Check for updates

Adv. Appl. Probab. **57,** 1321–1359 (2025)
*doi:*10.1017/apr.2025.13

STEADY-STATE DIRICHLET APPROXIMATION OF THE WRIGHT–FISHER MODEL USING THE PRELIMIT GENERATOR COMPARISON APPROACH OF STEIN'S METHOD

ANTON BRAVERMAN,* *Northwestern University*
HAN L. GAN,** *University of Waikato*

Abstract

The Wright–Fisher model, originating in Wright (1931) is one of the canonical probabilistic models used in mathematical population genetics to study how genetic type frequencies evolve in time. In this paper we bound the rate of convergence of the stationary distribution for a finite population Wright–Fisher Markov chain with parent-independent mutation to the Dirichlet distribution. Our result improves the rate of convergence established in Gan *et al.* (2017) from $\mathrm{O}(1/\sqrt{N})$ to $\mathrm{O}(1/N)$. The results are derived using Stein's method, in particular, the prelimit generator comparison method.

Keywords: Dirichlet approximation; Wright–Fisher model; Stein's method

2020 Mathematics Subject Classification: Primary 60F05
Secondary 60K99

1. Introduction

We focus on approximating the stationary distribution for a finite Wright–Fisher Markov chain with parent-independent mutation, where the population has fixed size N and a fixed number of allele types K. We represent this model as a discrete-time Markov chain (DTMC) $\{\boldsymbol{U}(t) : t \in \mathbb{Z}_+\}$ taking values in

$$\nabla^K = \left\{ \boldsymbol{u} \in \delta\mathbb{Z}^{K-1} : u_i \geq 0, \sum_{i=1}^{K-1} u_i \leq 1 \right\},$$

where $\delta = 1/N$ is a scaling parameter, $U_i(t)$ denotes the fraction of genes that are of type $1 \leq i \leq K-1$, and $1 - \sum_{i=1}^{K-1} U_i(t)$ is the fraction of genes of type K. For any $\boldsymbol{u}, \boldsymbol{y} \in \nabla^K$ and probabilities $p_1, \ldots, p_K$ such that $\sum_{i=1}^{K} p_i \leq 1$, the transition probabilities $P_{\boldsymbol{u}}(\boldsymbol{y}) = \mathbb{P}(U(1) = \boldsymbol{y} | U(0) = \boldsymbol{u})$ of this process satisfy

$$P_{\boldsymbol{u}}(\boldsymbol{y}) = \binom{N}{Ny_1, \ldots, Ny_{K-1}} \prod_{j=1}^{K-1} \left(u_j \left(1 - \sum_{i=1}^{K} p_i\right) + p_j \right)^{Ny_j}. \tag{1.1}$$

Received 9 January 2024; accepted 13 February 2025.
* Postal address: Kellogg School of Management, 2211 Campus Drive, Evanston, IL 60208, USA. Email: anton.braverman@kellogg.northwestern.edu
** Postal address: Department of Mathematics, Private Bag 3105, Hamilton 3240, New Zealand. Email: han.gan@waikato.ac.nz

© The Author(s), 2025. Published by Cambridge University Press on behalf of Applied Probability Trust. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

Our DTMC is irreducible, aperiodic, and positive recurrent because its state space is finite, and we let $\boldsymbol{U}$ denote the vector having the unique stationary distribution.

As it will be useful later, we informally give a common interpretation of the model and how it models changes in allele types over time. In each generation of fixed size N, given the parent generation, each individual in the offspring population independently chooses a parent, uniformly at random. In addition to this random genealogy structure, a random mutation structure is added such that each individual offspring independently has a probability of mutating to type i with probability p_i. Otherwise, with probability $1 - \sum_{i=1}^{K} p_i$, the offspring does not mutate and takes on the type of their parent. Note that in this structure each child could mutate to the same type as their respective parent.

The Wright–Fisher DTMC, while easy to formulate, is known to be relatively intractable, and hence, the limiting diffusion process as $N \to \infty$ is of particular interest. The limiting diffusion in this case is the well known Wright–Fisher diffusion, defined later in (1.8). Using the tools of [21], one can show that the Markov chain converges weakly to the diffusion process. Rates of convergence to the diffusion have been studied as well. In [11] the authors obtained an explicit bound of order $\mathrm{O}(1/N)$ when $K = 2$, but required bounded sixth derivatives on test functions. A similar result can be found in the more recent work of [16], where the author builds on the approach for Gaussian process approximation via Stein's method in [1] to establish a rate of convergence for the Moran model (a continuous-time cousin of the Wright–Fisher DTMC) to the limiting Wright–Fisher diffusion. The result in [16] is for the transient distribution of the Moran model and, like [11], requires $K = 2$. The rate of convergence obtained there is of order $\mathrm{O}(1/N^{1/4})$.

In [13] bounds for quantities of the form $|\mathbb{E}h(\boldsymbol{U}) - \mathbb{E}h(\boldsymbol{Z})|$ are derived, where $\boldsymbol{Z}$ is an appropriately chosen Dirichlet random variable, and h is any general test function with two bounded derivatives and a bounded Lipschitz constant on the second derivative. Under the typical assumption for these models that the mutation probabilities are rare, in the sense that $p_i = \mathrm{O}(1/N)$, [13] establish an upper bound on $|\mathbb{E}h(\boldsymbol{U}) - \mathbb{E}h(\boldsymbol{Z})|$ that is of order $\mathrm{O}(1/\sqrt{N})$. It has been anecdotally conjectured to the authors that this bound may not be of the optimal order, and that the correct order may be of order $\mathrm{O}(1/N)$, as suggested, for example, by the results of [11]. In this paper we derive a bound of order $\mathrm{O}(1/N)$ for approximating the stationary distribution assuming four bounded derivatives. Before we present the main result, we first define the Dirichlet distribution and our approximating metric.

Let $\bar{\nabla}^K = \{\boldsymbol{u} \in \mathbb{R}^{K-1} : x \geq 0, \sum_{i=1}^{K-1} x_i \leq 1\}$. We define the Dirichlet distribution with parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)$, where $\beta_1 > 0, \ldots, \beta_K > 0$, to be supported on the $(K-1)$-dimensional open simplex

$$\mathrm{int}(\bar{\nabla}^K) = \left\{ \boldsymbol{x} = (x_1, \ldots, x_{K-1}) : x_1 > 0, \ldots, x_{K-1} > 0, \sum_{i=1}^{K-1} x_i < 1 \right\} \subset \mathbb{R}^{K-1}.$$

The Dirichlet distribution has density

$$\psi_{\beta}(x_1, \ldots x_{K-1}) = \frac{\Gamma(s)}{\prod_{i=1}^{K} \Gamma(\beta_i)} \prod_{i=1}^{K} x_i^{\beta_i - 1}, \quad \boldsymbol{x} \in \mathrm{int}(\bar{\nabla}^K), \tag{1.2}$$

where $s = \sum_{i=1}^{K} \beta_i$, and we set $x_K = 1 - \sum_{i=1}^{K-1} x_i$. We assume that our mutation probabilities p_i satisfy

$$p_i = \frac{\beta_i}{2N}, \quad 1 \leq i \leq K, \tag{1.3}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

for some fixed $\boldsymbol{\beta}$ and all $N > 0$.

The metric we will be using is the Lipschitz type metric defined as follows. For any vector $\boldsymbol{a} \in \mathbb{Z}^{K-1}$ consisting of nonnegative integer values and a function $f : \mathbb{R}^{K-1} \to \mathbb{R}$, let

$$D^{\boldsymbol{a}} f(\boldsymbol{x}) = \frac{\partial^{a_{K-1}}}{\partial x_{K-1}^{a_{K-1}}} \cdots \frac{\partial^{a_1}}{\partial x_1^{a_1}} f(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^{K-1}, \tag{1.4}$$

and define

$$\mathcal{M}_j = \big\{ h : \mathbb{R}^{K-1} \to \mathbb{R},\ \sup_{\boldsymbol{x}} |D^{\boldsymbol{a}} h(\boldsymbol{x})| \le 1,\ 1 \le \|\boldsymbol{a}\|_1 \le j,\ \boldsymbol{a} \ge 0 \big\}. \tag{1.5}$$

Then for any random vectors $\boldsymbol{V}, \boldsymbol{V}' \in \mathbb{R}^{K-1}$, set

$$d_{\mathcal{M}_j}(\boldsymbol{V}, \boldsymbol{V}') = \sup_{h \in \mathcal{M}_j} \big| \mathbb{E} h(\boldsymbol{V}) - \mathbb{E} h(\boldsymbol{V}') \big|.$$

Lemma 2.2 of [18] proves that $\mathcal{M}_3$ is a convergence-determining class; i.e. $d_{\mathcal{M}_3}(\boldsymbol{V}, \boldsymbol{V}') \to 0$ implies that $\boldsymbol{V}$ and $\boldsymbol{V}'$ converge in distribution. Their result can be readily extended to show that $\mathcal{M}_4$, the class of functions used in this paper, is also convergence determining.

Theorem 1.1. *Let $\boldsymbol{U}$ denote the random vector with stationary distribution for the transition function (1.1), assume that the mutation probabilities satisfy (1.3) for some $\boldsymbol{\beta} > 0$, set $s = \sum_{i=1}^{K} \beta_i$, and let $\boldsymbol{Z}$ be a Dirichlet random variable with parameter vector $\boldsymbol{\beta}$. Then for all N such that $s/2N < 1$ and $N > 100K^2$,*

$$\begin{aligned} d_{\mathcal{M}_4}(\boldsymbol{U}, \boldsymbol{Z}) \le C(K) \Bigg[& \frac{1}{N^4} + \frac{2K^2(1+s)^2 + 20K^3 1595^K (1+s)^3 + 2K^4 1595^K (1+s)^4}{Ns} \\ & + \frac{1}{N^{\beta_K/2-3}} \frac{K^3 (10K)^{\beta_K} (4+6s)\Gamma(s)}{s\Gamma(s-\beta_K)\Gamma(\beta_K)\beta_K} \Big(1 + \frac{1}{(1-1/\sqrt{N})^{|s-\beta_K-1|}} \Big) \\ & + \frac{16 \cdot 1595^K K^5 (4K+1) 2^{4K} N^{8K-5}}{s(s-\beta_K)^{4K}} \Big(1 - \frac{10K(\sqrt{N}-1)(1-s/2N)}{N} \Big)^N \Bigg] \\ & + C(K)^2 \Bigg[\frac{Ks + 6K^2(1+s^2)}{N^2 s} + \frac{K}{N^3} \Bigg], \end{aligned}$$

where $C(K) \le (K^4 + 7975^K)(4 \times 12^K + K \cdot 11404^5 1595^K + 1)$ and $\Gamma(x) = \int_0^\infty t^{x-1} \mathrm{e}^{-t} dt$.

We make two remarks. First, most of the terms are $\mathrm{O}(1/N)$ except for the one involving $N^{3-\beta_K/2}$. The latter is also $\mathrm{O}(1/N)$ when $\beta_K \ge 8$ (without loss of generality, we can assume that β_K is the largest of all the betas). We believe it to be an artifact of our methodology and expect that it can be removed, although at considerable effort. Second, the constants of the form C^K (such as 1595^K, not to be confused with $C(K)$) are again a result of our prelimit generator approach and are not a fundamental part of the Wright–Fisher diffusion approximation error. Though tracking error terms more carefully would result in a modest decrease of the value of C in C^K, the presence of these terms is an inevitable side effect of our prelimit approach. We expand on both remarks at the end of this section after introducing the prelimit approach.

The primary tool used in this paper to prove the main results is Stein's method. Stein's method is a powerful tool in probability theory that is used to derive an explicit bound for

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

the difference between two probability distributions. Typically, one aims to use it to find an upper bound on the errors incurred when approximating an intractable target distribution with a commonly used simple reference distribution. It was first developed for the normal distribution in [20] to bound the approximation errors when applying the central limit theorem, and it has since been developed for numerous distributions, such as Poisson [4, 9], beta [10, 15], Dirichlet [13], Poisson-Dirichlet [14], negative binomial [7], exponential [12] to just name a few. For many more examples and applications, see, for example, the surveys or monographs [3, 8, 17, 19]. In the following we give a brief introduction to Stein's method.

To successfully apply Stein's method, one of the main approaches is what is known as the generator method, first pioneered in [2]. Below we give a brief description of Stein's method in general, with a particular focus on the generator method, and details of our approach. In this brief description we discuss the univariate case, but note that the multivariate case is analogous. Our goal is to bound the difference between the typically unknown or intractable law of our target random variable X with the law of a well understood and simple reference random variable Z. Stein's method can usually be summarised in the following three main steps.

1. Identify a characterising operator $\mathcal{G}_Z$ or identity that is satisfied only by the distribution of the reference random variable Z. In the generator method, the characterising operator is a generator of a Markov chain or diffusion process, and the reference distribution is the associated stationary distribution. The generator characterises its associated stationary distribution through the identity that $\mathbb{E}\mathcal{G}_Z f(Z) = 0$ for all suitable functions f if and only if Z follows the stationary distribution.

2. For any arbitrary function h, solve for the function f_h that satisfies

$$\mathcal{G}_Z f_h(x) = h(x) - \mathbb{E}h(Z). \tag{1.6}$$

   Then by setting $x = X$ and taking expectations,

$$|\mathbb{E}\mathcal{G}_Z f_h(X)| = |\mathbb{E}h(X) - \mathbb{E}h(Z)|. \tag{1.7}$$

   Properties of the function f_h turn out to be crucial to derive a good bound with Stein's method. Typically, one will require good bounds on f_h and its derivatives or differences if Z is continuous or discrete. Using the generator method, one can usually express f_h in terms of the semi-group of the process defined by $\mathcal{G}_Z$, and exploit properties and couplings of the process that found such bounds.

3. The goal is to bound (1.7) for all h from a rich enough family of test functions $\mathcal{H}$, where $\mathcal{H}$ is typically a convergence-determining class. Rather than directly bounding $|\mathbb{E}h(X) - \mathbb{E}h(Z)|$, which would typically require knowledge of the unknown density or distribution function of X, the final step is to derive a bound for $|\mathbb{E}\mathcal{G}_Z f_h(X)|$ that is more tractable. Standard approaches often involve Taylor expansions and couplings.

In [13] the above approach is used where $\mathcal{G}_{\mathbf{Z}}$ is the generator of the Wright–Fisher diffusion,

$$\mathcal{G}_{\mathbf{Z}} f(\mathbf{z}) := \frac{1}{2}\left[\sum_{i,j=1}^{K-1} z_i(\delta_{ij} - z_j)\frac{\partial f}{\partial z_i \partial z_j}(\mathbf{z}) + \sum_{i=1}^{K-1}(\beta_i - s z_i)\frac{\partial f}{\partial z_i}(\mathbf{z})\right], \quad \mathbf{z} \in \operatorname{int}(\bar{\nabla}^K), \tag{1.8}$$

where $s = \sum_{i=1}^{K} \beta_i$ and δ_{ij} denotes the Kronecker delta function. The stationary distribution associated to this generator is the Dirichlet distribution on $\operatorname{int}(\bar{\nabla}^K)$ with parameters $\beta_1, \ldots, \beta_K$.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

Recalling the definition of $\boldsymbol{U}$ as the stationary distribution associated with (1.1), letting $\boldsymbol{Z} \sim \mathrm{Dir}(\beta_1, \ldots, \beta_K)$, a bound for $|\mathbb{E}h(\boldsymbol{U}) - \mathbb{E}h(\boldsymbol{Z})|$ is derived by finding a bound for $|\mathbb{E}\mathcal{G}_{\boldsymbol{Z}} f_h(\boldsymbol{U})|$. The following (Stein) lemma formalises the link between $\mathcal{G}_{\boldsymbol{Z}}$ and $\boldsymbol{Z}$.

Lemma 1.1. *The random vector $\boldsymbol{Z} \sim \mathrm{Dir}(\beta_1, \ldots, \beta_K)$ if and only if, for all $f \in C^2(\bar{\nabla}^K)$ with bounded partial derivatives up to the second order and Lipschitz continuous second-order partial derivatives,*

$$\mathbb{E}\mathcal{G}_{\boldsymbol{Z}} f(\boldsymbol{Z}) = 0.$$

In [5] a variation of the generator method was innovated, namely the *prelimit generator comparison approach*, which is the approach we use in this paper. We briefly describe the general idea of the approach and how it compares with the traditional generator method. The full details of our approach are contained in the appendices. The traditional generator comparison approach works by noting that (1.7) yields

$$|\mathbb{E}h(X) - \mathbb{E}h(Z)| = |\mathbb{E}\mathcal{G}_Z f_h(X) - \mathbb{E}\mathcal{G}_X f_h(X)|,$$

where $\mathcal{G}_X$ is a characterising operator/generator for X, and hence, $\mathbb{E}\mathcal{G}_X f_h(X) = 0$. Typically, X will be a discrete object and Z will be its continuous limit. The generator comparison approach then follows the intuition that if X is approximately equal to Z then the operators $\mathcal{G}_X$ and $\mathcal{G}_Z$ should also be approximately equal, and the distributional distance between X and Z can be quantified by the differences in $\mathcal{G}_X$ and $\mathcal{G}_Z$. In a sense, this approach takes the discrete object X and evaluates it with respect to the continuous operator $\mathcal{G}_Z$. The prelimit generator comparison approach swaps the roles of the continuous and discrete terms.

Let $h : \delta\mathbb{Z}^d \mapsto \mathbb{R}$, where $\delta > 0$ is a test function defined on the lattice $\delta\mathbb{Z}^d$. For a random vector $\boldsymbol{U}$ that takes values on $\delta\mathbb{Z}^d$, suppose there is a characterising operator $\mathcal{G}_{\boldsymbol{U}}$ and given h one can find the solution to the Stein equation

$$\mathcal{G}_{\boldsymbol{U}} f_h(\boldsymbol{u}) = h(\boldsymbol{u}) - \mathbb{E}h(\boldsymbol{U}). \tag{1.9}$$

Then, naively speaking at least, for some continuous $\boldsymbol{Z}$ and its characterising operator $\mathcal{G}_{\boldsymbol{Z}}$,

$$|\mathbb{E}h(\boldsymbol{Z}) - \mathbb{E}h(\boldsymbol{U})| = |\mathbb{E}\mathcal{G}_{\boldsymbol{U}} f_h(\boldsymbol{Z}) - \mathbb{E}\mathcal{G}_{\boldsymbol{Z}} f_h(\boldsymbol{Z})|. \tag{1.10}$$

The general approach remains similar to the typical generator approach, the bound is reliant upon $\mathcal{G}_{\boldsymbol{U}}$ and $\mathcal{G}_{\boldsymbol{Z}}$ being close. One can now see however that in comparison to the standard generator comparison method, we are putting the continuous object $\boldsymbol{Z}$ into the discrete (state-space) generator for $\boldsymbol{U}$. This can be advantageous if the solution f_h to the Stein equation (1.9) with respect to the discrete generator is tractable. Unfortunately, (1.10) is not that straightforward as $\mathcal{G}_{\boldsymbol{U}}$ is only defined on the lattice $\delta\mathbb{Z}^d$, but we wish to input the continuous object $\boldsymbol{Z}$. To address this issue, a smoothing interpolation operator for $\mathcal{G}_{\boldsymbol{U}}$ is required, which will be described in detail in this paper.

Although the task of interpolating $\mathcal{G}_{\boldsymbol{U}}$ is conceptually straightforward, its execution is non-trivial. Thus, examples of the prelimit approach in practice are crucial for its proliferation. To date, only two examples exist. A simple single-server queue ($M/M/1$ model) is considered in [5], while [6] considers a much more involved queueing model—a load-balancing model under the join-the-shortest-queue policy. The former is a simple one-dimensional model that does not fully illustrate the challenges of interpolating $\mathcal{G}_{\boldsymbol{U}}$, while the latter example is extremely involved due to the complicated dynamics of the queueing model. The Wright–Fisher model

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

we consider falls nicely between these two examples in terms of difficulty—our model is multi-dimensional, highlighting all the challenges of interpolating $\mathcal{G}_{\boldsymbol{U}}$, whilst the task of bounding the Stein factors, which is unrelated to interpolation, is relatively straightforward. Additionally, this paper refines the original implementation of the prelimit generator approach in [5]. We present several results in Appendix A that simplify working with the interpolating operator. Specifically, we state and prove Lemma A.1, Lemma A.2, Proposition A.1, and Corollary A.1. We anticipate that these results will help future users of the approach.

The approach used in [13], follows the traditional generator method described earlier, whereas we use the prelimit generator comparison method in this paper. There are advantages and disadvantages to either approach. If we consider the main three steps in applying Stein's method discussed earlier, steps 1 and 3 are more or less the same in both approaches. Step 1 is in the traditional approach involves finding a diffusion operator that characterises the Dirichlet distribution. In the prelimit approach, in addition to the same operator, we also require an operator for the discrete population Wright–Fisher Markov chain stationary distribution, which is not difficult. Step 3 in the traditional approach in [13] uses an exchangeable pair coupling, and ultimately the main work involves a series of moment calculations for the discrete stationary distribution. In the prelimit approach, near identical calculations are required. The primary difference between the two approaches lies in step 2. In [13], solving the Stein equation and bounding the derivatives of the solution, known as the Stein factors, is a lengthy process and requires knowledge of coalescent theory and a dual process representation of the Wright–Fisher diffusion process governed by (1.8). These bounds are one of the primary contributions of [13]. In contrast, using the prelimit approach, we instead require Stein factors for the Markov chain associated with (1.1), which we bound in Lemma 2.3. The bounds are simple, elegant and require only a short proof using elementary Markov chain knowledge and couplings. The price we pay to use this simpler approach for the Stein factors is the requirement for an interpolation operator, which leads to numerous technical difficulties, and this is the main trade off between the two approaches.

Lastly, expanding on the two remarks we made following Theorem 1.1, the $N^{3-\beta_K/2}$ appears because we use an interpolation operator based on forward differences and, as a result, we have to treat the case when $\boldsymbol{Z}$ is close to the 'right' boundary of $\bar{\nabla}^K$ separately. Getting rid of the $N^{3-\beta_K/2}$ term would require modifying our interpolation operator to use forward and backward differences when close to the 'left' and 'right' boundaries of $\bar{\nabla}^K$, respectively, and central differences 'in the middle' to smoothly transition between the forward and backward differences. The C^K terms are an inevitable consequence of extending a function defined on the grid $\delta\mathbb{Z}^K$ to the continuum. For example, see Appendix A.1 where we prove inequalities (A.4) and (A.8) of Theorem A.1.

The remainder of the paper will be as follows. In Section 2 we outline the proof to Theorem 1.1, then provide a number of technical lemmas, and given these lemmas, we prove Theorem 1.1. The proofs of the technical lemmas are relegated to the two appendices.

2. Proof of the main theorem

2.1. Notation

1. For any function $f : \nabla^K \to \mathbb{R}$, and a nonnegative integer valued vector $\boldsymbol{a}$, let

$$B_i(f) = \sup\{|\Delta^{\boldsymbol{a}} f(\boldsymbol{u})| : \|\boldsymbol{a}\|_1 = i, \boldsymbol{u} \in \nabla^K, \text{ and } \boldsymbol{u} + \delta\boldsymbol{a} \in \nabla^K\}, \tag{2.1}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

where $\Delta^{\boldsymbol{a}}$ refers to the forwards difference operator with step size δ, and analogous to the definition of (1.4), $\boldsymbol{a}$ indicates in what directions the forward differences are taken. Note that given $\|\boldsymbol{a}\|_1 = i$, $\Delta^{\boldsymbol{a}}$ is the composition of i forward differences in the directions indicated by the entries of $\boldsymbol{a}$ and not a single forward difference in the direction of $\boldsymbol{a}$ treated as a whole. For example, if $\boldsymbol{a} = (1, 0, \ldots, 0)$ then $\Delta^{\boldsymbol{a}} f(\boldsymbol{u}) = f(\boldsymbol{u} + \delta \boldsymbol{e}_1) - f(\boldsymbol{u})$ and if $\boldsymbol{a} = (1, 1, 0, \ldots, 0)$ then $\Delta^{\boldsymbol{a}} f(\boldsymbol{u}) = f(\boldsymbol{u} + \delta \boldsymbol{e}_1 + \delta \boldsymbol{e}_2) - f(\boldsymbol{u} + \delta \boldsymbol{e}_1) - f(\boldsymbol{u} + \delta \boldsymbol{e}_2) + f(\delta \boldsymbol{u})$, where $\boldsymbol{e}_i$ denotes the usual standard basis vector with 1 in the ith component.

2. We reserve the variables $\boldsymbol{u}$ (and $\boldsymbol{U}$) to emphasise when a function is defined on the lattice ∇^K, and $\boldsymbol{x}$ (and $\boldsymbol{X}$) when the function is defined on the continuous simplex $\bar{\nabla}^K$.

3. The vector $\boldsymbol{e}$ is reserved to denote a K-dimensional vector of ones, that is, $\boldsymbol{e} = (1, 1, \ldots, 1)$. Furthermore, any inequalities with respect to $\boldsymbol{e}$ are intended to be element-by-element wise. That is, if $\boldsymbol{x} \leq \boldsymbol{e}$ then $x_i \leq e_i$ for all i.

4. We use $\Sigma = \sum_{i=1}^{K} p_i$ to denote the sum of the mutation probabilities in (1.1).

2.2. Outline of the proof

Recall that our goal is to bound $d_{\mathcal{M}_4}(\boldsymbol{U}, \boldsymbol{Z}) = \sup_{h \in \mathcal{M}_4} |\mathbb{E}h(\boldsymbol{U}) - \mathbb{E}h(\boldsymbol{Z})|$. We achieve this bound in three main steps.

1. *Solve the Stein equation:* identify a charaterising operator for $\mathcal{G}_{\boldsymbol{U}}$ for $\boldsymbol{U}$ and then, for any function $h \in \mathcal{M}_4$, solve for f_h that satisfies the Stein equation

$$\mathcal{G}_{\boldsymbol{U}} f_h(\boldsymbol{u}) = h(\boldsymbol{u}) - \mathbb{E}h(\boldsymbol{U}). \tag{2.2}$$

2. *The interpolation operator:* we would like to simply substitute $\boldsymbol{u} = \boldsymbol{Z}$ in (2.2) and take expectations, but $\mathcal{G}_{\boldsymbol{U}}$ is not well defined for continuous objects as it characterises $\boldsymbol{U}$ that is discrete. We therefore extend $f_h(\boldsymbol{u})$ to take arguments from $\bar{\nabla}^K$ using a carefully chosen interpolation operator A that satisfies $Af_h(\boldsymbol{x}) = f_h(\boldsymbol{u})$ for all $\boldsymbol{x} = \boldsymbol{u} \in \nabla^K$, and A applied to a constant equals that constant. Then by applying A again to (2.2),

$$A(\mathcal{G}_{\boldsymbol{U}}(Af_h))(\boldsymbol{x}) = Ah(\boldsymbol{x}) - \mathbb{E}h(\boldsymbol{U}). \tag{2.3}$$

3. *Generator expansion:* noting that $\mathbb{E}\mathcal{G}_{\boldsymbol{Z}} Af_h(\boldsymbol{Z}) = 0$, then by setting $\boldsymbol{x} = \boldsymbol{Z}$ in (2.3) and taking expectations,

$$\mathbb{E}[A\mathcal{G}_{\boldsymbol{U}} Af_h(\boldsymbol{Z}) - \mathcal{G}_{\boldsymbol{Z}} Af_h(\boldsymbol{Z})] = \mathbb{E}Ah(\boldsymbol{Z}) - \mathbb{E}h(\boldsymbol{U}). \tag{2.4}$$

We therefore need to carefully bound the left-hand side. We achieve this via Taylor expansion of $A\mathcal{G}_{\boldsymbol{U}} Af_h(\boldsymbol{x})$. The choice of A plays a crucial role here as we will have specifically chosen A in such a manner that the derivatives of Af_h correspond to the finite differences of f_h up to the fourth order.

2.2.1. *Solving the Stein equation* As the concept of a generator for a DTMC is relatively uncommon, for the benefit of readers, we spend some time to define the generator and the general form of the solution to its Stein equation.

Definition 2.1 *Let $\boldsymbol{U}(t)$ be the Wright–Fisher Markov chain with parent-independent mutation, rescaled to take values on $\nabla^K \subset \delta\mathbb{Z}^d$. For any function f from a suitable class of test functions $\mathcal{F}$, we define the Markov chain generator of this process $\mathcal{G}_{\boldsymbol{U}}$ as*

$$\mathcal{G}_{\boldsymbol{U}} f(\boldsymbol{u}) = \mathbb{E}[f(\boldsymbol{U}(1)) | \boldsymbol{U}(0) = \boldsymbol{u}] - \mathbb{E}f(\boldsymbol{u}), \quad \boldsymbol{u} \in \nabla^K.$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

Note that we will use subscript notation on generator operators, for example, $\mathcal{G}_{\boldsymbol{U}}$, to associate a generator with its stationary distribution $\boldsymbol{U}$.

Lemma 2.1. *Let $\boldsymbol{U}(t)$ denote a Markov chain governed by the generator $\mathcal{G}_{\boldsymbol{U}}$, then for all $h \in \mathcal{M}_4$, the function*

$$f_h(\boldsymbol{u}) = \sum_{t=0}^{\infty} \left[\mathbb{E}(h(\boldsymbol{U}(t))|\boldsymbol{U}(0) = \boldsymbol{u}) - \mathbb{E}h(\boldsymbol{U})\right], \quad \boldsymbol{u} \in \nabla^K \tag{2.5}$$

is well defined, and is the solution to

$$\mathcal{G}_{\boldsymbol{U}} f_h(\boldsymbol{u}) = h(\boldsymbol{u}) - \mathbb{E}h(\boldsymbol{U}), \quad \boldsymbol{u} \in \nabla^K.$$

Proof. This can be shown by adapting Lemma 2 of [5] to the discrete-time setting (see also Lemma 1 of [2]). □

2.2.2. *The interpolation operator* Our proof relies on the ability to extend any function $f : \delta\mathbb{Z}^d \to \mathbb{R}$ to $\mathbb{R}^d$ in a way that the derivatives of the extension correspond to the finite differences of f. Many such extensions are possible, but we use an interpolating seventh-order Hermite polynomial spline. The spline is a linear operator A acting on functions $f : \delta\mathbb{Z}^d \to \mathbb{R}$ and returning an extension $Af : \mathbb{R}^d \to \mathbb{R}$. When $d = 1$,

$$Af(x) = \sum_{i=0}^{4} \alpha_{k(x)+i}^{k(x)}(x) f(\delta(k(x) + i)), \quad x \in \mathbb{R}, \tag{2.6}$$

where $k(x) = \lfloor x/\delta \rfloor$ and $\alpha_{k+i}^{k} : \mathbb{R} \to \mathbb{R}$ are weights defined for all $k \in \mathbb{Z}$ and $i = 0, \ldots, 4$, making $A f(x)$ a weighted sum of the five points $f(\delta k(x)), \ldots, f(\delta(k(x) + 4))$. We use five points so that the derivatives of $A f(x)$ capture the finite differences of $f(x)$ up to the fourth order.

The details about $\alpha_{k+i}^{k}(x)$ and the definition of $A f(x)$ for $d > 1$ are left to Appendix A. For the purposes of this section, it suffices to know that A is a linear operator, $Af(\boldsymbol{x}) = f(\boldsymbol{u})$ for $\boldsymbol{x} = \boldsymbol{u} \in \delta\mathbb{Z}^d$, $Af(\boldsymbol{x})$ is twice continuously differentiable, and that A applied to a constant equals that constant.

2.2.3. *Generator expansion* We first define a discrete analog of $\mathcal{M}_j$. Let

$$\mathcal{M}_{\text{disc},j} = \left\{h : \delta\mathbb{Z}^{K-1} \to \mathbb{R},\ |\Delta^{\boldsymbol{a}} h(\delta\boldsymbol{k})| \le \delta^{\|\boldsymbol{a}\|_1},\ 1 \le \|\boldsymbol{a}\|_1 \le j,\ \delta\boldsymbol{k} \in \delta\mathbb{Z}^{K-1}\right\}, \quad j \ge 1.$$

The following lemma, which we prove in Appendix B, shows that the Taylor expansion of $A(\mathcal{G}_{\boldsymbol{U}}(Af_h))(\boldsymbol{x})$ equals $\mathcal{G}_{\boldsymbol{Z}} Af_h(\boldsymbol{x})$ plus an error term on a subset of $\bar{\nabla}^K$. To state it, we define

$$\nabla_N^K = \{\boldsymbol{u} \in \nabla^K : u_K \ge 10K/\sqrt{N}\} = \{\boldsymbol{u} \in \nabla^K : \sum_{i=1}^{K-1} u_i \le 1 - 10K/\sqrt{N}\}, \tag{2.7}$$

and let $\bar{\nabla}_N^K = \text{Conv}(\nabla_N^K)$. Note that $\nabla_N^K \neq \emptyset$ if $1 - 10K/\sqrt{N} > 0$, or $N > 100K^2$, which we assume going forward.

Lemma 2.2. *The extension $A(\mathcal{G}_{\boldsymbol{U}}(Af_h))(\boldsymbol{x})$ is well defined for all $\boldsymbol{x} \in \bar{\nabla}_N^K$. Furthermore, if the mutation probabilities satisfy (1.3) for some $\boldsymbol{\beta} > 0$ then*

$$A(\mathcal{G}_{\boldsymbol{U}}(Af_h))(\boldsymbol{x}) = \delta \mathcal{G}_{\boldsymbol{Z}} Af_h(\boldsymbol{x}) + \epsilon_{\mathcal{G}}(\boldsymbol{x}), \quad \boldsymbol{x} \in \bar{\nabla}_N^K,$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

where $\mathcal{G}_{\boldsymbol{Z}}Af_h(\boldsymbol{x})$ is defined in (1.8) and, for all $N > 0$,

$$\begin{aligned}|\epsilon_{\mathcal{G}}(\boldsymbol{x})| \le{}& K^2 2(1+s)^2 C_{A.4}(K)B_2(f_h)\\ &+ K\delta s C_{A.4}(K)\big(C_{A.10}(K)B_2(f_h) + C_{A.11}(K)\delta^{-1}B_4(f_h)\big)\\ &+ K^2(1+s)^2 C_{A.4}(K)\big(5C_{A.10}(K)B_3(f_h) + C_{A.11}(K)\delta^{-1}B_4(f_h)\big)\\ &+ K^3 1595^K 20(1+s)^3 C_{A.4}(K)\delta^{-1}B_3(f_h)\\ &+ 1595^K K^4 C_{A.4}(K)\bigg((2+s)^4\delta^{-2}B_4(f_h)\\ &\qquad + \frac{(4K+1)N^{4K}}{(\Sigma - p_K)^{4K}}\Big(1 - \frac{10K(\sqrt{N}-1)}{N}(1-\Sigma)\Big)^N \delta^{-5}16KB_1(f_h)\bigg).\end{aligned}$$

The final main ingredient needed to prove Theorem 1.1 are bounds for the Stein factors $B_i(f_h)$.

Lemma 2.3. *Let $f_h(\boldsymbol{u})$ be defined as in (2.5), then for all $h \in \mathcal{M}_{\mathrm{disc},4}$,*

$$B_i(f_h) \le \frac{\delta^i}{1-(1-\Sigma)^i}.$$

2.3. Proof of Theorem 1.1

We recall that to prove Theorem 1.1 we need to bound $d_{\mathcal{M}_4}(U, Y)$. Recall that $\delta = 1/N$. The following lemma relates $\mathcal{M}_j$ to $\mathcal{M}_{\mathrm{disc},j}$. We prove it in Appendix A after stating Proposition A.1.

Lemma 2.4. *There exists $C(d) > 0$ such that, for any $\boldsymbol{V} \in \delta\mathbb{Z}^d$ and $\boldsymbol{V}' \in \mathbb{R}^d$,*

$$d_{\mathcal{M}_4}(\boldsymbol{V}, \boldsymbol{V}') \le \sup_{h \in \mathcal{M}_{\mathrm{disc},4}} |\mathbb{E}h(\boldsymbol{V}) - \mathbb{E}Ah(\boldsymbol{V}')| + C_{A.11}(d)\delta^4, \tag{2.8}$$

where $C_{A.11}(d)$ is as in Proposition A.1.

We require the following auxiliary lemma, which is proved in Appendix A.

Lemma 2.5. *Assume that the mutation probabilities satisfy (1.3) for some $\boldsymbol{\beta} > 0$. Then for any $h \in \mathcal{M}_{\mathrm{disc},4}$ and $\boldsymbol{u} \in \nabla^K$,*

$$\begin{aligned}|h(\boldsymbol{u})| &\le |h(0)| + 1,\\ |Ah(\boldsymbol{Z})| &\le |h(0)| + C_{A.4}(K),\\ |\mathcal{G}_{\boldsymbol{Z}}Af_h(\boldsymbol{Z})| &\le C_{A.4}(K)(2+s)\delta^{-3}K^3B_1(f_h).\end{aligned}$$

We are now ready to prove the main theorem.

Proof of Theorem 1.1. Fix $h \in \mathcal{M}_{\mathrm{disc},4}$. As a consequence of Lemma 2.4, it suffices to bound $|\mathbb{E}h(\boldsymbol{U}) - \mathbb{E}Ah(\boldsymbol{Z})|$ to prove Theorem 1.1. Recall the Stein equation (1.9), or

$$\mathcal{G}_{\boldsymbol{U}}f_h(\boldsymbol{u}) = h(\boldsymbol{u}) - \mathbb{E}h(\boldsymbol{U}), \quad \boldsymbol{u} \in \nabla^K.$$

Define $f_h(\boldsymbol{u}) = 0$ for $\boldsymbol{u} \in \delta\mathbb{Z}^{K-1} \setminus \nabla^K$ (this choice is made for convenience), which allows us to define $Af_h(x)$ for all $\boldsymbol{x} \in \mathbb{R}^{K-1}$. Since $Af_h(\boldsymbol{x}) = f_h(\boldsymbol{u})$ for $\boldsymbol{x} = \boldsymbol{u} \in \nabla^K$, the Stein equation yields

$$\mathcal{G}_{\boldsymbol{U}}(Af_h)(\boldsymbol{u}) = h(\boldsymbol{u}) - \mathbb{E}h(\boldsymbol{U}), \quad \boldsymbol{u} \in \nabla^K.$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

Note that $\mathcal{G}_U(Af_h)(\boldsymbol{u})$ is only defined for $\boldsymbol{u} \in \nabla^K$ even though $Af_h(\boldsymbol{x})$ is defined on $\mathbb{R}^{K-1}$.

Applying A to $h(\boldsymbol{u}) - \mathbb{E}h(\boldsymbol{X})$ and recalling from Lemma 2.2 that $A(\mathcal{G}_U(Af_h))(\boldsymbol{x})$ is well defined for $\boldsymbol{x} \in \bar{\nabla}_N^K$, we see that, for any $\boldsymbol{x} \in \mathbb{R}^{K-1}$,

$$\begin{aligned} Ah(\boldsymbol{x}) - \mathbb{E}h(\boldsymbol{U}) &= A\big(\mathbb{E}h(\boldsymbol{U}) - h\big)(\boldsymbol{x})1(\boldsymbol{x} \in \bar{\nabla}_N^K) + A\big(\mathbb{E}h(\boldsymbol{U}) - h\big)(\boldsymbol{x})1(\boldsymbol{x} \notin \bar{\nabla}_N^K) \\ &= A(\mathcal{G}_U(Af_h))(\boldsymbol{x})1(\boldsymbol{x} \in \bar{\nabla}_N^K) + \big(\mathbb{E}h(\boldsymbol{U}) - Ah(\boldsymbol{x})\big)1(\boldsymbol{x} \notin \bar{\nabla}_N^K). \end{aligned}$$

We claim that $Af_h(\boldsymbol{x})$ satisfies the conditions of Lemma 1.1, implying that $\mathbb{E}\mathcal{G}_{\boldsymbol{Z}}Af_h(\boldsymbol{Z}) = 0$. Our claim holds because $Af_h(\boldsymbol{x})$ is twice continuously differentiable, and $Af_h(\boldsymbol{x})$ has compact support, which follows by Theorem A.1 of Appendix A since $f_h(\boldsymbol{u}) = 0$ for $\boldsymbol{u} \in \delta\mathbb{Z}^{K-1} \setminus \nabla^K$. Setting $\boldsymbol{x} = \boldsymbol{Z}$ and taking expected values yields

$$\begin{aligned} \mathbb{E}Ah(\boldsymbol{Z}) - \mathbb{E}h(\boldsymbol{U}) &= \mathbb{E}\big(A(\mathcal{G}_U(Af_h))(\boldsymbol{Z}) - \delta\mathcal{G}_{\boldsymbol{Z}}Af_h(\boldsymbol{Z})\big)1(\boldsymbol{Z} \in \bar{\nabla}_N^K) \\ &\quad + \mathbb{E}\big(\mathbb{E}h(\boldsymbol{U}) - Ah(\boldsymbol{Z}) - \delta\mathcal{G}_{\boldsymbol{Z}}Af_h(\boldsymbol{Z})\big)1(\boldsymbol{Z} \notin \bar{\nabla}_N^K) \\ &= \mathbb{E}\epsilon_{\mathcal{G}}(\boldsymbol{Z})1(\boldsymbol{Z} \in \bar{\nabla}_N^K) + \mathbb{E}\big(\mathbb{E}h(\boldsymbol{U}) - Ah(\boldsymbol{Z}) - \delta\mathcal{G}_{\boldsymbol{Z}}Af_h(\boldsymbol{Z})\big)1(\boldsymbol{Z} \notin \bar{\nabla}_N^K). \end{aligned}$$

To bound the first term, we recall assumption (1.3), which implies that $\Sigma = \delta s/2$. Combining Lemma 2.2 with the Stein factor bounds in Lemma 2.3 then yields

$$\begin{aligned} |\epsilon_{\mathcal{G}}(\boldsymbol{x})| &\le K^2 2(1+s)^2 C_{A.4}(K)\frac{\delta^2}{1-(1-\Sigma)^2} \\ &\quad + K\delta s C_{A.4}(K)\Big(C_{A.10}(K)\frac{\delta^2}{1-(1-\Sigma)^2} + C_{A.11}(K)\frac{\delta^3}{1-(1-\Sigma)^4}\Big) \\ &\quad + K^2(1+s)^2 C_{A.4}(K)\Big(5C_{A.10}(K)\frac{\delta^3}{1-(1-\Sigma)^3} + C_{A.11}(K)\frac{\delta^3}{1-(1-\Sigma)^4}\Big) \\ &\quad + K^3 1595^K 20(1+s)^3 C_{A.4}(K)\frac{\delta^2}{1-(1-\Sigma)^3} \\ &\quad + 1595^K K^4 C_{A.4}(K)\Big((2+s)^4\frac{\delta^2}{1-(1-\Sigma)^4} \\ &\qquad + \frac{(4K+1)N^{4K}}{(\Sigma - p_K)^{4K}}\Big(1 - \frac{10K(\sqrt{N}-1)}{N}(1-\Sigma)\Big)^N \delta^{-5} 16K\Sigma^{-1}\Big). \end{aligned}$$

Note that $(1-(10K(\sqrt{N}-1)/N)(1-\Sigma))^N$ behaves like $\mathrm{e}^{-10K\sqrt{N}}$ for large N. Let us bound the second term. Recall that $\boldsymbol{Z}$ has density given by (1.2), which implies that $Z_K \sim \mathrm{Beta}(\beta_K, s-\beta_K)$. Therefore,

$$\begin{aligned} \mathbb{P}(\boldsymbol{Z} \notin \bar{\nabla}_N^K) &= \mathbb{P}(Z_K \le 10K/\sqrt{N}) \\ &= \int_0^{10K/\sqrt{N}} \frac{\Gamma(s)}{\Gamma(s-\beta_K)\Gamma(\beta_K)}(1-x_K)^{s-\beta_K-1}x_K^{\beta_K-1}dx_K \\ &\le \frac{\Gamma(s)}{\Gamma(s-\beta_K)\Gamma(\beta_K)}\frac{1}{\beta_K}\Big(\frac{10K}{\sqrt{N}}\Big)^{\beta_K}\Big(1 + \frac{1}{(1-1/\sqrt{N})^{|s-\beta_K-1|}}\Big) \\ &\le C_{2.10}\delta^{\beta_K/2}, \end{aligned} \tag{2.9}$$

where

$$C_{2.10} = \frac{\Gamma(s)}{\Gamma(s-\beta_K)\Gamma(\beta_K)}\frac{(10K)^{\beta_K}}{\beta_K}\Big(1 + \frac{1}{(1-1/\sqrt{N})^{|s-\beta_K-1|}}\Big). \tag{2.10}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

Without loss of generality, we may assume that $h(0) = 0$. Otherwise, we can replace $h(u)$ by $h(u) - h(0)$ without affecting the value of $\mathbb{E}h(U) - \mathbb{E}Ah(Y)$. Observe that

$$\begin{aligned}
&\mathbb{E}\big(\big|\mathbb{E}h(\boldsymbol{U}) - Ah(\boldsymbol{Z}) - \delta\mathcal{G}_{\boldsymbol{Z}}Af_h(\boldsymbol{Z})\big|1(\boldsymbol{Z} \notin \bar{\nabla}_N^K)\big) \\
&\quad \le \mathbb{E}\Big(\big(\mathbb{E}|h(\boldsymbol{U})| + |Ah(\boldsymbol{Z})| + \delta|\mathcal{G}_{\boldsymbol{Z}}Af_h(\boldsymbol{Z})|\big)1(\boldsymbol{Z} \notin \bar{\nabla}_N^K)\big)\Big) \\
&\quad \le \big(1 + C_{A.4}(K) + C_{A.4}(K)(2+s)\delta^{-3}K^3B_1(f_h)\big)\mathbb{P}(\boldsymbol{Z} \notin \bar{\nabla}_N^K) \\
&\quad \le \big(1 + C_{A.4}(K) + C_{A.4}(K)(2+s)K^3 2/s\big)\delta^{-3}C_{2.10}\delta^{\beta_K/2}.
\end{aligned}$$

In the second inequality we used Lemma 2.5. In the last inequality we used (2.9), the fact that $\delta < 1$, the Stein factor bound $B_1(f_h) \le \delta/\Sigma$ from Lemma 2.3, and the fact that $\delta/\Sigma = 2/s$, which follows from (1.3) and the definitions of s and Σ.

The final bound now follows from choosing $C(K)$ to be a uniform upper bound for $C_{A.4}(K)$, $C_{A.10}(K)$, and $C_{A.11}(K)$ and some routine simplification and rearrangement. □

Appendix A. The interpolator A

The operator A discussed in this appendix is identical to the one introduced in Appendix A of [5]. We repeat some of its key properties, originally presented in [5], as they are needed for the proof of Theorem 1.1. We also present several useful technical results about A that are not found in [5]. Namely, Lemma A.1, Lemma A.2, Proposition A.1, and Corollary A.1.

Building on the discussion in Section 2.2.2, for a one-dimensional function $f : \delta\mathbb{Z} \to \mathbb{R}$, we define

$$Af(x) = \sum_{i=0}^{4} \alpha_{k(x)+i}^{k(x)}(x) f(\delta(k(x) + i)),$$

where $k(x) = \lfloor x/\delta \rfloor$ and $\alpha_{k+i}^{k} : \mathbb{R} \to \mathbb{R}$ are weights. Braverman [5] described how to choose these weights to make $Af(x)$ coincide with $f(\cdot)$ on grid points, and also to make it a differentiable function whose derivatives behave like the corresponding finite differences of $f(\cdot)$. The idea can be applied to multidimensional grid-valued functions as well. The following result builds on Theorem A.1 of [5]. We use this as an interface that contains the important properties of A without delving into the low-level details behind its construction.

Theorem A.1. *Given a convex set $K \subset \mathbb{R}^d$, define*

$$K_4 = \{\boldsymbol{x} \in K \cap \delta\mathbb{Z}^d : \delta(k(\boldsymbol{x}) + \boldsymbol{i}) \in K \cap \delta\mathbb{Z}^d \text{ for all } 0 \le \boldsymbol{i} \le 4\boldsymbol{e}\},$$

where $k(\boldsymbol{x})$ by $k_j(\boldsymbol{x}) = \lfloor x_j/\delta \rfloor$. Let $\mathrm{Conv}(K_4)$ be the convex hull of K_4 and for $\boldsymbol{x} \in \mathbb{R}^d$. There exist weights $\{\alpha_{k+i}^{k} : \mathbb{R} \to \mathbb{R},\ k \in \mathbb{Z},\ i = 0, 1, 2, 3, 4\}$ such that, for any $f : K \cap \delta\mathbb{Z}^d \to \mathbb{R}$, the function

$$\begin{aligned}
Af(\boldsymbol{x}) &= \sum_{i_d=0}^{4} \alpha_{k_d(\boldsymbol{x})+i_d}^{k_d(\boldsymbol{x})}(x_d) \cdots \sum_{i_1=0}^{4} \alpha_{k_1(\boldsymbol{x})+i_1}^{k_1(\boldsymbol{x})}(x_1) f(\delta(k(\boldsymbol{x}) + \boldsymbol{i})) \\
&= \sum_{i_1,\ldots,i_d=0}^{4} \Big(\prod_{j=1}^{d} \alpha_{k_j(\boldsymbol{x})+i_j}^{k_j(\boldsymbol{x})}(x_j)\Big) f(\delta(k(\boldsymbol{x}) + \boldsymbol{i})), \quad \boldsymbol{x} \in \mathrm{Conv}(K_4)
\end{aligned} \tag{A.1}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

satisfies $Af(\boldsymbol{x}) \in C^3(\mathrm{Conv}(K_4))$, *where* $\boldsymbol{i} = (i_1, \ldots, i_d)$ *in (A.1). Additionally,* $Af(\boldsymbol{x})$ *is infinitely differentiable almost everywhere on* $\mathrm{Conv}(K_4)$,

$$Af(\delta \boldsymbol{k}) = f(\delta \boldsymbol{k}), \quad \delta \boldsymbol{k} \in K_4, \tag{A.2}$$

and for $0 \le \|\boldsymbol{a}\|_1 \le 3$,

$$\left|D^{\boldsymbol{a}} Af(\boldsymbol{x})\right| \le C_{A.4}(d)\delta^{-\|\boldsymbol{a}\|_1} \max_{0 \le \boldsymbol{i} \le 4\boldsymbol{e}-\boldsymbol{a}} |\Delta^{\boldsymbol{a}} f(\delta(k(\boldsymbol{x}) + \boldsymbol{i}))|, \quad \boldsymbol{x} \in \mathrm{Conv}(K_4), \tag{A.3}$$

where

$$C_{A.4}(d) \le 11404^5 1595^d. \tag{A.4}$$

When $\|\boldsymbol{a}\|_1 = 4$, *(A.3) also holds for almost all* $\boldsymbol{x} \in \mathrm{Conv}(K_4)$. *Additionally, the weights* $\{\alpha^k_{k+i} : \mathbb{R} \to \mathbb{R},\ k \in \mathbb{Z},\ i = 0, 1, 2, 3, 4\}$ *are degree-7 polynomials in* $(x - \delta k)/\delta$ *whose coefficients do not depend on* k *or* δ. *For any* $x \in \mathbb{R}$, *they satisfy*

$$\alpha^k_k(\delta k) = 1 \quad \text{and} \quad \alpha^k_{k+i}(\delta k) = 0, \quad k \in \mathbb{Z},\ i = 1, 2, 3, 4, \tag{A.5}$$

$$\sum_{i=0}^{4} \alpha^k_{k+i}(x) = 1, \quad k \in \mathbb{Z}, \tag{A.6}$$

$$\alpha^{k+j}_{k+j+i}(x + \delta j) = \alpha^k_{k+i}(x), \quad i, j, k \in \mathbb{Z}, \tag{A.7}$$

$$|\alpha^{k(x)}_{k(x)+i}(x)| \le 319. \tag{A.8}$$

We prove (A.4) and (A.8) in Appendix A.1. The rest of Theorem A.1 is identical to Theorem A.1 of [5]. We now present three useful properties of A. While Theorem A.1 only guarantees that $Af(\boldsymbol{x})$ is thrice continuously differentiable, we need to control the fourth-order remainder term in the Taylor expansion of $Af(\boldsymbol{x})$. The following lemma, which would have been trivial if $Af(\boldsymbol{x})$ were four-times continuously differentiable, helps with this. Define

$$\mathcal{D}^d = \left\{ f : \mathbb{R}^d \to \mathbb{R} :\ \text{for all } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d,\ |f(\boldsymbol{x}) - f(\boldsymbol{y})| \le \|\boldsymbol{x} - \boldsymbol{y}\|_1 \sup_{\substack{\min(\boldsymbol{x},\boldsymbol{y}) \le z \le \max(\boldsymbol{x},\boldsymbol{y}) \\ \|\boldsymbol{a}\|_1 = 1}} |D^{\boldsymbol{a}} f(z)| \right\}, \tag{A.9}$$

where the minimum and maximum are taken elementwise.

Lemma A.1. *For any* $f : \delta\mathbb{Z}^d \to \mathbb{R}$, *let* $Af(\boldsymbol{x})$ *be as defined in (A.1). Then* $D^{\boldsymbol{a}} Af(\boldsymbol{x}) \in \mathcal{D}^d$ *for any* $\boldsymbol{a} > 0$ *with* $\|\boldsymbol{a}\|_1 = 3$.

The second lemma is a useful identity for applying A to products of functions.

Lemma A.2. *Given* $f, g : \mathbb{Z}^d \to \mathbb{R}$ *let* $h(\boldsymbol{u}) = f(\boldsymbol{u})g(\boldsymbol{u})$. *There exists* $\epsilon_p : \mathbb{R}^d \to \mathbb{R}$ *and a constant* $C(d)$ *such that*

$$\begin{aligned} Ah(\boldsymbol{x}) &= Af(\boldsymbol{x})Ag(\boldsymbol{x}) + \epsilon_p(\boldsymbol{x}), \\ |\epsilon_p(\boldsymbol{x})| &\le C_{A.10}(d) \max_{\substack{\|\boldsymbol{a}\|_1 = 1 \\ 0 \le i \le 4\boldsymbol{e}-\boldsymbol{a}}} |\Delta^{\boldsymbol{a}} g(\delta(k(\boldsymbol{x}) + \boldsymbol{i}))| \max_{\substack{\|\boldsymbol{a}\|_1 = 1 \\ 0 \le i \le 4\boldsymbol{e}-\boldsymbol{a}}} |\Delta^{\boldsymbol{a}} f(\delta(k(\boldsymbol{x}) + \boldsymbol{i}))|, \\ C_{A.10}(d) &\le 7975^d (5^d + d11404^5 1595^d). \end{aligned} \tag{A.10}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

For our third result, let $f(\boldsymbol{x})$ be a function defined for all $\boldsymbol{x} \in \mathbb{R}^d$ and let $f(\boldsymbol{u})$ denote its restriction to $\delta\mathbb{Z}^d$. Proposition A.1 provides an upper bound on how well $Af(\boldsymbol{x})$ approximates $f(\boldsymbol{x})$. The smoother the function $f(\boldsymbol{x})$, the higher the exponent of δ in the error bound.

Proposition A.1. *Suppose that $f \in C^{s-1}(\mathbb{R}^d)$ for some $s \in \{1, 2, 3, 4\}$ and that $D^{\boldsymbol{a}} f(\boldsymbol{x}) \in \mathcal{D}^d$ when $\|\boldsymbol{a}\|_1 = s - 1$. Then*

$$|f(\boldsymbol{x}) - Af(\boldsymbol{x})| \le C_{A.11}(d)\delta^s \max_{\|\boldsymbol{a}\|_1 = s} \sup_{0 \le z \le 5\delta \boldsymbol{e}} |D^{\boldsymbol{a}} f(\delta k(\boldsymbol{x}) + z)|, \quad \boldsymbol{x} \in \mathbb{R}^d,$$

$$C_{A.11}(d) \le d^4(4 \times 12^d + 11404^5 1595^d + 1). \tag{A.11}$$

The following corollary plays an important role later on in the proof of Lemma B.2.

Corollary A.1. *If $f : \mathbb{R}^d \to \mathbb{R}$ is a polynomial of degree at most three then $Af(\boldsymbol{x}) = f(\boldsymbol{x})$.*

Proof of Corollary A.1. If $f(x)$ is a polynomial up to the third order, then its fourth-order derivatives are zero. The result follows from applying Proposition A.1 with $s = 4$. □

We now prove Lemmas A.1 and A.2, and Proposition A.1, in that order.

Proof of Lemma A.1. Given $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, define

$$\boldsymbol{x}^{(j)} = (x_1, \ldots, x_{j-1}, y_j, \ldots, y_d), \quad 1 \le j \le d,$$

and note that $\boldsymbol{x}^{(1)} = \boldsymbol{y}$ and $\boldsymbol{x}^{(d)} = \boldsymbol{x}$. Fix $\boldsymbol{a} \in \mathbb{Z}^d$ with $\boldsymbol{a} > 0$ and $\|\boldsymbol{a}\|_1 = 3$. Then

$$D^{\boldsymbol{a}} Af(\boldsymbol{x}) - D^{\boldsymbol{a}} Af(\boldsymbol{y}) = D^{\boldsymbol{a}} Af(\boldsymbol{x}^{(d)}) - D^{\boldsymbol{a}} Af(\boldsymbol{x}^{(1)}) = \sum_{j=1}^{d-1} D^{\boldsymbol{a}} Af(\boldsymbol{x}^{(j+1)}) - D^{\boldsymbol{a}} Af(\boldsymbol{x}^{(j)}).$$

We now show that

$$|D^{\boldsymbol{a}} Af(\boldsymbol{x}^{(j+1)}) - D^{\boldsymbol{a}} Af(\boldsymbol{x}^{(j)})| \le |\boldsymbol{x}_j - \boldsymbol{y}_j| \sup_{\substack{\min(\boldsymbol{x},\boldsymbol{y}) \le z \\ z \le \max(\boldsymbol{x},\boldsymbol{y})}} |D^{\boldsymbol{a}} Af(z)|.$$

Suppose that $j = d - 1$; the argument is similar for other values of j. Note that

$$\begin{aligned}
&D^{\boldsymbol{a}} Af(\boldsymbol{x}) \\
&= \partial_d^{a_d} \cdots \partial_1^{a_1} \left(\sum_{i_d=0}^{4} \alpha_{k_d(\boldsymbol{x})+i_d}^{k_d(\boldsymbol{x})}(x_d) \cdots \sum_{i_1=0}^{4} \alpha_{k_1(\boldsymbol{x})+i_1}^{k_1(\boldsymbol{x})}(x_1) f(\delta(k(\boldsymbol{x}) + \boldsymbol{i})) \right) \\
&= \partial_d^{a_d} \sum_{i_d=0}^{4} \alpha_{k_d(x_d)+i_d}^{k_d(x_d)}(x_d) \\
&\qquad \times \left[\partial_{d-1}^{a_{d-1}} \left(\sum_{i_{d-1}=0}^{4} \alpha_{k_{d-1}(x_{d-1})+i_{d-1}}^{k_{d-1}(x_{d-1})}(x_{d-1}) \cdots \partial_1^{a_1} \left(\sum_{i_1=0}^{4} \alpha_{k_1(x_1)+i_1}^{k_1(x_1)}(x_1) f(\delta(k(\boldsymbol{x}) + \boldsymbol{i})) \right) \right) \right],
\end{aligned}$$

where the first equality follows the definition of $Af(\boldsymbol{x})$ in (A.1). Treating $x_1, \ldots, x_{d-1}$ as fixed, let us consider the term inside the square brackets as a one-dimensional function in the dth dimension. To be precise, we define $g_{x_1,\ldots,x_{d-1}} : \delta\mathbb{Z} \to \mathbb{R}$ as

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

$$
\begin{aligned}
&g_{x_1,\dots,x_{d-1}}(\delta\ell)\\
&\quad=\partial_{d-1}^{a_{d-1}}\Bigg(\sum_{i_{d-1}=0}^{4}\alpha_{k_{d-1}(x_{d-1})+i_{d-1}}^{k_{d-1}(x_{d-1})}(x_{d-1})\cdots\partial_1^{a_1}\Bigg(\sum_{i_1=0}^{4}\alpha_{k_1(x_1)+i_1}^{k_1(x_1)}(x_1)\\
&\qquad\qquad\times f\big(\delta(k(x_1,\dots,x_{d-1},0)+(i_1,\dots,i_{d-1},0))+\delta\ell\big)\Bigg)\Bigg),\qquad \ell\in\mathbb{Z}.
\end{aligned}
$$

Then

$$
D^{\boldsymbol{a}}Af(\boldsymbol{x})=\partial_d^{a_d}\sum_{i_d=0}^{4}\alpha_{k_d(x_d)+i_d}^{k_d(x_d)}(x_d)g_{x_1,\dots,x_{d-1}}(\delta(k_d(x_d)+i_d))=\partial^{a_d}Ag_{x_1,\dots,x_{d-1}}(x_d).
$$

Theorem A.1 states that $Ag_{x_1,\dots,x_{d-1}}(x_d)$ is infinitely differentiable almost everywhere. Thus,

$$
\begin{aligned}
&|D^{\boldsymbol{a}}Af(\boldsymbol{x}^{(d)})-D^{\boldsymbol{a}}Af(\boldsymbol{x}^{(d-1)})|\\
&\quad=|\partial^{a_d}Ag_{x_1,\dots,x_{d-1}}(x_d)-\partial^{a_d}Ag_{x_1,\dots,x_{d-1}}(y_d)|\\
&\quad=|\int_{y_d}^{x_d}\partial^{a_d+1}Ag_{x_1,\dots,x_{d-1}}(x')dx'|\\
&\quad\le|x_d-y_d|\sup_{(x_d\wedge y_d)\le z_d\le(x_d\vee y_d)}|\partial^{a_d+1}Ag_{x_1,\dots,x_{d-1}}(z_d)|\\
&\quad\le|x_d-y_d|\max_{\|\boldsymbol{a}'\|_1=\|\boldsymbol{a}\|_1+1}\sup_{\substack{\min(\boldsymbol{x},\boldsymbol{y})\le z\\ z\le\max(\boldsymbol{x},\boldsymbol{y})}}|D^{\boldsymbol{a}'}Af(z)|,
\end{aligned}
$$

where in the last inequality we used $\partial^{a_d+1}Ag_{x_1,\dots,x_{d-1}}(x_d)=\partial_dD^{\boldsymbol{a}}Af(\boldsymbol{x})$. □

Proof of Lemma A.2. Using (A.1) of Theorem A.1, for $\boldsymbol{x}\in\mathbb{R}^d$,

$$
\begin{aligned}
Ah(\boldsymbol{x})&=\sum_{i_1,\dots,i_d=0}^{4}\prod_{j=1}^{d}\alpha_{k_j(\boldsymbol{x})+i_j}^{k_j(\boldsymbol{x})}(x_j)f(\delta(k(\boldsymbol{x})+\boldsymbol{i}))g(\delta(k(\boldsymbol{x})+\boldsymbol{i}))\\
&=\sum_{i_1,\dots,i_d=0}^{4}\prod_{j=1}^{d}\alpha_{k_j(\boldsymbol{x})+i_j}^{k_j(\boldsymbol{x})}(x_j)Af(\boldsymbol{x})g(\delta(k(\boldsymbol{x})+\boldsymbol{i}))\\
&\quad+\sum_{i_1,\dots,i_d=0}^{4}\prod_{j=1}^{d}\alpha_{k_j(\boldsymbol{x})+i_j}^{k_j(\boldsymbol{x})}(x_j)\big(f(\delta(k(\boldsymbol{x})+\boldsymbol{i}))-Af(\boldsymbol{x})\big)g(\delta(k(\boldsymbol{x})+\boldsymbol{i})).
\end{aligned}
$$

By the definition of A,

$$
\sum_{i_1,\dots,i_d=0}^{4}\prod_{j=1}^{d}\alpha_{k_j(\boldsymbol{x})+i_j}^{k_j(\boldsymbol{x})}(x_j)Af(\boldsymbol{x})g(\delta(k(\boldsymbol{x})+\boldsymbol{i}))=Af(\boldsymbol{x})Ag(\boldsymbol{x}).
$$

Furthermore,

$$
\begin{aligned}
&\sum_{i_1,\dots,i_d=0}^{4}\prod_{j=1}^{d}\alpha_{k_j(\boldsymbol{x})+i_j}^{k_j(\boldsymbol{x})}(x_j)\big(f(\delta(k(\boldsymbol{x})+\boldsymbol{i}))-Af(\boldsymbol{x})\big)g(\delta(k(\boldsymbol{x})+\boldsymbol{i}))\\
&\quad=\sum_{i_1,\dots,i_d=0}^{4}\prod_{j=1}^{d}\alpha_{k_j(\boldsymbol{x})+i_j}^{k_j(\boldsymbol{x})}(x_j)\big(f(\delta(k(\boldsymbol{x})+\boldsymbol{i}))-Af(\boldsymbol{x})\big)\big(g(\delta(k(\boldsymbol{x})+\boldsymbol{i}))-g(\delta k(\boldsymbol{x}))\big),
\end{aligned}
$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

because again, by the definition of A,

$$\sum_{i_1,\dots,i_d=0}^{4}\prod_{j=1}^{d}\alpha_{k_j(\boldsymbol{x})+i_j}^{k_j(\boldsymbol{x})}(x_j)\big(f(\delta(k(\boldsymbol{x})+\boldsymbol{i}))-Af(\boldsymbol{x})\big)g(\delta k(\boldsymbol{x}))=g(\delta k(\boldsymbol{x}))(Af(\boldsymbol{x})-Af(\boldsymbol{x}))=0.$$

Setting

$$\epsilon_p(\boldsymbol{x})=\sum_{i_1,\dots,i_d=0}^{4}\prod_{j=1}^{d}\alpha_{k_j(\boldsymbol{x})+i_j}^{k_j(\boldsymbol{x})}(x_j)\big(f(\delta(k(\boldsymbol{x})+\boldsymbol{i}))-Af(\boldsymbol{x})\big)\big(g(\delta(k(\boldsymbol{x})+\boldsymbol{i}))-g(\delta k(\boldsymbol{x}))\big),$$

we have shown that $Ah(\boldsymbol{x})=Af(\boldsymbol{x})Ag(\boldsymbol{x})+\epsilon_p(\boldsymbol{x})$. Since

$$|\alpha_{k_j(\boldsymbol{x})+i_j}^{k_j(\boldsymbol{x})}(x_j)|\le 319$$

by (A.8), it follows that

$$|\epsilon_p(\boldsymbol{x})|\le 5^d 319^d\max_{0\le\boldsymbol{i}\le 4\boldsymbol{e}}\big(|f(\delta(k(\boldsymbol{x})+\boldsymbol{i}))-Af(\boldsymbol{x})||g(\delta(k(\boldsymbol{x})+\boldsymbol{i}))-g(\delta k(\boldsymbol{x}))|\big).$$

To conclude the proof, we argue that

$$|g(\delta(k(\boldsymbol{x})+\boldsymbol{i}))-g(\delta k(\boldsymbol{x}))|\le 5^d\max_{\substack{\|\boldsymbol{a}\|_1=1\\0\le\boldsymbol{j}\le 4\boldsymbol{e}-\boldsymbol{a}}}|\Delta^{\boldsymbol{a}}g(\delta(k(\boldsymbol{x})+\boldsymbol{j})|,$$

$$|f(\delta(k(\boldsymbol{x})+\boldsymbol{i}))-Af(\boldsymbol{x})|\le(5^d+dC_{A.4}(d))\max_{\substack{\|\boldsymbol{a}\|_1=1\\0\le j\le 4\boldsymbol{e}-\boldsymbol{a}}}|\Delta^{\boldsymbol{a}}f(\delta(k(\boldsymbol{x})+\boldsymbol{j})|,$$

which, when combined with the bound on $C_{A.4}(d)$ in (A.4), yields $C_{A.10}(d)\le 7975^d(5^d+dC_{A.4}(d))\le 7975^d(5^d+d11404^5 1595^d)$. The first inequality follows by rewriting $g(\delta(k(\boldsymbol{x})+\boldsymbol{i}))-g(\delta k(\boldsymbol{x}))$ as a telescoping series and using the fact that $\boldsymbol{i}\le 4\boldsymbol{e}$. Similarly,

$$\begin{aligned}|f(\delta(k(\boldsymbol{x})+\boldsymbol{i}))-Af(\boldsymbol{x})|&\le|f(\delta(k(\boldsymbol{x})+\boldsymbol{i}))-f(\delta k(\boldsymbol{x}))|+|f(\delta k(\boldsymbol{x}))-Af(\boldsymbol{x})|\\&\le 5^d\max_{\substack{\|\boldsymbol{a}\|_1=1\\0\le\boldsymbol{j}\le 4\boldsymbol{e}-\boldsymbol{a}}}|\Delta^{\boldsymbol{a}}f(\delta(k(\boldsymbol{x})+\boldsymbol{j})|+|f(\delta k(\boldsymbol{x}))-Af(\boldsymbol{x})|,\end{aligned}$$

and

$$\begin{aligned}|f(\delta k(\boldsymbol{x}))-Af(\boldsymbol{x})|=|Af(\delta k(\boldsymbol{x}))-Af(\boldsymbol{x})|&\le\|\boldsymbol{x}-\delta k(\boldsymbol{x})\|_1\sup_{\substack{z\in[\delta k(\boldsymbol{x}),\boldsymbol{x}]\\\|\boldsymbol{a}\|_1=1}}|D^{\boldsymbol{a}}Af(z)|\\&\le\delta d\sup_{\substack{z\in[\delta k(\boldsymbol{x}),\boldsymbol{x}]\\\|\boldsymbol{a}\|_1=1}}|D^{\boldsymbol{a}}Af(z)|\\&\le dC_{A.4}(d)\max_{\substack{\|\boldsymbol{a}\|_1=1\\0\le i\le 4\boldsymbol{e}-\boldsymbol{a}}}|\Delta^{\boldsymbol{a}}f(\delta(k(\boldsymbol{x})+\boldsymbol{i}))|,\end{aligned}$$

where the last inequality is due to (A.3) of Theorem A.1. □

Fix a function $f(\boldsymbol{x})$ satisfying the conditions of Proposition A.1. The following is a key auxiliary lemma, which we prove after proving Proposition A.1.

Lemma A.3. *For any $\delta<1$, $s\in\{1,2,3,4\}$, and $1\le\|\boldsymbol{a}\|_1\le s-1$,*

$$D^{\boldsymbol{a}}Af(\delta k(\boldsymbol{x}))=D^{\boldsymbol{a}}f(\delta k(\boldsymbol{x}))+12^d\delta^{s-\|\boldsymbol{a}\|_1}E^{(s)}\big([\delta k(\boldsymbol{x}),\delta(k(\boldsymbol{x})+3\boldsymbol{e})]\big),\quad\boldsymbol{x}\in\mathbb{R}^d,$$

where, for any $A\subset\mathbb{R}^d$, $E^{(s)}(A)$ denotes a quantity that satisfies

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

$$|E^{(s)}(A)| \le \max_{\|\boldsymbol{a}\|_1=s} \sup_{z\in A} |D^{\boldsymbol{a}} f(z)|. \tag{A.12}$$

Proof of Proposition A.1. Throughout the proof we let $E^{(s)}(A)$ denote a generic quantity that satisfies (A.12), but whose value may change from line to line. Expanding both $f(\boldsymbol{x})$ and $Af(\boldsymbol{x})$ around $\delta k(\boldsymbol{x})$ yields

$$\begin{aligned} Af(\boldsymbol{x}) - f(\boldsymbol{x}) = &\sum_{j=1}^{s-1} \frac{1}{j!} \sum_{\boldsymbol{a}\,:\,\|\boldsymbol{a}\|_1=j} \left(\prod_{i=1}^{d} (x_i - \delta k_i(\boldsymbol{x}))^{a_i} \right) \left(D^{\boldsymbol{a}} Af(\delta k(\boldsymbol{x})) - D^{\boldsymbol{a}} f(\delta k(\boldsymbol{x})) \right) \\ &+ \frac{1}{(s-1)!} \sum_{\boldsymbol{a}\,:\,\|\boldsymbol{a}\|_1=s-1} \left(\prod_{i=1}^{d} (x_i - \delta k_i(\boldsymbol{x}))^{a_i} \right) \left(D^{\boldsymbol{a}} Af(\boldsymbol{\xi}_1) - D^{\boldsymbol{a}} Af(\delta k(\boldsymbol{x})) \right) \\ &+ \frac{1}{(s-1)!} \sum_{\boldsymbol{a}\,:\,\|\boldsymbol{a}\|_1=s-1} \left(\prod_{i=1}^{d} (x_i - \delta k_i(\boldsymbol{x}))^{a_i} \right) \left(D^{\boldsymbol{a}} f(\boldsymbol{\xi}_2) - D^{\boldsymbol{a}} f(\delta k(\boldsymbol{x})) \right), \end{aligned}$$

where $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2 \in [\delta k(\boldsymbol{x}), \boldsymbol{x}] \subset [\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + \boldsymbol{e}))$. Consider separately each of the three terms on the right-hand side. Since $|x_i - \delta k_i(\boldsymbol{x})| \le \delta$, then

$$\begin{aligned} &\sum_{j=1}^{s-1} \frac{1}{j!} \sum_{\boldsymbol{a}\,:\,\|\boldsymbol{a}\|_1=j} \left(\prod_{i=1}^{d} (x_i - \delta k_i(\boldsymbol{x}))^{a_i} \right) \left(D^{\boldsymbol{a}} Af(\delta k(\boldsymbol{x})) - D^{\boldsymbol{a}} f(\delta k(\boldsymbol{x})) \right) \\ &\quad \le \sum_{j=1}^{s-1} \frac{1}{j!} \sum_{\boldsymbol{a}\,:\,\|\boldsymbol{a}\|_1=j} \delta^{\|\boldsymbol{a}\|_1} \delta^{s-\|\boldsymbol{a}\|_1} 12^d E^{(s)}\left([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 3\boldsymbol{e})] \right) \\ &\quad \le \sum_{j=1}^{s-1} \frac{1}{j!} d^j \delta^s 12^d E^{(s)}\left([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 3\boldsymbol{e})] \right) \\ &\quad \le 4d^4 12^d \delta^s E^{(s)}\left([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 3\boldsymbol{e})] \right). \end{aligned}$$

The first inequality is due to Lemma A.3, the second is due to the fact that there are at most d^j permutations that yield $\|\boldsymbol{a}\|_1 = j$, and the final inequality follows from $j < s \le 4$.

For the second term we note that

$$\begin{aligned} &\frac{1}{(s-1)!} \sum_{\boldsymbol{a}\,:\,\|\boldsymbol{a}\|_1=s-1} \left(\prod_{i=1}^{d} (x_i - \delta k_i(\boldsymbol{x}))^{a_i} \right) (D^{\boldsymbol{a}} Af(\boldsymbol{\xi}_1) - D^{\boldsymbol{a}} Af(\delta k(\boldsymbol{x}))) \\ &\quad \le d^{s-1} \delta^{s-1} \left| D^{\boldsymbol{a}} Af(\boldsymbol{\xi}_1) - D^{\boldsymbol{a}} Af(\delta k(\boldsymbol{x})) \right| \\ &\quad \le d^{s-1} \delta^{s-1} \|\boldsymbol{\xi}_1 - \delta k(\boldsymbol{x})\|_1 \sup_{\substack{\delta k(\boldsymbol{x}) \le z \\ z \le \delta(k(\boldsymbol{x})+\boldsymbol{e}) \\ \|\boldsymbol{a}\|_1 = s}} |D^{\boldsymbol{a}} Af(z)| \\ &\quad \le d^s \delta^s C_{A.4}(d) \delta^{-\|\boldsymbol{a}\|_1} \sup_{\substack{\delta k(\boldsymbol{x}) \le z \\ z \le \delta(k(\boldsymbol{x})+\boldsymbol{e}) \\ \|\boldsymbol{a}\|_1 = s}} \max_{0 \le \boldsymbol{i} \le 4\boldsymbol{e} - \boldsymbol{a}} |\Delta^{\boldsymbol{a}} f(\delta(k(z) + \boldsymbol{i}))| \\ &\quad \le d^s \delta^s C_{A.4}(d) \sup_{\substack{\delta k(\boldsymbol{x}) \le z \\ z \le \delta(k(\boldsymbol{x})+5\boldsymbol{e}) \\ \|\boldsymbol{a}\|_1 = s}} |D^{\boldsymbol{a}} f(z)| \\ &\quad \le d^s \delta^s C_{A.4}(d) E^{(s)}([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 5\boldsymbol{e})]). \end{aligned}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

The second inequality holds because we assumed that $D^{\boldsymbol{a}}f(\boldsymbol{x}) \in \mathcal{D}^d$ when $\|\boldsymbol{a}\|_1 = s-1$. The third holds because $\|\boldsymbol{\xi}_1 - \delta k(\boldsymbol{x})\|_1 \le d\delta$ (since $\delta k(\boldsymbol{x}) \le \boldsymbol{\xi}_1 \le \boldsymbol{x} < \delta(k(\boldsymbol{x}) + \boldsymbol{e})$), and by (A.3). The second-last inequality is a consequence of the fundamental theorem of calculus: suppose that $\Delta^{\boldsymbol{a}} f(\delta(k(\boldsymbol{z}) + \boldsymbol{i})) = \Delta_{j_s} \cdots \Delta_{j_1} f(\delta(k(\boldsymbol{z}) + \boldsymbol{i}))$ for some $1 \le j_1, \dots, j_s \le d$. By our assumptions that $f \in C^{s-1}(\mathbb{R}^d)$ and that $D^{\boldsymbol{a}} f(\boldsymbol{x}) \in \mathcal{D}^d$ when $\|\boldsymbol{a}\|_1 = s-1$, it follows that

$$\begin{aligned}
&|\Delta_{j_s} \cdots \Delta_{j_1} f(\delta(k(\boldsymbol{z}) + \boldsymbol{i}))| \\
&\quad = \left| \Delta_{j_s} \int_0^\delta \cdots \int_0^\delta \partial_{j_{s-1}} \cdots \partial_{j_1} f(\delta(k(\boldsymbol{z}) + \boldsymbol{i}) + \gamma_1 e_{j_1} + \cdots + \gamma_{s-1} e_{j_{s-1}}) d\gamma_1 \cdots d\gamma_{s-1} \right| \\
&\quad \le \delta^s \sup_{\substack{\delta(k(\boldsymbol{z})+\boldsymbol{i}) \le \boldsymbol{z}' \\ \boldsymbol{z}' \le \delta(k(\boldsymbol{z})+\boldsymbol{i}+\boldsymbol{a})}} |D^{\boldsymbol{a}} f(\boldsymbol{z}')|. \qquad \text{(A.13)}
\end{aligned}$$

Finally, for the third term, since we assumed that $D^{\boldsymbol{a}} f(\boldsymbol{x}) \in \mathcal{D}^d$ when $\|\boldsymbol{a}\|_1 = s-1$,

$$\begin{aligned}
&\frac{1}{(s-1)!} \sum_{\boldsymbol{a} : \|\boldsymbol{a}\|_1 = s-1} \left(\prod_{i=1}^d (x_i - \delta k_i(\boldsymbol{x}))^{a_i} \right) \left(D^{\boldsymbol{a}} f(\boldsymbol{\xi}_2) - D^{\boldsymbol{a}} f(\delta k(\boldsymbol{x})) \right) \\
&\qquad \le d^{s-1} \delta^{s-1} \|\boldsymbol{\xi}_1 - \delta k(\boldsymbol{x})\|_1 \sup_{\substack{\delta k(\boldsymbol{x}) \le \boldsymbol{z} \\ \boldsymbol{z} \le \delta(k(\boldsymbol{x})+\boldsymbol{e}) \\ \|\boldsymbol{a}\|_1 = s}} |D^{\boldsymbol{a}} A f(\boldsymbol{z})| \\
&\qquad \le d^{s-1} \delta^{s-1} d\delta E^{(s)}\big([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + \boldsymbol{e})] \big).
\end{aligned}$$

Adding up the bounds on the three terms and using the fact that $s \le 4$ yields

$$C_{A.11}(d) \le d^4(4 \times 12^d + C_{A.4}(d) + 1) \le d^4(412^d + 11404^5 1595^d + 1).$$

□

To prove Lemma A.3, for $f : \mathbb{R}^d \to \mathbb{R}$, $1 \le i \le d$, and $\boldsymbol{x} \in \mathbb{R}^d$, we define $\Delta_i^0 f(\boldsymbol{x}) = f(\boldsymbol{x})$, $\widetilde{\Delta}_i^{(0)} f(\boldsymbol{x}) = f(\boldsymbol{x})$,

$$\begin{aligned}
\Delta_i f(\boldsymbol{x}) &= f(\boldsymbol{x} + \delta \boldsymbol{e}_i) - f(\boldsymbol{x}), \\
\widetilde{\Delta}_i^{(1)} f(\boldsymbol{x}) &= \left(\Delta_i - \frac{1}{2}\Delta_i^2 + \frac{1}{3}\Delta_i^3 \right) f(\boldsymbol{x}), \\
\widetilde{\Delta}_i^{(2)} f(\boldsymbol{x}) &= \left(\Delta_i^2 - \Delta_i^3 \right) f(\boldsymbol{x}), \quad \widetilde{\Delta}_i^{(3)} f(\boldsymbol{x}) = \Delta_i^3 f(\boldsymbol{x}). \qquad \text{(A.14)}
\end{aligned}$$

Lemma A.4. *Given $f : \mathbb{R}^d \to \mathbb{R}$ and the corresponding $Af(\boldsymbol{x})$, for $1 \le \|\boldsymbol{a}\|_1 \le 3$,*

$$D^{\boldsymbol{a}} A f(\delta k(\boldsymbol{x})) = \delta^{-\|\boldsymbol{a}\|_1} \widetilde{\Delta}_1^{(a_1)} \cdots \widetilde{\Delta}_d^{(a_d)} f(\delta k(\boldsymbol{x})), \quad \boldsymbol{x} \in \mathbb{R}^d.$$

We now prove Lemmas A.3 and A.4.

Proof of Lemma A.3. Fix $s \in \{1, 2, 3, 4\}$ and $\boldsymbol{a}$ with $1 \le \|\boldsymbol{a}\|_1 \le s-1$. For any $A \subset \mathbb{R}^d$, we write $E_i^{(s)}(A)$ to denote any quantity satisfying

$$|E_i^{(s)}(A)| \le \sup_{z \in A} |\partial_i^s f(z)|,$$

and recall from (A.12) that $E^{(s)}(A)$ denotes a quantity satisfying

$$|E^{(s)}(A)| \le \max_{\|\boldsymbol{a}\|_1 = s} \sup_{z \in A} |D^{\boldsymbol{a}} f(z)|.$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

Suppose we have shown that, for $1 \le i \le d$ and $1 \le j \le s-1$,

$$\widetilde{\Delta}_i^{(j)} f(\boldsymbol{x}) = \delta^j \partial_i^j f(\boldsymbol{x}) + 5\delta^s E_i^{(s)}([\boldsymbol{x}, \boldsymbol{x} + 3\delta \boldsymbol{e}_i]). \tag{A.15}$$

Combining (A.15) with Lemma A.4 yields

$$\begin{aligned}
&D^{\boldsymbol{a}} A f(\delta k(\boldsymbol{x})) \\
&\quad = \delta^{-\|\boldsymbol{a}\|_1} \widetilde{\Delta}_1^{(a_1)} \cdots \widetilde{\Delta}_d^{(a_d)} f(\delta k(\boldsymbol{x})) \\
&\quad = \delta^{-\|\boldsymbol{a}\|_1} \widetilde{\Delta}_1^{(a_1)} \cdots \widetilde{\Delta}_{d-1}^{(a_{d-1})} (\delta^{a_d} \partial_d^{a_d} f(\delta k(\boldsymbol{x})) + 5\delta^s E_d^{(s)}([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 3\boldsymbol{e}_d)])) \\
&\quad = \delta^{-\|\boldsymbol{a}\|_1} \delta^{a_d} \widetilde{\Delta}_1^{(a_1)} \cdots \widetilde{\Delta}_{d-1}^{(a_{d-1})} \partial_d^{a_d} f(\delta k(\boldsymbol{x})) \\
&\qquad + 5\delta^{s-\|\boldsymbol{a}\|_1} \widetilde{\Delta}_1^{(a_1)} \cdots \widetilde{\Delta}_{d-1}^{(a_{d-1})} E_d^{(s)}([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 3\boldsymbol{e}_d)]) \\
&\quad = \delta^{-\|\boldsymbol{a}\|_1} \delta^{a_d} \widetilde{\Delta}_1^{(a_1)} \cdots \widetilde{\Delta}_{d-2}^{(a_{d-2})} (\delta^{a_{d-1}} \partial_{d-1}^{a_{d-1}} \partial_d^{a_d} f(\delta k(\boldsymbol{x})) + 5\delta^s E_{d-1}^{(s)}([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 3\boldsymbol{e}_{d-1})])) \\
&\qquad + 5\delta^{s-\|\boldsymbol{a}\|_1} \widetilde{\Delta}_1^{(a_1)} \cdots \widetilde{\Delta}_{d-1}^{(a_{d-1})} E_d^{(s)}([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 3\boldsymbol{e}_d)]).
\end{aligned}$$

Continuing recursively, we arrive at

$$\begin{aligned}
D^{\boldsymbol{a}} A f(\delta k(\boldsymbol{x})) &= \delta^{-\|\boldsymbol{a}\|_1} \delta^{a_1} \cdots \delta^{a_d} D^{\boldsymbol{a}} f(\delta k(\boldsymbol{x})) \\
&\quad + 5\delta^{s-\|\boldsymbol{a}\|_1} \delta^{a_d} \cdots \delta^{a_2} E_1^{(s)}([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 3\boldsymbol{e}_1)]) \\
&\quad + \cdots \\
&\quad + 5\delta^{s-\|\boldsymbol{a}\|_1} \widetilde{\Delta}_1^{(a_1)} \cdots \widetilde{\Delta}_{d-1}^{(a_{d-1})} E_d^{(s)}([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 3\boldsymbol{e}_d)])
\end{aligned} \tag{A.16}$$

We now argue that, for any $1 \le j \le d-1$,

$$\widetilde{\Delta}_1^{(a_1)} \cdots \widetilde{\Delta}_{j-1}^{(a_{j-1})} E_j^{(s)}([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 3\boldsymbol{e}_j)]) = 12^{j-1} E^{(s)}([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 3\boldsymbol{e})]), \tag{A.17}$$

because applying (A.17) to (A.16) and using the fact that $\delta < 1$ yields

$$\begin{aligned}
D^{\boldsymbol{a}} A f(\delta k(\boldsymbol{x})) &= D^{\boldsymbol{a}} f(\delta k(\boldsymbol{x})) + 5\delta^{s-\|\boldsymbol{a}\|_1} \Big(\sum_{j=1}^{d-1} 12^{j-1} \Big) E^{(s)}([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 3\boldsymbol{e})]) \\
&= D^{\boldsymbol{a}} f(\delta k(\boldsymbol{x})) + 12^d \delta^{s-\|\boldsymbol{a}\|_1} E^{(s)}([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 3\boldsymbol{e})])
\end{aligned}$$

where the second equality follows from $5 \sum_{j=1}^{d-1} 12^{j-1} \le 12^d$. To prove (A.17), we set $j = d$ and evaluate $\widetilde{\Delta}_1^{(a_1)} \cdots \widetilde{\Delta}_{d-1}^{(a_{d-1})} E_d^{(s)}([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 3\boldsymbol{e}_d)])$. We remind the reader that, for any $1 \le \ell \le d-1$,

$$\begin{aligned}
&\Delta_\ell E_d^{(s)}([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 3\boldsymbol{e}_d)]) \\
&\qquad = E_d^{(s)}([\delta k(\boldsymbol{x} + \delta \boldsymbol{e}_\ell),\, \delta(k(\boldsymbol{x} + \delta \boldsymbol{e}_\ell) + 3\boldsymbol{e}_d)]) - E_d^{(s)}([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 3\boldsymbol{e}_d)]) \\
&\qquad = 2E_d^{(s)}([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + \boldsymbol{e}_\ell + 3\boldsymbol{e}_d)]),
\end{aligned}$$

where in the second equality we used the fact that $k(\boldsymbol{x} + \delta \boldsymbol{e}_\ell) = k(\boldsymbol{x}) + \boldsymbol{e}_\ell$, which follows from the definition of $k(\boldsymbol{x})$. Similarly,

$$\begin{aligned}
\Delta_\ell^2 E_d^{(s)}([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 3\boldsymbol{e}_d)]) &= 4E_d^{(s)}([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 2\boldsymbol{e}_\ell + 3\boldsymbol{e}_d)]), \\
\Delta_\ell^3 E_d^{(s)}([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 3\boldsymbol{e}_d)]) &= 8E_d^{(s)}([\delta k(\boldsymbol{x}),\, \delta(k(\boldsymbol{x}) + 3\boldsymbol{e}_\ell + 3\boldsymbol{e}_d)]).
\end{aligned}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

These expressions, together with the definition of $\widetilde{\Delta}_{d-1}^{(a_{d-1})} f(\boldsymbol{x})$ in (A.14) and the fact that $0 \le a_{d-1} \le 3$, imply that

$$\widetilde{\Delta}_{d-1}^{(a_{d-1})} E_d^{(s)}([\delta k(\boldsymbol{x}), \delta(k(\boldsymbol{x}) + 3\boldsymbol{e}_d)]) = 12 E_d^{(s)}([\delta k(\boldsymbol{x}), \delta(k(\boldsymbol{x}) + 3\boldsymbol{e}_{d-1} + 3\boldsymbol{e}_d)])$$

and, similarly,

$$\begin{aligned}\widetilde{\Delta}_1^{(a_1)} \cdots \widetilde{\Delta}_{d-1}^{(a_{d-1})} E_d^{(s)}([\delta k(\boldsymbol{x}), \delta(k(\boldsymbol{x}) + 3\boldsymbol{e}_d)]) &= 12^{d-1} E_d^{(s)}([\delta k(\boldsymbol{x}), \delta(k(\boldsymbol{x}) + 3\boldsymbol{e})]) \\ &= 12^{d-1} E^{(s)}([\delta k(\boldsymbol{x}), \delta(k(\boldsymbol{x}) + 3\boldsymbol{e})]),\end{aligned}$$

proving (A.17).

It remains to prove (A.15). Suppose that $s = 4$, Taylor expansion yields

$$\begin{aligned}\Delta_i f(\boldsymbol{x}) &= \delta \partial_i f(\boldsymbol{x}) + \frac{1}{2}\delta^2 \partial_i^2 f(\boldsymbol{x}) + \frac{1}{6}\delta^3 \partial_i^3 f(\boldsymbol{x}) + \frac{1}{24}\delta^4 E_i^{(4)}([\boldsymbol{x}, \boldsymbol{x} + \delta\boldsymbol{e}_i]), \\ \Delta_i^2 f(\boldsymbol{x}) &= \delta^2 \partial_i^2 f(\boldsymbol{x}) + \delta^3 \partial_i^3 f(\boldsymbol{x}) + \frac{2}{3}\delta^4 E_i^{(4)}([\boldsymbol{x}, \boldsymbol{x} + 2\delta\boldsymbol{e}_i]), \\ \Delta_i^3 f(\boldsymbol{x}) &= \delta^3 \partial_i^3 f(\boldsymbol{x}) + \frac{17}{6}\delta^4 E_i^{(4)}([\boldsymbol{x}, \boldsymbol{x} + 3\delta\boldsymbol{e}_i]),\end{aligned} \tag{A.18}$$

The second equality is obtained by applying Δ_i to both sides of the first equality and observing that

$$\Delta_i E_i^{(4)}([\boldsymbol{x}, \boldsymbol{x} + \delta\boldsymbol{e}_i]) = E_i^{(4)}([\boldsymbol{x} + \delta\boldsymbol{e}_i, \boldsymbol{x} + 2\delta\boldsymbol{e}_i]) - E_i^{(4)}([\boldsymbol{x}, \boldsymbol{x} + \delta\boldsymbol{e}_i]) = 2E_i^{(4)}([\boldsymbol{x}, \boldsymbol{x} + 2\delta\boldsymbol{e}_i]).$$

Similarly, the third equality is obtained by applying Δ_i to both sides of the second equality. To prove (A.15), we combine (A.18) with the definition of $\widetilde{\Delta}_i^{(j)} f(\boldsymbol{x})$ to obtain

$$\begin{aligned}\widetilde{\Delta}_i^{(1)} f(\boldsymbol{x}) = {} & \delta \partial_i f(\boldsymbol{x}) + \frac{1}{24}\delta^4 E_i^{(4)}([\boldsymbol{x}, \boldsymbol{x} + \delta\boldsymbol{e}_i]) - \frac{1}{2} \times \frac{2}{3}\delta^4 E_i^{(4)}([\boldsymbol{x}, \boldsymbol{x} + 2\delta\boldsymbol{e}_i]) \\ & + \frac{1}{3} \times \frac{17}{6}\delta^4 E_i^{(4)}([\boldsymbol{x}, \boldsymbol{x} + 3\delta\boldsymbol{e}_i]) \\ = {} & \delta \partial_i f(\boldsymbol{x}) + 2\delta^4 E_i^{(4)}([\boldsymbol{x}, \boldsymbol{x} + 3\delta\boldsymbol{e}_i]),\end{aligned}$$

and

$$\begin{aligned}\widetilde{\Delta}_i^{(2)} f(\boldsymbol{x}) &= \delta^2 \partial_i^2 f(\boldsymbol{x}) + 4\delta^4 E_i^{(4)}([\boldsymbol{x}, \boldsymbol{x} + 3\delta\boldsymbol{e}_i]), \\ \widetilde{\Delta}_i^{(3)} f(\boldsymbol{x}) &= \delta^3 \partial_i^3 f(\boldsymbol{x}) + 3\delta^4 E_i^{(4)}([\boldsymbol{x}, \boldsymbol{x} + 3\delta\boldsymbol{e}_i]).\end{aligned}$$

For completeness, we outline the main steps needed to prove (A.15) when $s < 4$. Suppose that $s = 3$, in which case Taylor expansion yields

$$\begin{aligned}\Delta_i f(\boldsymbol{x}) &= \delta \partial_i f(\boldsymbol{x}) + \frac{1}{2}\delta^2 \partial_i^2 f(\boldsymbol{x}) + \frac{1}{6}\delta^3 E_i^{(3)}([\boldsymbol{x}, \boldsymbol{x} + \delta\boldsymbol{e}_i]), \\ \Delta_i^2 f(\boldsymbol{x}) &= \delta^2 \partial_i^2 f(\boldsymbol{x}) + \frac{4}{3}\delta^3 E_i^{(3)}([\boldsymbol{x}, \boldsymbol{x} + 2\delta\boldsymbol{e}_i]), \\ \Delta_i^3 f(\boldsymbol{x}) &= \frac{11}{3}\delta^3 E_i^{(3)}([\boldsymbol{x}, \boldsymbol{x} + 3\delta\boldsymbol{e}_i]),\end{aligned}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

which implies that

$$\widetilde{\Delta}_i^{(1)} f(\boldsymbol{x}) = \delta \partial_i f(\boldsymbol{x}) + 3\delta^3 E_i^{(3)}([\boldsymbol{x}, \boldsymbol{x} + 3\delta \boldsymbol{e}_i]),$$
$$\widetilde{\Delta}_i^{(2)} f(\boldsymbol{x}) = \delta^2 \partial_i^2 f(\boldsymbol{x}) + 5\delta^3 E_i^{(3)}([\boldsymbol{x}, \boldsymbol{x} + 3\delta \boldsymbol{e}_i]).$$

When $s = 2$,

$$\Delta_i f(\boldsymbol{x}) = \delta \partial_i f(\boldsymbol{x}) + \frac{1}{2}\delta^2 E_i^{(2)}([\boldsymbol{x}, \boldsymbol{x} + \delta \boldsymbol{e}_i]),$$
$$\Delta_i^2 f(\boldsymbol{x}) = 2\delta^2 E_i^{(2)}([\boldsymbol{x}, \boldsymbol{x} + 2\delta \boldsymbol{e}_i]), \quad \text{and} \quad \Delta_i^3 f(\boldsymbol{x}) = 4\delta^2 E_i^{(2)}([\boldsymbol{x}, \boldsymbol{x} + 3\delta \boldsymbol{e}_i]),$$

implying that

$$\widetilde{\Delta}_i^{(1)} f(\boldsymbol{x}) = \delta \partial_i f(\boldsymbol{x}) + 3\delta^3 E_i^{(2)}([\boldsymbol{x}, \boldsymbol{x} + 3\delta \boldsymbol{e}_i]).$$

□

Proof of Lemma A.4. Let us first consider the case when $d = 1$. According to the original definition of $A f(x)$ in Theorem 1 of [5],

$$Af(x) = P_{k(x)}(x), \quad x \in \mathbb{R},$$

where

$$\begin{aligned} P_k(x) = f(\delta k) &+ \Big(\frac{x - \delta k}{\delta}\Big)(\Delta - \frac{1}{2}\Delta^2 + \frac{1}{3}\Delta^3) f(\delta k) \\ &+ \frac{1}{2}\Big(\frac{x - \delta k}{\delta}\Big)^2 (\Delta^2 - \Delta^3) f(\delta k) + \frac{1}{6}\Big(\frac{x - \delta k}{\delta}\Big)^3 \Delta^3 f(\delta k) \\ &- \frac{23}{3}\Big(\frac{x - \delta k}{\delta}\Big)^4 \Delta^4 f(\delta k) + \frac{41}{2}\Big(\frac{x - \delta k}{\delta}\Big)^5 \Delta^4 f(\delta k) \\ &- \frac{55}{3}\Big(\frac{x - \delta k}{\delta}\Big)^6 \Delta^4 f(\delta k) + \frac{11}{2}\Big(\frac{x - \delta k}{\delta}\Big)^7 \Delta^4 f(\delta k), \quad x \in \mathbb{R}. \end{aligned} \tag{A.19}$$

We can also write $A f(x)$ as the weighted sum

$$Af(x) = \sum_{i=0}^{4} \alpha_{k(x)+i}^{k(x)}(x) f(\delta(k(x) + i)),$$

where the weights $\alpha_{k+i}^{k}(x)$ are defined by the corresponding polynomial $P_k(x)$. Now $Af \in C^3(\mathbb{R})$ by Theorem A.1, meaning that $\partial^j Af(\delta k(x))$ equals the corresponding right derivative of $P_{k(x)}(x)$ at $x = \delta k(x)$ for $1 \le j \le 3$. The result follows immediately by inspecting the derivatives of (A.20).

When $d > 1$, the proof is similar. Let us write $\boldsymbol{k}$ instead of $k(\boldsymbol{x})$ for notational convenience. Since $Af(\boldsymbol{x}) \in C^3(\mathbb{R}^d)$ by Theorem A.1, we fix $\boldsymbol{a}$ with $1 \le \|\boldsymbol{a}\|_1 \le 3$ and consider

$$\begin{aligned} D^{\boldsymbol{a}} Af(\boldsymbol{x}) &= D^{\boldsymbol{a}} \left(\sum_{i_d=0}^{4} \alpha_{k_d+i_d}^{k_d}(x_d) \cdots \sum_{i_1=0}^{4} \alpha_{k_1+i_1}^{k_1}(x_1) f(\delta(\boldsymbol{k} + \boldsymbol{i})) \right) \\ &= \left(\sum_{i_d=0}^{4} \partial_d^{a_d} \alpha_{k_d+i_d}^{k_d}(x_d) \cdots \left(\sum_{i_2=0}^{4} \partial_2^{a_2} \alpha_{k_2+i_2}^{k_2}(x_2) \left(\sum_{i_1=0}^{4} \partial_1^{a_1} \alpha_{k_1+i_1}^{k_1}(x_1) f(\delta(\boldsymbol{k} + \boldsymbol{i})) \right) \right) \right). \end{aligned}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

The first equality follows from (A.1) of Theorem A.1. If we think of the innermost sum as a one-dimensional function in x_1 only, it follows that

$$\sum_{i_1=0}^{4} \partial_1^{a_1} \alpha_{k_1+i_1}^{k_1}(\delta k_1) f(\delta(\boldsymbol{k}+\boldsymbol{i})) = \delta^{-a_1} \widetilde{\Delta}_1^{(a_1)} f\big(\delta(\boldsymbol{k}+(0, i_2, \ldots, i_d))\big).$$

Proceeding identically along each of the remaining dimensions yields the result. □

Before concluding this section, we prove Lemmas 2.4 and 2.5.

Proof of Lemma 2.4. Fix $h \in \mathcal{M}_4$ and note that

$$|\mathbb{E}h(\boldsymbol{V}) - \mathbb{E}h(\boldsymbol{V}')| \le |\mathbb{E}h(\boldsymbol{V}) - \mathbb{E}Ah(\boldsymbol{V}')| + |\mathbb{E}Ah(\boldsymbol{V}') - \mathbb{E}h(\boldsymbol{V}')|,$$

where $Ah(\cdot)$ is understood to be the operator A applied to the restriction of $h(\boldsymbol{x})$ to $\delta\mathbb{Z}^d$. Proposition A.1 and the fact that $h \in \mathcal{M}_4$ imply that

$$|\mathbb{E}Ah(\boldsymbol{V}') - \mathbb{E}h(\boldsymbol{V}')| \le C_{A.11}(d)\delta^4 \max_{\|\boldsymbol{a}\|_1=4} \|D^{\boldsymbol{a}} h\| \le C_{A.11}(d)\delta^4.$$

Furthermore, (A.13) implies that the restriction of $h(\boldsymbol{x})$ to $\delta\mathbb{Z}^d$ belongs to $\mathcal{M}_{\mathrm{disc},4}$ and, therefore,

$$|\mathbb{E}h(\boldsymbol{V}) - \mathbb{E}Ah(\boldsymbol{V}')| \le \sup_{h\in\mathcal{M}_{\mathrm{disc},4}} |\mathbb{E}h(\boldsymbol{V}) - \mathbb{E}Ah(\boldsymbol{V}')|.$$

□

Proof of Lemma 2.5. Since $h \in \mathcal{M}_{\mathrm{disc},4}$ implies that $|h(\boldsymbol{u}+\delta\boldsymbol{e}_i) - h(\boldsymbol{u})| \le \delta$ for all $1 \le i \le K-1$, it follows that

$$|h(\boldsymbol{u})| \le |h(0)| + \|\boldsymbol{u}\|_1, \quad \boldsymbol{u} \in \delta\mathbb{Z}^{K-1}.$$

Furthermore, (A.3) of Theorem A.1 and the fact that $h \in \mathcal{M}_{\mathrm{disc},4}$ imply that the first-order partial derivatives of $Ah(\boldsymbol{x})$ are bounded by $C_{A.4}(d)$, and therefore,

$$|Ah(\boldsymbol{Z})| \le |Ah(0)| + C_{A.4}(d)\|\boldsymbol{Z}\|_1 \le |h(0)| + C_{A.4}(d),$$

where the last inequality follows from $Ah(0) = h(0)$ and $\boldsymbol{Z} \in \bar{\nabla}^K$. To bound $|\mathcal{G}_{\boldsymbol{Z}} A f_h(\boldsymbol{Z})|$, we recall that

$$\mathcal{G}_{\boldsymbol{Z}} A f_h(\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{K-1} (\beta_i - s x_i)\partial_i A f_h(\boldsymbol{x}) + \frac{1}{2}\sum_{i,j=1}^{K-1} x_i(\delta_{ij} - x_j)\partial_i\partial_j A f_h(\boldsymbol{x}), \quad \boldsymbol{x} \in \bar{\nabla}^K.$$

Note that $|\beta_i - s x_i| \le s$, $|x_i(\delta_{ij} - x_j)| \le 1$, and that (A.3) of Theorem A.1 implies that

$$|\partial_i A f_h(\boldsymbol{x})| \le C_{A.4}(d)\delta^{-1} \max_{\|\boldsymbol{a}\|_1=1} \|\Delta^{\boldsymbol{a}} f_h\| \le 2C_{A.4}(K)\delta^{-1}\|f_h\|,$$
$$|\partial_i\partial_j A f_h(\boldsymbol{x})| \le C_{A.4}(d)\delta^{-2} \max_{\|\boldsymbol{a}\|_1=2} \|\Delta^{\boldsymbol{a}} f_h\| \le 4C_{A.4}(K)\delta^{-2}\|f_h\|.$$

Recalling from (B.7) that $\|f_h\| \le \delta^{-1}KB_1(f_h)$, we combine the bounds to arrive at

$$\begin{aligned}
|\mathcal{G}_{\boldsymbol{Z}} A f_h(\boldsymbol{x})| &\le \frac{1}{2} Ks 2C_{A.4}(K)\delta^{-1}\|f_h\| + \frac{1}{2}K^2 4C_{A.4}(K)\delta^{-2}\|f_h\| \\
&\le C_{A.4}(K)\big(Ks\delta^{-1} + 2K^2\delta^{-2}\big)\delta^{-1}KB_1(f_h) \\
&\le C_{A.4}(K)(2+s)\delta^{-3}K^3 B_1(f_h),
\end{aligned}$$

where in the last inequality we used $\delta \le 1$. □

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

Appendix A. Explicit constants in Theorem A.1

Proof of (A.4) and (A.8). We show that

$$C_{A.4}(d) \le \prod_{j=1}^{d} (1595 \times 1(a_j = 0) + 3408 \times 1(a_j = 1) + 7677 \times 1(a_j = 2) + 9510 \times 1(a_j = 3) + 11404 \times 1(a_j = 4)),$$

which implies the result. Consider first the case when $d = 1$; we want a bound on $Af(x)$ and its first four derivatives. For any $f : \mathbb{Z} \to \mathbb{R}$, (A.1) of [5] states that

$$Af(x) = \sum_{i=0}^{4} \alpha_{k(x)+i}^{k(x)}(x) f(\delta(k(x) + i)) = P_{k(x)}(x),$$

where

$$\begin{aligned} P_k(x) = f(\delta k) &+ \Big(\frac{x - \delta k}{\delta}\Big)(\Delta - \frac{1}{2}\Delta^2 + \frac{1}{3}\Delta^3) f(\delta k) \\ &+ \frac{1}{2}\Big(\frac{x - \delta k}{\delta}\Big)^2 (\Delta^2 - \Delta^3) f(\delta k) + \frac{1}{6}\Big(\frac{x - \delta k}{\delta}\Big)^3 \Delta^3 f(\delta k) \\ &- \frac{23}{3}\Big(\frac{x - \delta k}{\delta}\Big)^4 \Delta^4 f(\delta k) + \frac{41}{2}\Big(\frac{x - \delta k}{\delta}\Big)^5 \Delta^4 f(\delta k) \\ &- \frac{55}{3}\Big(\frac{x - \delta k}{\delta}\Big)^6 \Delta^4 f(\delta k) + \frac{11}{2}\Big(\frac{x - \delta k}{\delta}\Big)^7 \Delta^4 f(\delta k), \quad x \in \mathbb{R}. \end{aligned} \tag{A.20}$$

To bound $|Af(x)|$, we need to bound each of the terms on the right-hand side of (A.20). For instance,

$$\begin{aligned} (\Delta - \frac{1}{2}\Delta^2 + \frac{1}{3}\Delta^3) f(\delta k) = &- f(\delta k) + f(\delta(k+1)) \\ &- \frac{1}{2}\big(f(\delta k) - 2f(\delta(k+1)) + f(\delta(k+2))\big) \\ &+ \frac{1}{3}\big(-f(\delta k) + 3f(\delta(k+1)) - 3f(\delta(k+2)) + f(\delta(k+3))\big), \end{aligned}$$

implying that

$$|(\Delta - \frac{1}{2}\Delta^2 + \frac{1}{3}\Delta^3) f(\delta k)| \le 3 \times 5 \max_{0 \le i \le 4} |f(k(x) + i)|.$$

Proceeding similarly with each of the terms on the right-hand side of (A.20), we see that $|Af(x)|$ is bounded by

$$\begin{aligned} &\Big(1 + 3\Big(\frac{x - \delta k(x)}{\delta}\Big) + \frac{5}{2}\Big(\frac{x - \delta k(x)}{\delta}\Big)^2 + \frac{1}{2}\Big(\frac{x - \delta k(x)}{\delta}\Big)^3 + \frac{23}{3} \times 6\Big(\frac{x - \delta k(x)}{\delta}\Big)^4 \\ &+ \frac{41}{2} \times 6\Big(\frac{x - \delta k(x)}{\delta}\Big)^5 + \frac{55}{3} \times 6\Big(\frac{x - \delta(x)}{\delta}\Big)^6 + \frac{11}{2} \times 6\Big(\frac{x - \delta k(x)}{\delta}\Big)^7\Big) \max_{0 \le i \le 4} |f(k(x) + i)| \\ &\quad \le 319 \times 5 \max_{0 \le i \le 4} |f(k(x) + i)|. \end{aligned}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

Note that (A.8) follows from a similar argument because $\alpha^k_{k+i}(x)$ is the coefficient corresponding to $f(\delta(k+i))$. To bound $|\partial Af(x)|$, we differentiate (A.20) once to get

$$\begin{aligned}\partial P_k(x) = \frac{1}{\delta}\Big(&(1-\frac{1}{2}\Delta+\frac{1}{3}\Delta^2)\Delta f(\delta k)\\ &+\frac{1}{2}\times 2\Big(\frac{x-\delta k}{\delta}\Big)(\Delta-\Delta^2)(\Delta f(\delta k))+\frac{1}{6}\times 3\Big(\frac{x-\delta k}{\delta}\Big)^2\Delta^2(\Delta f(\delta k))\\ &-\frac{23}{3}\times 4\Big(\frac{x-\delta k}{\delta}\Big)^3\Delta^3(\Delta f(\delta k))+\frac{41}{2}\times 5\Big(\frac{x-\delta k}{\delta}\Big)^4\Delta^3(\Delta f(\delta k))\\ &-\frac{55}{3}\times 6\Big(\frac{x-\delta k}{\delta}\Big)^5\Delta^3(\Delta f(\delta k))+\frac{11}{2}\times 7\Big(\frac{x-\delta k}{\delta}\Big)^6\Delta^3(\Delta f(\delta k))\Big),\quad x\in\mathbb{R}.\end{aligned}$$

Bounding the right-hand side the same way as before yields

$$\begin{aligned}|\partial Af(x)| &\le \frac{1}{\delta}\Big((1+1/2+2/3)+(1+2)\Big(\frac{x-\delta k(x)}{\delta}\Big)+\frac{3}{6}\times 2\Big(\frac{x-\delta k(x)}{\delta}\Big)^2\\ &\quad+\frac{23}{3}\times 4\times 3\Big(\frac{x-\delta k(x)}{\delta}\Big)^3+\frac{41}{2}\times 5\times 3\Big(\frac{x-\delta k(x)}{\delta}\Big)^4\\ &\quad+\frac{55}{3}\times 6\times 3\Big(\frac{x-\delta(x)}{\delta}\Big)^5+\frac{11}{2}\times 7\times 3\Big(\frac{x-\delta k(x)}{\delta}\Big)^6\Big)\times\max_{0\le i\le 3}|\Delta f(k(x)+i)|\\ &\le \frac{1}{\delta}852\times 4\max_{0\le i\le 3}|\Delta f(k(x)+i)|.\end{aligned}$$

Similarly,

$$\begin{aligned}\partial^2 P_k(x) = \frac{1}{\delta^2}\Big(&(1-\Delta)(\Delta^2 f(\delta k))+\Big(\frac{x-\delta k}{\delta}\Big)\Delta(\Delta^2 f(\delta k))\\ &-23\times 4\Big(\frac{x-\delta k}{\delta}\Big)^2\Delta^2(\Delta^2 f(\delta k))+81\times 5\Big(\frac{x-\delta k}{\delta}\Big)^3\Delta^2(\Delta^2 f(\delta k))\\ &-\frac{55}{3}\times 6\times 5\Big(\frac{x-\delta k}{\delta}\Big)^4\Delta^2(\Delta^2 f(\delta k))\\ &+\frac{11}{2}\times 7\times 6\Big(\frac{x-\delta k}{\delta}\Big)^5\Delta^2(\Delta^2 f(\delta k))\Big),\quad x\in\mathbb{R},\end{aligned}$$

$$\begin{aligned}\partial^3 P_k(x) = \frac{1}{\delta^3}\Big(&\Delta^3 f(\delta k)-23\times 4\times 2\Big(\frac{x-\delta k}{\delta}\Big)\Delta(\Delta^3 f(\delta k))+81\times 5\times 3\Big(\frac{x-\delta k}{\delta}\Big)^2\Delta(\Delta^3 f(\delta k))\\ &-\frac{55}{3}\times 6\times 5\times 4\Big(\frac{x-\delta k}{\delta}\Big)^3\Delta(\Delta^3 f(\delta k))\\ &+\frac{11}{2}\times 7\times 6\times 5\Big(\frac{x-\delta k}{\delta}\Big)^4\Delta(\Delta^3 f(\delta k))\Big),\quad x\in\mathbb{R},\end{aligned}$$

and

$$\begin{aligned}\partial^4 P_k(x) = \frac{1}{\delta^4}\Big(&-23\times 4\times 2\Delta^4 f(\delta k)+81\times 5\times 3\times 2\Big(\frac{x-\delta k}{\delta}\Big)\Delta^4 f(\delta k)\\ &-\frac{55}{3}\times 6\times 5\times 4\times 3\Big(\frac{x-\delta k}{\delta}\Big)^2\Delta^4 f(\delta k)\\ &+\frac{11}{2}\times 7\times 6\times 5\times 4\Big(\frac{x-\delta k}{\delta}\Big)^4\Delta^4 f(\delta k)\Big),\quad x\in\mathbb{R},\end{aligned}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

yield the bounds

$$|\partial^2 Af(x)| \le \frac{1}{\delta^2} 2559 \times 3 \max_{0\le i\le 2} |\Delta^2 f(k(x)+i)|,$$
$$|\partial^3 Af(x)| \le \frac{1}{\delta^3} 4775 \times 2 \max_{0\le i\le 1} |\Delta^3 f(k(x)+i)|,$$
$$|\partial^4 Af(x)| \le \frac{1}{\delta^4} 11404 |\Delta^4 f(\delta k(x))|.$$

When $d > 1$, we recall from (A.1) that

$$Af(\boldsymbol{x}) = \sum_{i_d=0}^{4} \alpha_{k_d(\boldsymbol{x})+i_d}^{k_d(\boldsymbol{x})}(x_d) \cdots \sum_{i_1=0}^{4} \alpha_{k_1(\boldsymbol{x})+i_1}^{k_1(\boldsymbol{x})}(x_1) f(\delta(k(\boldsymbol{x})+\boldsymbol{i})).$$

The result follows by applying the bounds from the $d = 1$ case along each dimension. □

Appendix B. Generator expansion and Stein factors

This appendix is dedicated to proving Lemmas 2.2 and 2.3. Recall that, for $\boldsymbol{x} \in \mathbb{R}^{K-1}$, we define $k(\boldsymbol{x})$ by $k_j(\boldsymbol{x}) = \lfloor x_j/\delta \rfloor$, that $\delta = 1/N$, ∇_N^K is defined in (2.7), and that $\bar{\nabla}_N^K = \mathrm{Conv}(\nabla_N^K)$. We may assume without loss of generality that $100K^2 < N$, which implies that $\nabla_N^K \ne \emptyset$, because the claim in Lemma 2.2 holds trivially for all $0 < N \le 100K^2$ (since this covers a finite number of N). We claim that

$$\text{if } \boldsymbol{x} \in \bar{\nabla}_N^K \text{ then } \delta(k(\boldsymbol{x})+\boldsymbol{i}) \in \nabla^K \text{ for all } 0 \le \boldsymbol{i} \le 10\boldsymbol{e}. \tag{B.1}$$

We argue this as follows. Since $0 \le \delta k(\boldsymbol{x}) \le \boldsymbol{x}$, then $\delta k(\boldsymbol{x}) \in \nabla_N^K$ for any $\boldsymbol{x} \in \bar{\nabla}_N^K$, because $\delta \sum_{j=1}^{K-1} k_j(\boldsymbol{x}) \le \sum_{j=1}^{K-1} x_j$. Thus, for all $0 \le i \le 10\boldsymbol{e}$,

$$\delta \sum_{j=1}^{K-1} (k_j(\boldsymbol{x}) + i_j) \le 1 - 10K/\sqrt{N} + 10K/N < 1,$$

implying (B.1). Combining (B.1) with (A.1) of Theorem A.1 implies that $A(\mathcal{G}_{\boldsymbol{U}} Af_h)(\boldsymbol{x})$ is well defined if $\boldsymbol{x} \in \bar{\nabla}_N^K$. To derive an expression for $A(\mathcal{G}_{\boldsymbol{U}} Af_h)(\boldsymbol{x})$, we define

$$b_i(\boldsymbol{u}) = \mathbb{E}_u(U_i(1) - u_i),$$
$$a_{ij}(\boldsymbol{u}) = \mathbb{E}_u(U_i(1) - u_i)(U_j(1) - u_j),$$
$$c_{ijk}(\boldsymbol{u}) = \mathbb{E}_u(U_i(1) - u_i)(U_j(1) - u_j)(U_k(1) - u_k),$$
$$\bar{d}_{ijk\ell}(\boldsymbol{u}) = \mathbb{E}_u|(U_i(1) - u_i)(U_j(1) - u_j)(U_k(1) - u_k)(U_\ell(1) - u_\ell)|, \quad \boldsymbol{u} \in \nabla^K.$$

Since $Af_h \in C^3(\mathbb{R}^{K-1})$, by Theorem A.1, we know that, for any $\boldsymbol{u} \in \nabla^K$,

$$\begin{aligned}\mathcal{G}_{\boldsymbol{U}}(Af_h)(\boldsymbol{u}) &= \mathbb{E}_{\boldsymbol{u}} Af_h(\boldsymbol{U}(1)) - Af_h(\boldsymbol{u}) \\ &= \sum_{i=1}^{K-1} b_i(\boldsymbol{u})\partial_i Af_h(\boldsymbol{u}) + \frac{1}{2}\sum_{i,j=1}^{K-1} a_{ij}(\boldsymbol{u})\partial_i\partial_j Af_h(\boldsymbol{u}) \\ &\quad + \frac{1}{6}\sum_{i,j,k=1}^{K-1} c_{ijk}(\boldsymbol{u})\partial_i\partial_j\partial_k Af_h(\boldsymbol{u}) + \epsilon(\boldsymbol{u}),\end{aligned}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

where

$$\epsilon(\boldsymbol{u}) = \frac{1}{6}\sum_{i,j,k=1}^{K-1} \mathbb{E}_{\boldsymbol{u}}\Big((U_i(1)-u_i)(U_j(1)-u_j)(U_k(1)-u_k)\big(\partial_i\partial_j\partial_k Af_h(\boldsymbol{\xi}) - \partial_i\partial_j\partial_k Af_h(\boldsymbol{u})\big)\Big),$$

and $\boldsymbol{\xi}$ is between $\boldsymbol{u}$ and $\boldsymbol{U}(1)$. Since A is a linear operator, it follows that, for any $\boldsymbol{x} \in \bar{\nabla}_N^K$,

$$\begin{aligned} A(\mathcal{G}_{\boldsymbol{U}} Af_h)(\boldsymbol{x}) &= \sum_{i=1}^{K-1} A(b_i\partial_i Af_h)(\boldsymbol{x}) + \frac{1}{2}\sum_{i,j=1}^{K-1} A(a_{ij}\partial_i\partial_j Af_h)(\boldsymbol{x}) \\ &\quad + \frac{1}{6}\sum_{i,j,k=1}^{K-1} A(c_{ijk}\partial_i\partial_j\partial_k Af_h)(\boldsymbol{x}) + A\epsilon(\boldsymbol{x}). \end{aligned} \tag{B.2}$$

We present several lemmas needed to analyze the right-hand side. Recall assumption (1.3) or that $p_i = \beta_i/(2N) = \delta\beta_i/2$ and, subsequently, $\Sigma = \delta s/2$. Also recall the definition of $B_i(\cdot)$ from (2.1).

Lemma B.1. *Fix $\boldsymbol{u} \in \nabla^K$ and let $\bar{u}_i = u_i\Sigma - p_i$. Then for $1 \le i, j \le K-1$,*

$$b_i(\boldsymbol{u}) = -\bar{u}_i \quad \text{and} \quad a_{ij}(\boldsymbol{u}) = \bar{u}_i\bar{u}_j + \frac{1}{N}\big(u_i - \bar{u}_i\big)\big(\delta_{ij} - (u_j - \bar{u}_j)\big),$$
$$B_1(b_i) \le \delta^2 s \quad \text{and} \quad B_1(a_{ij}) \le \delta^2 5(1+s)^2.$$

Furthermore, for any $1 \le i, j, k, \ell \le K-1$,

$$|c_{ijk}(\boldsymbol{u})| \le 20\delta^2(1+s)^3 \quad \text{and} \quad \bar{d}_{ijk\ell}(\boldsymbol{u}) \le \delta^2(2+s)^4, \quad \boldsymbol{u} \in \nabla^K.$$

The proof of Lemma B.1 is left to later in this appendix as it essentially has many elementary but tedious moment calculations. Since $b_i(\boldsymbol{u})$ and $a_{ij}(\boldsymbol{u})$ are polynomials in $\boldsymbol{u}$, we let $b_i(\boldsymbol{x})$ and $a_{ij}(\boldsymbol{x})$ be their natural extensions to $\bar{\nabla}^K$. Furthermore, Corollary A.1 implies that $Ab_i(\boldsymbol{x}) = b_i(\boldsymbol{x})$ and $Aa_{ij}(\boldsymbol{x}) = a_{ij}(\boldsymbol{x})$.

Lemma B.2. *For any $\boldsymbol{x} \in \bar{\nabla}_N^K$,*

$$\begin{aligned} A(b_i\partial_i Af_h)(\boldsymbol{x}) &= b_i(\boldsymbol{x})\partial_i Af_h(\boldsymbol{x}) + \tilde{\epsilon}_i(\boldsymbol{x}), \\ A(a_{ij}\partial_i Af_h)(\boldsymbol{x}) &= a_{ij}(\boldsymbol{x})\partial_i\partial_j Af_h(\boldsymbol{x}) + \tilde{\epsilon}_{ij}(\boldsymbol{x}), \\ A(c_{ijk}\partial_i\partial_j\partial_k Af_h)(\boldsymbol{x}) &= \tilde{\epsilon}_{ijk}(\boldsymbol{x}), \end{aligned}$$

where

$$|\tilde{\epsilon}_i(\boldsymbol{x})| \le \delta s C_{A.4}(K)\big(C_{A.10}(K)B_2(f_h) + C_{A.11}(K)\delta^{-1}B_4(f_h)\big), \tag{B.3}$$
$$|\tilde{\epsilon}_{ij}(\boldsymbol{x})| \le (1+s)^2 C_{A.4}(K)\big(5C_{A.10}(K)B_3(f_h) + C_{A.11}(K)\delta^{-1}B_4(f_h)\big), \tag{B.4}$$
$$|\tilde{\epsilon}_{ijk}(\boldsymbol{x})| \le 1595^K 20(1+s)^3 C_{A.4}(K)\delta^{-1}B_3(f_h). \tag{B.5}$$

Furthermore,

$$\begin{aligned} |A\epsilon(\boldsymbol{x})| \le 1595^K K^4 C_{A.4}(K)\Big(&(2+s)^4\delta^{-2}B_4(f_h) \\ &+ \frac{(4K+1)N^{4K}}{(\Sigma - p_K)^{4K}}\Big(1 - \frac{10K(\sqrt{N}-1)}{N}(1-\Sigma)\Big)^N \delta^{-5}16KB_1(f_h)\Big). \end{aligned} \tag{B.6}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

Proof of Lemma 2.2. Assumption (1.3) states that $p_i = \beta_i/(2N)$, which implies that $\Sigma = s/(2N)$. Thus, $\mathcal{G}_{\mathbf{Z}} f(\boldsymbol{x})$, which is defined in (1.8), satisfies

$$\begin{aligned}\delta\mathcal{G}_{\mathbf{Z}} f(\boldsymbol{x}) &= \sum_{i=1}^{K-1}(p_i - \Sigma x_i)\partial_i f(\boldsymbol{x}) + \frac{1}{2N}\sum_{i,j=1}^{K-1} x_i(\delta_{ij} - x_j)\partial_i\partial_j f(\boldsymbol{x}) \\ &= \sum_{i=1}^{K-1} b_i(\boldsymbol{x})\partial_i f(\boldsymbol{x}) + \frac{1}{2N}\sum_{i,j=1}^{K-1} x_i(\delta_{ij} - x_j)\partial_i\partial_j f(\boldsymbol{x}), \quad \boldsymbol{x} \in \bar{\nabla}^K,\end{aligned}$$

where the second equality is due to Lemma B.1. Combining Lemma B.2 with (B.2) yields

$$\begin{aligned}A(\mathcal{G}_U A f_h)(\boldsymbol{x}) = &\sum_{i=1}^{K-1} b_i(\boldsymbol{x})\partial_i A f_h(\boldsymbol{x}) + \frac{1}{2}\sum_{i,j=1}^{K-1} a_{ij}(\boldsymbol{x})\partial_i\partial_j A f_h(\boldsymbol{x}) \\ &+ \sum_{i=1}^{K-1}\tilde{\epsilon}_i(\boldsymbol{x}) + \frac{1}{2}\sum_{i,j=1}^{K-1}\tilde{\epsilon}_{ij}(\boldsymbol{x}) + \frac{1}{6}\sum_{i,j,k=1}^{K-1}\tilde{\epsilon}_{ijk}(\boldsymbol{x}) + A\epsilon(\boldsymbol{x}).\end{aligned}$$

Let us compare the first two terms on the right-hand side to $\delta\mathcal{G}_{\mathbf{Z}} A f_h(\boldsymbol{x})$. Set $\bar{x}_i = x_i\Sigma - p_i$ and observe by Lemma B.1 that

$$\begin{aligned}a_{ij}(\boldsymbol{x}) &= \bar{x}_i\bar{x}_j + \frac{1}{N}\big(x_i - \bar{x}_i\big)\big(\delta_{ij} - (x_j - \bar{x}_j)\big) \\ &= \frac{1}{N}x_i(\delta_{ij} - x_j) - \frac{1}{N}\Big(\bar{x}_i(\delta_{ij} - x_j) - x_i\bar{x}_j - \bar{x}_i\bar{x}_j\Big) + \bar{x}_i\bar{x}_j.\end{aligned}$$

Since $\delta = 1/N$, $|\bar{x}_i| \le \Sigma \le \delta s$, and $|x_i| \le 1$, it follows that

$$\begin{aligned}&\Big|\sum_{i=1}^{K-1} b_i(\boldsymbol{x})\partial_i A f_h(\boldsymbol{x}) + \frac{1}{2}\sum_{i,j=1}^{K-1} a_{ij}(x)\partial_i\partial_j A f_h(\boldsymbol{x}) - \delta\mathcal{G}_{\mathbf{Z}} A f_h(\boldsymbol{x})\Big| \\ &\quad= \Big|\frac{1}{2}\sum_{i,j=1}^{K-1}\Big(-\frac{1}{N}\Big(\bar{x}_i(\delta_{ij} - x_j) - x_i\bar{x}_j - \bar{x}_i\bar{x}_j\Big) + \bar{x}_i\bar{x}_j\Big)\partial_i\partial_j A f_h(\boldsymbol{x})\Big| \\ &\quad\le \frac{1}{2}\sum_{i,j=1}^{K-1}\big(\delta(2\Sigma + \Sigma^2) + \Sigma^2\big)C_{A.4}(K)\delta^{-2}B_2(f_h) \\ &\quad\le K^2 2(1+s)^2 C_{A.4}(K)B_2(f_h).\end{aligned}$$

Combining this bound with the bounds on $|\tilde{\epsilon}_i(\boldsymbol{x})|, \ldots, |A\epsilon(\boldsymbol{x})|$ from Lemma B.2 yields

$$\begin{aligned}|\epsilon_{\mathcal{G}}(\boldsymbol{x})| \le\ & K^2 2(1+s)^2 C_{A.4}(K)B_2(f_h) \\ &+ K\delta s C_{A.4}(K)\big(C_{A.10}(K)B_2(f_h) + C_{A.11}(K)\delta^{-1}B_4(f_h)\big) \\ &+ K^2(1+s)^2 C_{A.4}(K)\big(5C_{A.10}(K)B_3(f_h) + C_{A.11}(K)\delta^{-1}B_4(f_h)\big) \\ &+ K^3 1595^K 20(1+s)^3 C_{A.4}(K)\delta^{-1}B_3(f_h) \\ &+ 1595^K K^4 C_{A.4}(K)\Big((2+s)^4\delta^{-2}B_4(f_h) \\ &\qquad + \frac{(4K+1)N^{4K}}{(\Sigma - p_K)^{4K}}\Big(1 - \frac{10K(\sqrt{N}-1)}{N}(1-\Sigma)\Big)^N\delta^{-5}16KB_1(f_h)\Big).\end{aligned}$$

□

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

We first state and prove an auxiliary lemma, followed by the proof of Lemma B.2.

Lemma B.3. *For any $\boldsymbol{u} \in \nabla^K$ and any integer $M \geq 0$,*

$$\mathbb{P}_{\boldsymbol{u}}(U_K(1) \leq M/N) \leq \frac{(M+1)N^M}{(\Sigma - p_K)^M}(1 - u_K(1-\Sigma))^N.$$

Proof of Lemma B.3. Recall that from (1.1), $NU_K(1)|u_K \sim \text{Binomial}(N, u_K - \bar{u}_K)$, where $\bar{u}_K = -\Sigma u_K + p_K$. For any $0 \leq j \leq M$,

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{u}}(U_K(1) = j/N) &= \frac{N!}{j!(N-j)!}(u_K(1-\Sigma) + p_K)^j(1 - u_K(1-\Sigma) - p_K)^{N-j} \\
&\leq N^j(1 - u_K(1-\Sigma) - p_K)^N \frac{1}{(1 - u_K(1-\Sigma) - p_K)^j} \\
&\leq N^M(1 - u_K(1-\Sigma))^N \frac{1}{(\Sigma - p_K)^j} \\
&\leq \frac{N^M}{(\Sigma - p_K)^M}(1 - u_K(1-\Sigma))^N,
\end{aligned}$$

where the second inequality follows from $1 - u_K(1-\Sigma) - p_K \geq \Sigma - p_K$ since $u_K \leq 1$. Thus,

$$\mathbb{P}_{\boldsymbol{u}}(U_K(1) \leq M/N) = \sum_{j=0}^{M} \mathbb{P}_{\boldsymbol{u}}(U_K(1) = j/N) \leq \frac{(M+1)N^M}{(\Sigma - p_K)^M}(1 - u_K(1-\Sigma))^N.$$

□

Proof of Lemma B.2. Let us prove (B.3). Note that $Ab_i(\boldsymbol{x}) = b_i(\boldsymbol{x})$ by Corollary A.1. Lemma A.2 then gives us

$$\begin{aligned}
A(b_i\partial_i A f_h)(\boldsymbol{x}) &= Ab_i(x)A(\partial_i A f_h)(\boldsymbol{x}) + \epsilon_i(\boldsymbol{x}) \\
&= b_i(\boldsymbol{x})\partial_i A f_h(\boldsymbol{x}) + \epsilon_i(\boldsymbol{x}) + b_i(\boldsymbol{x})\big(A(\partial_i A f_h)(\boldsymbol{x}) - \partial_i A f_h(\boldsymbol{x})\big),
\end{aligned}$$

where

$$\begin{aligned}
|\epsilon_i(\boldsymbol{x})| &\leq C_{A.10}(K) \max_{\substack{\|\boldsymbol{a}\|_1 = 1 \\ 0 \leq \boldsymbol{j} \leq 4\boldsymbol{e} - \boldsymbol{a}}} |\Delta^{\boldsymbol{a}} b_i(\delta(k(\boldsymbol{x}) + \boldsymbol{j}))| \max_{\substack{\|\boldsymbol{a}\|_1 = 1 \\ 0 \leq \boldsymbol{j} \leq 4\boldsymbol{e} - \boldsymbol{a}}} |\Delta^{\boldsymbol{a}} \partial_i A f_h(\delta(k(\boldsymbol{x}) + \boldsymbol{j}))| \\
&\leq C_{A.10}(K)\delta^2 s \max_{\substack{\|\boldsymbol{a}\|_1 = 1 \\ 0 \leq \boldsymbol{j} \leq 4\boldsymbol{e} - \boldsymbol{a}}} |\Delta^{\boldsymbol{a}} \partial_i A f_h(\delta(k(\boldsymbol{x}) + \boldsymbol{j}))|.
\end{aligned}$$

The constant $C_{A.10}(K)$ is as in Lemma A.2 and the second inequality is due to the bound on $B_1(b_i)$ from Lemma B.1. Now

$$\begin{aligned}
\max_{\substack{\|\boldsymbol{a}\|_1 = 1 \\ 0 \leq \boldsymbol{j} \leq 4\boldsymbol{e} - \boldsymbol{a}}} |\Delta^{\boldsymbol{a}} \partial_i A f_h(\delta(k(\boldsymbol{x}) + \boldsymbol{j}))| &\leq \sup_{\substack{\|\boldsymbol{a}\|_1 = 2 \\ 0 \leq z \leq 5\delta\boldsymbol{e}}} |\delta D^{\boldsymbol{a}} A f_h(\delta k(\boldsymbol{x}) + z)| \\
&\leq C_{A.4}(K) \max_{\substack{\|\boldsymbol{a}\|_1 = 2 \\ 0 \leq \boldsymbol{j} \leq 9\boldsymbol{e} - \boldsymbol{a}}} |\delta\delta^{-2}\Delta^{\boldsymbol{a}} f_h(\delta(k(\boldsymbol{x}) + \boldsymbol{j}))| \\
&\leq C_{A.4}(K)\delta^{-1} B_2(f_h),
\end{aligned}$$

where the second-last inequality is due to (A.3) of Theorem A.1. Thus,

$$|\epsilon_i(\boldsymbol{x})| \leq C_{A.10}(K)\delta s C_{A.4}(K) B_2(f_h).$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

To complete the proof of (B.3), note that

$$\begin{aligned}|b_i(\boldsymbol{x})\big(A(\partial_i Af_h)(\boldsymbol{x})-\partial_i Af_h(\boldsymbol{x})\big)| &\le \delta s|A(\partial_i Af_h)(\boldsymbol{x})-\partial_i Af_h(x)|\\ &\le \delta s C_{A.11}(K)\delta^3 \max_{\|\boldsymbol{a}\|_1=4}\sup_{0\le z\le 5\delta \boldsymbol{e}}|D^{\boldsymbol{a}}Af_h(\delta k(\boldsymbol{x})+z)|\\ &\le \delta s C_{A.11}(K)C_{A.4}(K)\delta^{-1}\max_{\substack{\|\boldsymbol{a}\|_1=4\\ 0\le \boldsymbol{j}\le 9\boldsymbol{e}-\boldsymbol{a}}}|\Delta^{\boldsymbol{a}}f_h(\delta(k(\boldsymbol{x})+\boldsymbol{j}))|\\ &\le sC_{A.11}(K)C_{A.4}(K)B_4(f_h).\end{aligned}$$

The first inequality follows from $|b_i|(\boldsymbol{u})\le\Sigma\le\delta s$; see Lemma B.1. The second inequality follows from Proposition A.1 with $f(\boldsymbol{x})=\partial_i Af_h(\boldsymbol{x})$ and $s=3$ there, and the third inequality follows from (A.3) of Theorem A.1. The proof of (B.4) is similar to that of (B.3). Lemma A.2 and the fact that $Aa_{ij}(\boldsymbol{x})=a_{ij}(\boldsymbol{x})$ imply that

$$\begin{aligned}A(a_{ij}\partial_i\partial_j Af_h)(\boldsymbol{x}) &= Aa_{ij}(\boldsymbol{x})A(\partial_i\partial_j Af_h)(x)+\epsilon_{ij}(\boldsymbol{x})\\ &= a_{ij}(x)\partial_i\partial_j Af_h(\boldsymbol{x})+\epsilon_{ij}(\boldsymbol{x})+a_{ij}(\boldsymbol{x})\big(A(\partial_i\partial_j Af_h)(\boldsymbol{x})-\partial_i\partial_j Af_h(\boldsymbol{x})\big),\end{aligned}$$

where

$$|\epsilon_{ij}(\boldsymbol{x})|\le C_{A.10}(K)\max_{\substack{\|\boldsymbol{a}\|_1=1\\ 0\le\boldsymbol{j}\le 4\boldsymbol{e}-\boldsymbol{a}}}|\Delta^{\boldsymbol{a}}a_{ij}(\delta(k(\boldsymbol{x})+\boldsymbol{j}))|\max_{\substack{\|\boldsymbol{a}\|_1=1\\ 0\le\boldsymbol{j}\le 4\boldsymbol{e}-\boldsymbol{a}}}|\Delta^{\boldsymbol{a}}\partial_i\partial_j Af_h(\delta(k(\boldsymbol{x})+\boldsymbol{j}))|.$$

We know that $B_1(a_{ij})\le\delta^2 5(1+s)^2$ by Lemma B.1, so that repeating the arguments used to bound $\epsilon_i(\boldsymbol{x})$ yields

$$\begin{aligned}|\epsilon_{ij}(\boldsymbol{x})| &\le C_{A.10}(K)\delta^2 5(1+s)^2\max_{\substack{\|\boldsymbol{a}\|_1=1\\ 0\le\boldsymbol{j}\le 4\boldsymbol{e}-\boldsymbol{a}}}|\Delta^{\boldsymbol{a}}\partial_i\partial_j Af_h(\delta(k(\boldsymbol{x})+\boldsymbol{j}))|\\ &\le C_{A.10}(K)\delta^2 5(1+s)^2C_{A.4}(K)\delta\delta^{-3}B_3(f_h)\\ &= C_{A.10}(K)5(1+s)^2C_{A.4}(K)B_3(f_h).\end{aligned}$$

Furthermore, since $|a_{ij}(\boldsymbol{x})|\le\Sigma^2+\delta\le\delta(\delta s^2+1)$,

$$\begin{aligned}&|a_{ij}(\boldsymbol{x})\big(A(\partial_i\partial_j Af_h)(\boldsymbol{x})-\partial_i\partial_j Af_h(\boldsymbol{x})\big)|\\ &\quad\le\delta(\delta s^2+1)|A(\partial_i\partial_j Af_h)(\boldsymbol{x})-\partial_i\partial_j Af_h(\boldsymbol{x})|\\ &\quad\le\delta(\delta s^2+1)C_{A.11}(K)\delta^2\max_{\|\boldsymbol{a}\|_1=4}\sup_{0\le z\le 5\delta\boldsymbol{e}}|D^{\boldsymbol{a}}Af_h(\delta k(\boldsymbol{x})+z)|\\ &\quad\le\delta(\delta s^2+1)C_{A.11}(K)C_{A.4}(K)\delta^2\delta^{-4}\max_{\substack{\|\boldsymbol{a}\|_1=4\\ 0\le\boldsymbol{j}\le 9\boldsymbol{e}}}|\Delta^{\boldsymbol{a}}f_h(\delta(k(\boldsymbol{x})+\boldsymbol{j}))|\\ &\quad\le(1+s)^2C_{A.11}(K)C_{A.4}(K)\delta^{-1}B_4(f_h),\end{aligned}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

which proves (B.4). Let us prove (B.5). Note that

$$\begin{aligned}
&|A(c_{ijk}\partial_i\partial_j\partial_k Af_h)(\boldsymbol{x})| \\
&\quad= \Bigg| \sum_{i_1,\dots,i_{K-1}=0}^{4} \Bigg(\prod_{j=1}^{K-1} \alpha_{k_j(\boldsymbol{x})+i_j}^{k_j(\boldsymbol{x})}(x_j) \Bigg) c_{ijk}(\delta(k(\boldsymbol{x})+\boldsymbol{i}))\partial_i\partial_j\partial_k Af_h(\delta(k(\boldsymbol{x})+\boldsymbol{i})) \Bigg| \\
&\quad\le 319^K 20\delta^2(1+s)^3 \sum_{i_1,\dots,i_{K-1}=0}^{4} |\partial_i\partial_j\partial_k Af_h(\delta(k(\boldsymbol{x})+\boldsymbol{i}))| \\
&\quad\le 319^K 20\delta^2(1+s)^3 C_{A.4}(K)\delta^{-3} \sum_{i_1,\dots,i_{K-1}=0}^{4} \max_{\substack{\|\boldsymbol{a}\|_1=3 \\ 0\le \boldsymbol{j}\le 4\boldsymbol{e}-\boldsymbol{a}}} |\Delta^{\boldsymbol{a}} f_h(\delta(k(\boldsymbol{x})+\boldsymbol{i}+\boldsymbol{j}))| \\
&\quad\le 319^K 20\delta^2(1+s)^3 5^K C_{A.4}(K)\delta^{-3} B_3(f_h) \\
&\quad= 1595^K 20\delta^2(1+s)^3 C_{A.4}(K)\delta^{-3} B_3(f_h),
\end{aligned}$$

where the first inequality is due to (A.8) and Lemma B.1, and the second inequality is due to (A.3) of Theorem A.1, and the final inequality is due to (B.1). Lastly, to prove (B.6), we bound $|A\epsilon(\boldsymbol{x})|$ by bounding $\epsilon(\boldsymbol{u})$, $\boldsymbol{u}\in\nabla^K$ and using (A.8).

From (A.1) of Theorem A.1, we see that $A\epsilon(\boldsymbol{x})$ depends only on $\epsilon(\delta(k(\boldsymbol{x})+\boldsymbol{i}))$ for $0\le\boldsymbol{i}\le 4\boldsymbol{e}$, so going forward we fix $\boldsymbol{u}=\delta(k(\boldsymbol{x})+\boldsymbol{i})$. Recall that

$$\epsilon(\boldsymbol{u}) = \frac{1}{6}\sum_{i,j,k=1}^{K-1} \mathbb{E}_{\boldsymbol{u}}\Big((U_i(1)-u_i)(U_j(1)-u_j)(U_k(1)-u_k)\big(\partial_i\partial_j\partial_k Af_h(\boldsymbol{\xi}) - \partial_i\partial_j\partial_k Af_h(\boldsymbol{u})\big)\Big),$$

where $\boldsymbol{\xi}$ is between $\boldsymbol{u}$ and $\boldsymbol{U}(1)$. We know that $\partial_i\partial_j\partial_k Af_h\in\mathcal{D}^{K-1}$ by Lemma A.1, implying that

$$|\partial_i\partial_j\partial_k Af_h(\boldsymbol{\xi}) - \partial_i\partial_j\partial_k Af_h(\boldsymbol{u})| \le \|\boldsymbol{U}(1)-\boldsymbol{u}\|_1 \sup_{\substack{\min(\boldsymbol{u},\boldsymbol{U}(1))\le\boldsymbol{z} \\ \boldsymbol{z}\le\max(\boldsymbol{u},\boldsymbol{U}(1)) \\ \|\boldsymbol{a}\|_1=4}} |D^{\boldsymbol{a}} Af_h(\boldsymbol{z})|.$$

Consider separately the cases when $U_K(1) > 4K/N$ and $U_K(1)\le 4K/N$. If $U_K(1) > 4K/N$ then

$$\begin{aligned}
\sup_{\substack{\min(\boldsymbol{u},\boldsymbol{U}(1))\le\boldsymbol{z} \\ \boldsymbol{z}\le\max(\boldsymbol{u},\boldsymbol{U}(1)) \\ \|\boldsymbol{a}\|_1=4}} |D^{\boldsymbol{a}} Af_h(\boldsymbol{z})| &\le C_{A.4}(K)\delta^{-4} \sup_{\substack{\min(\boldsymbol{u},\boldsymbol{U}(1))\le\boldsymbol{z} \\ \boldsymbol{z}\le\max(\boldsymbol{u},\boldsymbol{U}(1)) \\ \|\boldsymbol{a}\|_1=4}} \max_{0\le\boldsymbol{i}\le 4\boldsymbol{e}-\boldsymbol{a}} |\Delta^{\boldsymbol{a}} f_h(\delta(k(\boldsymbol{z})+\boldsymbol{i}))| \\
&\le C_{A.4}(K)\delta^{-4} B_4(f_h).
\end{aligned}$$

For the last inequality, we used the facts that $U(1)+\delta\boldsymbol{a}\in\nabla^K$ and $\boldsymbol{u}+\delta\boldsymbol{a}\in\nabla^K$ for $\|a\|_1\le 4$, because $U_K(1) > 4K/N$ and $\boldsymbol{x}\in\bar{\nabla}_N^K$, respectively. If $U_K(1)\le 4K/N$, we use the coarse bound of

$$\begin{aligned}
\sup_{\substack{\min(\boldsymbol{u},\boldsymbol{U}(1))\le\boldsymbol{z} \\ \boldsymbol{z}\le\max(\boldsymbol{u},\boldsymbol{U}(1)) \\ \|\boldsymbol{a}\|_1=4}} |D^{\boldsymbol{a}} Af_h(\boldsymbol{z})| &\le C_{A.4}(K)\delta^{-4} \sup_{\substack{\min(\boldsymbol{u},\boldsymbol{U}(1))\le\boldsymbol{z} \\ \boldsymbol{z}\le\max(\boldsymbol{u},\boldsymbol{U}(1)) \\ \|\boldsymbol{a}\|_1=4}} \max_{0\le\boldsymbol{i}\le 4\boldsymbol{e}-\boldsymbol{a}} |\Delta^{\boldsymbol{a}} f_h(\delta(k(\boldsymbol{z})+\boldsymbol{i}))| \\
&\le C_{A.4}(K)\delta^{-4} 16\|f_h\| \\
&\le C_{A.4}(K)\delta^{-5} 16KB_1(f_h).
\end{aligned}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

The last inequality is due to the fact that

$$\|f_h\| \le KNB_1(f_h) = K\delta^{-1}B_1(f_h), \tag{B.7}$$

which we verify at the end of the proof. Combining the bounds on $|D^{\boldsymbol{a}}Af_h(\boldsymbol{z})|$ with the definition of $\epsilon(\boldsymbol{u})$ yields

$$\begin{aligned}
|\epsilon(\boldsymbol{u})| &\le \frac{1}{6}\sum_{i,j,k=1}^{K-1}\mathbb{E}_{\boldsymbol{u}}\Big(\big|(U_i(1)-u_i)(U_j(1)-u_j)(U_k(1)-u_k)\|\boldsymbol{U}(1)-\boldsymbol{u}\|_1\big|\Big)C_{A.4}(K)\delta^{-4}B_4(f_h)\\
&\quad+\frac{1}{6}\sum_{i,j,k=1}^{K-1}\mathbb{E}_{\boldsymbol{u}}\Big(\big|(U_i(1)-u_i)(U_j(1)-u_j)(U_k(1)-u_k)\|\boldsymbol{U}(1)-\boldsymbol{u}\|_1\big|\\
&\qquad\qquad\times C_{A.4}(K)1(U_K(1)\le 4K/N)\delta^{-5}16KB_1(f_h)\Big)\\
&\le K^4\max_{1\le i,j,k,\ell\le K-1}\bar{d}_{ijk\ell}(\boldsymbol{u})C_{A.4}(K)\delta^{-4}B_4(f_h)+K^4\mathbb{P}_{\boldsymbol{u}}(U_K(1)\le 4K/N)C_{A.4}(K)\delta^{-5}16KB_1(f_h)\\
&\le K^4C_{A.4}(K)\bigg(\delta^2(2+s)^4\delta^{-4}B_4(f_h)+\frac{(4K+1)N^{4K}}{(\Sigma-p_K)^{4K}}(1-u_K(1-\Sigma))^N\delta^{-5}16KB_1(f_h)\bigg).
\end{aligned}$$

In the second-last inequality we used the coarse bound of $|U_i(1)-u_i|\le 1$ to bound the second term, and in the last inequality we used the bound on $\bar{d}_{ijk\ell}(\boldsymbol{u})$ from Lemma B.1. Furthermore, $\boldsymbol{x}\in\bar{\nabla}_N^K$ implies that, for any $0\le\boldsymbol{i}\le 4\boldsymbol{e}$,

$$1-\sum_{j=1}^{K-1}\delta(k_j(\boldsymbol{x})+i_j)\ge 1-\sum_{j=1}^{K-1}\delta k_j(\boldsymbol{x})-4K/N\ge 10K\sqrt{N}/N-4K/N\ge 10K(\sqrt{N}-1)/N,$$

and, therefore, $(1-u_K(1-\Sigma))^N$ on the right-hand side is bounded by $(1-(10K(\sqrt{N}-1)/N)(1-\Sigma))^N$.

Finally, (A.8) yields the upper bound of

$$|A\epsilon(\boldsymbol{x})|=\bigg|\sum_{i_1,\dots,i_{K-1}=0}^{4}\bigg(\prod_{j=1}^{d}\alpha_{k_j(\boldsymbol{x})+i_j}^{k_j(\boldsymbol{x})}(x_j)\bigg)\epsilon(\delta(k(\boldsymbol{x})+\boldsymbol{i}))\bigg|\le 5^K319^K\max_{0\le\boldsymbol{i}\le 4\boldsymbol{e}}|\epsilon(\delta(k(\boldsymbol{x})+\boldsymbol{i}))|,$$

which proves (B.6). It remains to prove (B.7). Since we chose $f_h(\boldsymbol{u})=0$ for $\boldsymbol{u}$ outside of ∇^K, we need to show that

$$|f_h(\boldsymbol{u})|\le NKB_1(f_h),\quad \boldsymbol{u}\in\nabla^K.$$

Letting $\pi(\boldsymbol{u})=\mathbb{P}(\boldsymbol{U}=\boldsymbol{u})$, we have

$$\begin{aligned}
f_h(\boldsymbol{u})=\sum_{n=0}^{\infty}\big(\mathbb{E}_u h(\boldsymbol{U}(n))-\mathbb{E}h(\boldsymbol{U})\big)&=\sum_{n=0}^{\infty}\sum_{\boldsymbol{u}'\in\nabla^K}\pi(\boldsymbol{u}')\big(\mathbb{E}_{\boldsymbol{u}}h(\boldsymbol{U}(n))-\mathbb{E}_{\boldsymbol{u}'}h(\boldsymbol{U}(n))\big)\\
&=\sum_{\boldsymbol{u}'\in\nabla^K}\sum_{n=0}^{\infty}\pi(\boldsymbol{u}')\big(\mathbb{E}_{\boldsymbol{u}}h(\boldsymbol{U}(n))-\mathbb{E}_{\boldsymbol{u}'}h(\boldsymbol{U}(n))\big),
\end{aligned}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

where the interchange is justified by the Fubini–Tonelli theorem because $\{\boldsymbol{U}(n)\}$ is geometrically ergodic. Since

$$\Delta_i f_h(\boldsymbol{u}) = \sum_{n=0}^{\infty} \big(\mathbb{E}_{\boldsymbol{u}+\delta \boldsymbol{e}_i} h(U(n)) - \mathbb{E}_{\boldsymbol{u}} h(\boldsymbol{U}(n))\big),$$

it follows that

$$\sum_{n=0}^{\infty} |\mathbb{E}_{\boldsymbol{u}} h(\boldsymbol{U}(n)) - \mathbb{E}_{\boldsymbol{u}'} h(\boldsymbol{U}(n))| \le \|(\boldsymbol{u}-\boldsymbol{u}')/\delta\|_1 B_1(f_h) \le NKB_1(f_h),$$

proving (B.7). □

We now prove Lemma B.1. Recall that

$$((NU_1(1), \ldots, NU_K(1)) | U(0) = \boldsymbol{u}) \sim \text{Multinomial}(N, \big(u_1 - \bar{u}_1, \ldots, u_K - \bar{u}_K)),$$

where $\bar{u}_i = u_i \Sigma - p_i$. We also require the following result about the moments of the multinomial distribution.

Lemma B.4. *Let* $X \sim \text{Multinomial}(N, (p_1, \ldots, p_k))$. *Then for all* $i \ne j \ne k$,

$$\mathbb{E}(X_i) = Np_i, \qquad \mathbb{E}(X_i^2) = N(N-1)p_i^2 + Np_i, \qquad \mathbb{E}(X_iX_j) = N(N-1)p_ip_j,$$
$$\mathbb{E}(X_iX_jX_k) = N(N-1)(N-2)p_ip_jp_k,$$
$$\mathbb{E}(X_i^2X_j) = N(N-1)(N-2)p_i^2p_j + N(N-1)p_ip_j,$$
$$\mathbb{E}(X_i^3) = N(N-1)(N-2)p_i^3 + 3N(N-1)p_i^2 - 2Np_i.$$

Proof of Lemma B.1. For convenience, we write $\mathbb{E}(\cdot)$ instead of $\mathbb{E}_u(\cdot)$ and U_i instead of $U_i(1)$; e.g. we write $\mathbb{E}U_i$ instead of $\mathbb{E}_u U_i(1)$. Since $\mathbb{E}NU_i = N(u_i - \bar{u}_i)$, it follows from Lemma B.4 that

$$b_i(\boldsymbol{u}) = \mathbb{E}(U_i - u_i) = -\bar{u}_i.$$

Next, we observe that

$$a_{ij}(\boldsymbol{u}) = \mathbb{E}(U_i - u_i)(U_j - u_j) = \mathbb{E}(U_iU_j) - u_i\mathbb{E}U_j - u_j\mathbb{E}U_i + u_iu_j.$$

Lemma B.4 implies that $\mathbb{E}U_j = (u_j - \bar{u}_j)$ and

$$\begin{aligned}\mathbb{E}(U_iU_j) &= (1 - 1/N)(u_i - \bar{u}_i)(u_j - \bar{u}_j)\\ &= u_iu_j - u_i\bar{u}_j - u_j\bar{u}_i + \bar{u}_i\bar{u}_j - \frac{1}{N}(u_i - \bar{u}_i)(u_j - \bar{u}_j),\end{aligned}$$

from which it follows that

$$a_{ij}(\boldsymbol{u}) = \bar{u}_i\bar{u}_j - \frac{1}{N}(u_i - \bar{u}_i)(u_j - \bar{u}_j). \tag{B.8}$$

Similarly,

$$a_{ii}(\boldsymbol{u}) = \mathbb{E}(U_i - u_i)^2 = \mathbb{E}(U_i^2) - 2u_i\mathbb{E}U_i + u_i^2 = \mathbb{E}(U_i^2) - 2u_i(u_i - \bar{u}_i) + u_i^2.$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

By Lemma B.4, we know that

$$\begin{aligned}\mathbb{E}(U_i^2) &= (1-1/N)(u_i-\bar{u}_i)^2 + \frac{1}{N}(u_i-\bar{u}_i)\\ &= (u_i^2 - 2u_i\bar{u}_i + \bar{u}_i^2) + \frac{1}{N}\big((u_i-\bar{u}_i) - (u_i-\bar{u}_i)^2\big),\end{aligned}$$

and we conclude that

$$a_{ii}(u) = \bar{u}_i^2 + \frac{1}{N}\big((u_i-\bar{u}_i) - (u_i-\bar{u}_i)^2\big). \tag{B.9}$$

The form of $b_i(\boldsymbol{u})$ immediately implies that $B_1(b_i) = \delta\Sigma \le \delta^2 s$. Furthermore, one can check that

$$\begin{aligned}a_{ii}(\boldsymbol{u}) = u_i^2\big(\Sigma^2 - (1/N)(1-\Sigma)^2\big) + u_i\big(-2\Sigma p_i + (1/N)(1-\Sigma)(1-2p_i)\big)\\ + p_i^2 + (1/N)p_i(1-p_i),\end{aligned}$$

from which it follows that

$$\begin{aligned}|\Delta_i a_{ii}(\boldsymbol{u})| &= \big|(2u_i\delta+\delta^2)\big(\Sigma^2 - (1/N)(1-\Sigma)^2\big) + \delta\big(-2\Sigma p_i + (1/N)(1-\Sigma)(1-2p_i)\big)\big|\\ &\le \delta^2 5(1+s)^2.\end{aligned}$$

The bound is crude but simple, and follows from the facts that $u_i \le 1$, $\delta = 1/N \le 1$, and $p_i \le \Sigma \le \delta s$. Similarly, one can show that

$$|\Delta_i a_{ij}(\boldsymbol{u})|, |\Delta_j a_{ij}(\boldsymbol{u})| \le \delta^2 5(1+s)^2, \quad i \ne j,$$

implying the bound on $B_1(a_{ij})$.

To prove $|c_{ijk}(\boldsymbol{u})| \le 20\delta^2(1+s)^3$, we compute and then bound $c_{iii}(\boldsymbol{u})$, $c_{iij}(\boldsymbol{u})$, and $c_{ijk}(\boldsymbol{u})$, for $i \ne j \ne k$. We begin with

$$\begin{aligned}c_{iii}(\boldsymbol{u}) &= \mathbb{E}(U_i-u_i)^3 = \mathbb{E}U_i(U_i-u_i)^2 - u_i\mathbb{E}(U_i-u_i)^2\\ &= \mathbb{E}(U_i^3 - 2u_iU_i^2 + u_i^2U_i) - u_i\mathbb{E}(U_i-u_i)^2.\end{aligned}$$

From (B.9), we know that

$$-u_i\mathbb{E}(U_i-u_i)^2 = -u_i\bar{u}_i^2 - \frac{1}{N}u_i\big((u_i-\bar{u}_i) - (u_i-\bar{u}_i)^2\big).$$

For the remaining terms, we use Lemma B.4 to see that

$$\begin{aligned}\mathbb{E}U_i^3 &= (1-1/N)(1-2/N)(u_i-\bar{u}_i)^3 + \frac{3}{N}(1-1/N)(u_i-\bar{u}_i)^2 - \frac{2}{N^2}(u_i-\bar{u}_i)\\ &= (u_i-\bar{u}_i)^3 - \frac{3}{N}(u_i-\bar{u}_i)^3 + \frac{2}{N^2}(u_i-\bar{u}_i)^3 + \frac{3}{N}(u_i-\bar{u}_i)^2 - \frac{3}{N^2}(u_i-\bar{u}_i)^2 - \frac{2}{N^2}(u_i-\bar{u}_i),\\ -2u_i\mathbb{E}U_i^2 &= -2u_i\Big((1-1/N)(u_i-\bar{u}_i)^2 + \frac{1}{N}(u_i-\bar{u}_i)\Big)\\ &= -2u_i(u_i-\bar{u}_i)^2 + 2u_i\frac{1}{N}(u_i-\bar{u}_i)^2 - 2u_i\frac{1}{N}(u_i-\bar{u}_i),\\ u_i^2\mathbb{E}U_i &= u_i^2(u_i-\bar{u}_i).\end{aligned}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

We begin with the terms with a $1/N^2$ factor. Since $0 \le u_i - \bar{u}_i \le 1$, it follows that

$$\left| \frac{2}{N^2}(u_i - \bar{u}_i)^3 - \frac{3}{N^2}(u_i - \bar{u}_i)^2 - \frac{2}{N^2}(u_i - \bar{u}_i) \right| \le \frac{5}{N^2}.$$

Let us now consider only those containing $1/N$ in front of them. Namely,

$$\begin{aligned}
&-\frac{3}{N}(u_i - \bar{u}_i)^3 + \frac{3}{N}(u_i - \bar{u}_i)^2 + 2u_i\frac{1}{N}(u_i - \bar{u}_i)^2 - 2u_i\frac{1}{N}(u_i - \bar{u}_i) - \frac{1}{N}u_i\big((u_i - \bar{u}_i) - (u_i - \bar{u}_i)^2\big) \\
&\quad = \frac{3}{N}\Big(-(u_i - \bar{u}_i)^3 + (u_i - \bar{u}_i)^2 + u_i(u_i - \bar{u}_i)^2 - u_i(u_i - \bar{u}_i)\Big) \\
&\quad = \frac{3}{N}\Big(\bar{u}_i(u_i - \bar{u}_i)^2 + (u_i - \bar{u}_i)^2 - u_i(u_i - \bar{u}_i)\Big) \\
&\quad = \frac{3}{N}\Big(\bar{u}_i(u_i - \bar{u}_i)^2 - \bar{u}_i(u_i - \bar{u}_i)\Big) \\
&\quad = \frac{3}{N}\bar{u}_i(u_i - \bar{u}_i)\Big((u_i - \bar{u}_i) - 1\Big),
\end{aligned}$$

and note that

$$\left| \frac{3}{N}\bar{u}_i(u_i - \bar{u}_i)\Big((u_i - \bar{u}_i) - 1\Big) \right| \le \frac{3\Sigma}{4N} = \frac{3s}{8N^2}.$$

Lastly, we collect all the terms without $1/N$ or $1/N^2$ in front of them. Their sum equals

$$\begin{aligned}
&(u_i - \bar{u}_i)^3 - 2u_i(u_i - \bar{u}_i)^2 + u_i^2(u_i - \bar{u}_i) - u_i\bar{u}_i^2 \\
&\quad = u_i^3 - 3\bar{u}_i u_i^2 + 3\bar{u}_i^2 u_i - \bar{u}_i^3 - 2u_i^3 + 4u_i^2\bar{u}_i - 2u_i\bar{u}_i^2 + u_i^3 - u_i^2\bar{u}_i - u_i\bar{u}_i^2 \\
&\quad = -3\bar{u}_i u_i^2 + 3\bar{u}_i^2 u_i - \bar{u}_i^3 + 4u_i^2\bar{u}_i - 2u_i\bar{u}_i^2 - u_i^2\bar{u}_i - u_i\bar{u}_i^2 \\
&\quad = 3\bar{u}_i^2 u_i - \bar{u}_i^3 - 2u_i\bar{u}_i^2 - u_i\bar{u}_i^2 \\
&\quad = -\bar{u}_i^3.
\end{aligned}$$

Since $|\bar{u}_i^3| \le \Sigma^3 = s^3/(8N^3)$, we have shown that $|c_{iii}(u)| \le 5/N^2 + 3s/(8N^2) + s^3/(8N^3) \le 20\delta^2(1+s)^3$. Next, we bound

$$\begin{aligned}
c_{iij}(\boldsymbol{u}) &= \mathbb{E}(U_i - u_i)^2(U_j - u_j) \\
&= \mathbb{E}U_i(U_i - u_i)(U_j - u_j) - u_i\mathbb{E}(U_i - u_i)(U_j - u_j) \\
&= \mathbb{E}U_i^2 U_j - u_j\mathbb{E}U_i^2 - u_i\mathbb{E}U_i U_j + u_i u_j\mathbb{E}U_i - u_i\mathbb{E}(U_i - u_i)(U_j - u_j).
\end{aligned}$$

From (B.8), we know that

$$-u_i\mathbb{E}(U_i - u_i)(U_j - u_j) = -u_i\Big(\bar{u}_i\bar{u}_j - \frac{1}{N}(u_i - \bar{u}_i)(u_j - \bar{u}_j)\Big),$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

and, for the rest of the terms, we use Lemma B.4 to obtain

$$\begin{aligned}
\mathbb{E}U_i^2U_j &= (1-1/N)(1-2/N)(u_i-\bar{u}_i)^2(u_j-\bar{u}_j)+\frac{1}{N}(1-1/N)(u_i-\bar{u}_i)(u_j-\bar{u}_j)\\
&= (u_i-\bar{u}_i)^2(u_j-\bar{u}_j)-\frac{3}{N}(u_i-\bar{u}_i)^2(u_j-\bar{u}_j)+\frac{2}{N^2}(u_i-\bar{u}_i)^2(u_j-\bar{u}_j)\\
&\quad+\frac{1}{N}(u_i-\bar{u}_i)(u_j-\bar{u}_j)-\frac{1}{N^2}(u_i-\bar{u}_i)(u_j-\bar{u}_j),\\
-u_j\mathbb{E}U_i^2 &= -u_j\Big((1-1/N)(u_i-\bar{u}_i)^2+\frac{1}{N}(u_i-\bar{u}_i)\Big)\\
&= -u_j\Big((u_i-\bar{u}_i)^2-\frac{1}{N}(u_i-\bar{u}_i)^2+\frac{1}{N}(u_i-\bar{u}_i)\Big),\\
-u_i\mathbb{E}U_iU_j &= -u_i\Big((u_i-\bar{u}_i)(u_j-\bar{u}_j)-\frac{1}{N}(u_i-\bar{u}_i)(u_j-\bar{u}_j)\Big),\\
u_iu_j\mathbb{E}U_i &= u_iu_j(u_i-\bar{u}_i).
\end{aligned}$$

Considering the $1/N^2$ terms,

$$\left|\frac{2}{N^2}(u_i-\bar{u}_i)^2(u_j-\bar{u}_j)-\frac{1}{N^2}(u_i-\bar{u}_i)(u_j-\bar{u}_j)\right|\le\frac{1}{N^2}.$$

Collecting all terms with $1/N$ in front, the result equals

$$\begin{aligned}
&\frac{1}{N}(-3(u_i-\bar{u}_i)^2(u_j-\bar{u}_j)+(u_i-\bar{u}_i)(u_j-\bar{u}_j)+u_j(u_i-\bar{u}_i)^2-u_j(u_i-\bar{u}_i)\\
&\quad+u_i(u_i-\bar{u}_i)(u_j-\bar{u}_j)+u_i(u_i-\bar{u}_i)(u_j-\bar{u}_j))\\
&\quad=\frac{1}{N}(u_i-\bar{u}_i)(-3(u_i-\bar{u}_i)(u_j-\bar{u}_j)+(u_j-\bar{u}_j)+u_j(u_i-\bar{u}_i)-u_j+2u_i(u_j-\bar{u}_j))\\
&\quad=\frac{1}{N}(u_i-\bar{u}_i)(-3u_iu_j+3u_i\bar{u}_j+3u_j\bar{u}_i-3\bar{u}_i\bar{u}_j\\
&\qquad\qquad+(u_j-\bar{u}_j)+u_ju_i-u_j\bar{u}_i-u_j+2u_iu_j-2u_i\bar{u}_j)\\
&\quad=\frac{1}{N}(u_i-\bar{u}_i)(3u_i\bar{u}_j+3u_j\bar{u}_i-3\bar{u}_i\bar{u}_j-\bar{u}_j-u_j\bar{u}_i-2u_i\bar{u}_j),
\end{aligned}$$

the absolute value of which can be bounded by $15\Sigma/N$, because $\bar{u}_i,\bar{u}_j\le\Sigma$. Next, collecting all terms without $1/N^2$ or $1/N$ yields

$$\begin{aligned}
&(u_i-\bar{u}_i)^2(u_j-\bar{u}_j)-u_j(u_i-\bar{u}_i)^2-u_i(u_i-\bar{u}_i)(u_j-\bar{u}_j)+u_iu_j(u_i-\bar{u}_i)-u_i\bar{u}_i\bar{u}_j\\
&\quad=-\bar{u}_j(u_i-\bar{u}_i)^2-u_i(u_i-\bar{u}_i)(u_j-\bar{u}_j)+u_iu_j(u_i-\bar{u}_i)-u_i\bar{u}_i\bar{u}_j\\
&\quad=-\bar{u}_j(u_i-\bar{u}_i)^2+u_i\bar{u}_j(u_i-\bar{u}_i)-u_i\bar{u}_i\bar{u}_j\\
&\quad=-\bar{u}_j((u_i-\bar{u}_i)^2-u_i(u_i-\bar{u}_i)+u_i\bar{u}_i)\\
&\quad=-\bar{u}_j(-\bar{u}_i(u_i-\bar{u}_i)+u_i\bar{u}_i)\\
&\quad=-\bar{u}_j\bar{u}_i^2,
\end{aligned}$$

which is bounded by Σ^3. Thus, we have shown that

$$|c_{iij}(\boldsymbol{u})|\le\frac{1}{N^2}+\frac{15\Sigma}{N}+\Sigma^3\le 20\delta^2(1+s)^3.$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

Lastly, we consider

$$\begin{aligned} c_{ijk}(\boldsymbol{u}) &= \mathbb{E}(U_i - u_i)(U_j - u_j)(U_k - u_k) \\ &= \mathbb{E}U_k(U_i - u_i)(U_j - u_j) - u_k\mathbb{E}(U_i - u_i)(U_j - u_j) \\ &= \mathbb{E}U_kU_iU_j - u_j\mathbb{E}U_kU_i - u_i\mathbb{E}U_kU_j + u_iu_j\mathbb{E}U_k - u_k\mathbb{E}(U_i - u_i)(U_j - u_j). \end{aligned}$$

From (B.8) we have

$$-u_k\mathbb{E}(U_i - u_i)(U_j - u_j) = -u_k\Big(\bar{u}_i\bar{u}_j - \frac{1}{N}(u_i - \bar{u}_i)(u_j - \bar{u}_j)\Big),$$

and from Lemma B.4 we have

$$\begin{aligned} \mathbb{E}U_kU_iU_j &= \Big(1 - \frac{3}{N} + \frac{2}{N^2}\Big)(u_i - \bar{u}_i)(u_j - \bar{u}_j)(u_k - \bar{u}_k), \\ -u_j\mathbb{E}U_kU_i &= -u_j\Big((u_k - \bar{u}_k)(u_i - \bar{u}_i) - \frac{1}{N}(u_k - \bar{u}_k)(u_i - \bar{u}_i)\Big), \\ -u_i\mathbb{E}U_kU_j &= -u_i\Big((u_k - \bar{u}_k)(u_j - \bar{u}_j) - \frac{1}{N}(u_k - \bar{u}_k)(u_j - \bar{u}_j)\Big). \end{aligned}$$

The combined $1/N^2$ terms can be bounded by $2/N^2$. Collecting terms with $1/N$ in front yields

$$\begin{aligned} &\frac{1}{N}(-3(u_i - \bar{u}_i)(u_j - \bar{u}_j)(u_k - \bar{u}_k) + u_j(u_k - \bar{u}_k)(u_i - \bar{u}_i) \\ &\quad + u_i(u_k - \bar{u}_k)(u_j - \bar{u}_j) + u_k(u_i - \bar{u}_i)(u_j - \bar{u}_j)) \\ &\quad = \frac{1}{N}(\bar{u}_j(u_i - \bar{u}_i)(u_k - \bar{u}_k) + \bar{u}_i(u_j - \bar{u}_j)(u_k - \bar{u}_k) + \bar{u}_k(u_j - \bar{u}_j)(u_i - \bar{u}_i)), \end{aligned}$$

and this term can be bounded by $3\Sigma/N$. Similarly, collecting all terms without $1/N^2$ and $1/N$ yields

$$\begin{aligned} &(u_i - \bar{u}_i)(u_j - \bar{u}_j)(u_k - \bar{u}_k) - u_j(u_k - \bar{u}_k)(u_i - \bar{u}_i) - u_i(u_k - \bar{u}_k)(u_j - \bar{u}_j) \\ &+ u_iu_j(u_k - \bar{u}_k) - u_k\bar{u}_i\bar{u}_j \\ &\quad = -\bar{u}_j(u_i - \bar{u}_i)(u_k - \bar{u}_k) - u_i(u_k - \bar{u}_k)(u_j - \bar{u}_j) + u_iu_j(u_k - \bar{u}_k) - u_k\bar{u}_i\bar{u}_j \\ &\quad = -\bar{u}_j(u_iu_k - u_k\bar{u}_i - u_i\bar{u}_k + \bar{u}_i\bar{u}_k) - u_i(u_ku_j - u_k\bar{u}_j - u_j\bar{u}_k + \bar{u}_j\bar{u}_k) + u_iu_ju_k \\ &\qquad - u_iu_j\bar{u}_k - u_k\bar{u}_i\bar{u}_j \\ &\quad = -\bar{u}_j(u_iu_k - u_k\bar{u}_i - u_i\bar{u}_k + \bar{u}_i\bar{u}_k) - u_i(-u_k\bar{u}_j + \bar{u}_j\bar{u}_k) - u_k\bar{u}_i\bar{u}_j \\ &\quad = -\bar{u}_j(-u_k\bar{u}_i - u_i\bar{u}_k + \bar{u}_i\bar{u}_k) - u_i\bar{u}_j\bar{u}_k - u_k\bar{u}_i\bar{u}_j, \end{aligned}$$

and this term can be bounded by $5\Sigma^2$. This completes the bound for

$$|c_{ijk}(\boldsymbol{u})| \le \frac{2}{N^2} + \frac{3\Sigma}{N} + 5\Sigma^3 \le 20\delta^2(1+s)^3.$$

Finally, we show that

$$\bar{d}_{ijk\ell}(\boldsymbol{u}) \le \left(\frac{2}{\sqrt{N}} + \Sigma\right)^4 \le \delta^2(2+s)^4.$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

Using the Cauchy–Schwarz inequality twice,

$$\begin{aligned}&\mathbb{E}|(U_i-u_i)(U_j-u_j)(U_k-u_k)(U_\ell-u_\ell)|\\&\quad\le\left[\mathbb{E}(U_i-u_i)^4\right]^{\frac14}\left[\mathbb{E}_u(U_j-u_j)^4\right]^{\frac14}\left[\mathbb{E}(U_k-u_k)^4\right]^{\frac14}\left[\mathbb{E}_u(U_\ell-u_\ell)^4\right]^{\frac14}.\end{aligned}$$

Recalling that $NU_i|u_i\sim\mathrm{Bin}(N,u_i-\bar{u}_i)$ and then using Minkowski's inequality,

$$\begin{aligned}\left[\mathbb{E}(U_i-u_i)^4\right]^{\frac14}&=\frac{1}{N}\left[\mathbb{E}_u(NU_i-N(u_i-\bar{u}_i)-N\bar{u}_i)^4\right]^{\frac14}\\&\le\frac{1}{N}\left[\mathbb{E}_u(NU_i-N(u_i-\bar{u}_i))^4\right]^{\frac14}+\frac{1}{N}N\bar{u}_i.\end{aligned}$$

Noting that, for $Y\sim\mathrm{Bin}(n,p)$,

$$\begin{aligned}\mathbb{E}(Y-np)^4&=3(np(1-p))^2+np(1-p)(1-6p(1-p))\\&\le3(np(1-p)^2+np(1-p)\le4n^2,\end{aligned}$$

we conclude that

$$\left[\mathbb{E}(U_i-u_i)^4\right]^{\frac14}\le\frac{1}{N}\left((4N^2)^{\frac14}+N\bar{u}_i\right)\le\frac{2}{\sqrt{N}}+\Sigma,$$

and the final bound is now clear. □

Finally we complete this appendix with the proof of the Stein factor bounds in Lemma 2.3.

Proof of Lemma 2.3. For notational clarity, given $\boldsymbol{a}$ such that $\|\boldsymbol{a}\|=k$, decompose $\boldsymbol{a}$ into its components such that $\boldsymbol{a}=\boldsymbol{a}_1+\ldots+\boldsymbol{a}_k$, where $\boldsymbol{a}_i$ is a standard basis unit vector. Furthermore, for the remainder of this proof, let $\boldsymbol{U}_{\boldsymbol{u}}(t)$ denote the process $\boldsymbol{U}(t)$ started at $\boldsymbol{U}(0)=\boldsymbol{u}$. We first start with the case where $\|\boldsymbol{a}\|=1$. Starting from (2.5), given $h\in\mathcal{M}_{\mathrm{disc},4}(C)$,

$$|\Delta^{\boldsymbol{a}_1}f_h(\boldsymbol{u})|=\left|\sum_{t=0}^{\infty}\left[\mathbb{E}h(\boldsymbol{U}_{\boldsymbol{u}+\delta\boldsymbol{a}_1}(t))-\mathbb{E}h(\boldsymbol{U}_{\boldsymbol{u}}(t))\right]\right|\le\sum_{t=0}^{\infty}C\mathbb{E}\|\boldsymbol{U}_{\boldsymbol{u}+\delta\boldsymbol{a}_1}(t)-\boldsymbol{U}_{\boldsymbol{u}}(t)\|_1.\tag{B.10}$$

We couple the two processes $\boldsymbol{U}_{\boldsymbol{u}+\delta\boldsymbol{a}_1}(t)$ and $\boldsymbol{U}_{\boldsymbol{u}}(t)$ in the following manner. Index the parents so that the types for individuals in both processes match except for the one entry where the first process will have an individual of type depending on $\boldsymbol{a}_1$ and the second process has an individual of type K. (Recall we reserve the final type K to be the remainder.) Given the current generation, to generate the next generation, take a random sample of size N from the indices $\{1,\ldots N\}$, and use this common random sample to choose parents for the offspring of both processes. Given mutation is parent independent, we also couple the mutations identically across both processes in the obvious manner. Figure 1 illustrates the joint evolution of $\boldsymbol{U}_{\boldsymbol{u}+\delta\boldsymbol{a}_1}(t)$ and $\boldsymbol{U}_{\boldsymbol{u}}(t)$.

Given the starting configuration at time 0, let $\boldsymbol{V}(t)=(V_1(t),\ldots,V_N(t))$ denote the process that tracks the ancestry of the original configuration. That is, $V_j(t)$ tracks the number of individuals at time t that trace their ancestry directly back to individual j at time 0. Note that if a mutation occurs, it is removed from this process, hence, $\|\boldsymbol{V}(0)\|_1=N$, and $\lim_{t\to\infty}\boldsymbol{V}(t)=0$. For readers familiar with coalescent theory, this is analogous to a coalescent process looking forwards in time.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

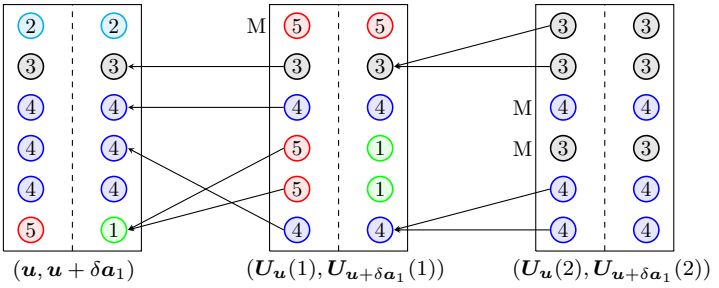


FIGURE 1. An illustration of our coupling at times $t = 0, 1, 2$ with population size $N = 6$ and $K = 5$ types. Each gene type is colour and number coded; e.g. type 1 is green, type 5 is red, etc. An 'M' next to a row represents a mutation, while arrows represent parental relationships; e.g. rows two and three (from the bottom) in the middle plot are children of the first row in the leftmost plot, while row six of the middle plot mutated. Coupling occurs at time $t = 2$ since rows two and three in the middle plot have no children.

Without loss of generality, we can therefore set

$$\|\boldsymbol{U}_{\boldsymbol{u}+\delta\boldsymbol{a}_1}(t) - \boldsymbol{U}_{\boldsymbol{u}}(t)\|_1 = \delta V_1(t).$$

In the context of Figure 1, $V_1(t)$ is tracking the number of replicates at time t of the pair (5,1) from the final row in the processes at time 0. Therefore, in this particular realisation, $V_1(0) = 1$, $V_1(1) = 2$, $V_1(2) = 0$. Given $V_1(t-1)$, $V_1(t)$ will be made up of the number of times one of the $V_1(t-1)$ individuals are chosen, which will occur with probability $\delta V_1(t-1)$, who also do not mutate. Hence, recalling that $\Sigma = \sum_{i=1}^{K} p_k$ denotes the probability of any mutation occurring,

$$V_1(t)|V_1(t-1) \sim \text{Bin}\Big(N, \delta V_1(t-1)(1-\Sigma)\Big).$$

Observing that $\mathbb{E}[V_1(1)] = \mathbb{E}[\mathbb{E}[V_1(1)|V_1(0)]] = \delta\mathbb{E}[V_1(0)](1-\Sigma) = \delta(1-\Sigma)$, and then applying this recursively we conclude that $\mathbb{E}[V_1(t)] = \delta(1-\Sigma)^t$ for all integers $t \geq 0$. Therefore, following on from (B.10), for $h \in \mathcal{M}_{\text{disc},4}(C)$,

$$|\Delta^{\boldsymbol{a}_1} f_h(\boldsymbol{u})| \leq \sum_{t=0}^{\infty} C\delta\mathbb{E}V_1(t) = \sum_{t=0}^{\infty} C\delta(1-\Sigma)^t = \frac{C\delta}{\Sigma}.$$

For the second-order difference where $\|\boldsymbol{a}\| = 2$, for $h \in \mathcal{M}_{\text{disc},4}(C)$,

$$\begin{aligned} |\Delta^{\boldsymbol{a}} f_h(\boldsymbol{u})| &= \left|\sum_{t=0}^{\infty} \mathbb{E}\left[h(\boldsymbol{U}_{\boldsymbol{u}+\delta\boldsymbol{a}_1+\delta\boldsymbol{a}_2}(t)) - h(\boldsymbol{U}_{\boldsymbol{u}+\delta\boldsymbol{a}_1}(t)) - h(\boldsymbol{U}_{\boldsymbol{u}+\delta\boldsymbol{a}_2}(t)) + h(\boldsymbol{U}_{\boldsymbol{u}}(t))\right]\right| \\ &\leq C\sum_{t=0}^{\infty} \mathbb{E}\left[\|\boldsymbol{U}_{\boldsymbol{u}+\boldsymbol{a}_1+\boldsymbol{a}_2}(t) - \boldsymbol{U}_{\boldsymbol{u}+\boldsymbol{a}_1}(t)\|_1 \|\boldsymbol{U}_{\boldsymbol{u}+\boldsymbol{a}_2}(t) - \boldsymbol{U}_{\boldsymbol{u}}(t)\|_1\right], \end{aligned} \tag{B.11}$$

where the inequality is due to the fact that, in general,

$$|f(\boldsymbol{x}+\boldsymbol{a}+\boldsymbol{b}) - f(\boldsymbol{x}+\boldsymbol{a}) - f(\boldsymbol{x}+\boldsymbol{b}) - f(\boldsymbol{x})| \leq \|\boldsymbol{a}\|_1 \|\boldsymbol{b}\|_1 \max_{\|\boldsymbol{a}'\|=2} \|\Delta^{\boldsymbol{a}'} f\|. \tag{B.12}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

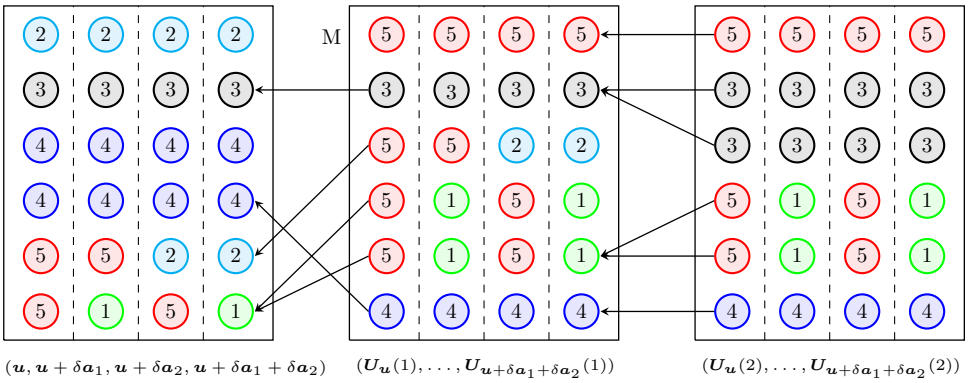


FIGURE 2. An illustration of our second-order difference coupling at times $t = 0, 1, 2$ with population size $N = 6$ and $K = 5$ types.

For one-dimensional functions $f : \mathbb{Z} \to \mathbb{R}$, inequality (B.12) follows from

$$f(x+a+b) - f(x+a) - f(x+b) + f(x) = \sum_{i=0}^{b-1} \sum_{j=1}^{a-1} \big(f(x+j+i) - 2f(x+j-1+i) + f(x+j-1+i-1)\big), \quad x, a, b \in \mathbb{Z}.$$

A similar idea can be used to justify (B.12) for multidimensional functions.

We couple the four processes on the right-hand side in an analogous manner to the first difference; i.e. parent selection and mutation are coupled to be identical for all four processes. Figure 2 illustrates their evolution. Similar to the first difference, we need to keep track of any differences between the four processes, which can be achieved by tracking genealogies using the process $\boldsymbol{V}(t)$ where we set $(V_1(t), V_2(t))$ to jointly track the propagation of the initial two rows in Figure 2.

Therefore, we can without loss of generality set

$$(\|\boldsymbol{U}_{\boldsymbol{u}+\boldsymbol{a}_1+\boldsymbol{a}_2}(t) - \boldsymbol{U}_{\boldsymbol{u}+\boldsymbol{a}_1}(t)\|_1, \|\boldsymbol{U}_{\boldsymbol{u}+\boldsymbol{a}_2}(t) - \boldsymbol{U}_{\boldsymbol{u}}(t)\|_1) = (\delta V_1(t), \delta V_2(t)).$$

Analogously to the first difference, we can set $V_1(t)$ and $V_2(t)$ jointly such that

$$((V_1(t), V_2(t)) | V_1(t-1), V_2(t-1)) \sim \text{Multinomial}\,(N, \{\delta V_1(t-1)(1-\Sigma), \delta V_2(t-1)(1-\Sigma)\}).$$

Given $V_1(0) = V_2(0) = 1$, we can recursively show, with the help of Lemma B.4, that $\mathbb{E}[N^2 V_1(t) V_2(t)] = N(N-1)\delta^2(1-\Sigma)^{2t}$. Hence, continuing from (B.11),

$$|\Delta^{\boldsymbol{a}} f_h(\boldsymbol{u})| \le \sum_{t=0}^{\infty} C\delta^2 \mathbb{E} V_1(t) V_2(t) \le \sum_{t=0}^{\infty} C\delta^2 (1-\Sigma)^{2t} = \frac{C\delta^2}{(1-(1-\Sigma)^2)}.$$

We omit the proof of the bounds when $\|\boldsymbol{a}\| = 3, 4$, as they follow from exactly the same proof methodology. □

Funding information

There are no funding bodies to thank relating to the creation of this paper.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this paper.

References

[1] Barbour, A. (1990). Stein's method for diffusion approximations. *Probab. Theory and Related Fields* **84**, 297–322.

[2] BARBOUR, A. D. (1988). Stein's method and Poisson process convergence. J. Appl. Probab. 175–184. A celebration of applied probability.

[3] BARBOUR, A. D. AND CHEN, L. H. Y. (eds) (2005). *An Introduction to Stein's Method*. (Lecture Notes *Series*, Institute for Mathematical Sciences, National University of Singapore 4). Singapore University Press, Singapore; World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ. Lectures from the Meeting on Stein's Method and Applications: a Program in Honor of Charles Stein held at the National University of Singapore, Singapore, July 28–August **31**, 2003.

[4] BARBOUR, A. D., HOLST, L. AND JANSON, S. (1992). *Poisson Approximation (Oxford Studies in Probability 2)*. The Clarendon Press, Oxford University Press, Oxford Science Publications, New York.

[5] BRAVERMAN, A. (2022). The prelimit generator comparison approach of Stein's method. *Stochastic Systems* **12**, 181–204.

[6] BRAVERMAN, A. (2023). The join-the-shortest-queue system in the Halfin-Whitt regime: Rates of convergence to the diffusion limit. *Stochastic Systems* **13**, 1–39.

[7] BROWN, T. C. AND PHILLIPS, M. J. (1999). Negative binomial approximation with Stein's method. *Methodol. Comput. Appl. Probab.* **1**, 407–421.

[8] CHATTERJEE, S. (2014). A short survey of Stein's method. In *Proceedings of the International Congress of Mathematicians—Seoul 2014*, Vol. IV. Kyung Moon Sa, Seoul. pp. 1–24.

[9] CHEN, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3**, 534–545.

[10] DÖBLER, C. (2015). Stein's method of exchangeable pairs for the beta distribution and generalizations. *Electron. J. Probab.* **20**, no. 109, 34.

[11] ETHIER, S. N. AND NORMAN, M. F. (1977). Error estimate for the diffusion approximation of the Wright–Fisher model. *Proceedings of the National Academy of Sciences* **74**, 5096–5098.

[12] FULMAN, J. AND ROSS, N. (2013). Exponential approximation and Stein's method of exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.* **10**, 1–13.

[13] GAN, H. L., RÖLLIN, A. AND ROSS, N. (2017). Dirichlet approximation of equilibrium distributions in Cannings models with mutation. *Adv. in Appl. Probab.* **49**, 927–959.

[14] GAN, H. L. AND ROSS, N. (2021). Stein's method for the Poisson-Dirichlet distribution and the Ewens sampling formula, with applications to Wright-Fisher models. *Ann. Appl. Probab.* **31**, 625–667.

[15] GOLDSTEIN, L. AND REINERT, G. (2013). Stein's method for the beta distribution and the Pólya-Eggenberger urn. *J. Appl. Probab.* **50**, 1187–1205.

[16] KASPRZAK, M. J. (2017). *Diffusion approximations via Stein's method and time changes*. Preprint. Available at https://arxiv.org/abs/1701.07633.

[17] LEY, C., REINERT, G. AND SWAN, Y. (2017). Stein's method for comparison of univariate distributions. *Probability Surveys* **14**, 1–52.

[18] MACKEY, L. AND GORHAM, J. (2016). Multivariate Stein factors for a class of strongly log-concave distributions. *Electron. Commun. Probab.* **21**, 14.

[19] ROSS, N. (2011). Fundamentals of Stein's method. *Probab. Surv.* **8**, 210–293.

[20] STEIN, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (University of California, Berkeley, CA, 1970/1971), Vol. II: Probability theory*. University California Press, Berkeley, CA, pp. 583–602.

[21] STROOCK, D. W. AND VARADHAN, S. R. S. (1979). *Multidimensional Diffusion Processes*. Springer, New York.

[22] WRIGHT, S. (1931). Evolution in Mendelian populations. *Genetics* **16**, 97.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:33:10, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/apr.2025.13