V^{\otimes j};W)} \leq r_0$. For each $l \in \{0,\ldots,n\}$ let $R_l^F : \Gamma \times \Gamma \to \mathcal{L}(V^{\otimes l};W)$ be defined for $x,y \in \Gamma$ and $v \in V^{\otimes l}$ by

$$R_l^F(x,y)[v] := F^{(l)}(y)[v] - \sum_{j=0}^{n-l} \frac{1}{j!} F^{(j+l)}(x)\left[v \otimes (y-x)^{\otimes j}\right]. \tag{C.17}$$

An application of Lemma C.1, with A, r_0, ρ, n and θ here playing the same roles there, yields that for each $l \in \{0,\ldots,n\}$ and any $x \in \Sigma$ we have (cf. (C.3) for $q = n$)

$$\left|\left|F^{(l)}(x)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \min\left\{A\,,\ A\,\mathrm{dist}(x,p)^{\rho-l} + r_0 \sum_{s=0}^{n-l} \frac{1}{s!}\mathrm{dist}(x,p)^s\right\} \tag{C.18}$$

and (cf. (C.2))

$$||R_l^F(x,y)||_{\mathcal{L}(V^{\otimes l};W)} \leq A\left(\mathrm{dist}(x,p) + \mathrm{dist}(y,p)\right)^{\rho-\theta} ||y-x||_V^{\theta-l}. \tag{C.19}$$

Now let $\delta \in [0,1]$ and define $\Omega := \overline{\mathbb{B}}_V(z,\delta) \cap \Gamma \subset \Gamma$. Then given any $l \in \{0,\ldots,n\}$ and any $x,y \in \Omega$, (C.18) tells us that

$$\left|\left|F^{(l)}(x)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \min\left\{A\,,\ A\delta^{\rho-l} + r_0\sum_{s=0}^{n-l}\frac{\delta^s}{s!}\right\} \leq \min\left\{A\,,\ A\delta^{\rho-n} + r_0 e^{\delta}\right\}, \tag{C.20}$$

since $\delta \in [0,1]$ means $\delta^{\rho-l} \leq \delta^{\rho-n}$ for every $l \in \{0,\ldots,n\}$, whilst (C.19) tells us that

$$||R_l^F(x,y)||_{\mathcal{L}(V^{\otimes l};W)} \leq A\,(2\delta)^{\rho-\theta}\,||y-x||_V^{\theta-l}. \tag{C.21}$$

The estimates (C.20) and (C.21) allow us to conclude that $F \in \mathrm{Lip}(\eta,\Omega,W)$ with

$$||F||_{\mathrm{Lip}(\eta,\Omega,W)} \leq \max\left\{(2\delta)^{\rho-\theta}A\,,\ \min\left\{A\,,\ A\delta^{\rho-n} + r_0e^{\delta}\right\}\right\}$$

as claimed in (C.16). This completes the proof of Lemma C.2. ■

We can use Lemma C.2 to prove Theorem B.5 for the case that $\eta \in (k,\gamma)$.

Proof of Theorem B.5 when $\eta \in (k,\gamma)$. Assume that V and W are Banach spaces and that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Let $\Sigma \subset V$ be non-empty and closed. Let $\varepsilon, K_0, \gamma > 0$ with $k \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k,k+1]$, $0 \leq \varepsilon_0 < \min\{2K_0,\varepsilon\}$, and $\eta \in (k,\gamma)$. With a view to later applying Lemma C.2, define $\delta_0 = \delta_0(\varepsilon,\varepsilon_0,K_0,\gamma) > 0$ by

$$\delta_0 := \sup\left\{\theta \in (0,1]\ :\ 2K_0(2\theta)^{\gamma-\eta} \leq \min\{2K_0,\varepsilon\} \text{ and } 2K_0\theta^{\gamma-k} + \varepsilon_0 e^{\theta} \leq \min\{2K_0,\varepsilon\}\right\} > 0. \tag{C.22}$$

It initially appears that δ_0 additionally depends on k. However, k is determined by γ, thus any dependence on k is really dependence on γ. We now fix the value of $\delta_0 > 0$ for the remainder of the proof. We record that (C.22) ensures that $\delta_0 \leq 1$ and

$$\textbf{(I)}\quad 2K_0(2\delta_0)^{\gamma-\eta} \leq \min\{2K_0,\varepsilon\} \qquad \text{and} \qquad \textbf{(II)}\quad 2K_0\delta_0^{\gamma-k} + \varepsilon_0 e^{\delta_0} \leq \min\{2K_0,\varepsilon\}. \tag{C.23}$$

Now assume that $p \in \Sigma$ and that $\psi = \left(\psi^{(0)},\ldots,\psi^{(k)}\right)$ and $\varphi = \left(\varphi^{(0)},\ldots,\varphi^{(k)}\right)$ are elements in $\mathrm{Lip}(\gamma,\Sigma,W)$ with $||\psi||_{\mathrm{Lip}(\gamma,\Sigma,W)}, ||\varphi||_{\mathrm{Lip}(\gamma,\Sigma,W)} \leq K_0$. Further suppose that for every $l \in \{0,\ldots,k\}$ the difference $\psi^{(l)}(p) - \varphi^{(l)}(p) \in \mathcal{L}(V^{\otimes l};W)$ satisfies the bound

$$\left|\left|\psi^{(l)}(p) - \varphi^{(l)}(p)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \varepsilon_0. \tag{C.24}$$

Define $\Omega := \overline{\mathbb{B}}_V(p,\delta_0) \cap \Sigma$ and $F \in \mathrm{Lip}(\gamma,\Sigma,W)$ by $F := \psi - \varphi$, so that for every $j \in \{0,\ldots,k\}$ we have

$F^{(j)} = \psi^{(j)} - \varphi^{(j)}$. We apply Lemma C.2 to F, with $A := 2K_0$, $r_0 := \varepsilon_0$, $\rho := \gamma$, $n := k$, $z := p$, $\theta := \eta$ and $\delta := \delta_0$, to conclude both that $F \in \mathrm{Lip}(\eta, \Omega, W)$ and (cf. (C.16)) that

$$||F||_{\mathrm{Lip}(\eta,\Omega,W)} \leq \max\left\{2K_0(2\delta_0)^{\gamma-\eta} \ , \ \min\left\{2K_0 \ , \ 2K_0\delta_0^{\gamma-k} + \varepsilon_0 e^{\delta_0}\right\}\right\} \overset{\textbf{(II)} \text{ of (C.23)}}{\leq} \max\left\{2K_0(2\delta_0)^{\gamma-\eta} \ , \ \varepsilon\right\} \overset{\textbf{(I)} \text{ of (C.23)}}{\leq} \max\{\varepsilon \ , \ \varepsilon\} = \varepsilon.$$

Since $F = \psi - \varphi$ and $\Omega := \Sigma \cap \overline{\mathbb{B}}_V(p, \delta_0)$, this is precisely the estimate claimed in (B.6). This completes the proof of Theorem B.5 for $\eta \in (k, \gamma)$. ∎

Proving Theorem B.5 for the case $\eta \in (0, k]$ is more challenging. We first combine Lemma A.4 from Section A and Lemma C.2 to obtain the following local Lipschitz bounds.

Lemma C.3 (Local Lipschitz Bounds II). *Let V and W be Banach spaces, and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume that $\Gamma \subset V$ is a non-empty closed subset with $z \in \Gamma$. Let $A > 0$, $r_0 \in [0, A)$, $\rho > 1$ with $n \in \mathbb{Z}_{\geq 1}$ such that $\rho \in (n, n+1]$, and $\theta \in (0, n]$ with $q \in \{0, \ldots, n-1\}$ such that $\theta \in (q, q+1]$. Suppose that $F \in \mathrm{Lip}(\rho, \Gamma, W)$ satisfies that $||F||_{\mathrm{Lip}(\rho,\Gamma,W)} \leq A$, and that for every $j \in \{0, \ldots, n\}$ we have the bound $\left|\left|F^{(j)}(z)\right|\right|_{\mathcal{L}(V^{\otimes j};W)} \leq r_0$. Given any $\delta \in [0, 1]$ we have, for $\Omega := \overline{\mathbb{B}}_V(z, \delta) \cap \Gamma$, that*

$$\left|\left|F_{[q]}\right|\right|_{\mathrm{Lip}(\theta,\Omega,W)} \leq \max\left\{(2\delta)^{q+1-\theta} E_{n-b_q}, \min\left\{E_{n-b_q} \ , \ \delta E_{n-b_q} + r_0 e^{\delta}\right\}\right\} \tag{C.25}$$

where $F_{[q]} = \left(F^{(0)}, \ldots, F^{(q)}\right)$, $b_q := n - (q+1)$, and for $s \in \{0, \ldots, n-1\}$ E_{n-s} is inductively defined by

$$E_{n-s} := \begin{cases} \left(1 + (2\delta)^{\frac{\rho-n}{2}}\right) \max\left\{(2\delta)^{\frac{\rho-n}{2}} A \ , \ \min\left\{A \ , \ \delta^{\rho-n} A + r_0 e^{\delta}\right\}\right\} & \text{if } s = 0 \\ \left(1 + \sqrt{2\delta}\right) \max\left\{\sqrt{2\delta} E_{n-(s-1)} \ , \ \min\left\{E_{n-(s-1)} \ , \ \delta E_{n-(s-1)} + r_0 e^{\delta}\right\}\right\} & \text{if } s \geq 1. \end{cases} \tag{C.26}$$

Consequently, if $r_0 = 0$ we can conclude that

$$\left|\left|F_{[q]}\right|\right|_{\mathrm{Lip}(\theta,\Omega,W)} \leq \left(1 + (2\delta)^{\frac{\rho-n}{2}}\right)\left(1 + \sqrt{2\delta}\right)^{n-(q+1)} (2\delta)^{\frac{\rho-\theta}{2} + \frac{q+1-\theta}{2}} A. \tag{C.27}$$

If $0 < r_0 < A$ then there exists $\delta_* = \delta_*(A, r_0, \rho) > 0$ such that if we additionally impose that $\delta \in [0, \delta_*]$ then we may conclude that

$$\left|\left|F_{[q]}\right|\right|_{\mathrm{Lip}(\theta,\Omega,W)} \leq \max\left\{(2\delta)^{q+1-\theta} \mathcal{E} \ , \ \min\left\{\mathcal{E} \ , \ \delta\mathcal{E} + r_0 e^{\delta}\right\}\right\} \tag{C.28}$$

for $\mathcal{E} = \mathcal{E}(A, r_0, \rho, \theta, \delta) > 0$ defined by

$$\mathcal{E} := \left(1 + (2\delta)^{\frac{\rho-n}{2}}\right)\left(1 + \sqrt{2\delta}\right)^{n-(q+1)} \left(\delta^{\rho-(q+1)} A + r_0 \delta^{n-(q+1)} e^{\delta}\right) + \mathbf{X}_{n-(q+1)}(\delta) \tag{C.29}$$

where, for $t \in \{0, \ldots, n-1\}$, the quantity $\mathbf{X}_t(\delta)$ is defined by

$$\mathbf{X}_t(\delta) := \begin{cases} 0 & \text{if } t = 0 \\ \left(1 + \sqrt{2\delta}\right) r_0 e^{\delta} \sum_{j=0}^{t-1} \delta^j \left(1 + \sqrt{2\delta}\right)^j & \text{if } t \geq 1. \end{cases} \tag{C.30}$$

Proof of Lemma C.3. Let V and W be Banach spaces, and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume that $\Gamma \subset V$ is a non-empty closed subset with $z \in \Gamma$. Let $A > 0$, $r_0 \in [0, A)$, $\rho > 1$ with $n \in \mathbb{Z}_{\geq 1}$ such that $\rho \in (n, n+1]$, and $\theta \in (0, n]$ with $q \in \{0, \ldots, n-1\}$ such that $\theta \in (q, q+1]$. Suppose that $F \in \mathrm{Lip}(\rho, \Gamma, W)$ satisfies that $||F||_{\mathrm{Lip}(\rho,\Gamma,W)} \leq A$, and that for every $j \in \{0, \ldots, n\}$ we have the bound $\left|\left|F^{(j)}(z)\right|\right|_{\mathcal{L}(V^{\otimes j};W)} \leq r_0$. Fix $\delta \in [0, 1]$ and define $\Omega := \Gamma \cap \overline{\mathbb{B}}_V(z, \delta)$. For $s \in \{0, \ldots, n-1\}$

inductively define

$$E_{n-s} := \begin{cases} \left(1 + (2\delta)^{\frac{\rho-n}{2}}\right) \max\left\{(2\delta)^{\frac{\rho-n}{2}} A \ , \ \min\left\{A \ , \ A\delta^{\rho-n} + r_0 e^{\delta}\right\}\right\} & \text{if } s = 0 \\ \left(1 + \sqrt{2\delta}\right) \max\left\{\sqrt{2\delta} E_{n-(s-1)} \ , \ \min\left\{E_{n-(s-1)} \ , \ \delta E_{n-(s-1)} + r_0 e^{\delta}\right\}\right\} & \text{if } s \geq 1. \end{cases} \tag{C.31}$$

We first prove that each E_{n-s} is bounded from below by r_0. This is the content of the following claim.

Claim C.4. *For every* $s \in \{0, \ldots, n-1\}$ *we have*

$$E_{n-s} \geq r_0. \tag{C.32}$$

Proof of Claim C.4. The claim is proven via induction on $s \in \{0, \ldots, n-1\}$. For $s = 0$ we have

$$E_n \stackrel{(C.31)}{\geq} \max\left\{(2\delta)^{\frac{\rho-n}{2}} A \ , \ \min\left\{A \ , \ A\delta^{\rho-n} + r_0 e^{\delta}\right\}\right\} \geq \min\left\{A \ , \ A\delta^{\rho-n} + r_0 e^{\delta}\right\} \geq r_0 \tag{C.33}$$

where the last inequality uses that $r_0 \leq A$. If (C.32) is valid for $s \in \{0, \ldots, n-2\}$ then we compute that

$$\begin{aligned} E_{n-(s+1)} &\stackrel{(C.31)}{=} \left(1 + \sqrt{2\delta}\right) \max\left\{\sqrt{2\delta} E_{n-s} \ , \ \min\left\{E_{n-s} \ , \ \delta E_{n-s} + r_0 e^{\delta}\right\}\right\} \\ &\geq \max\left\{\sqrt{2\delta} E_{n-s} \ , \ \min\left\{E_{n-s} \ , \ \delta E_{n-s} + r_0 e^{\delta}\right\}\right\} \\ &\geq \min\left\{E_{n-s} \ , \ \delta E_{n-s} + r_0 e^{\delta}\right\} \geq r_0 \end{aligned}$$

where the last line uses that $E_{n-s} \geq r_0$ by the assumption that (C.32) is valid for s, and that $\delta E_{n-s} + r_0 e^{\delta} \geq r_0$ since $\delta \geq 0$. Thus we have established that the estimate (C.32) for $s \in \{0, \ldots, n-2\}$ yields that the estimate (C.32) is true for $s+1$. Since (C.33) establishes that (C.32) is true for $s = 0$, we may use induction to prove that (C.32) is in fact true for every $s \in \{0, \ldots, n-1\}$ as claimed. This completes the proof of Claim C.4. ■

We now prove that, for each $s \in \{0, \ldots, n-1\}$, the $\text{Lip}(n-s, \Omega, W)$-norm of $F_{[n-s-1]} = \left(F^{(0)}, \ldots, F^{(n-s-1)}\right)$ is bounded above by E_{n-s}. This is the content of the following claim.

Claim C.5. *For every* $s \in \{0, \ldots, n-1\}$ *we have that*

$$\left|\left|F_{[n-(s+1)]}\right|\right|_{\text{Lip}(n-s,\Omega,W)} \leq E_{n-s}. \tag{C.34}$$

Proof of Claim C.5. We will prove (C.34) via induction on $s \in \{0, \ldots, n-1\}$. We begin with the base case that $s = 0$. In this case, consider $\xi := \frac{\rho-n}{2} \in (0, \rho-n)$ so that $n + \xi \in (n, \rho)$ with $0 < \xi \leq 1/2$. An initial application of Lemma C.2, with Γ, A, r_0, ρ and n here playing the same role and with the θ in Lemma C.2 being $n + \xi$ here, yields that

$$||F||_{\text{Lip}(n+\xi,\Omega,W)} \leq \max\left\{(2\delta)^{\rho-n-\xi} A, \min\left\{A \ , \ A\delta^{\rho-n} + r_0 e^{\delta}\right\}\right\}. \tag{C.35}$$

We next apply Lemma A.4, with the Γ, ρ and θ of that result as Ω, $n+\xi$ and n here respectively, to obtain that

$$\left|\left|F_{[n-1]}\right|\right|_{\text{Lip}(n,\Omega,W)} \leq \min\{C_1, C_2\} ||F||_{\text{Lip}(n+\xi,\Omega,W)} \tag{C.36}$$

where (cf. (A.35) and recalling both that $\text{diam}(\Omega) \leq 2\delta \leq 2$ and that $0 < \xi \leq 1$)

$$C_1 = \max\left\{1 \ , \ \min\left\{1 + e \ , \ \text{diam}(\Omega)^{\xi} + \sum_{j=n}^{n} \frac{\text{diam}(\Omega)^{j-n}}{(j-(n-1))!}\right\}\right\} = 1 + \text{diam}(\Omega)^{\xi} \leq 1 + (2\delta)^{\xi}, \tag{C.37}$$

and (cf. (A.36))

$$C_2 = \max\left\{1, \min\left\{1 + e, \text{diam}(\Omega)^{n-n}\right\}\right\} \left(1 + \min\left\{e, \text{diam}(\Omega)\right\}^{\xi}\right) \left(1 + \min\left\{e, \text{diam}(\Omega)\right\}\right)^{n-n} \tag{C.38}$$

so that, since $\operatorname{diam}(\Omega) \leq 2\delta \leq 2$ and $0 < \xi \leq 1$, we have

$$C_2 = 1 + \operatorname{diam}(\Omega)^{\xi} \leq 1 + (2\delta)^{\xi} \tag{C.39}$$

The combination of (C.35), (C.36), (C.37), and (C.39) yields that

$$\begin{aligned} \left|\left|F_{[n-1]}\right|\right|_{\operatorname{Lip}(n,\Omega,W)} &\leq \left(1 + (2\delta)^{\xi}\right) \max\left\{(2\delta)^{\rho-n-\xi}A, \min\left\{A\ ,\ A\delta^{\rho-n} + r_0 e^{\delta}\right\}\right\} \\ &= \left(1 + (2\delta)^{\frac{\rho-n}{2}}\right) \max\left\{(2\delta)^{\frac{\rho-n}{2}}A, \min\left\{A\ ,\ A\delta^{\rho-n} + r_0 e^{\delta}\right\}\right\} \stackrel{(C.31)}{=} E_n. \end{aligned}$$

This completes the base case of our induction by verifying (C.34) when $s = 0$.

Now assume that $n-1 \geq 1$, that $s \in \{1, \ldots, n-1\}$, and that (C.34) is true for $s-1$. Consider $\xi := \frac{1}{2} \in (0,1)$ so that $n - s + \xi \in (n-s, n-(s-1))$. An initial application of Lemma C.2, with Γ, A, r_0, ρ and θ of that result as Ω, $E_{n-(s-1)}$, r_0, $n-(s-1)$ and $n-s+\xi$ here respectively, yields that

$$\left|\left|F_{[n-s]}\right|\right|_{\operatorname{Lip}(n-s+\xi,\Omega,W)} \leq \max\left\{(2\delta)^{1-\xi}E_{n-(s-1)}, \min\left\{E_{n-(s-1)}\ ,\ E_{n-(s-1)}\delta + r_0 e^{\delta}\right\}\right\}. \tag{C.40}$$

We next apply Lemma A.4, with the Γ, ρ and θ of that result as Ω, $n-s+\xi$ and $n-s$ here respectively, to obtain that

$$\left|\left|F_{[n-(s+1)]}\right|\right|_{\operatorname{Lip}(n-s,\Omega,W)} \leq \min\{D_1, D_2\} \left|\left|F_{[n-s]}\right|\right|_{\operatorname{Lip}(n-s+\xi,\Omega,W)} \tag{C.41}$$

where (cf. (A.35) and recalling that $\operatorname{diam}(\Omega) \leq 2\delta \leq 2$ and $\xi := \frac{1}{2} \leq 1$)

$$D_1 = \max\left\{1\ ,\ \min\left\{1 + e, \operatorname{diam}(\Omega)^{\xi} + \sum_{j=n-s}^{n-s} \frac{\operatorname{diam}(\Omega)^{j-(n-s)}}{(j-(n-s-1))!}\right\}\right\} = 1 + \operatorname{diam}(\Omega)^{\xi} \leq 1 + (2\delta)^{\xi}, \tag{C.42}$$

and (cf. (A.36))

$$D_2 = \max\left\{1\ ,\ \min\left\{1 + e, \operatorname{diam}(\Omega)^{0}\right\}\right\}\left(1 + \min\left\{e, \operatorname{diam}(\Omega)\right\}^{\xi}\right)\left(1 + \min\left\{e, \operatorname{diam}(\Omega)\right\}\right)^{0} \tag{C.43}$$

so that, since $\operatorname{diam}(\Omega) \leq 2\delta \leq 2$ and $\xi := \frac{1}{2} \leq 1$, we have

$$D_2 = 1 + \operatorname{diam}(\Omega)^{\xi} \leq 1 + (2\delta)^{\xi}. \tag{C.44}$$

The combination of (C.40), (C.41), (C.42), and (C.44) yields that

$$\begin{aligned} \left|\left|F_{[n-(s+1)]}\right|\right|_{\operatorname{Lip}(n-s,\Omega,W)} &\leq \left(1 + (2\delta)^{\xi}\right) \max\left\{(2\delta)^{1-\xi}E_{n-(s-1)}, \min\left\{E_{n-(s-1)}\ ,\ \delta E_{n-(s-1)} + r_0 e^{\delta}\right\}\right\} \\ &= \left(1 + \sqrt{2\delta}\right) \max\left\{\sqrt{2\delta}E_{n-(s-1)}, \min\left\{E_{n-(s-1)}\ ,\ \delta E_{n-(s-1)} + r_0 e^{\delta}\right\}\right\} \\ &\stackrel{(C.31)}{=} E_{n-s}. \end{aligned}$$

This completes the proof of the inductive step by establishing that if (C.34) is valid for $s-1$ with $s \in \{1, \ldots, n-1\}$, then (C.34) is in fact valid for s.

Using the base case and the inductive step allows us to conclude that (C.34) is valid for every $s \in \{0, \ldots, n-1\}$. This completes the proof of Claim C.5. ■

By appealing to Claim C.5, we conclude that, for every $s \in \{0, \ldots, n-1\}$ we have that

$$\left|\left|F_{[n-(s+1)]}\right|\right|_{\operatorname{Lip}(n-s,\Omega,W)} \leq E_{n-s}. \tag{C.45}$$

Let $b_q := n - (q+1) \in \{0, \ldots, n-1\}$ so that $q + 1 = n - b_q$. Then (C.45) for $s := b_q$ tells us that

$$\left|\left|F_{[q]}\right|\right|_{\operatorname{Lip}(q+1,\Omega,W)} \leq E_{n-b_q}. \tag{C.46}$$

A final application of Lemma C.2, with Γ, A, r_0, ρ and θ of that result as Ω, E_{n-b_q}, r_0, $q+1$ and θ here, yields that

$$||F_{[q]}||_{\mathrm{Lip}(\theta,\Omega,W)} \leq \max\left\{(2\delta)^{q+1-\theta}E_{n-b_q}, \min\left\{E_{n-b_q}\ ,\ \delta E_{n-b_q} + r_0 e^{\delta}\right\}\right\} \quad \text{(C.47)}$$

which is precisely the bound claimed in (C.25).

Now suppose that $r_0 = 0$. Then from (C.31), for $s \in \{0, \ldots, n-1\}$ we have that

$$E_{n-s} := \begin{cases} \left(1+(2\delta)^{\frac{\rho-n}{2}}\right)\max\left\{(2\delta)^{\frac{\rho-n}{2}}A\ ,\ \min\left\{A\ ,\ \delta^{\rho-n}A\right\}\right\} & \text{if } s=0 \\ \left(1+\sqrt{2\delta}\right)\max\left\{\sqrt{2\delta}E_{n-(s-1)}\ ,\ \min\left\{E_{n-(s-1)}\ ,\ \delta E_{n-(s-1)}\right\}\right\} & \text{if } s\geq 1. \end{cases} \quad \text{(C.48)}$$

Since $\delta \in [0,1]$ we have both that $\delta \leq \sqrt{\delta} \leq 1$ and $\delta^{\rho-n} \leq \delta^{\frac{\rho-n}{2}} \leq 1$ Consequently, (C.48) yields that

$$E_{n-s} := \begin{cases} \left(1+(2\delta)^{\frac{\rho-n}{2}}\right)(2\delta)^{\frac{\rho-n}{2}}A & \text{if } s=0 \\ \left(1+\sqrt{2\delta}\right)\sqrt{2\delta}E_{n-(s-1)} & \text{if } s\geq 1. \end{cases} \quad \text{(C.49)}$$

Proceeding inductively via (C.49), we establish that for any $s \in \{0, \ldots, n-1\}$ we have that

$$E_{n-s} = \left(1+(2\delta)^{\frac{\rho-n}{2}}\right)\left(1+\sqrt{2\delta}\right)^{s}(2\delta)^{\frac{\rho-n+s}{2}}A. \quad \text{(C.50)}$$

Observing that $\delta \in [0,1]$ means that $\delta \leq \delta^{q+1-\theta} \leq 1$, we may combine (C.47) and (C.50) for the choice $s := b_q = n-(q+1)$ to obtain that

$$||F_{[q]}||_{\mathrm{Lip}(\theta,\Omega,W)} \leq (2\delta)^{q+1-\theta}E_{n-b_q} \overset{\text{(C.50)}}{=} \left(1+(2\delta)^{\frac{\rho-n}{2}}\right)\left(1+\sqrt{2\delta}\right)^{b_q}(2\delta)^{\frac{\rho-n+b_q}{2}+q+1-\theta}A. \quad \text{(C.51)}$$

Since $\rho - n + b_q = \rho - (q+1)$ we see that (C.51) is precisely the estimate claimed in (C.27).

Now assume that $r_0 \in (0, A)$. We first let $\delta_* := 1$ To establish (C.28) we must reduce δ_* to a smaller constant. With the benefit of hindsight, it will suffice to reduce δ_*, depending only on A, r_0, and ρ, to ensure that whenever $\delta \in [0, \delta_*]$ we have the estimates

$$\begin{cases} \textbf{(I)} \quad \max\left\{1+\sqrt{2\delta}\ ,\ 1+(2\delta)^{\frac{\rho-n}{2}}\right\} < 2 \quad (\text{in particular } 2\delta < 1), \\ \textbf{(II)} \quad r_0 e^{\delta} \leq A\left(1-\delta^{\rho-n}\right), \\ \textbf{(III)} \quad \left(2^{\frac{\rho-n}{2}} - \delta^{\frac{\rho-n}{2}}\right)\delta^{\frac{\rho-n}{2}}A \leq r_0 e^{\delta}, \\ \textbf{(IV)} \quad 2\sqrt{2\delta}\left(\delta^{\rho-n}A + r_0 e^{\delta}\right) \leq r_0 e^{\delta}, \quad \text{and} \\ \textbf{(V)} \quad \sqrt{2\delta}\left[2^n\left(\delta^{\rho-n}A + r_0\delta e^{\delta}\right) + 2r_0 e^{\delta}\left(\dfrac{1-(2\delta)^n}{1-2\delta}\right)\right] \leq r_0 e^{\delta}. \end{cases} \quad \text{(C.52)}$$

We now consider a fixed choice of $\delta \in [0, \delta_*]$ and establish the estimate claimed in (C.28) for $\Omega := \overline{\mathbb{B}}_V(p,\delta) \cap \Gamma$. We begin by estimating the terms E_{n-s} for $s \in \{0, \ldots, n-1\}$. We first prove, for every $t \in \{0, \ldots, n-1\}$, that

$$E_{n-t} = \left(1+(2\delta)^{\frac{\rho-n}{2}}\right)\left(1+\sqrt{2\delta}\right)^{t}\left(\delta^{\rho-n+t}A + r_0\delta^t e^{\delta}\right) + \mathbf{X}_t(\delta) \quad \text{(C.53)}$$

where $\mathbf{X}_t(\delta)$ is the quantity defined in (C.30). That is,

$$\mathbf{X}_t(\delta) := \begin{cases} 0 & \text{if } t=0 \\ \left(1+\sqrt{2\delta}\right)r_0 e^{\delta}\sum_{j=0}^{t-1}\delta^j\left(1+\sqrt{2\delta}\right)^j & \text{if } t\geq 1. \end{cases} \quad \text{(C.54)}$$

We begin by considering $t := 0$. From (C.31) we have that

$$E_n = \left(1+(2\delta)^{\frac{\rho-n}{2}}\right)\max\left\{(2\delta)^{\frac{\rho-n}{2}}A\ ,\ \min\left\{A\ ,\ \delta^{\rho-n}A + r_0 e^{\delta}\right\}\right\}. \quad \text{(C.55)}$$

A consequence of **(II)** in (C.52) is that $\delta^{\rho-n}A + r_0e^{\delta} \leq A$ so that from (C.55) we see that

$$E_n = \left(1 + (2\delta)^{\frac{\rho-n}{2}}\right) \max\left\{(2\delta)^{\frac{\rho-n}{2}}A\ ,\ \delta^{\rho-n}A + r_0e^{\delta}\right\}. \tag{C.56}$$

A consequence of **(III)** in (C.52) is that $(2\delta)^{\frac{\rho-n}{2}}A \leq \delta^{\rho-n}A + r_0e^{\delta}$ so that from (C.56) we see that

$$E_n = \left(1 + (2\delta)^{\frac{\rho-n}{2}}\right)\left(\delta^{\rho-n}A + r_0e^{\delta}\right) \tag{C.57}$$

which is the estimate claimed in (C.53) for $t = 0$ since $\mathbf{X}_0(\delta) := 0$.

Now consider $t \geq 1$ and assume that (C.53) is true for $t-1$. From (C.31) we have that

$$E_{n-t} = \left(1 + \sqrt{2\delta}\right) \max\left\{\sqrt{2\delta}E_{n-(t-1)}\ ,\ \min\left\{E_{n-(t-1)}\ ,\ \delta E_{n-(t-1)} + r_0e^{\delta}\right\}\right\}. \tag{C.58}$$

Since (C.53) is valid for $t-1$ we have that

$$E_{n-(t-1)} = \left(1 + (2\delta)^{\frac{\rho-n}{2}}\right)\left(1 + \sqrt{2\delta}\right)^{t-1}\left(\delta^{\rho-n+t-1}A + r_0\delta^{t-1}e^{\delta}\right) + \mathbf{X}_{t-1}(\delta) \tag{C.59}$$

We claim that $\delta E_{n-(t-1)} + r_0e^{\delta} \leq E_{n-(t-1)}$.

If $t = 1$, then (C.57) ensures that $(1-\delta)E_n \geq (1-\delta)\left(1 + (2\delta)^{\frac{\rho-n}{2}}\right)r_0e^{\delta}$. If we are able to conclude that $(1-\delta)\left(1 + (2\delta)^{\frac{\rho-n}{2}}\right) \geq 1$ then our desired estimate $\delta E_n + r_0e^{\delta} \leq E_n$ is true. The required lower bound $(1-\delta)\left(1 + (2\delta)^{\frac{\rho-n}{2}}\right) \geq 1$ is equivalent to $(2\delta)^{\frac{\rho-n}{2}} - \delta - \delta(2\delta)^{\frac{\rho-n}{2}} \geq 0$. A consequence of **(I)** in (C.52) is that $(2\delta)^{\frac{\rho-n}{2}} < 1$. This tells us that $\delta + \delta(2\delta)^{\frac{\rho-n}{2}} \leq 2\delta \leq (2\delta)^{\frac{\rho-n}{2}}$ where the latter inequality is true since $2\delta < 1$ and $\frac{\rho-n}{2} < 1$. Hence $(1-\delta)\left(1 + (2\delta)^{\frac{\rho-n}{2}}\right) \geq 1$ and so we have that $\delta E_n + r_0e^{\delta} \leq E_n$ as required.

If $t > 1$ then (C.59) ensures that $(1-\delta)E_{n-(t-1)} \geq (1-\delta)\mathbf{X}_{t-1}(\delta) \geq (1-\delta)\left(1+\sqrt{2\delta}\right)r_0e^{\delta}$. If we are able to conclude that $(1-\delta)\left(1+\sqrt{2\delta}\right) \geq 1$ then our desired estimate $\delta E_{n-(t-1)} + r_0e^{\delta} \leq E_{n-(t-1)}$ is true. The required lower bound $(1-\delta)\left(1+\sqrt{2\delta}\right) \geq 1$ is equivalent to $\sqrt{2\delta} - \delta - \delta\sqrt{2\delta} \geq 0$. A consequence of **(I)** in (C.52) is that $\sqrt{2\delta} < 1$. This tells us that $\delta + \delta\sqrt{2\delta} \leq 2\delta \leq \sqrt{2\delta}$ where the latter inequality is true since $2\delta < 1$. Hence $(1-\delta)\left(1+\sqrt{2\delta}\right) \geq 1$ and so we have that $\delta E_{n-(t-1)} + r_0e^{\delta} \leq E_{n-(t-1)}$ as required.

Having established that $\delta E_{n-(t-1)} + r_0e^{\delta} \leq E_{n-(t-1)}$, (C.58) tells us that

$$E_{n-t} = \left(1 + \sqrt{2\delta}\right) \max\left\{\sqrt{2\delta}E_{n-(t-1)}\ ,\ \delta E_{n-(t-1)} + r_0e^{\delta}\right\}. \tag{C.60}$$

We claim that $\sqrt{2\delta}E_{n-(t-1)} \leq \delta E_{n-(t-1)} + r_0e^{\delta}$. If $t = 1$ then we compute, using (C.59) for $t = 1$, that

$$\begin{aligned}\left(\sqrt{2\delta} - \delta\right)E_n &= \sqrt{\delta}\left(\sqrt{2} - \sqrt{\delta}\right)\left(1 + (2\delta)^{\frac{\rho-n}{2}}\right)\left(\delta^{\rho-n}A + r_0e^{\delta}\right) \\ &\overset{\textbf{(I)} \text{ in (C.52)}}{\leq} 2\sqrt{2\delta}\left(\delta^{\rho-n}A + r_0e^{\delta}\right) \overset{\textbf{(IV)} \text{ in (C.52)}}{\leq} r_0e^{\delta}.\end{aligned}$$

Consequently we have $\sqrt{2\delta}E_n \leq \delta E_n + r_0 e^\delta$ as claimed. If $t > 1$ then we compute, using (C.59) for $t > 1$, that

$$\begin{aligned}
\left(\sqrt{2\delta} - \delta\right) E_{n-(t-1)} &= \left(\sqrt{2\delta} - \delta\right)\left(\left(1 + (2\delta)^{\frac{\rho-n}{2}}\right)\left(1 + \sqrt{2\delta}\right)^{t-1}\left(\delta^{\rho-n+t-1}A + r_0\delta^{t-1}e^\delta\right) + \mathbf{X}_{t-1}(\delta)\right) \\
&\overset{(\mathbf{I}) \text{ in (C.52)}}{\leq} 2^t\sqrt{2\delta}\left(\delta^{\rho-n+t-1}A + r_0\delta^{t-1}e^\delta\right) + \sqrt{2\delta}\mathbf{X}_{t-1}(\delta) \\
&= 2^t\sqrt{2\delta}\left(\delta^{\rho-n+t-1}A + r_0\delta^{t-1}e^\delta\right) + \sqrt{2\delta}\left(1+\sqrt{2\delta}\right) r_0 e^\delta \sum_{j=0}^{t-1}\delta^j\left(1+\sqrt{2\delta}\right)^j \\
&\overset{(\mathbf{I}) \text{ in (C.52)}}{\leq} 2^t\sqrt{2\delta}\left(\delta^{\rho-n}A + r_0\delta e^\delta\right) + 2\sqrt{2\delta}r_0e^\delta\sum_{j=0}^{t-1}(2\delta)^j \\
&\overset{(\mathbf{I}) \text{ in (C.52)}}{=} \sqrt{2\delta}\left[2^t\left(\delta^{\rho-n}A + r_0\delta e^\delta\right) + 2r_0e^\delta\left(\frac{1-(2\delta)^t}{1-2\delta}\right)\right] \\
&\leq \sqrt{2\delta}\left[2^n\left(\delta^{\rho-n}A + r_0\delta e^\delta\right) + 2r_0e^\delta\left(\frac{1-(2\delta)^n}{1-2\delta}\right)\right] \overset{(\mathbf{IV}) \text{ in (C.52)}}{\leq} r_0e^\delta.
\end{aligned}$$

Consequently we have that $\sqrt{2\delta}E_{n-(t-1)} \leq \delta E_{n-(t-1)} + r_0e^\delta$ as claimed.

Returning our attention to (C.60), the inequality $\sqrt{2\delta}E_{n-(t-1)} \leq \delta E_{n-(t-1)} + r_0e^\delta$ means that

$$\begin{aligned}
E_{n-t} &= \left(1+\sqrt{2\delta}\right)\left(\delta E_{n-(t-1)} + r_0e^\delta\right) \\
&\overset{(C.59)}{=} \left(1+\sqrt{2\delta}\right)\left(\left(1 + (2\delta)^{\frac{\rho-n}{2}}\right)\left(1+\sqrt{2\delta}\right)^{t-1}\left(\delta^{\rho-n+t}A + r_0\delta^te^\delta\right) + \delta\mathbf{X}_{t-1}(\delta) + r_0e^\delta\right) \\
&= \left(1 + (2\delta)^{\frac{\rho-n}{2}}\right)\left(1+\sqrt{2\delta}\right)^{t}\left(\delta^{\rho-n+t}A + r_0\delta^te^\delta\right) + \left(1+\sqrt{2\delta}\right)\left(r_0e^\delta + \delta\mathbf{X}_{t-1}(\delta)\right).
\end{aligned}$$

We observe that

$$\delta\left(1+\sqrt{2\delta}\right)\mathbf{X}_{t-1}(\delta) \overset{(C.54)}{=} \begin{cases} 0 & \text{if } t-1=0 \\ \left(1+\sqrt{2\delta}\right)r_0e^\delta\sum_{j=1}^t\delta^j\left(1+\sqrt{2\delta}\right)^j & \text{if } t-1\geq 1. \end{cases} \tag{C.61}$$

Via (C.61) we see that

$$\begin{aligned}
\left(1+\sqrt{2\delta}\right)\left(r_0e^\delta + \delta\mathbf{X}_{t-1}(\delta)\right) &= \begin{cases} \left(1+\sqrt{2\delta}\right)r_0e^\delta & \text{if } t-1=0 \\ \left(1+\sqrt{2\delta}\right)r_0e^\delta\sum_{j=0}^t\delta^j\left(1+\sqrt{2\delta}\right)^j & \text{if } t-1\geq 1 \end{cases} \\
&\overset{(C.54)}{=} \begin{cases} \mathbf{X}_1(\delta) & \text{if } t=1 \\ \mathbf{X}_t(\delta) & \text{if } t\geq 2 \end{cases} \\
&= \mathbf{X}_t(\delta).
\end{aligned}$$

Therefore we have established that

$$E_{n-t} = \left(1 + (2\delta)^{\frac{\rho-n}{2}}\right)\left(1+\sqrt{2\delta}\right)^t\left(\delta^{\rho-n+t}A + r_0\delta^te^\delta\right) + \mathbf{X}_t(\delta) \tag{C.62}$$

which is the estimate claimed in (C.53) for t. Induction now allows us to conclude that the estimate (C.53) is valid for every $t \in \{0,\ldots,n-1\}$ as claimed.

To conclude, recall that $\theta \in (q, q+1]$ and $b_q := n-(q+1) \in \{0,\ldots,n-1\}$. Then (C.47) yields that

$$\left|\left|F_{[q]}\right|\right|_{\mathrm{Lip}(\theta,\Omega,W)} \leq \max\left\{(2\delta)^{q+1-\theta}\mathcal{E}\ ,\ \min\left\{\mathcal{E}\ ,\ \delta\mathcal{E} + r_0e^\delta\right\}\right\} \tag{C.63}$$

where, via (C.53) for $t := b_q$, $\mathcal{E} = \mathcal{E}(A, r_0, \rho, \theta, \delta) := E_{n-b_q}$, i.e.

$$\mathcal{E} \overset{(C.53)}{=} \left(1 + (2\delta)^{\frac{\rho-n}{2}}\right)\left(1+\sqrt{2\delta}\right)^{n-(q+1)}\left(\delta^{\rho-(q+1)}A + r_0\delta^{n-(q+1)}e^{\delta}\right) + \mathbf{X}_{n-(q+1)}(\delta) \tag{C.64}$$

as claimed in (C.28) and (C.29). This completes the proof of Lemma C.3. ■

We can use Lemma C.3 to complete our proof of Theorem B.5 by dealing with the case that $\eta \in (0,k]$.

Proof of Theorem B.5 for $0 < \eta \leq k$. Assume that V and W are Banach spaces and that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Let $\Sigma \subset V$ be closed and non-empty. Let $\varepsilon, K_0 > 0, \gamma > 0$ with $k \in \mathbb{Z}_{\geq 1}$ such that $\gamma \in (k, k+1]$, and $\eta \in (0,k]$. Observe that this requires $1 \leq k < \gamma$. Let $q \in \{0, \ldots, k-1\}$ such that $\eta \in (q, q+1] \subset (0,k]$. Finally let $0 \leq \varepsilon_0 < \min\{2K_0, \varepsilon\}$.

Our strategy is to establish the desired $\mathrm{Lip}(\eta)$-norm bounds via an application of Lemma C.3. For this purpose we retrieve the constant δ_* arising in Lemma C.3 for $A := 2K_0$, $r_0 := \varepsilon_0$, $\rho := \gamma$ and $\theta := \eta$. Note that we are not actually applying Lemma C.3, but simply retrieving a constant in preparation for its future application.

Let $\delta_0 := \min\{1\ ,\ \delta_*\} > 0$, which depends only on K_0, ε_0, γ and η. In order to ensure that applying Lemma C.3 yields the desired $\mathrm{Lip}(\eta)$-norm estimate, we will allow ourselevs to (potentially) further reduce δ_0, additionally now depending on ε. With the benefit of hindsight, it will suffice alter δ_0 to ensure that

$$\begin{cases} \textbf{(A)} & \max\left\{1 + (2\delta_0)^{\frac{\gamma-k}{2}}\ ,\ 1+\sqrt{2\delta_0}\right\} < 2, \quad \text{(In particular, } 2\delta_0 < 1\text{)}, \\ \textbf{(B)} & (2\delta_0)^{\frac{\gamma-\eta}{2}+\frac{q+1-\eta}{2}} \leq \dfrac{\varepsilon}{2^{k-q+1}K_0}, \\ \textbf{(C)} & \left(1+(2\delta_0)^{\frac{\gamma-k}{2}}\right)\left(\delta_0^{\gamma-k}(2K_0) + \varepsilon_0 e^{\delta_0}\right) \leq \varepsilon, \\ \textbf{(D)} & 2^{k-q}\left(\delta_0^{\gamma-k}(2K_0) + \varepsilon_0\delta_0 e^{\delta_0}\right) + \left(\dfrac{1+\sqrt{2\delta_0}}{1-2\delta_0}\right)\varepsilon_0 e^{\delta_0} \leq \varepsilon, \quad \text{and} \\ \textbf{(E)} & \varepsilon_0 e^{\delta_0} \leq (1-\delta_0)\,\varepsilon. \end{cases} \tag{C.65}$$

We now fix the value of $\delta_0 = \delta_0(K_0, \gamma, \eta, \varepsilon_0, \varepsilon) > 0$ for the remainder of the proof.

Now assume $p \in \Sigma$ and that $\psi = \left(\psi^{(0)}, \ldots, \psi^{(k)}\right)$ and $\varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(k)}\right)$ are elements in $\mathrm{Lip}(\gamma, \Sigma, W)$ with $||\psi||_{\mathrm{Lip}(\gamma,\Sigma,W)}, ||\varphi||_{\mathrm{Lip}(\gamma,\Sigma,W)} \leq K_0$. Further suppose that for every $l \in \{0, \ldots, k\}$ the difference $\psi^{(l)}(p) - \varphi^{(l)}(p) \in \mathcal{L}(V^{\otimes l}; W)$ satisfies the bound

$$\left|\left|\psi^{(l)}(p) - \varphi^{(l)}(p)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \varepsilon_0. \tag{C.66}$$

Define $\Omega := \overline{\mathbb{B}}_V(p, \delta_0) \cap \Sigma$ and $F \in \mathrm{Lip}(\gamma, \Sigma, W)$ by $F := \psi - \varphi$ so that for every $j \in \{0, \ldots, k\}$ we have $F^{(j)} = \psi^{(j)} - \varphi^{(j)}$.

We begin with the case that $\varepsilon_0 = 0$. Since $\delta_0 \leq 1$, the bounds (C.66) allow us to apply Lemma C.3 to F, with $A := 2K_0$, $r_0 := \varepsilon_0$, $\rho := \gamma$, $\theta := \eta$ and $\delta := \delta_0$, to conclude that (cf. (C.27))

$$\left|\left|F_{[q]}\right|\right|_{\mathrm{Lip}(\eta,\Omega,W)} \leq \left(1+(2\delta_0)^{\frac{\gamma-k}{2}}\right)\left(1+\sqrt{2\delta_0}\right)^{k-(q+1)}(2\delta_0)^{\frac{\gamma-\eta}{2}+\frac{q+1-\eta}{2}}(2K_0). \tag{C.67}$$

We compute that

$$\begin{aligned} \left|\left|F_{[q]}\right|\right|_{\mathrm{Lip}(\eta,\Omega,W)} &\overset{(C.67)}{\leq} \left(1+(2\delta_0)^{\frac{\gamma-k}{2}}\right)\left(1+\sqrt{2\delta_0}\right)^{k-(q+1)}(2\delta_0)^{\frac{\gamma-\eta}{2}+\frac{q+1-\eta}{2}}(2K_0) \\ &\overset{\textbf{(A)} \text{ in (C.65)}}{\leq} 2^{k-q}(2\delta_0)^{\frac{\gamma-\eta}{2}+\frac{q+1-\eta}{2}}(2K_0) \overset{\textbf{(B)} \text{ in (C.65)}}{\leq} 2^{k-q}\frac{\varepsilon}{2^{k-q+1}K_0}(2K_0) = \varepsilon. \end{aligned}$$

Recalling that $\Omega := \Sigma \cap \overline{\mathbb{B}}_V(p, \delta_0)$ and $F := \psi - \varphi$, this is precisely the estimate claimed in (B.6), and our proof is complete for the case that $\varepsilon_0 = 0$.

Now consider the case that $\varepsilon_0 > 0$. Recalling how we chose δ_0, the bounds (C.66) allow us to apply Lemma C.3 to F, with $A := 2K_0$, $r_0 := \varepsilon_0$, $\rho := \gamma$, $\theta := \eta$ and $\delta := \delta_0$, to conclude via (C.28) that

$$\left|\left|F_{[q]}\right|\right|_{\mathrm{Lip}(\eta,\Omega,W)} \leq \max\left\{(2\delta_0)^{q+1-\eta}\mathcal{E}\ ,\ \min\left\{\mathcal{E}\ ,\ \delta_0\mathcal{E} + \varepsilon_0 e^{\delta_0}\right\}\right\} \tag{C.68}$$

for $\mathcal{E} = \mathcal{E}(K_0, \gamma, \eta, \varepsilon_0) > 0$ defined by (cf. (C.29))

$$\mathcal{E} := \left(1 + (2\delta_0)^{\frac{\gamma-k}{2}}\right)\left(1+\sqrt{2\delta_0}\right)^{k-(q+1)}\left(\delta_0^{\gamma-(q+1)}(2K_0) + \varepsilon_0\delta_0^{k-(q+1)}e^{\delta_0}\right) + \mathbf{X}_{k-(q+1)}(\delta_0) \tag{C.69}$$

where, for $t \in \{0, \ldots, k-1\}$, the quantity $\mathbf{X}_t(\delta)$ is defined by (cf. (C.30))

$$\mathbf{X}_t(\delta) := \begin{cases} 0 & \text{if } t = 0 \\ \left(1+\sqrt{2\delta_0}\right)\varepsilon_0 e^{\delta_0}\sum_{j=0}^{t-1}\delta_0^j\left(1+\sqrt{2\delta_0}\right)^j & \text{if } t \geq 1. \end{cases} \tag{C.70}$$

We first prove that $(1-\delta_0)\,\mathcal{E} \geq \varepsilon_0 e^{\delta_0}$.

If $k = q+1$, then (C.69) ensures that $(1-\delta_0)\,\mathcal{E} \geq (1-\delta_0)\left(1+(2\delta_0)^{\frac{\gamma-k}{2}}\right)\varepsilon_0 e^{\delta_0}$. If we are able to conclude that $(1-\delta_0)\left(1+(2\delta_0)^{\frac{\gamma-k}{2}}\right) \geq 1$ then our desired estimate $(1-\delta_0)\,\mathcal{E} \geq \varepsilon_0 e^{\delta_0}$ is true. The required lower bound $(1-\delta_0)\left(1+(2\delta_0)^{\frac{\gamma-k}{2}}\right) \geq 1$ is equivalent to $(2\delta_0)^{\frac{\gamma-k}{2}} - \delta_0 - \delta_0(2\delta_0)^{\frac{\gamma-k}{2}} \geq 0$. A consequence of (**A**) in (C.65) is that $(2\delta_0)^{\frac{\gamma-k}{2}} < 1$. This tells us that $\delta_0 + \delta_0(2\delta_0)^{\frac{\gamma-k}{2}} \leq 2\delta_0 \leq (2\delta_0)^{\frac{\gamma-k}{2}}$ where the latter inequality is true since $2\delta_0 < 1$ and $\frac{\gamma-k}{2} < 1$. Hence $(1-\delta_0)\left(1+(2\delta_0)^{\frac{\gamma-k}{2}}\right) \geq 1$ and so we have $(1-\delta_0)\,\mathcal{E} \geq \varepsilon_0 e^{\delta_0}$ as required.

If $k > q+1$, then (C.69) and (C.70) yield $(1-\delta_0)\,\mathcal{E} \geq (1-\delta_0)\,\mathbf{X}_{k-(q+1)}(\delta_0) \geq (1-\delta_0)\left(1+\sqrt{2\delta_0}\right)\varepsilon_0 e^{\delta_0}$. If we are able to conclude that $(1-\delta_0)\left(1+\sqrt{2\delta_0}\right) \geq 1$ then our desired estimate $(1-\delta_0)\,\mathcal{E} \geq \varepsilon_0 e^{\delta_0}$ is true. The required lower bound $(1-\delta_0)\left(1+\sqrt{2\delta_0}\right) \geq 1$ is equivalent to $\sqrt{2\delta_0} - \delta_0 - \delta_0\sqrt{2\delta_0} \geq 0$. A consequence of (**A**) in (C.65) is that $\sqrt{2\delta_0} < 1$. This tells us that $\delta_0 + \delta_0\sqrt{2\delta_0} \leq 2\delta_0 \leq \sqrt{2\delta_0}$ where the latter inequality is true since $2\delta_0 < 1$. Hence $(1-\delta_0)\left(1+\sqrt{2\delta_0}\right) \geq 1$ and so we have $(1-\delta_0)\,\mathcal{E} \geq \varepsilon_0 e^{\delta_0}$ as required.

Having established that $(1-\delta_0)\,\mathcal{E} \geq \varepsilon_0 e^{\delta_0}$ we observe that (C.68) becomes

$$\left|\left|F_{[q]}\right|\right|_{\mathrm{Lip}(\eta,\Omega,W)} \leq \max\left\{(2\delta_0)^{q+1-\eta}\mathcal{E}\ ,\ \delta_0\mathcal{E} + \varepsilon_0 e^{\delta_0}\right\} \tag{C.71}$$

We now prove the upper bound for $\mathcal{E}$ that $\mathcal{E} \leq \varepsilon$. For this purpose note that when $k = q+1$ (C.69) yields that

$$\mathcal{E} = \left(1+(2\delta_0)^{\frac{\gamma-k}{2}}\right)\left(\delta_0^{\gamma-k}(2K_0) + \varepsilon_0 e^{\delta_0}\right) \overset{(\mathbf{C})\text{ in (C.65)}}{\leq} \varepsilon \tag{C.72}$$

since $\mathbf{X}_0(\delta_0) = 0$ from (C.70). If, however, $k > q+1$ then $k-(q+1) \geq 1$ and so, recalling that (**A**) in (C.65) ensures that $2\delta_0 < 1$, we have

$$\mathbf{X}_{k-(q+1)}(\delta_0) \overset{\text{(C.70)}}{=} \left(1+\sqrt{2\delta_0}\right)\varepsilon_0 e^{\delta_0}\sum_{j=0}^{k-(q+1)}\delta_0^j\left(1+\sqrt{2\delta_0}\right)^j \leq \frac{\left(1+\sqrt{2\delta_0}\right)}{1-2\delta_0}\varepsilon_0 e^{\delta_0}. \tag{C.73}$$

Moreover, $2\delta_0 < 1$ ensures that $\delta_0^{k-(q+1)} \leq \delta_0$. Hence we can combine (C.69) and (C.73) to obtain that

$$\mathcal{E} \leq 2^{k-q}\left(\delta_0^{\gamma-k}(2K_0) + \varepsilon_0\delta_0 e^{\delta_0}\right) + \frac{\left(1+\sqrt{2\delta_0}\right)}{1-2\delta_0}\varepsilon_0 e^{\delta_0} \overset{(\mathbf{D})\text{ in (C.65)}}{\leq} \varepsilon.$$

Therefore in both the case that $k = q+1$ and the case that $k > q+1$ we obtain that

$$\mathcal{E} \leq \varepsilon. \tag{C.74}$$

We complete the proof by using the upper bound in (C.74) to control $\left|\left|F_{[q]}\right|\right|_{\mathrm{Lip}(\eta,\Omega,W)}$. Recalling that (**A**) in (C.65) means that $2\delta_0 < 1$, we have that

$$\textbf{(I)} \quad (2\delta_0)^{q+1-\eta}\,\mathcal{E} \leq \mathcal{E} \overset{\text{(C.74)}}{\leq} \varepsilon \quad \text{and} \quad \textbf{(II)} \quad \delta_0\mathcal{E} + \varepsilon_0 e^{\delta_0} \overset{\text{(C.74)}}{\leq} \delta_0\varepsilon + \varepsilon_0 e^{\delta_0} \overset{\textbf{(E)} \text{ in (C.65)}}{\leq} \varepsilon. \tag{C.75}$$

Thus

$$||F_{[q]}||_{\mathrm{Lip}(\eta,\Omega,W)} \overset{\text{(C.71)}}{\leq} \max\left\{(2\delta_0)^{q+1-\eta}\mathcal{E}, \delta_0\mathcal{E} + \varepsilon_0 e^{\delta_0}\right\} \overset{\textbf{(I)} \text{ in (C.75)}}{\leq} \max\left\{\varepsilon, \delta_0\mathcal{E} + \varepsilon_0 e^{\delta_0}\right\} \overset{\textbf{(II)} \text{ in (C.75)}}{=} \varepsilon. \tag{C.76}$$

Recalling that $\Omega := \Sigma \cap \overline{\mathbb{B}}_V(p,\delta_0)$ and $F := \psi - \varphi$, (C.76) is precisely the estimate claimed in (B.6), and our proof is complete for the case that $\varepsilon_0 > 0$. Having already established the conclusion for the case that $\varepsilon_0 = 0$, this completes the proof of Theorem B.5 for the case that $\eta \in (0,k]$. ∎

We now turn our attention to the full *Lipschitz Sandwich Theorem* B.1. Our strategy to prove this result will be to patch together the local Lipschitz bounds achieved in Theorem B.5. To be more precise, recall that $\Sigma \subset V$ is closed and $\gamma > 0$ with $k \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k,k+1]$. Let $\eta \in (0,\gamma)$, $K_0, \varepsilon > 0$, and $0 \leq \varepsilon_0 < \min\{2K_0,\varepsilon\}$. Retrieve the constant $\delta_0 = \delta_0(K_0,\varepsilon,\varepsilon_0,\gamma,\eta) > 0$ arising in the *Single-Point Lipschitz Sandwich Theorem* B.5. Assume that $B \subset \Sigma$ is a δ_0-cover of Σ in the sense that the δ_0-fattening of B contains Σ.

Suppose $\psi = \left(\psi^{(0)},\ldots,\psi^{(k)}\right) \in \mathrm{Lip}(\gamma,\Sigma,W)$ and $\varphi \in \left(\varphi^{(0)},\ldots,\varphi^{(k)}\right) \in \mathrm{Lip}(\gamma,\Sigma,W)$ both satisfy the norm bounds $||\psi||_{\mathrm{Lip}(\gamma,\Sigma,W)}, ||\varphi||_{\mathrm{Lip}(\gamma,\Sigma,W)} \leq K_0$. Further suppose that for every $j \in \{0,\ldots,k\}$ and every $x \in B$ the difference $\psi^{(j)}(x) - \varphi^{(j)}(x) \in \mathcal{L}(V^{\otimes j};W)$ satisfies $\left|\left|\psi^{(j)}(x) - \varphi^{(j)}(x)\right|\right|_{\mathcal{L}(V^{\otimes j};W)} \leq \varepsilon_0$. Then given any point $p \in B$, we can apply the *Single-Point Lipschitz Sandwich Theorem* B.5 to conclude that $||\psi_{[q]} - \varphi_{[q]}||_{\mathrm{Lip}(\eta,\Omega_p,W)} \leq \varepsilon$ for $\Omega_p := \Sigma \cap \overline{\mathbb{B}}_V(p,\delta_0)$ and $q \in \{0,\ldots,k\}$ such that $\eta \in (q,q+1]$.

It may initially appear that since $\Sigma = \cup_{p\in B}\Omega_p$ these local $\mathrm{Lip}(\eta)$-norm bounds should combine together to yield $||\psi_{[q]} - \varphi_{[q]}||_{\mathrm{Lip}(\eta,\Sigma,W)} \leq \varepsilon$. However, this is not necessarily true. For example, given any $\alpha \in (0,1)$, consider the function $F : [0,1] \cup [1+\alpha,2] \to \mathbb{R}$ defined by $F(x) := 0$ if $x \in [0,1]$ and $F(x) := \alpha$ if $x \in [1+\alpha,2]$. Then $F \in \mathrm{Lip}(1,[0,1]\cup[1+\alpha,2],\mathbb{R})$ and we have that $||F||_{\mathrm{Lip}(1,[0,1],\mathbb{R})} = 0$ and $||F||_{\mathrm{Lip}(1,[1+\alpha,2],\mathbb{R})} = \alpha$. But $|F(1+\alpha) - F(1)| = \alpha = |1+\alpha-1|$ and so $||F||_{\mathrm{Lip}(1,[0,1]\cup[1+\alpha,2],\mathbb{R})} = 1 > \alpha$.

The main content of our proof of the *Lipschitz Sandwich Theorem* B.1 is to overcome this problem. We prove that, by requiring the constant ε_0 to be sufficiently small, depending only on ε, K_0, γ and η, rather than an arbitrary real number in the interval $[0,\min\{2K_0,\varepsilon\})$, we *can* patch together local Lipschitz estimates resulting from an application of the *Single-Point Lipschitz Sandwich Theorem* B.5 to yield global Lipschitz estimates throughout Σ.

Proof of Theorem B.1. Let V and W be Banach spaces, and assume that the tensor powers of V are all equipped with admissible norms (cf. Definition A.1). Assume that $\Sigma \subset V$ is non-empty and closed. Let $K_0, \varepsilon, \gamma > 0$ with $k \in \mathbb{Z}_{\geq 0}$ such that $\gamma \in (k,k+1]$. Further let $\eta \in (0,\gamma)$ with $q \in \{0,\ldots,k\}$ such that $\eta \in (q,q+1]$. It suffices to prove the theorem under the additional assumption that $\varepsilon \leq 2K_0$; the conclusion (B.4) being valid for ε immediately means it is also valid for any constant $\varepsilon' \geq \varepsilon$.

Define $\theta := \frac{1}{2(1+e)} > 0$ and retrieve the constant $\delta_0 > 0$ arising from Theorem B.5 for the same constants K_0, γ and η as here respectively, and with the choices of $\theta\varepsilon$ and $\frac{\theta}{2}\varepsilon$ here as the constants ε and ε_0 in Theorem B.5 respectively. Note that we are not actually applying Theorem B.5, but simply retrieving a constant in preparation for its future application. Examining the dependencies in Theorem B.5 reveals that $\delta_0 > 0$ depends only on ε, K_0, γ and η. If necessary, we reduce δ_0, without additional dependencies, so that $\delta_0 \leq 1$. Further, we replace δ_0 by $\delta_0/2$.

Our choice of $\delta_0 > 0$ means that if $\psi = \left(\psi^{(0)},\ldots,\psi^{(k)}\right)$, $\varphi = \left(\varphi^{(0)},\ldots,\varphi^{(k)}\right) \in \mathrm{Lip}(\gamma,\Sigma,W)$ with $||\psi||_{\mathrm{Lip}(\gamma,\Sigma,W)}, ||\varphi||_{\mathrm{Lip}(\gamma,\Sigma,W)} \leq K_0$, and if for a point $p \in \Sigma$ and every $l \in \{0,\ldots,k\}$ we have the estimate $\left|\left|\psi^{(l)}(p) - \varphi^{(l)}(p)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \frac{\theta}{2}\varepsilon$, then an application of Theorem B.5 would allow us to conclude the estimate that $\left|\left|\psi_{[q]} - \varphi_{[q]}\right|\right|_{\mathrm{Lip}(\eta,\Omega_p,W)} \leq \theta\varepsilon$ for $\Omega_p := \overline{\mathbb{B}}_V(p,2\delta_0) \cap \Sigma$.

We now fix the value of $\delta_0 > 0$ for the remainder of the proof. Having done so, we define $\varepsilon_0 > 0$ by

$$\varepsilon_0 := \min\left\{\theta\,,\, \frac{\delta_0^{\eta}}{e^{\delta_0}\left(1+e^{\delta_0}\right)}\right\}\frac{\varepsilon}{2} > 0. \tag{C.77}$$

Examining the dependencies in (C.77) reveals that ε_0 depends only on ε, K_0, γ and η. We may now fix the value of $\varepsilon_0 > 0$ for the the remainder of the proof.

Let $B \subset \Sigma$ satisfy that

$$\Sigma \subset \bigcup_{x \in B} \overline{\mathbb{B}}_V(x, \delta_0). \tag{C.78}$$

Suppose $\psi = \left(\psi^{(0)}, \ldots, \psi^{(k)}\right), \varphi = \left(\varphi^{(0)}, \ldots, \varphi^{(k)}\right) \in \mathrm{Lip}(\gamma, \Sigma, W)$ with $||\psi||_{\mathrm{Lip}(\gamma,\Sigma,W)}, ||\varphi||_{\mathrm{Lip}(\gamma,\Sigma,W)} \leq K_0$. Further assume that whenever $l \in \{0, \ldots, k\}$ and $x \in B$ we have the estimate $\left|\left|\psi^{(l)}(x) - \varphi^{(l)}(x)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \varepsilon_0$. Let $p \in B$. Recalling how we chose the constant $\delta_0 > 0$ and that (C.77) means that $\varepsilon_0 \leq \frac{\theta}{2}\varepsilon$, we may appeal to Theorem B.5 to conclude that

$$\left|\left|\psi_{[q]} - \varphi_{[q]}\right|\right|_{\mathrm{Lip}(\eta,\Omega_p,W)} \leq \theta\varepsilon \tag{C.79}$$

where $\Omega_p := \Sigma \cap \overline{\mathbb{B}}_V(p, 2\delta_0)$. The arbitrariness of $p \in B$ allows us to conclude that the estimate (C.79) is valid for every $p \in B$.

We complete the proof of Theorem B.1 by establishing that having the bounds (C.79) for every $p \in B$ allows us to conclude that $\left|\left|\psi_{[q]} - \varphi_{[q]}\right|\right|_{\mathrm{Lip}(\eta,\Sigma,W)} \leq \varepsilon$. This is proven in the following claim.

Claim C.6. *If $F = \left(F^{(0)}, \ldots, F^{(k)}\right) \in \mathrm{Lip}(\gamma, \Sigma, W)$ satisfies, for every $l \in \{0, \ldots, k\}$ and every $z \in B$, that $\left|\left|F^{(l)}(z)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \varepsilon_0$ and $\left|\left|F_{[q]}\right|\right|_{\mathrm{Lip}(\eta,\Omega_z,W)} \leq \theta\varepsilon$, where $\Omega_z := \Sigma \cap \overline{\mathbb{B}}_V(z, 2\delta_0)$, then we have*

$$\left|\left|F_{[q]}\right|\right|_{\mathrm{Lip}(\eta,\Sigma,W)} \leq \varepsilon. \tag{C.80}$$

Proof of Claim C.6. For each $l \in \{0, \ldots, k\}$ let $R_l^F : \Sigma \times \Sigma \to \mathcal{L}(V^{\otimes l}; W)$ denote the remainder term associated to $F^{(l)}$. Therefore whenever $l \in \{0, \ldots, k\}$, $x, y \in \Sigma$ and $v \in V^{\otimes l}$, we have that (cf. (A.5))

$$R_l^F(x, y)[v] := F^{(l)}(y)[v] - \sum_{s=0}^{k-l} \frac{1}{s!} F^{(l+s)}(x)\left[v \otimes (y-x)^{\otimes j}\right]. \tag{C.81}$$

If $q = k$ then we may work with the unaltered remainder terms defined in (C.81). But if $q < k$ then we must first appropriately alter the remainder terms. For this purpose, for each $l \in \{0, \ldots, q\}$ we define $\hat{R}_l^F : \Sigma \times \Sigma \to \mathcal{L}(V^{\otimes l}; W)$ for $x, y \in \Sigma$ and $v \in V^{\otimes l}$ by

$$\hat{R}_l^F(x, y)[v] := \begin{cases} R_l^F(x, y)[v] & \text{if } q = k \\ R_l^F(x, y)[v] + \sum_{s=q+1-l}^{k-l} \frac{1}{s!} F^{(l+s)}(x)\left[v \otimes (y-x)^{\otimes s}\right] & \text{if } q < k. \end{cases} \tag{C.82}$$

It follows from (C.81) and (C.82) that whenever $l \in \{0, \ldots, q\}$, $x, y \in \Sigma$ and $v \in V^{\otimes l}$ we have

$$F^{(l)}(y)[v] = \sum_{s=0}^{q-l} F^{(l+s)}(x)\left[v \otimes (y-x)^{\otimes s}\right] + \hat{R}_l^F(x, y)[v]. \tag{C.83}$$

For each $z \in B$, the assumption that $||F_{[q]}||_{\mathrm{Lip}(\eta,\Omega_z,W)} \leq \theta\varepsilon$ for $\Omega_z := \Sigma \cap \overline{\mathbb{B}}_V(z, 2\delta_0)$ tells us that for every $l \in \{0, \ldots, q\}$ and any $x, y \in \Sigma \cap \overline{\mathbb{B}}_V(z, 2\delta_0)$ we have

$$\textbf{(I)} \quad \left|\left|F^{(l)}(x)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \theta\varepsilon \quad \text{and} \quad \textbf{(II)} \quad \left|\left|\hat{R}_l^F(x, y)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \theta\varepsilon||y-x||_V^{\eta-l}. \tag{C.84}$$

Consider $p \in \Sigma$. From (C.78) we know that $p \in \Sigma \cap \overline{\mathbb{B}}_V(z, \delta_0)$ for some $z \in B$. Consequently the bound **(I)** in (C.84) holds for $x := p$. Since $p \in \Sigma$ was arbitrary, we conclude that for any $p \in \Sigma$ we have

$$\left|\left|F^{(l)}(p)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \theta\varepsilon. \tag{C.85}$$

Consider $l \in \{0, \ldots, q\}$ and $p, w \in \Sigma$. If there exists $z \in B$ for which $p, w \in \overline{\mathbb{B}}_V(z, 2\delta_0)$ then **(II)** in (C.84) yields that

$$\left|\left|\hat{R}_l^F(p, w)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \theta\varepsilon||w-p||_V^{\eta-l}. \tag{C.86}$$

Now suppose that no single ball $\overline{\mathbb{B}}_V(z, 2\delta_0)$ contains both p and w. From (C.78) we know that $p \in \Sigma \cap \overline{\mathbb{B}}_V(z_i, \delta_0)$ and $w \in \Sigma \cap \overline{\mathbb{B}}_V(z_j, \delta_0)$ for some $z_i, z_j \in B$ which must be distinct. In fact, since $w \notin \Sigma \cap \overline{\mathbb{B}}_V(z_i, 2\delta_0)$, we can conclude that

$$||w - p||_V \geq \delta_0. \tag{C.87}$$

Observe that from (C.83) we have, for any $v \in V^{\otimes l}$, that

$$\hat{R}^F_l(p, w)[v] = F^{(l)}(w)[v] - \sum_{s=0}^{q-l} \frac{1}{s!} F^{(l+s)}(p) \left[v \otimes (w-p)^{\otimes s}\right]. \tag{C.88}$$

We may further use (C.83) to compute that

$$F^{(l)}(w)[v] = \sum_{u=0}^{q-l} \frac{1}{u!} F^{(l+u)}(z_j) \left[v \otimes (w - z_j)^{\otimes u}\right] + \hat{R}^F_l(z_j, w)[v]. \tag{C.89}$$

Since $w \in \overline{\mathbb{B}}_V(z_j, \delta_0)$ we may use (C.86) to conclude that

$$\left|\left|\hat{R}^F_l(z_j, w)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \theta\varepsilon ||w - z_j||_V^{\eta - l} \leq \theta\varepsilon\delta_0^{\eta - l}. \tag{C.90}$$

Additionally, since $z_j \in B$ we may compute that

$$\sum_{u=0}^{q-l} \frac{1}{u!} \left|\left|F^{(l+u)}(z_j) \left[v \otimes (w - z_j)^{\otimes u}\right]\right|\right|_W \leq \varepsilon_0 \sum_{u=0}^{q-l} \frac{1}{u!} ||w - z_j||_V^u ||v||_{V^{\otimes l}} \leq \varepsilon_0 e^{\delta_0} ||v||_{V^{\otimes l}}. \tag{C.91}$$

Combining (C.89), (C.90), and (C.91) yields the estimate

$$\left|\left|F^{(l)}(w)[v]\right|\right|_W \leq \left(\theta\varepsilon\delta_0^{\eta - l} + \varepsilon_0 e^{\delta_0}\right) ||v||_{V^{\otimes l}}. \tag{C.92}$$

Turning our attention to the second term in (C.88), note that for any $s \in \{0, \ldots, q - l\}$ we have via (C.83), for $v' := v \otimes (w - p)^{\otimes s} \in V^{\otimes(l+s)}$, that

$$F^{(l+s)}(p)\left[v'\right] = \sum_{u=0}^{q-l-s} \frac{1}{u!} F^{(l+s+u)}(z_i) \left[v' \otimes (p - z_i)^{\otimes u}\right] + \hat{R}^F_{l+s}(z_i, p)\left[v'\right]. \tag{C.93}$$

Since $p \in \overline{\mathbb{B}}_V(z_i, \delta_0)$ we may use (C.86) to conclude that

$$\left|\left|\hat{R}^F_{l+s}(z_i, p)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \theta\varepsilon ||p - z_i||_V^{\eta - (l+s)} \leq \theta\varepsilon\delta_0^{\eta - (l+s)}. \tag{C.94}$$

Via similar computations to those used to establish (C.91), the fact that $z_i \in B$ allows us to compute that

$$\sum_{u=0}^{q-l-s} \frac{1}{u!} \left|\left|F^{(l+u+s)}(z_j) \left[v' \otimes (p - z_i)^{\otimes u}\right]\right|\right|_W \leq \varepsilon_0 e^{\delta_0} ||v'||_{V^{\otimes(l+s)}}. \tag{C.95}$$

Combining (C.93), (C.94), and (C.95) yields that

$$\left|\left|F^{(l+s)}(p)\left[v'\right]\right|\right|_W \leq \left(\varepsilon_0 e^{\delta_0} + \theta\varepsilon\delta_0^{\eta-(l+s)}\right) ||v'||_{V^{\otimes(l+s)}} = \left(\varepsilon_0 e^{\delta_0} + \theta\varepsilon\delta_0^{\eta-(l+s)}\right) ||w - p||_V^s ||v||_{V^{\otimes l}} \tag{C.96}$$

where the last equality uses that $v' := v \otimes (w - p)^{\otimes s}$ and that the tensor powers of V are equipped with admissible norms (cf. Definition A.1). A consequence of (C.96) is that

$$\sum_{s=0}^{q-l} \frac{1}{s!} \left|\left|F^{(l+s)}(p)\left[v'\right]\right|\right|_W \leq \sum_{s=0}^{q-l} \frac{1}{s!} \left(\varepsilon_0 e^{\delta_0} + \theta\varepsilon\delta_0^{\eta-(l+s)}\right) ||w-p||_V^s ||v||_{V^{\otimes l}}. \tag{C.97}$$

Since from (C.87) we have that $\delta_0 \leq ||w-p||_V$, we may multiply each term in the sum on the RHS of (C.97) by $||w-p||_V^{\eta-(l+s)}\delta_0^{-(\eta-(l+s))} \geq 1$ to conclude that

$$\sum_{s=0}^{q-l} \frac{1}{s!} \left|\left|F^{(l+s)}(p)\left[v'\right]\right|\right|_W \leq \sum_{s=0}^{q-l} \frac{1}{s!} \left(\varepsilon_0 e^{\delta_0} + \theta\varepsilon\delta_0^{\eta-(l+s)}\right) \delta_0^{-(\eta-l-s)} ||w-p||_V^{\eta-l} ||v||_{V^{\otimes l}}. \tag{C.98}$$

Combining (C.88), (C.92), and (C.98) allows us to deduce that

$$\left|\left|\hat{R}_l^F(p,w)[v]\right|\right|_W \leq \left(\theta\varepsilon\delta_0^{\eta-l} + \varepsilon_0 e^{\delta_0} + \sum_{s=0}^{q-l}\frac{1}{s!}\left(\varepsilon_0 e^{\delta_0}\delta_0^{-(\eta-l-s)} + \theta\varepsilon\right)||w-p||_V^{\eta-l}\right)||v||_{V^{\otimes l}}. \tag{C.99}$$

Observe that

$$\theta\varepsilon\delta_0^{\eta-l} \overset{\text{(C.87)}}{\leq} \theta\varepsilon||w-p||_V^{\eta-l}, \tag{C.100}$$

$$\varepsilon_0 e^{\delta_0} \overset{\text{(C.87)}}{\leq} \varepsilon_0\delta_0^{-(\eta-l)} e^{\delta_0}||w-p||_V^{\eta-l}, \tag{C.101}$$

$$\varepsilon_0 e^{\delta_0}\delta_0^{-(\eta-l)}\sum_{s=0}^{q-l}\frac{1}{s!}\delta_0^s \leq \varepsilon_0\delta_0^{-(\eta-l)}e^{2\delta_0}, \quad \text{and} \tag{C.102}$$

$$\theta\varepsilon\sum_{s=0}^{q-l}\frac{1}{s!} \leq \theta e\varepsilon. \tag{C.103}$$

Combining (C.100), (C.101), (C.102), and (C.103) with (C.99) yields

$$\left|\left|\hat{R}_l^{\varphi}(p,w)[v]\right|\right|_W \leq \left(\theta\varepsilon(1+e) + \varepsilon_0\delta_0^{-(\eta-l)}e^{\delta_0}\left(1+e^{\delta_0}\right)\right)||w-p||_V^{\eta-l}||v||_{V^{\otimes l}}. \tag{C.104}$$

Taking the supremum over $v \in V^{\otimes l}$ with unit $V^{\otimes l}$-norm in (C.104) yields

$$\left|\left|\hat{R}_l^F(p,w)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \left(\theta\varepsilon(1+e) + \varepsilon_0\delta_0^{-\eta}e^{\delta_0}\left(1+e^{\delta_0}\right)\right)||w-p||_V^{\eta-l} \tag{C.105}$$

since $\delta_0 \leq 1$ means $\delta_0^{-(\eta-l)} \leq \delta_0^{-\eta}$ for every $l \in \{0,\ldots,q\}$.

Together (C.86), (C.105) and the inequality $\theta\varepsilon < \theta\varepsilon(1+e) + \varepsilon_0\delta_0^{-\eta}e^{\delta_0}\left(1+e^{\delta_0}\right)$ mean that for any $l \in \{0,\ldots,q\}$ and any $p,w \in \Sigma$ we have

$$\left|\left|\hat{R}_l^F(p,w)\right|\right|_{\mathcal{L}(V^{\otimes l};W)} \leq \left(\theta\varepsilon(1+e) + \varepsilon_0\frac{e^{\delta_0}\left(1+e^{\delta_0}\right)}{\delta_0^\eta}\right)||w-p||_V^{\eta-l}. \tag{C.106}$$

The definitions (C.83), the bounds (C.85), and the Hölder estimates (C.106) tell us that

$$\begin{aligned}||F_{[q]}||_{\mathrm{Lip}(\eta,\Sigma,W)} &\leq \theta\varepsilon(1+e) + \varepsilon_0\frac{e^{\delta_0}\left(1+e^{\delta_0}\right)}{\delta_0^\eta} \\ &\overset{\text{(C.77)}}{\leq} \frac{1}{2(1+e)}\varepsilon(1+e) + \frac{\varepsilon}{2}\frac{\delta_0^\eta}{e^{\delta_0}\left(1+e^{\delta_0}\right)}\frac{e^{\delta_0}\left(1+e^{\delta_0}\right)}{\delta_0^\eta} = \frac{\varepsilon}{2}+\frac{\varepsilon}{2} = \varepsilon\end{aligned}$$

as claimed in (C.80). This completes the proof of Claim C.6 ■

Returning to the proof of Theorem B.1 itself, we define $F := \psi - \varphi \in \mathrm{Lip}(\gamma,\Sigma,W)$ so that for every $j \in \{0,\ldots,k\}$

we have $F^{(j)} = \psi^{(j)} - \varphi^{(j)}$. Then, by assumption, we have for every $j \in \{0, \ldots, k\}$ and every $z \in B$ that $\left|\left|F^{(j)}(z)\right|\right|_{\mathcal{L}(V^{\otimes j};W)} \leq \varepsilon_0$. Moreover, (C.79) tells us that whenever $z \in B$ we have that $\left|\left|F_{[q]}\right|\right|_{\mathrm{Lip}(\eta,\Omega_z,W)} \leq \theta\varepsilon$ for $\Omega_z := \Sigma \cap \overline{\mathbb{B}}_V(z, 2\delta_0)$. Therefore we can apply Claim C.6 to F and conclude that $\left|\left|F_{[q]}\right|\right|_{\mathrm{Lip}(\eta,\Sigma,W)} \leq \varepsilon$. Since $F := \psi - \varphi$ this gives the estimate claimed in (B.4) and completes the proof of Theorem B.1. ■

References

[ABDHP21] D. Alistarh, T. Ben-Nun, N. Dryden, T. Hoefler and A. Peste, *Sparsity in Deep Learning: Pruning and growth for efficient inference and training in neural networks*, J. Mach. Learn. Res., **22**(241), 1-124, 2021.

[ABCGMM18] J.-P. Argaud, B. Bouriquet, F. de Caso, H. Gong, Y. Maday and O. Mula, *Sensor Placement in Nuclear Reactors Based on the Generalized Empirical Interpolation Method*, J. Comput. Phys., vol. **363**, pp. 354-370, 2018.

[ABGMM16] J.-P. Argaud, B. Bouriquet, H. Gong, Y. Maday and O. Mula, *Stabilization of (G)EIM in Presence of Measurement Noise: Application to Nuclear Reactor Physics*, In Spectral and High Order Methods for Partial Differential Equations-ICOSAHOM 2016, vol. **119** of Lect. Notes Comput. Sci. Eng., pp.133-145, Springer, Cham, 2017.

[AK13] M. G. Augasts and T. Kathirvalavakumar, *Pruning algorithms of neural networks - A comparative study*, Open Computer Science **3**, p.105-115, 2013.

[BLL15] Francis Bach, Simon Lacoste-Julien and Fredrik Lindsten, *Sequential Kernel Herding: Frank-Wolfe Optimization for Particle Filtering*, In Artificial Intelligence and Statistics, pages 544-552, PMLR, 2015.

[BMPSZ08] F. R. Bach, J. Mairal, J. Ponce, G. Sapiro and A. Zisserman, *Discriminative learned dictionaries for local image analysis*, Compute Vision and Pattern Recognition (CVPR), pp.1-8, IEEE 2008.

[BMPSZ09] F. R. Bach, J. Mairal, J. Ponce, G. Sapiro and A. Zisserman, *Supervised dictionary learning*, Advances in Neural Information Processing Systems, pp.1033-1040, 2009.

[BMNP04] M. Barrault, Y. Maday, N. C. Nguyen and A. T. Patera, *An Empirical Interpolation Method: Application to Efficient Reduced-Basis Discretization of Partial Differential Equations*, C. R. Acad. Sci. Paris, Série I., **339**, pp.667-672, 2004

[BS18] A. G. Barto and R. S. Sutton, *Reinforcement learning: an introduction*, (2nd). Cambridge, MA, USA: MIT Press.

[BT11] Alain Belinet and Christine Thomas-Agnan, *Reproducing kernel Hilbert spaces in probabiility and statistics*, Springer Science & Business Media, 2011.

[BCGMO18] F. -X. Briol, W. Y. Chen, J. Gorham, L. Mackey and C. J. Oates, *Stein Points*, In Proceedings of the 35th International Conference on Machine Learning, Volume **80**, pp. 843-852, PMLR, 2018.

[CT05] E. Candés and T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory, **51**, 2005.

[CRT06] E. Candés, J. Romberg and T. Tao, *Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information*, IEEE Trans. Inform. Theory, **52**, pp. 489-509, 2006.

[Car81] B. Carl, *Entropy numbers of diagonal operators with an application to eigenvalue problems*, J. Approx. Theory, **32**, pp.135-150, 1981.

[CS90] B. Carl and I. Stephani, *Entropy, Compactness, and the Approximation of Operators*, Cambridge University Press, Cambridge, UK, 1990.

[CHXZ20] L. Chen, A. Huang, S. Xu and B. Zhang, *Convolutional Neural Network Pruning: A Survey*, 39th Chinese Control Conference (CCC), IEEE, p.7458-7463, 2020.

[CSW11] Zhixiang Chen, Baohuai Sheng and Jianli Wang, *The Covering Number for Some Mercer Kernel Hilbert Spaces on the Unit Sphere*, Taiwanese J. Math. **15**(3), p.1325-1340, 2011.

[CSW10] Y. Chen, A. Smola and M. Welling, *Super-Samples from Kernel Herding*, In Proceedings of the Conference on Uncertainty in Artificial Intelligence, 2010.

[CS08] A. Christmann and I. Steinwart, *Support Vector Machines*, Springer Science & Business Media, 2008.

[CM73] Jon F. Claerbout and Francis Muir, *Robust Modeling with Erratic Data*, Geophysics, Vol. **38**, So. 5, pp. 826-844, (1973)

[Don06] D. Donoho, *Compressed Sensing*, IEEE Trans. Inform. Theory, **52**, pp. 1289-1306, 2006.

[DE03] D. Donoho and M. Elad, *Optimally Sparse Representation in General (Nonorthogonal) Dictionaries via l^1 Minimization*, Proc. Natl. Acad. Sci. USA, **100**, pp.2197-2202, 2003.

[DGOY08] J. Darbon, D. Goldfarb, S. Osher and W. Yin, *Bregman Iterative Algorithms for L1 Minimization with Applications to Compressed Sensing*, SIAM J. Imaging Sci. **1**, pp.143-168, 2008.

[DM21a] Raaz Dwivedi and Lester Mackey, *Kernel Thinning*, `https://arxiv.org/abs/2105.05842` [stat.ML], 13 November 2021.

[DM21b] Raaz Dwivedi and Lester Mackey, *Generalized Kernel Thinning*, `https://arxiv.org/abs/2110.01593` [stat.ML], 16 November 2021.

[DMS21] Raaz Dwivedi, Lester Mackey and Abhishek Shetty, *Distribution Compression in Near-Linear Time*, `https://arxiv.org/abs/2111.07941`. [stat.ML], 17 November 2021.

[DF15] Chris Dyer and Manaal Faruqui, *Non-distributional Word Vector Representations*, Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Association for Computational Linguistics, Beijing, China, Volume **2**: Short Papers, pp.464-469, 2015.

[ET96] D. E. Edmunds and H. Triebel, *Function Spaces, Entropy Numbers, and Differential Operators*, Cambridge University Press, Cambridge, UK, 1996.

[EMS08] M. Elad, J. Mairal and G. Sapiro, *Sparse representation for color image restoration*, IEEE Transactions on Image Processing, **17**(1), 53-69, 2008.

[EPP00] T. Evgeniou, M. Pontil and T. Poggio, *Regularization networks and support vector machines*, Adv. Comput. Math., **13**, pp.1-50, 2000.

[FS21] S. Fischer and I. Steinwart, *A closer look at covering number bounds for Gaussian kernels*, Journal of Complexity, Volume **62**, 2021.

[FMMT14] Alona Fyshe, Tom M. Mitchell, Brian Murphy and Partha P. Talukdar, *Interpretable semantic Vectors from a joint model of brain-and text-based meaning*, Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Baltimore, Maryland, Volume **1**: Long Papers, pp.489-499, 2014.

[GMSWY09] A. Ganesh, Y. Ma, S. S. Sastry, J. Wright and A. Yang, *Robust face recognition via sparse representation*, TPAMI **31**(2), 210-227, 2009.

[GLSWZ21] John Glossner, Tailin Liang, Shaobo Shi, Lei Wang and Xiaotong Zhang, *Pruning and quantization for deep neural network acceleration: A survey*, Neurocomputing, Volume **461**, pages 370-403, 2021.

[GO09] T. Goldstein and S. Osher, *The Split Bregman Method for L1-Regularized Problems*, SIAM J. Imaging Sci. **2**, pp.323-343, 2009.

[GMNP07] M. A. Grepl, Y. Maday, N. C. Nguyen and A. T. Patera, *Efficient Reduced-Basis Treatment of Nonaffine and Nonlinear Partial Differential Equations*, M2AN (Math. Model. Numer. Anal.), 2007.

[HLO21] Satoshi Hayakawa, Terry Lyons and Harald Oberhauser, *Positively weighted kernel Quadrature via Subsampling*, arXiv preprint, `https://arxiv.org/abs/2107.09597`

[HS19] J. Hernandez-Garcia and R. Sutton, *Learning Sparse Representations Incrementally in Deep Reinforcement Learning*, `https://arxiv.org/abs/1912.04002`, [cs.LG], 9 Dec 2019.

[HLLL18] J. Huang, H. Lian, H. Lin and S. Lv, *Oracle Inequalitites for Sparse Additive Quantile Regression in Reproducing Kernel Hilbert Space*, Annals of Statistics **46**, p.781-813, 2018.

[JJWWY20] C. Jin, M. I. Jordan, M. Wang, Z. Wang and Z. Yang, *On Function Approximation in Reinforcement Learning: Optimism in the Face of Large State Spaces*, 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada, 2020

[KK20] M. Kalinić and P. Kómár, *Denoising DNA encoded library screens with sparse learning*, ACS Combinatorial Science, Vol. **22**, no. 8, pp.410-421, 2020.

[Kol56] A. N. Kolmogorov, *Asymptotic characteristics of some completely bounded metric spaces*, Dokl. Akad. Nauk. SSSR, **108**, pp.585-589, 1956.

[Kuh11] T. Kühn, *Covering numbers of Gaussian reproducing kernel Hilbert spaces*, J. Complexity, **27**, pp.489-499, 2011.

[LY07] Y. Lin and M. Yuan, *On the Non-Negative Garrote Estimator*, J. R. Stat. Soc. Ser. B **69**, pp. 143-161, 2007

[LL16] W. Lee and T. Lyons, *The Adaptive Patched Cubature Filter and its Implementation*, Communications in Mathematical Sciences, **14**(3), pp.799-829, 2016.

[LL99] W. V. Li and W. Linde, *Approximation, metric entropy and small ball estimates for Gaussian measures*, Ann. Probab., **27**, pp.1556-1578, 1999.

[LL12] C. Litterer and T. Lyons, *High order recombination and an application to cubature on wiener space*, The Annals of Applied Probability, Vol. **22**, No. 4, 1301-1327, 2012.

[LP04] H. Luschgy and G. Pagés, *Sharp asymptotics of the Kolmogorov entropy for Gaussian measures*, J. Funct. Anal., **212**, pp.89-120, 2004.

[MM13] Y. Maday, O. Mula, *A generalized empirical interpolation method: Application of reduced basis techniques to data assimilation*, F. Brezzi, P. Colli Franzone, U. Gianazza, G. Gilardi (Eds.), Analysis and Numerics of Partial Differential Equations, Vol. 4 of Springer INdAM Series, Springer Milan, 2013, pp. 221–235.

[MMT14] Y. Maday, O. Mula and G. Turinici, *Convergence analysis of the Generalized Empirical Interpolation Method*, SIAM J. Numer. Anal., **54(3)** 1713-1731, 2014.

[MNPP09] Y. Maday, N. C. Nguyen, A. T. Patera and G. S. H. Pau, *A General Multipurpose Interpolation Procedure: The Magic Points*, Commun. Pure Appl. Anal., **81**, pp. 383-404, 2009.

[McS34] Edward James McShane, *Extension of range of functions*, Bull. Amer. Math. Soc., **40**, p.837-842, 1934.

[MRT12] Mehryar Mohri, Afshin Rostamizadeh and Ameet Talwalkar, *Foundations of Machine Learning*, Massachusetts: MIT Press, USA, 2012.

[NPS22] M. Nikdast, S. Pasricha and F. Sunny, *SONIC: A Sparse Neural Network Inference Accelerator with Silicon Photonics for Energy-Efficient Deep Learning*, 27th Asia and South Pacific Design Automation Conference (ASP-DAC), pp. 214-219, 2022.

[PTT22] Sebastian Pokutta, Ken'ichiro Tanaka and Kazuma Tsuji, *Sparser Kernel Herding with Pairwise Conditional Gradients without Swap Steps*, `https://arxiv.org/abs/2110.12650`, [math.OC], 8 February 2022.

[Ree93] R. Reed, *Pruning Algorithms - A Survey*, IEEE Transactions on Neural Networks **4**, p.74-747, 1993.

[Rya13] Raymond A Ryan, *Introduction to tensor products of Banach spaces*, Springer Science & Business Media, 2013.

[SS86] Fadil Santosaf and William W. Symes, *Linear Inversion of Band-Limited Reflection Seismograms*, SIAM J. ScI. STAT. COMPUT. Vol. **7**, No. 4, 1986.

[SSW01] B. Schölkopf, A. J. Smola and R. C. Williamson, *Generalization performance of regularization networks and support vector machines via entrop numbers of compact operators*, IEEE Trans. Inform. Theory, **47**, pp.2516-2532, 2001.

[Ste70] E. M. Stein, *Singular Integrals and Differentiability Properties of Finctions*, Princeton Mathamatical Series, vol. **30**, Princeton Univeristy Press, Princeton, 1970.

[Ste03] I. Steinwart, *Entropy numbers of convex hulls and an application to learning algorithms*, Arch. Math., **80**, pp.310-318, 2003.

[SS07] I. Steinwart and C. Scovel, *Fast rates for support vector machines using Gaussian kernels*, Ann. Statist. **35** (2), 2007.

[SS12] I. Steinwart and C. Scovel, *Mercer's theorem on general domains: On the interaction between measures, kernels, and RKHSs*, Constructive Approximation, **35**(3):363-417, 2012.

[Str71] A. H. Stroud, *Approximate calculation of multiple integrals*, Series in Automatic Computation, Englewood Cliffs, NJ: Prentice-Hall, 1971.

[Suz18] T. Suzuki, *Fast Learning Rate of Non-Sparse Multiple Kernel Learning and Optimal Regularization Strategies*, Electronic Journal of Statistics **12**, p.2141-2192, 2018.

[SS13] M. Sugiyama and T. Suzuki, *Fast Learning Rate of Multiple Kernel Learning: Tradeoff Between Sparsity and Smoothness*, Annals of Statistics **41**, p.1381-1405, 2013.

[TT21] Ken'ichiro Tanaka and Kazuma Tsuji, *Acceleration of the Kernel Herding Algorithm by Improved Gradient Approximation*, `https://arxiv.org/abs/2105.07900`, [math.NA], 17 May 2021.

[TW19] GL. Tian and M. Wang, *Adaptive Group LASSO for High-Dimensional Generalized Linear Models*, Stat Papers **60**, pp. 1469-1486, 2019.

[Tib96] Robert Tibshirani, *Regression Shrinkage and Selection via the Lasso*, Journal of the Royal Statistical Society. Series B (Methodological), Vol. **58**, No. 1, pp. 267-288, 1996.

[Wel09a] M. Welling, *Herding Dynamical Weights to Learn*, In Proceedings of the 21st International Conference on Machine Learning, Montreal, Quebec, CAN, 2009.

[Wel09b] M. Welling, *Herding Dynamic Weights for Partially Observed Random Field Models*, In Proc. of the Conf. on Uncertainty in Artificial Intelligence, Montreal, Quebec, CAN, 2009.

[Whi34] H. Whitney, *Analytic extensions of differentiable functions defined in closed sets*, Trans. Amer. Math. Soc. vol. **36** (1934) pp. 63-89, 1934.

[Whi44] H. Whitney, *On the extension of differentiable functions*, Bull. Amer. Math. Soc. **50(2)**: 76-81, 1944.

[XZ16] Y. Xiang and C. Zhang, *On the Oracle Property of Adaptive Group LASSO in High-Dimensional Linear Models*, Stat. Pap. **57**, pp. 249-265, 2016.

[Zho02] D. -X. Zhou, *The covering number in learning theory*, J. Complexity, **18**, pp.739-767, 2002.

University of Oxford, Radcliffe Observatory, Andrew Wiles Building, Woodstock Rd, Oxford, OX2 6GG, UK.

TL: tlyons@maths.ox.ac.uk
`https://www.maths.ox.ac.uk/people/terry.lyons`

AM: andrew.mcleod@maths.ox.ac.uk
`https://www.maths.ox.ac.uk/people/andrew.mcleod`