

Polynomial diagrams for microstructure modelling

David P. Bourne¹

*Maxwell Institute for Mathematical Sciences and Department of Mathematics,
Heriot-Watt University, Edinburgh, EH14 4AS, UK*

Maciej Buze²

*MARS: Mathematics for AI in Real-world Systems, School of Mathematical Sciences,
Lancaster University, Lancaster, LA1 4YF, UK*

Thomas Gallouët³

*Université Paris-Saclay, CNRS, Inria, ParMA Team,
Laboratoire de Mathématiques d'Orsay, 91405 Orsay, France*

Quentin Mérigot⁴

*Université Paris-Saclay, CNRS, Inria,
Laboratoire de Mathématiques d'Orsay, 91405 Orsay, France
DMA, École normale supérieure, Université PSL, CNRS, 75005 Paris, France*

May 20, 2026

Abstract

We formulate a framework of *polynomial diagrams*, which are a generalisation of power diagrams (PDs) and anisotropic power diagrams (APDs) allowing for boundaries between cells to be algebraic curves of a prescribed degree. We show that they arise naturally from rephrasing PDs (APDs) as first-degree (second-degree) instances of *linear parametrised minimisation diagrams*. We also develop an efficient GPU-accelerated framework for fitting polynomial diagrams to image data using Legendre polynomials and by maximising a regularised *concave* objective function adapted from classical logistic regression literature. A largely self-contained analysis of the optimisation algorithm is also provided, including identification of scale and gauge invariances and the limiting objective function as the regularisation parameter vanishes. We apply the algorithm to fit polynomial diagrams to electron backscatter diffraction images of steel.

1 Introduction

The modelling of polycrystalline materials is a central challenge in materials science and is of high relevance to industrial applications. The macroscopic response of a metal is strongly

¹d.bourne@hw.ac.uk

²m.buze@lancaster.ac.uk

³thomas.gallouet@inria.fr

⁴quentin.merigot@universite-paris-saclay.fr

mediated by its grain structure: grain boundaries obstruct slip and promote strain heterogeneity, while grain size, morphology, and crystallographic texture enter directly into classical strength and hardening laws and into modern microstructure-sensitive constitutive modelling [24, 34, 45].

Experimental microstructure data is commonly obtained from electron backscatter diffraction (EBSD) measurements, and complementary datasets arise as discretised outputs of grain-growth, crystal-plasticity and FEM simulations [8, 36, 45]. In all cases the data is inherently discrete, given as pixels/voxels on a set Ω approximating an ambient domain $\bar{\Omega}$. It can then be post-processed, e.g., using the open-source software MTEX [11], resulting in a grain map $G: \Omega \rightarrow [N]$ (where $[N] := \{1, \dots, N\}$), which is a partition of Ω into N grains, where $G(x) = k$ if the point x belongs to grain k .

For analysis, storage, and fast geometric queries it is advantageous to replace a pixel/voxel level segmentation by a compact parametric representation whose induced cells approximate the observed grains. Classical choices include Voronoi diagrams [10], power diagrams [22, 33, 43, 21, 14, 28, 15], and, more recently, anisotropic power diagrams [22, 29, 33, 43], which remain a topic of active research [7, 3, 47, 48, 50, 39, 5, 4, 26] – see also [27] for recent work on extending these approaches to account for crystallographic orientation of grains and [41] on fast computation of Apollonius diagrams on a GPU. These diagrams have geometrically interpretable parameters, but their expressive power is limited when experimental boundaries depart from the assumed algebraic class.

In this work we formulate a common framework for such representations termed *parametrised minimisation diagrams* (PMDs): given a set of functions $\{f_1, \dots, f_N\}$ defined on the ambient space $\bar{\Omega}$, each point $x \in \Omega$ is *arg-min assigned*, that is, assigned to the index attaining the minimum of $\{f_i(x)\}_{i \in [N]}$. We focus on the parametrised case

$$f_i(x) = h(\theta_i, x), \quad h: \mathbb{R}^K \times \bar{\Omega} \rightarrow \mathbb{R}$$

and on fitting $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N) \in \mathbb{R}^{K \times N}$ so that the resulting arg-min assignment matches a given grain map G as much as possible.

Directly maximising pixel/voxel accuracy is, in general, a non-smooth nonlinear programming problem in $\boldsymbol{\theta}$. This was solved using stochastic optimisation in [49]. For the case of APDs, it was recently shown that this is a high-dimensional linear programming problem [5]. To obtain a smooth, concave maximisation problem we use the well-known ε -scale soft assignment based on a log-sum-exp normalisation, leading to the standard multinomial logistic regression problem [17, 25, 37]. See [39] for a recent use of similar ideas in the context of microstructure modelling, where the problem is formulated as a non-convex logistic regression problem (see Remark 2.2). For a recent survey of algorithms for fitting (anisotropic) power diagrams to image data, see [6]. The problem of fitting diagrams to *incomplete* data, where the full grain map G is not available and only the volumes and centroids of the grains are known, is studied for example in [40, 43, 16, 13].

A foundational observation for the present work is that for the *linear* class

$$h_{\theta_i}(x) = \theta_i \cdot \eta(x), \quad \text{where } \eta: \Omega \rightarrow \mathbb{R}^K \text{ is some } \textit{design function},$$

the resulting objective function is strictly concave after removing a natural gauge invariance (common shifts of all θ_i). This yields a numerically robust optimisation problem even at large scale, and it allows us to analyse the behaviour of optimisers and near-optimisers, including the ε -scaling, the sharp $\varepsilon \rightarrow 0$ limit, and the classical non-attainment phenomenon under

perfect reconstruction (typically referred to as *linear separation* in the logistic regression literature [2, 46]).

To increase expressivity while preserving concavity, we extend the design function η beyond the standard low-degree cases. Power diagrams and anisotropic power diagrams are recovered as degree one and two *polynomial diagrams*, and we propose a general polynomial diagram framework with the design function η containing all polynomial terms up to a degree d . In particular, we advocate rescaling to $[-1, 1]^2$ and using a Legendre product basis, which spans the same polynomial space but typically improves the conditioning of the *design matrix* $\eta(\Omega) \in \mathbb{R}^{K \times |\Omega|}$. We solve the resulting optimisation problem on a GPU using our Python library PYAPD [18, 1]. We report numerical experiments on (i) synthetic data generated by power/anisotropic power diagrams and (ii) EBSD-derived grain maps of steel, illustrating the practical accuracy–complexity trade-off as the polynomial degree increases.

Outline of the paper. In Section 2 we introduce the general framework of parametrised minimisation diagrams, formulate the fitting problem, and explain why the linear parametrisation class yields a concave multinomial logistic objective; standard power diagrams and anisotropic power diagrams are recovered in Section 2.3.1 as first- and second-degree instances of this framework. In Section 3 we define polynomial diagrams of arbitrary degree, first in a monomial basis and then in a Legendre basis, and discuss their relation to the classical cases; see in particular Section 3.3. Section 4 contains a largely self-contained analysis of the optimisation problem for linear PMDs, including the gradient and Hessian formulas, strict concavity after gauge fixing, the role of the ε -parameter, the $\varepsilon \rightarrow 0$ limit, and the existence/non-existence of maximisers depending on whether perfect reconstruction is possible. In Section 5 we present GPU-based numerical experiments, first on synthetic power and anisotropic power diagram data (Section 5.4) and then on EBSD grain maps of steel (Section 5.5), followed by a discussion of degree selection, initialisation, recovery of physically meaningful parameters, and the practical role of ε in Section 5.6. Finally, Section 6 summarises the main findings and outlines several directions for future work.

Acknowledgements DPB and MB would like to thank the UK Engineering and Physical Sciences Research Council (EPSRC) for financial support via the grant EP/V00204X/1. MB is currently supported by Research England under the Expanding Excellence in England (E3) funding stream, which was awarded to MARS: Mathematics for AI in Real-world Systems in the School of Mathematical Sciences at Lancaster University. QM is supported by Agence nationale de la recherche, through the PEPR PDE-AI project (ANR- 23-PEIA-0004). The authors thank Tata Steel Netherlands for providing EBSD datasets and Shadeform for providing access to cloud GPU resources.

2 Parametrised minimisation diagrams

2.1 Setting and notation

We consider an ambient space $\overline{\Omega} \subseteq \mathbb{R}^d$. To simplify the presentation, we take $d = 2$ and

$$\overline{\Omega} := [-1, 1]^2,$$

although our results are not limited to two dimensions. We take a discretisation of $\bar{\Omega}$ given by a set of pixels

$$\Omega := \left\{ (-1, -1) + \frac{1}{M} \left(k_1 - \frac{1}{2}, k_2 - \frac{1}{2} \right) \mid k_i \in [2M] \right\},$$

where $M \in \mathbb{N}_+$ is the pixel resolution parameter and $[2M] := \{1, \dots, 2M\}$. To be precise, the points in Ω are the centres of pixels, but we refer to them as pixels for brevity. Overall we have $|\Omega| = (2M)^2 = 4M^2$ pixels. More generally, one may replace Ω by any finite set of sample points in $\bar{\Omega}$, for instance an unstructured pixel list.

A *grain map* is an assignment map $G: \Omega \rightarrow [N]$, where $N > 1$, which induces a tessellation of Ω into N clusters, each given by $G_i := G^{-1}(\{i\})$. We assume that G is given.

A *minimisation diagram* generated by a function $\mathbf{f} = (f_1, \dots, f_N): \bar{\Omega} \rightarrow \mathbb{R}^N$ is a tessellation of $\bar{\Omega}$ into N cells:

$$L_i := \left\{ x \in \bar{\Omega} \mid f_i(x) < f_j(x) \ \forall j < i, \quad f_i(x) \leq f_j(x) \ \forall j > i \right\}. \quad (1)$$

That is, ties are broken by assigning x to the smallest index attaining the minimum. In particular,

$$\bar{\Omega} = \bigcup_i L_i, \quad L_i \cap L_j = \emptyset \quad \text{for } i \neq j.$$

We restrict our attention to *parametrised minimisation diagrams (PMDs)*, namely

$$f_i(\cdot) = h(\theta_i, \cdot), \quad h: \mathbb{R}^K \times \bar{\Omega} \rightarrow \mathbb{R}, \quad (\text{short-hand notation } h(\theta_i, \cdot) \equiv h_{\theta_i}(\cdot)). \quad (2)$$

Thus $\theta_i \in \mathbb{R}^K$ is a set of parameters defining f_i and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N) \in \mathbb{R}^{K \times N}$ is the overall set of parameters determining the diagram.

Remark 2.1 (Bold font and subscript convention). We reserve bold font for highest level collections of objects such as the collection $\mathbf{f} = (f_1, \dots, f_N)$ of functions generating a minimisation diagram or a collection of parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$. Depending on context a generic constituent function will be referred to as either f or f_i and a generic constituent parameter vector as θ or θ_i . Lower level ordering for vectors $\theta, \theta_i \in \mathbb{R}^K$ will be separated by a comma, leading to notation $\theta_{i,j} \in \mathbb{R}$ and $\theta_j \in \mathbb{R}$. The same convention applies to pixels $\Omega = \{x_1, \dots, x_{|\Omega|}\}$, with $x_{i,j} \in [-1, 1]$ and the notation for a generic point $x = (x_1, x_2) \in \bar{\Omega}$, or standard parameters defining (anisotropic) power diagrams (see Section 2.3.1).

The PMD-induced grain map is

$$F_{\boldsymbol{\theta}}: \bar{\Omega} \rightarrow [N], \quad F_{\boldsymbol{\theta}}(x) := \arg \min_{i \in [N]} h_{\theta_i}(x).$$

If the minimiser is not unique, we select the smallest index (c.f., (1)) and such diagrams are referred to as *ambiguous*. An important special case is a *fully ambiguous* diagram, which arises when $h_{\theta_i}(x) = h_{\theta_j}(x)$ for all $i, j \in [N]$ and all $x \in \bar{\Omega}$.

The characteristic function of the i th cell L_i is given by

$$(p_{\boldsymbol{\theta}})_i: \bar{\Omega} \rightarrow \{0, 1\}, \quad (p_{\boldsymbol{\theta}})_i(x) := \delta(i, F_{\boldsymbol{\theta}}(x)), \quad (3)$$

where δ is the Kronecker delta, so $\delta(k, j) = 1$ if $k = j$ and zero otherwise. For $\varepsilon > 0$, its smeared-out counterpart is

$$(p_{\boldsymbol{\theta}}^{\varepsilon})_i: \bar{\Omega} \rightarrow (0, 1), \quad (p_{\boldsymbol{\theta}}^{\varepsilon})_i(x) := \frac{\exp\left(-\frac{1}{\varepsilon}h_{\theta_i}(x)\right)}{S_{\boldsymbol{\theta}}^{\varepsilon}(x)}, \quad S_{\boldsymbol{\theta}}^{\varepsilon}(x) := \sum_{j \in [N]} \exp\left(-\frac{1}{\varepsilon}h_{\theta_j}(x)\right). \quad (4)$$

We can think of $(p_{\boldsymbol{\theta}}^{\varepsilon})_i(x)$ as being the probability that pixel x belongs to grain i . If $\boldsymbol{\theta}$ is such that the resulting diagram is fully ambiguous, then $(p_{\boldsymbol{\theta}}^{\varepsilon})_i(x) = \frac{1}{N}$ for all $i \in [N]$ and all $x \in \bar{\Omega}$.

In what follows, we will investigate several possible choices of parameterisation in (2) and discuss optimisation approaches so that the PMD-induced grain map $F_{\boldsymbol{\theta}}$ is as close to the given grain map G as possible.

2.2 Optimisation approaches

For now assume that the parameterisation function h in (2) is fixed. The most natural measure of the discrepancy between G and $F_{\boldsymbol{\theta}}$ is to simply count the number of pixels being assigned to correct grains, which gives the overall accuracy as

$$\text{Acc}^G(\boldsymbol{\theta}) := \frac{1}{|\Omega|} \sum_{x \in \Omega} (p_{\boldsymbol{\theta}})_{G(x)}(x) \quad (5)$$

where $(p_{\boldsymbol{\theta}})_i$ is the characteristic function defined in (3). By construction, $\text{Acc}^G(\boldsymbol{\theta}) \in [0, 1]$, where 0 corresponds to a *complete mismatch* (no pixels assigned correctly) and 1 to *perfect reconstruction* (all pixels assigned correctly). This leads to the optimisation problem

$$\max_{\boldsymbol{\theta} \in \mathbb{R}^{K \times N}} \text{Acc}^G(\boldsymbol{\theta}). \quad (6)$$

The mapping $\boldsymbol{\theta} \mapsto \text{Acc}^G(\boldsymbol{\theta})$ is typically discontinuous, which renders solving (6) directly tricky. One way to address this is to replace $p_{\boldsymbol{\theta}}$ with its smeared-out counterpart $p_{\boldsymbol{\theta}}^{\varepsilon}$ defined in (4), leading to

$$\max_{\boldsymbol{\theta} \in \mathbb{R}^{K \times N}} \text{Acc}_{\varepsilon}^G(\boldsymbol{\theta}), \quad \text{Acc}_{\varepsilon}^G(\boldsymbol{\theta}) := \frac{1}{|\Omega|} \sum_{x \in \Omega} (p_{\boldsymbol{\theta}}^{\varepsilon})_{G(x)}(x). \quad (7)$$

The exponential function in the definition of $p_{\boldsymbol{\theta}}^{\varepsilon}$ can lead to numerical overflow in finite-precision arithmetic when ε is small. A numerically stable counterpart of (7) is the optimisation problem

$$\max_{\boldsymbol{\theta} \in \mathbb{R}^{K \times N}} \Phi_{\varepsilon}^G(\boldsymbol{\theta}), \quad \Phi_{\varepsilon}^G(\boldsymbol{\theta}) := \frac{1}{|\Omega|} \sum_{x \in \Omega} \log\left((p_{\boldsymbol{\theta}}^{\varepsilon})_{G(x)}(x)\right). \quad (8)$$

We recognise that this is the well-known multinomial logistic regression approach to clustering, which is standard in the statistics and machine-learning literature [17, 25, 37].

The problem in (8) is particularly nice when the objective function Φ_{ε}^G is concave, as then every local maximiser is also a global maximiser. This will be analysed in detail in Section 4. But first, in the next section, we will establish a broad class of PMDs for which the concavity of the objective function Φ_{ε}^G in (8) can be guaranteed.

2.3 Linear parameterisation ensures concavity

Recall (2), fix some $K \in \mathbb{N}_+$ and suppose that $h : \mathbb{R}^K \times \bar{\Omega} \rightarrow \mathbb{R}$ is an arbitrary parameterisation giving rise to functions $\{h_{\theta_i}\}_{i \in [N]}$. We first rewrite the objective function Φ_ε^G from (8) as

$$\Phi_\varepsilon^G(\boldsymbol{\theta}) = \frac{1}{|\Omega|} \sum_{x \in \Omega} \left(-\frac{1}{\varepsilon} h_{\theta_{G(x)}}(x) - \log S_{\boldsymbol{\theta}}^\varepsilon(x) \right) \quad (9)$$

and we recall from (4) that

$$S_{\boldsymbol{\theta}}^\varepsilon(x) = \sum_{j \in [N]} \exp \left(-\frac{1}{\varepsilon} h_{\theta_j}(x) \right).$$

It is a standard assertion relying on Jensen's inequality [17, 12] that, for fixed $\varepsilon > 0$ and $x \in \Omega$, the function $\boldsymbol{\theta} \mapsto -\log S_{\boldsymbol{\theta}}^\varepsilon(x)$ is concave if $\boldsymbol{\theta} \mapsto -h(\boldsymbol{\theta}, x)$ is convex. But in this case

$$\left(-\frac{1}{\varepsilon} h_{\theta_{G(x)}}(x) - \log S_{\boldsymbol{\theta}}^\varepsilon(x) \right) \quad (10)$$

is a sum of a convex function and a concave function. A sufficient condition to ensure concavity is thus that the mapping $\boldsymbol{\theta} \mapsto h(\boldsymbol{\theta}, x)$ is linear, so of the form

$$h(\boldsymbol{\theta}, x) = \boldsymbol{\theta} \cdot \boldsymbol{\eta}(x) \quad (11)$$

where $\boldsymbol{\eta} : \bar{\Omega} \rightarrow \mathbb{R}^K$ is a *design function*, giving rise to the well-known notion of a *design matrix*:

$$\boldsymbol{\eta}(\Omega) \in \mathbb{R}^{K \times |\Omega|}, \quad (\boldsymbol{\eta}(\Omega))_{ij} := \eta_i(x_j), \quad i \in [K], j \in [|\Omega|].$$

This is exactly the well-known ‘‘feature map’’ approach in statistical learning: $\boldsymbol{\eta}(x)$ is the feature (or basis) vector associated with input x , and $\boldsymbol{\eta}(\Omega)$ is the corresponding *design matrix* used in linear and generalised linear models; see, e.g., [25, 37]. Equivalently, $\boldsymbol{\eta}$ can be viewed as an explicit embedding of the input space into a (possibly higher-dimensional) feature space, in which the model is linear.

The fact that a linear parametrisation does indeed ensure concavity is well-known and will be discussed in Proposition 4.4.

2.3.1 Relation to standard diagrams

We will now show that two standard minimisation diagrams can be recovered as special cases of the linear parametrisation class (11) for specific choices of $\boldsymbol{\eta}$.

Power diagrams. The i th cell of the power diagram generated by the N seed points $\mathbf{y} = \{y_1, \dots, y_N\}$, $y_i \in \mathbb{R}^2$, and weights $\mathbf{w} = \{w_1, \dots, w_N\}$, $w_i \in \mathbb{R}$, is given by

$$L_i = \{x \in \bar{\Omega} \mid \|x - y_i\|^2 - w_i \leq \|x - y_j\|^2 - w_j, \forall j \in [N]\}.$$

After simplifying, this can be equivalently rewritten as a PMD with $K = 3$ and

$$h(\boldsymbol{\theta}, x) = \boldsymbol{\theta} \cdot \boldsymbol{\eta}(x), \quad \boldsymbol{\eta}(x) = (x_{,1}, x_{,2}, 1).$$

We also have a relation

$$\boldsymbol{\theta} = (-2y_{,1}, -2y_{,2}, y_{,1}^2 + y_{,2}^2 - w)$$

and conversely

$$y_{,1} = -\frac{1}{2}\boldsymbol{\theta}_{,1}, \quad y_{,2} = -\frac{1}{2}\boldsymbol{\theta}_{,2}, \quad w = \frac{1}{4}(\boldsymbol{\theta}_{,1})^2 + \frac{1}{4}(\boldsymbol{\theta}_{,2})^2 - \boldsymbol{\theta}_{,3}. \quad (12)$$

Anisotropic power diagrams. The i th cell of the anisotropic power diagram generated by the N seed points $\mathbf{y} = \{y_1, \dots, y_N\}$, $y_i \in \mathbb{R}^2$, weights $\mathbf{w} = \{w_1, \dots, w_N\}$, $w_i \in \mathbb{R}$, and anisotropy matrices $\mathbf{A} = (\mathbf{A}_1, \dots, \mathbf{A}_N)$, \mathbf{A}_i a 2x2 symmetric positive definite matrix, is given by

$$L_i = \{x \in \bar{\Omega} \mid \|x - y_i\|_{\mathbf{A}_i}^2 - w_i \leq \|x - y_j\|_{\mathbf{A}_j}^2 - w_j, \forall j \in [N]\}, \quad \|z\|_{\mathbf{A}}^2 := z \cdot \mathbf{A}z. \quad (13)$$

Expanding $\|x - y\|_{\mathbf{A}}^2 - w = x \cdot \mathbf{A}x - 2(\mathbf{A}y) \cdot x + (y \cdot \mathbf{A}y - w)$ shows this is a PMD with $K = 6$ and

$$h(\theta, x) = \theta \cdot \eta(x), \quad \eta(x) = (x_1^2, x_1 x_2, x_2^2, x_1, x_2, 1). \quad (14)$$

We also have a relation

$$\theta = \left(\mathbf{A}_{11}, 2\mathbf{A}_{12}, \mathbf{A}_{22}, -2(\mathbf{A}_{11}y_{,1} + \mathbf{A}_{12}y_{,2}), -2(\mathbf{A}_{12}y_{,1} + \mathbf{A}_{22}y_{,2}), y \cdot \mathbf{A}y - w \right),$$

and conversely

$$\mathbf{A} = \begin{pmatrix} \theta_{,1} & \frac{1}{2}\theta_{,2} \\ \frac{1}{2}\theta_{,2} & \theta_{,3} \end{pmatrix}, \quad y = -\frac{1}{2}\mathbf{A}^{-1} \begin{pmatrix} \theta_{,4} \\ \theta_{,5} \end{pmatrix}, \quad w = y \cdot \mathbf{A}y - \theta_{,6} = \frac{1}{4} \begin{pmatrix} \theta_{,4} \\ \theta_{,5} \end{pmatrix} \cdot \mathbf{A}^{-1} \begin{pmatrix} \theta_{,4} \\ \theta_{,5} \end{pmatrix} - \theta_{,6}. \quad (15)$$

For an arbitrary $\theta \in \mathbb{R}^{6 \times N}$, the corresponding anisotropy matrices \mathbf{A}_i may not be positive definite. It turns out, however, that the constraint of positive definiteness can be enforced while solving the *unconstrained* optimisation problem $\max_{\theta \in \mathbb{R}^{6 \times N}} \Phi_{\varepsilon}^G(\theta)$; see Section 5.6.3.

Remark 2.2 (Prior related work). A similar discussion on transforming the standard APD description in terms of $(\mathbf{w}, \mathbf{y}, \mathbf{A})$ from (13) into a linear one based on (14) and the substitution (15) was also considered in [5] in the context of formulating a linear programming problem to solve the unregularised problem (6). In [39] on the other hand, the authors do work with the smeared-out characteristic function p_{θ}^{ε} defined in (4), but using the standard APD description in terms of $(\mathbf{w}, \mathbf{y}, \mathbf{A})$. Furthermore, instead of Φ_{ε}^G , a binary cross-entropy objective function is considered, which in our linear θ description amounts to solving

$$\max_{\theta \in \mathbb{R}^{K \times N}} \Phi_{\varepsilon}^{\text{BCE}}(\theta), \quad \Phi_{\varepsilon}^{\text{BCE}}(\theta) := \frac{1}{|\Omega|} \sum_{i=1}^N \left(\sum_{x \in G_i} \log((p_{\theta}^{\varepsilon})_i(x)) + \sum_{x \notin G_i} \log(1 - (p_{\theta}^{\varepsilon})_i(x)) \right). \quad (16)$$

This is a well-known ‘‘one-vs-all’’ approach to multi-class classification problems [44], which is routinely used as an alternative to the multinomial logistic regression approach. While solving (16) has its advantages, with gradient information arguably richer, the big disadvantage is that the loss of concavity is unavoidable. This follows from

$$\log(1 - (p_{\theta}^{\varepsilon})_i(x)) = \log \left(\sum_{\substack{j \in [N] \\ j \neq i}} \exp \left(-\frac{1}{\varepsilon} h_{\theta_j}(x) \right) \right) - \log \left(\sum_{j \in [N]} \exp \left(-\frac{1}{\varepsilon} h_{\theta_j}(x) \right) \right),$$

which, similarly to the discussion following (10), is a difference of two convex functions. Since the underlying problem is one of exclusive multi-class classification (each pixel belongs to exactly one grain), a ‘‘one-vs-all’’ binary approach should probably be avoided, as (i) it

introduces no additional information (membership in one class automatically determines non-membership in all others), (ii) the computational cost is much increased (the complement of the set G_i contains most pixels), and (iii) the concavity is lost. The computational cost issue can be partially mitigated by employing batched stochastic optimisation techniques, as is done in [39], but this of course makes the gradient information less rich.

Remark 2.3 (Voronoi, multiplicative Voronoi, and Apollonius diagrams are not linear). Voronoi diagrams are the special case of power diagrams where $w_i = 0$ for all i . In this case, we observe that while we can still write a Voronoi diagram as a PMD with

$$h(\theta, x) = \theta \cdot \eta(x), \quad \eta(x) = (x_1, x_2, 1), \quad \theta = (-2y_1, -2y_2, \|y\|^2),$$

Voronoi diagrams are *not* linear because the parameter space is constrained by

$$\theta_3 - \frac{1}{4}(\theta_1^2 + \theta_2^2) = 0.$$

The same is true for multiplicatively weighted Voronoi diagrams, with cell formula

$$L_i = \{x \in \bar{\Omega} \mid w_i \|x - y_i\|^2 \leq w_j \|x - y_j\|^2, \forall j \in [N]\},$$

leading to an incomplete degree two monomial basis (a complete treatment of such polynomial diagrams will be given Section 3),

$$\tilde{\eta}^{(2)}(x) = (x_1^2, x_2^2, x_1, x_2, 1), \quad \theta = (w, w, -2w^2y_1, -2w^2y_2, w^2\|y\|^2),$$

and with a nonlinear constraint

$$\theta_3^2 + \theta_4^2 = 4\theta_1\theta_5.$$

Note that one typically also has the constraint $\theta_{i,1} = \theta_{i,2} = w_i > 0$ for all i . Similarly, an Apollonius diagram, with cell formula

$$L_i = \{x \in \bar{\Omega} \mid \|x - y_i\| - w_i \leq \|x - y_j\| - w_j, \forall j \in [N]\},$$

cannot be expressed as a linear parametrisation since the seed y_i enters inside the norm and squaring both sides of the inequality does not lead to a linear parametrisation either.

3 Polynomial diagrams

The linearisation of standard diagrams presented in Section 2.3.1 shows that a power diagram is a *first degree* diagram, since its design function includes all zeroth- and first-order polynomial terms in (x_1, x_2) . Likewise, an anisotropic power diagram is a *second degree* diagram. In this section we propose a natural generalisation to a class of *polynomial diagrams* of an arbitrary degree $d > 0$. Note that we avoid the term *higher-order diagram*, as this is a well-known term for a different type of generalisation, such as higher-order Voronoi diagrams [30, 10] and, e.g., their more recent extension to line segments [38].

3.1 Multi-index notation

For $\alpha = (\alpha_1, \alpha_2) \in \mathbb{N}_0^2$ and $x = (x_1, x_2) \in \overline{\Omega}$, write

$$|\alpha| := \alpha_1 + \alpha_2, \quad x^\alpha := x_1^{\alpha_1} x_2^{\alpha_2}.$$

Fix $d \in \mathbb{N}_+$ and define the set of multi-indices up to total degree d by

$$\mathcal{A}_d := \{\alpha \in \mathbb{N}_0^2 : |\alpha| \leq d\}, \quad K_d := |\mathcal{A}_d| = \frac{(d+1)(d+2)}{2}.$$

We fix once and for all an ordering of \mathcal{A}_d (e.g. graded lexicographic), and implicitly identify vectors in \mathbb{R}^{K_d} with multi-indices $\alpha \in \mathcal{A}_d$.

3.2 Monomial basis

Define the degree d monomial design function

$$\eta^{(d)} : \overline{\Omega} \rightarrow \mathbb{R}^{K_d}, \quad \eta^{(d)}(x) := (x^\alpha)_{\alpha \in \mathcal{A}_d}.$$

A *monomial degree d minimisation diagram* is a polynomial diagram of the linear form (11) with $\eta = \eta^{(d)}$, i.e.,

$$h(\theta, x) = \theta \cdot \eta^{(d)}(x) = \sum_{\alpha \in \mathcal{A}_d} \theta_{,\alpha} x^\alpha, \quad \theta \in \mathbb{R}^{K_d}. \quad (17)$$

For $\theta = (\theta_1, \dots, \theta_N) \in \mathbb{R}^{K_d \times N}$, the induced grain map is

$$F_\theta(x) = \arg \min_{i \in [N]} h_{\theta_i}(x),$$

with the same tie-breaking convention as in Section 2.1, and the corresponding cells are

$$L_i := \left\{ x \in \overline{\Omega} \mid h_{\theta_i}(x) < h_{\theta_j}(x) \ \forall j < i, \quad h_{\theta_i}(x) \leq h_{\theta_j}(x) \ \forall j > i \right\}. \quad (18)$$

The associated design matrix is

$$\eta^{(d)}(\Omega) \in \mathbb{R}^{K_d \times |\Omega|}, \quad (\eta^{(d)}(\Omega))_{\alpha,j} := x_j^\alpha, \quad \alpha \in \mathcal{A}_d, \quad j \in [|\Omega|].$$

A monomial basis approach is presented primarily as the most intuitive extension of first-degree (power) and second-degree (anisotropic power) diagrams. For practical computations, we instead work with Legendre polynomials, which we will discuss now.

3.3 Legendre basis

Since $\overline{\Omega} = [-1, 1]^2$ (and any other rectangular domain can be scaled to $[-1, 1]^2$), it is natural to replace monomials by (products of univariate) Legendre polynomials [51, Chapter 17]. Let $\{P_m\}_{m \geq 0}$ denote the standard Legendre polynomials on $[-1, 1]$, normalised by $P_0 \equiv 1$, $P_1(t) = t$ and satisfying the three-term recurrence

$$(m+1)P_{m+1}(t) = (2m+1)tP_m(t) - mP_{m-1}(t).$$

Legendre polynomials are orthogonal on $[-1, 1]$ with respect to uniform weight,

$$\int_{-1}^1 P_m(t) P_n(t) dt = \frac{2}{2n+1} \delta_{mn}.$$

This typically leads to the associated columns of the design matrix being less correlated, improving conditioning of the design matrix and stability of gradient-based optimisation, especially for larger d .

Define the Legendre product indexed by $\alpha = (\alpha_1, \alpha_2) \in \mathcal{A}_d$ as

$$\psi_\alpha(x) := P_{\alpha_1}(x_{,1}) P_{\alpha_2}(x_{,2}).$$

Then the degree d Legendre design function is

$$\eta_{\mathbb{L}}^{(d)} : \bar{\Omega} \rightarrow \mathbb{R}^{K_d}, \quad \eta_{\mathbb{L}}^{(d)}(x) := (\psi_\alpha(x))_{\alpha \in \mathcal{A}_d} = (P_{\alpha_1}(x_{,1}) P_{\alpha_2}(x_{,2}))_{\alpha \in \mathcal{A}_d}.$$

A *Legendre degree d minimisation diagram* is a polynomial diagram of the linear form (11) with $\eta = \eta_{\mathbb{L}}^{(d)}$, i.e.,

$$h(\theta, x) = \theta \cdot \eta_{\mathbb{L}}^{(d)}(x) = \sum_{\alpha \in \mathcal{A}_d} \theta_{,\alpha} P_{\alpha_1}(x_{,1}) P_{\alpha_2}(x_{,2}), \quad \theta \in \mathbb{R}^{K_d}. \quad (19)$$

The induced grain map F_θ and cells L_i are defined as before. The associated design matrix is

$$\eta_{\mathbb{L}}^{(d)}(\Omega) \in \mathbb{R}^{K_d \times |\Omega|}, \quad (\eta_{\mathbb{L}}^{(d)}(\Omega))_{\alpha,j} := P_{\alpha_1}(x_{j,1}) P_{\alpha_2}(x_{j,2}), \quad \alpha \in \mathcal{A}_d, \quad j \in [|\Omega|].$$

Remark 3.1 (Consistency with standard cases). For $d = 1$, the monomial design gives

$$\eta^{(1)}(x) = (x_{,1}, x_{,2}, 1),$$

recovering the power-diagram linearisation. For $d = 2$, the monomial design gives

$$\eta^{(2)}(x) = (x_{,1}^2, x_{,1}x_{,2}, x_{,2}^2, x_{,1}, x_{,2}, 1),$$

recovering (14). For the Legendre variant, note that $P_0 \equiv 1$ and $P_1(t) = t$, hence $\eta_{\mathbb{L}}^{(1)}(x)$ spans the same linear feature space as the monomial case (up to re-ordering), while higher d provide an alternative (often better conditioned) polynomial basis. Since both monomial and Legendre bases span the same polynomial space $\mathbb{P}_d(\mathbb{R}^2)$ (total degree $\leq d$), there exists a unique and easily computable invertible matrix $T^{(d)} \in \mathbb{R}^{K_d \times K_d}$ allowing us to conveniently convert between them.

Remark 3.2 (Geometric interpretation of cell boundaries). For any $i \neq j$, the interface between cells L_i and L_j is contained in

$$\{x \in \bar{\Omega} : h_{\theta_i}(x) = h_{\theta_j}(x)\} = \{x \in \bar{\Omega} : (\theta_i - \theta_j) \cdot \eta^{(d)}(x) = 0\},$$

which is an algebraic curve of degree at most d (independent of the chosen basis).

Extensive numerical tests showcasing the fitting of Legendre polynomial diagrams to grain maps will be presented in Section 5. Let us first, in the next section, provide a mathematical analysis of optimising abstract linear PMDs. In particular, the motivation for using Legendre polynomials is discussed in Remark 4.6.

4 Analysis of optimising linear PMDs

Throughout this section we assume the linear parametrisation

$$h_{\theta_i}(x) = \theta_i \cdot \eta(x), \quad x \in \overline{\Omega}, \quad i \in [N], \quad (20)$$

with design function $\eta : \overline{\Omega} \rightarrow \mathbb{R}^K$. For $\varepsilon > 0$ define the soft assignment

$$(p_{\boldsymbol{\theta}}^\varepsilon)_i(x) := \frac{\exp(-\frac{1}{\varepsilon}h_{\theta_i}(x))}{\sum_{j=1}^N \exp(-\frac{1}{\varepsilon}h_{\theta_j}(x))}$$

and recall that the objective function is

$$\Phi_\varepsilon^G(\boldsymbol{\theta}) := \frac{1}{|\Omega|} \sum_{x \in \Omega} \log((p_{\boldsymbol{\theta}}^\varepsilon)_{G(x)}(x)) \leq 0. \quad (21)$$

We also recall the induced hard assignment map

$$F_{\boldsymbol{\theta}}(x) := \arg \min_{i \in [N]} h_{\theta_i}(x),$$

with the same tie-breaking rule as in Section 2.1. Finally, define the fraction of misassigned pixels

$$\text{Err}^G(\boldsymbol{\theta}) := \frac{1}{|\Omega|} \left| \{x \in \Omega : F_{\boldsymbol{\theta}}(x) \neq G(x)\} \right|.$$

Note that by construction $\text{Err}^G(\boldsymbol{\theta}) = 1 - \text{Acc}^G(\boldsymbol{\theta})$ (defined in (5)).

We now present a short, self-contained analysis of the optimisation problem $\max \Phi_\varepsilon^G$, which is equivalent to softmax (multinomial logistic) regression with log-sum-exp normalisation. We do not claim novelty, as these results are standard in the statistics and machine-learning literature [17, 25, 37], including the classical discussion of (non-)existence of maximisers under separation, which to the best of our knowledge can be traced back to [2, 46].

We begin by showing that since the parametrisation is linear, the parameter $\varepsilon > 0$ can always be absorbed into $\boldsymbol{\theta}$.

Lemma 4.1 (ε -parameter scaling). *For any $\varepsilon > 0$ and $\boldsymbol{\theta} \in \mathbb{R}^{K \times N}$, define $\boldsymbol{\vartheta} := \boldsymbol{\theta}/\varepsilon$ (i.e., $\vartheta_i = \theta_i/\varepsilon$ for all i). Then, for all $x \in \Omega$,*

$$F_{\boldsymbol{\vartheta}}(x) = F_{\boldsymbol{\theta}}(x), \quad (p_{\boldsymbol{\vartheta}}^\varepsilon)_i(x) = (p_{\boldsymbol{\theta}}^1)_i(x) \quad i \in [N]$$

and in particular $\Phi_\varepsilon^G(\boldsymbol{\theta}) = \Phi_1^G(\boldsymbol{\vartheta})$.

Proof. This follows by a direct computation. □

Next we turn to the so-called gauge invariance and fixing.

Lemma 4.2 (Gauge invariance). *For any $c \in \mathbb{R}^K$, define $\boldsymbol{\theta}' = (\theta'_1, \dots, \theta'_N)$ by $\theta'_i := \theta_i + c$ for all i . Then for all $x \in \Omega$,*

$$F_{\boldsymbol{\theta}'}(x) = F_{\boldsymbol{\theta}}(x), \quad (p_{\boldsymbol{\theta}'}^\varepsilon)_i(x) = (p_{\boldsymbol{\theta}}^\varepsilon)_i(x), \quad i \in [N],$$

and hence $\Phi_\varepsilon^G(\boldsymbol{\theta}') = \Phi_\varepsilon^G(\boldsymbol{\theta})$.

Proof. This follows by a direct computation. \square

Remark 4.3 (Gauge fixing). We can factor out the gauge invariance by K independent linear constraints removing the common-shift direction. The simplest one is to treat one of the cells as the reference one and, e.g., fix $\theta_N \equiv 0$.

Proposition 4.4 (Concavity of Φ_ε^G and gradient and Hessian formulas). *Fix $\varepsilon > 0$. Under (20), the objective function Φ_ε^G in (21) is concave. Moreover, for each $i \in [N]$,*

$$\nabla_{\theta_i} \Phi_\varepsilon^G(\boldsymbol{\theta}) = -\frac{1}{\varepsilon|\Omega|} \sum_{x \in \Omega} \left(\delta_{iG(x)} - (p_{\boldsymbol{\theta}}^\varepsilon)_i(x) \right) \eta(x). \quad (22)$$

Similarly, for $i, k \in [N]$, the (i, k) -block of the Hessian is

$$\nabla_{\theta_k} \nabla_{\theta_i} \Phi_\varepsilon^G(\boldsymbol{\theta}) = -\frac{1}{\varepsilon^2|\Omega|} \sum_{x \in \Omega} (p_{\boldsymbol{\theta}}^\varepsilon)_i(x) (\delta_{ik} - (p_{\boldsymbol{\theta}}^\varepsilon)_k(x)) (\eta(x) \otimes \eta(x)), \quad (23)$$

and the Hessian is negative semidefinite (with a null direction corresponding to the gauge shift in Lemma 4.2).

Proof. Formulas (22) and (23) follow from a direct computation. The negative semidefiniteness of the Hessian follows from the proof of Proposition 4.5 below. \square

Proposition 4.5 (Strict concavity of Φ_ε^G). *Assume that the set $\{\eta(x) : x \in \Omega\}$ spans \mathbb{R}^K . Then the nullspace of the Hessian consists only of vectors of the form $\mathbf{v} = (c, \dots, c)$ for some $c \in \mathbb{R}^K$, and the function Φ_ε^G is strictly concave on the orthogonal complement of this nullspace.*

Proof. To improve readability, set $p_i(x) := (p_{\boldsymbol{\theta}}^\varepsilon)_i(x)$, and for every $x \in \Omega$ denote by

$$P_x = (p_1(x), \dots, p_N(x))$$

the associated probability distribution. Let $\mathbf{v} = (v_1, \dots, v_N)$ be a block vector, where $v_i \in \mathbb{R}^K$. The quadratic form associated with the Hessian matrix of $\Phi_\varepsilon^G(\boldsymbol{\theta})$ is

$$\begin{aligned} \mathbf{v}^\top \nabla^2 \Phi_\varepsilon^G(\boldsymbol{\theta}) \mathbf{v} &= -\frac{1}{\varepsilon^2|\Omega|} \sum_{x \in \Omega} \left[\sum_{i=1}^N p_i(x) (v_i \cdot \eta(x))^2 - \left(\sum_{i=1}^N p_i(x) (v_i \cdot \eta(x)) \right)^2 \right] \\ &= -\frac{1}{\varepsilon^2|\Omega|} \sum_{x \in \Omega} \text{Var}_{P_x}(i \mapsto v_i \cdot \eta(x)), \end{aligned}$$

where $\text{Var}_{P_x}(\cdot)$ denotes the variance with respect to the probability distribution P_x . Since the variance is non-negative, the quadratic form is non-positive, proving concavity of Φ_ε^G .

The formula above shows that \mathbf{v} belongs to the nullspace of the Hessian of Φ_ε^G if and only if for all x , the variance of $i \mapsto v_i \cdot \eta(x)$ vanishes. Since the probabilities $p_i(x) > 0$ are all strictly positive, this implies that $i \mapsto v_i \cdot \eta(x)$ is constant for all x . In particular, for each i the vector $v_i - v_1$ is orthogonal to $\eta(x)$ for all $x \in \Omega$. Since by assumption the vectors $\eta(x)$, $x \in \Omega$, span the entire space \mathbb{R}^K , we get $v_i \equiv v_1$ for all i , which corresponds exactly to the gauge shift invariance. \square

Remark 4.6 (Conditioning and choice of basis). The Hessian in (23) is built from weighted second-moment matrices of the form

$$\sum_{x \in \Omega} (p_{\boldsymbol{\theta}}^\varepsilon)_i(x) \eta(x) \otimes \eta(x),$$

and the quadratic-form representation in the proof of Proposition 4.5 shows that their conditioning directly influences the conditioning of the optimisation problem. This motivates choosing a basis η for which these matrices are better conditioned and, in favourable cases, close to diagonal. In particular, after rescaling the domain to $[-1, 1]^2$, tensor-product Legendre polynomials are a natural choice: in the continuum they are orthogonal with respect to the uniform measure [51, Chapter 17], and on sufficiently regular discrete grids the corresponding discrete Gram matrices are typically much better conditioned, and often closer to diagonal, than those arising from the monomial basis [23]. Although the weights $(p_{\boldsymbol{\theta}}^\varepsilon)_i(x)$ mean that the matrices above are not, in general, exactly diagonal in the Legendre basis, one still expects a substantial conditioning advantage over monomials whenever the induced discrete measures are not too far from uniform. This is one of the motivations for the Legendre-based parametrisation discussed in Section 3.3.

Next we investigate what happens as $\varepsilon \rightarrow 0$.

Proposition 4.7 (Uniform $\varepsilon \rightarrow 0$ limit of the rescaled objective). *Define $\mathcal{E}_\varepsilon(\boldsymbol{\theta}) := -\varepsilon \Phi_\varepsilon^G(\boldsymbol{\theta})$. Then for every $\boldsymbol{\theta} \in \mathbb{R}^{K \times N}$ and every $\varepsilon > 0$,*

$$0 \leq \mathcal{E}_\varepsilon(\boldsymbol{\theta}) - \mathcal{E}_0(\boldsymbol{\theta}) \leq \varepsilon \log N, \quad (24)$$

where the limit functional is

$$\mathcal{E}_0(\boldsymbol{\theta}) := \frac{1}{|\Omega|} \sum_{x \in \Omega} \left(h_{\theta_{G(x)}}(x) - \min_{j \in [N]} h_{\theta_j}(x) \right) \geq 0. \quad (25)$$

In particular, $\mathcal{E}_\varepsilon \rightarrow \mathcal{E}_0$ uniformly on $\mathbb{R}^{K \times N}$ as $\varepsilon \rightarrow 0$. Moreover, if perfect reconstruction of the given grain map G is possible (i.e., if there exists $\boldsymbol{\theta}$ such that $F_{\boldsymbol{\theta}}(x) = G(x)$ for all $x \in \Omega$), then $\min \mathcal{E}_0 = 0$.

Proof. For each $x \in \Omega$, write $a_j(x) := -h_{\theta_j}(x)$. Then

$$\varepsilon \log \sum_j e^{a_j(x)/\varepsilon}$$

is the standard log-sum-exp smooth approximation of $\max_j a_j(x)$, with the sharp bounds (see, e.g., [12])

$$\max_j a_j(x) \leq \varepsilon \log \sum_j e^{a_j(x)/\varepsilon} \leq \max_j a_j(x) + \varepsilon \log N.$$

Substituting $a_j(x) = -h_{\theta_j}(x)$, rearranging, and averaging over x yields

$$0 \leq \frac{1}{|\Omega|} \sum_{x \in \Omega} \varepsilon \log \sum_j \exp\left(-\frac{h_{\theta_j}(x)}{\varepsilon}\right) + \min_{j \in [N]} h_{\theta_j}(x) \leq \varepsilon \log N.$$

Then (24) follows immediately from the definitions of \mathcal{E}_ε and \mathcal{E}_0 . For perfect reconstruction, one has $h_{\theta_{G(x)}}(x) = \min_{j \in [N]} h_{\theta_j}(x)$ for all $x \in \Omega$, and hence $\min \mathcal{E}_0 = 0$, as claimed. \square

The final set of results concerns the question of (non-)existence of maximisers of Φ_ε^G , which is known to depend on whether the diagram representation can perfectly recover the given grain map G [2, 46].

Proposition 4.8 (Perfect reconstruction implies the supremum is not attained). *Assume that $N > 1$ and there exists θ^0 such that $\text{Err}^G(\theta^0) = 0$ (so $F_{\theta^0}(x) = G(x)$ for all $x \in \Omega$) and $G(x)$ is the strict minimiser of $i \mapsto h_{\theta_i^0}(x)$ for all $x \in \Omega$. Then, for any fixed $\varepsilon > 0$,*

$$\sup_{\theta} \Phi_\varepsilon^G(\theta) = 0,$$

and the supremum is not attained (i.e., there is no maximiser in $\mathbb{R}^{K \times N}$, even after gauge fixing).

Proof. Let $\theta^t := t\theta^0$. For each x , since $G(x)$ is a strict minimiser of $h_{\theta_i^0}(x)$ among $i \in [N]$, one has

$$(p_{\theta^t}^\varepsilon)_{G(x)}(x) = \frac{1}{1 + \sum_{j \neq G(x)} \left(\frac{\exp(-h_{\theta_j^0}(x)/\varepsilon)}{\exp(-h_{\theta_{G(x)}^0}(x)/\varepsilon)} \right)^t} \rightarrow 1 \quad \text{as } t \rightarrow \infty,$$

hence $\Phi_\varepsilon^G(\theta^t) \rightarrow 0$. On the other hand, for any finite θ and any x , one has $(p_\theta^\varepsilon)_{G(x)}(x) < 1$ because $N > 1$, so $\Phi_\varepsilon^G(\theta) < 0$ and thus the supremum cannot be attained. \square

Lemma 4.9 (Pixel misassignment has a cost). *Assume that $N > 1$. For any θ and any $x \in \Omega$, if $F_\theta(x) \neq G(x)$ then*

$$(p_\theta^\varepsilon)_{G(x)}(x) \leq \frac{1}{2} \quad \text{and hence} \quad \log((p_\theta^\varepsilon)_{G(x)}(x)) \leq -\log 2.$$

Consequently,

$$\Phi_\varepsilon^G(\theta) \leq -(\log 2) \text{Err}^G(\theta). \quad (26)$$

Proof. If $F_\theta(x) \neq G(x)$, then

$$(p_\theta^\varepsilon)_{G(x)}(x) \leq \frac{\exp(-h_{\theta_{G(x)}}(x)/\varepsilon)}{\exp(-h_{\theta_{G(x)}}(x)/\varepsilon) + \exp(-h_{\theta_{F_\theta(x)}}(x)/\varepsilon)} \leq \frac{1}{2}$$

as required. \square

The following is a direct consequence.

Corollary 4.10 (Near-optimality implies (near-)perfect reconstruction). *For any $\delta > 0$,*

$$\Phi_\varepsilon^G(\theta) > -\delta \quad \implies \quad \text{Err}^G(\theta) < \frac{\delta}{\log 2}.$$

In particular, if $\delta < \frac{\log 2}{|\Omega|}$, then

$$\Phi_\varepsilon^G(\theta) > -\delta$$

implies $\text{Err}^G(\theta) = 0$, i.e., $F_\theta(x) = G(x)$ for all $x \in \Omega$.

Note that this in particular implies that if a perfect reconstruction is possible, then any approximately optimal θ achieves it.

Proposition 4.11 (Imperfect reconstruction case: existence of a maximiser). *If there is no θ such that $\text{Err}^G(\theta) = 0$ (which is equivalent to $F_\theta(x) = G(x)$ for all $x \in \Omega$), then Φ_ε^G admits a maximiser (which is not unique due to the gauge invariance). Assume in addition that the set $\{\eta(x) : x \in \Omega\}$ spans \mathbb{R}^K . Then Φ_ε^G has a unique maximiser on the subspace $\{(c, \dots, c) : c \in \mathbb{R}^K\}^\perp$. Consequently Φ_ε^G also has a unique maximiser on the gauge-fixed subspace $\{\theta \in \mathbb{R}^{K \times N} : \theta_N = 0\}$.*

Proof. See [2, 46] and Proposition 4.5. □

5 Numerical experiments

5.1 Implementation and computational setting

In all the experiments we optimise the multinomial logistic objective function Φ_ε^G defined in (21) under the linear parametrisation (20), using Legendre polynomial diagrams (Section 3) of degrees $d \in \{1, \dots, 7\}$. We use the gauge fixing from Remark 4.3 by setting $\theta_N \equiv 0$ throughout.

The implementation is part of the GPU-oriented library PYAPD [18, 1]. The evaluation of Φ_ε^G and its gradient uses a PYKEOPS backend [19], avoiding explicit formation of the full cost matrix

$$(h_{\theta_i}(x))_{i \in [N], x \in \Omega} \in \mathbb{R}^{N \times |\Omega|},$$

whose (i, x) -entry is the cost of assigning pixel x to grain i . All runs were performed on a single NVIDIA A100 GPU using double precision arithmetic. As discussed in [18], for a combination of N (number of grains), $|\Omega|$ (number of pixels) and d (degree of the diagram) large enough, our implementation ensures near 100% utilisation of the GPU.

All data used and generated in the numerical experiments is available in a reproducible manner on GitHub through our library, PyAPD [1].

5.2 Optimisation protocol

To showcase the robustness of the method, we first initialise at the fully ambiguous state $\theta^{(0)} = \mathbf{0}$, which implies that $(p_{\theta^{(0)}}^\varepsilon)_i \equiv 1/N$. A second initialisation choice comes from the heuristic guess from [50] – see Section 5.6.2 for details.

We maximise Φ_ε^G using the L-BFGS method [32]. In all cases we rely on the automatic differentiation available in PYTORCH to compute derivatives. This is known to be faster than using the explicit formulas from Proposition 4.4 [18]. Reflecting the fact that Φ_ε^G becomes very flat near the optimum, we always run the optimisation for a fixed moderate number of iterations, either 1000 or 2000, and fix the regularisation parameter to be $\varepsilon = 10^{-2}$. A robust *annealing schedule* [52] of systematically decreasing ε and a bespoke stopping criterion will be considered in future work (see also the last paragraph of Section 5.6.4 for a related discussion).

5.3 Datasets and common discretisation

We always take $\bar{\Omega} = [-1, 1]^2$ with discretisation Ω , either as in Section 2.1 or provided by the EBSD dataset. A grain map $G : \Omega \rightarrow [N]$ is provided either synthetically (from a known power diagram or anisotropic power diagram), with discretisation parameter M chosen as suggested in [18, Table 1], or from experimental EBSD data (pre-processing in MTEX [11]).

5.4 Reconstruction of (anisotropic) power diagrams

To show that the method works as expected, we first consider synthetic grain maps G generated by (A) a power diagram (PD) and (B)–(E) anisotropic power diagrams (APD) with increasing anisotropy (elongation of the grains). These experiments are in the perfect reconstruction regime: there exists θ^* such that $F_{\theta^*}(x) = G(x)$ for all $x \in \Omega$, hence $\sup \Phi_\varepsilon^G = 0$ is not attained (Proposition 4.8) and we look for a near optimum achieving perfect reconstruction (Corollary 4.10). In practice, to counter the increasing cost of L-BFGS iterates as $\|\theta\|$ grows, we impose a fixed budget of 2000 iterations and accept a (negligibly) small value Err^G . The results are presented in Figure 1.

5.5 Fitting to EBSD grain maps

To showcase the true strength of the proposed framework, we next fit Legendre polynomial diagrams to EBSD-derived grain maps. In contrast to the synthetic setting of Section 5.4, the EBSD grain boundaries are not expected to lie exactly in the algebraic class induced by a fixed degree d , so the problem only allows for imperfect reconstruction and a maximiser exists (Proposition 4.11). We work with two EBSD datasets provided to us by our industrial partner, Tata Steel Netherlands. The first, referred to as the *small EBSD dataset*, yields a grain map $G : \Omega \rightarrow [N]$ with $N = 245$ grains and $|\Omega| = 63,252$ pixels. The second, referred to as the *big EBSD dataset*, results in a grain map with $N = 4686$ and $|\Omega| = 1,033,376$ pixels. The results are presented in Figure 2 and Figure 3.

5.6 Discussion

We now report on several important aspects illuminated in our numerical experiments.

5.6.1 Degree d .

Increasing d increases the number of parameters per grain to $K_d = \frac{(d+1)(d+2)}{2}$ and permits more complex algebraic interfaces (Remark 3.2). Empirically, the optimal value of the objective function Φ_ε^G improves monotonically with d , while the reconstruction accuracy on EBSD data improves up to a problem-dependent saturation degree. The improvement is mathematically obvious, since any lower-degree fit is an admissible higher-degree fit (with additional parameters set to zero).

A simple quantity that helps assess the trade-off between model complexity and data size is the *compression ratio*

$$\text{Compr}(d, N, |\Omega|) := \frac{K_d N}{3|\Omega|}. \quad (27)$$

Interpreting a labelled pixel set naively as storing, per pixel, two coordinates and one label, a grain map requires $3|\Omega|$ scalar values, whereas a fitted degree- d diagram is specified by $K_d N$ coefficients. This convention is also reasonable in practice for EBSD data, where grain maps are often handled as unstructured pixel lists (for instance after removing small grains or leaving unindexed regions as gaps), so that the pixel locations are stored explicitly together with the grain labels. We report on the compression ratio for our EBSD experiments in Table 1.

Compr data set	degree d							
	1	2	3	4	5	6	7	8
Small EBSD	0.39%	0.77%	1.29%	1.94%	2.71%	3.62%	4.65%	5.81%
Big EBSD	0.45%	0.91%	1.51%	2.27%	3.17%	4.23%	5.44%	6.80%

Table 1: The compression ratio (27) for the small and big EBSD datasets when reconstructed with a polynomial diagram of degree d for $d \in [8]$.

5.6.2 Moment-based heuristic initialisation for PDs and APDs.

For the cases $d = 1$ and $d = 2$, an alternative to the fully ambiguous initialisation $\boldsymbol{\theta}^{(0)} = \mathbf{0}$ is to use the well-known moment-based heuristic from [50], which was derived for the standard description of power diagrams in terms of (\mathbf{w}, \mathbf{y}) and anisotropic power diagrams in terms of $(\mathbf{w}, \mathbf{y}, \mathbf{A})$ (c.f. Section 2.3.1), rather than in terms of $\boldsymbol{\theta}$. For a given grain map $G : \Omega \rightarrow [N]$, it can be computed as follows. First we compute first moments (centroids)

$$y'_i := \frac{1}{|G_i|} \sum_{x \in G_i} x \quad (28)$$

and the (central) second-moment matrices

$$(B'_i)_{kl} := \frac{1}{|G_i|} \sum_{x \in G_i} (x - y'_i)_k (x - y'_i)_l, \quad k, l \in [2]. \quad (29)$$

If B'_i is invertible, the anisotropy heuristic suggested in [50] is

$$A'_i := (B'_i)^{-1}, \quad (30)$$

and the weight heuristic is

$$w'_i = \frac{\sqrt{\det A'_i} |G_i|}{|\Omega| \pi} \quad (31)$$

(in the case of a PD, we take $A'_i = \text{Id}$ in (31)). Note that w'_i is the ratio of the ‘area’ of the grain G_i divided by the area of the ellipse corresponding to B'_i . These heuristics give rise to an initial guess $(\mathbf{w}', \mathbf{y}')$ for a PD and $(\mathbf{w}', \mathbf{y}', \mathbf{A}')$ for an APD, which can then be converted to the linear parameter representation $\boldsymbol{\theta}'$ via the transformations in Section 2.3.1 and also to the Legendre basis as discussed in Remark 3.1.

We advocate using this heuristic guess as an initialisation (even for $d \geq 3$ by setting higher degree coefficients to zero), provided that ε is chosen small enough (see Section 5.6.4). The benefit is clearly seen in Figure 3, where for $d \geq 2$ starting from the heuristic guess leads to a much reduced error in pixel assignment.

5.6.3 Recovering a physically-relevant standard description.

When $d \in \{1, 2\}$ we may optimise in the linear coefficient representation $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$ (with a gauge fix $\theta_N \equiv 0$), and then convert a posteriori to the standard geometric parameters (\mathbf{w}, \mathbf{y}) (power diagrams) or $(\mathbf{w}, \mathbf{y}, \mathbf{A})$ (anisotropic power diagrams) via equations (12) and

(15). One should keep in mind that this conversion is *not canonical*: by Lemma 4.2 the diagram depends only on the equivalence class $\boldsymbol{\theta} \sim \boldsymbol{\theta} + (c, \dots, c)$, since adding the same $c \in \mathbb{R}^K$ shifts all scores by the same function $x \mapsto c \cdot \eta(x)$ and hence does not change the argmin. Consequently, the recovered physical parameters are defined only up to such a common shift. Note that the diagram is also invariant under multiplicative rescaling, $\boldsymbol{\theta} \sim \lambda \boldsymbol{\theta}$, for any $\lambda > 0$, but this transformation does not leave the objective function Φ_ε^G invariant (unless ε is also rescaled, see Lemma 4.1). This is discussed further in Section 5.6.4.

Recall that for $d = 2$ (APDs) the anisotropy matrices A_i are positive definite. However, after imposing the optimisation gauge $\theta_N = 0$, the matrix $A_N = \mathbf{0}$ is not positive definite, even though the induced diagram $\{L_i\}$ (defined in (18)) is perfectly meaningful. Similarly, maximising Φ_ε^G over all of $\mathbb{R}^{6 \times N}$ without gauge-fixing may lead to some A_i that are not positive definite. However, the gauge freedom can be used to enforce a physically-admissible anisotropy (i.e., A_i positive definite for all i) in post-processing. Indeed, in the monomial basis (14) the matrix A_i is read off from the quadratic coefficients as follows:

$$A_i = \begin{pmatrix} \theta_{i,1} & \frac{1}{2}\theta_{i,2} \\ \frac{1}{2}\theta_{i,2} & \theta_{i,3} \end{pmatrix}.$$

Adding to every θ_i a vector $c \in \mathbb{R}^6$ of the form $c = (c_1, c_2, c_3, 0, 0, 0)$ amounts to replacing every A_i by $A_i + C$, where C is the same symmetric matrix for all grains, while leaving the diagram $\{L_i\}$ (defined in (18)) unchanged. Choosing $C = \lambda I$ with

$$\lambda > -\min_{i \in [N]} \lambda_{\min}(A_i),$$

where $\lambda_{\min}(A_i)$ denotes the smallest eigenvalue of A_i , ensures that $A_i + C$ is positive definite for all i . One then recomputes (y_i, w_i) from (15) using the shifted coefficients. In summary, this procedure leaves the diagram $\{L_i\}$ unchanged while enforcing the constraint that the anisotropy matrices are positive definite, even though we solve the *unconstrained* optimisation problem $\max \Phi_\varepsilon^G$. This observation was made in [5]. If optimisation is carried out in a different polynomial basis, e.g. Legendre, the same procedure applies after converting θ to the monomial basis.

For $d = 1$ (PDs), the same phenomenon appears in a simpler form: a common shift of the linear coefficients corresponds to a common translation of all recovered seeds y_i (and a corresponding adjustment of the weights), again without changing the diagram. This invariance is well-known in the optimal transport literature; see, e.g., [35, Prop. 6].

5.6.4 Practical role of the regularisation parameter ε .

Lemma 4.1 shows that ε can be absorbed into a rescaling of parameters:

$$\Phi_\varepsilon^G(\boldsymbol{\theta}) = \Phi_1^G(\boldsymbol{\theta}/\varepsilon),$$

and hence (assuming a maximiser exists) maximisers satisfy

$$\bar{\boldsymbol{\theta}} \in \arg \max \Phi_1^G \implies \varepsilon \bar{\boldsymbol{\theta}} \in \arg \max \Phi_\varepsilon^G.$$

Separately, the induced hard assignment satisfies:

$$F_{\varepsilon \boldsymbol{\theta}}(x) = F_{\boldsymbol{\theta}}(x) \quad (\varepsilon > 0), \quad \text{and hence} \quad \text{Err}^G(\varepsilon \boldsymbol{\theta}) = \text{Err}^G(\boldsymbol{\theta}).$$

Thus, for the exact problem in exact arithmetic, ε carries no modelling content.

In computation, however, ε has a pronounced effect on the optimisation trajectory. Empirically, the heuristic initialisation $\boldsymbol{\theta}^{(0)} = \boldsymbol{\theta}'$ from Section 5.6.2 may yield a good hard reconstruction (small $\text{Err}^G(\boldsymbol{\theta}')$) while the corresponding values $h_{\theta'_i}(x)$ exhibit only weak separation between the winning index and its competitors on substantial subsets of pixels. With $\varepsilon \sim 1$ this typically leads to early L-BFGS iterates that substantially rearrange interfaces, and Err^G can increase markedly before decreasing again; in all the numerical experiments considered, this transient deterioration effectively removes the advantage of the heuristic relative to the fully ambiguous start $\boldsymbol{\theta}^{(0)} = \mathbf{0}$.

By contrast, taking $\varepsilon \ll 1$ makes the initial iterate behave more deterministically and we observe a much more stable evolution of Err^G , frequently with near-monotone decrease, albeit sometimes with slower progress in Φ_ε^G once the reconstruction is already good. Importantly, this behaviour can be reproduced while keeping $\varepsilon = 1$ by rescaling the initial parameters: since $\Phi_\varepsilon^G(\boldsymbol{\theta}) = \Phi_1^G(\boldsymbol{\theta}/\varepsilon)$, running at small ε from $\boldsymbol{\theta}'$ is equivalent (for the objective and its derivatives) to running at $\varepsilon = 1$ from the enlarged initialisation $\boldsymbol{\theta}'/\varepsilon$. Designing robust annealing strategies based on this scaling (and stopping criteria aligned with Err^G) is an algorithmic issue and will be pursued in future work.

6 Conclusions and outlook

We formulated a framework of parametrised minimisation diagrams (PMDs) and focused on the linear class $h_{\theta_i}(x) = \theta_i \cdot \eta(x)$, for which the logistic regression objective function Φ_ε^G is concave, and strictly concave after gauge fixing. Standard power diagrams and anisotropic power diagrams arise as low-degree instances of this framework, and polynomial diagrams provide a systematic extension in which cell interfaces are algebraic curves of controlled degree, which is the main novelty of this work in the context of microstructure modelling.

On the computational side, the concave objective function admits reliable large-scale optimisation. In the perfect reconstruction regime (synthetic PD/APD ground truth) the supremum of Φ_ε^G is not attained, but nevertheless near-optimal approximate maximisers yield a perfect reconstruction (no misassigned pixels). In the imperfect reconstruction regime, typically arising when dealing with EBSD data, a maximiser exists, and increasing the polynomial degree improves accuracy but with diminishing returns.

Several directions for further work appear natural. First, while degrees $d \geq 3$ empirically fit complex experimental boundaries well, their geometric and physical interpretation is less clear than for $d = 1$ (PD) and $d = 2$ (APD); understanding what classes of boundary regularity and anisotropy are efficiently represented by higher-degree diagrams is an open question. Second, the linear framework is not restricted to polynomial features: one may enrich the design function η with additional basis functions (e.g., trigonometric or radial functions) to capture oscillatory or multiscale boundaries, while preserving the concavity of Φ_ε^G . Finally, incorporating a physically-motivated regularisation term in the objective function (e.g., penalising curvature or encouraging mild shape constraints on the cells, such as approximate convexity or star-shapedness) could improve robustness on noisy EBSD segmentations and facilitate the interpretation of fitted parameters.

Another possible perspective is to view the present fitting problem as a semi-discrete inverse optimal transport problem. Indeed, in semi-discrete optimal transport, Laguerre cells arise from transporting an absolutely continuous source measure to a discrete target measure

under a parametrised cost. In our setting, one may heuristically interpret the labelled pixels as samples from such an optimal partition and the optimisation over PMD parameters as learning, within a restricted parametrised class, both the effective cost and the induced target masses from these samples. This is closely related in spirit to inverse optimal transport, where one seeks to infer the ground cost from samples of an observed coupling or matching [31, 20, 9, 42]. Making this connection precise, and clarifying what may be gained from the inverse-transport viewpoint both theoretically and computationally, would be an interesting direction for future work.

References

- [1] *PyAPD: A Python library for computing anisotropic power diagrams using GPU acceleration, v0.2.0*. <https://github.com/mbuze/PyAPD>, 2026.
- [2] A. ALBERT AND J. A. ANDERSON, *On the existence of maximum likelihood estimates in logistic regression models*, *Biometrika*, 71 (1984), pp. 1–10.
- [3] A. ALPERS, A. BRIEDEN, P. GRITZMANN, A. LYCKEGAARD, AND H. F. POULSEN, *Generalized balanced power diagrams for 3D representations of polycrystals*, *Philosophical Magazine*, 95 (2015), pp. 1016–1028, doi:10.1080/14786435.2015.1015469.
- [4] A. ALPERS, M. FIEDLER, P. GRITZMANN, AND F. KLEMM, *Dynamic grain models via fast heuristics for diagram representations*, *Philosophical Magazine*, 103 (2023), pp. 948–968, doi:10.1080/14786435.2023.2180679.
- [5] A. ALPERS, M. FIEDLER, P. GRITZMANN, AND F. KLEMM, *Turning grain maps into diagrams*, *SIAM Journal on Imaging Sciences*, 16 (2023), pp. 223–249, doi:10.1137/22M1491988.
- [6] A. ALPERS, O. FURAT, C. JUNG, M. NEUMANN, C. REDENBACH, A. SAKEN, AND V. SCHMIDT, *Comparative analysis of algorithms for the fitting of tessellations to 3D image data*, arXiv:2507.14268, (2025), <https://arxiv.org/abs/2507.14268>.
- [7] H. ALTENDORF, F. LATOURTE, D. JEULIN, M. FAESSEL, AND L. SAINTOYANT, *3D reconstruction of a multiscale microstructure by anisotropic tessellation models*, *Image Analysis and Stereology*, 33 (2014), pp. 121–130, doi:10.5566/ias.v33.p121-130.
- [8] M. P. ANDERSON, D. J. SROLOVITZ, G. S. GRETT, AND P. S. SAHNI, *Computer simulation of grain growth—i. kinetics*, *Acta Metallurgica*, 32 (1984), pp. 783–791, doi:10.1016/0001-6160(84)90151-2.
- [9] F. ANDRADE, G. PEYRÉ, AND C. POON, *Sparsistency for inverse optimal transport*, in *International Conference on Learning Representations*, 2024.
- [10] F. AURENHAMMER, *Voronoi diagrams—a survey of a fundamental geometric data structure*, *ACM Computing Surveys*, 23 (1991), pp. 345–405, doi:10.1145/116873.116880.
- [11] F. BACHMANN, R. HIELSCHER, AND H. SCHAEBEN, *Grain detection from 2D and 3D EBSD data—specification of the MTEX algorithm*, *Ultramicroscopy*, 111 (2011), pp. 1720–1733, doi:10.1016/j.ultramicro.2011.08.002.

- [12] P. BLANCHARD, D. J. HIGHAM, AND N. J. HIGHAM, *Accurately computing the log-sum-exp and softmax functions*, IMA Journal of Numerical Analysis, 41 (2021), pp. 2311–2330.
- [13] D. P. BOURNE, T. O. GALLOUËT, Q. MÉRIGOT, AND A. NATALE, *Semi-discrete convex order and Laguerre tessellation fitting*, hal-05008630, (2025), <https://inria.hal.science/hal-05008630v1>.
- [14] D. P. BOURNE, P. J. J. KOK, S. M. ROPER, AND W. D. T. SPANJER, *Laguerre tessellations and polycrystalline microstructures: a fast algorithm for generating grains of given volumes*, Philosophical Magazine, 100 (2020), pp. 2677–2707, doi:10.1080/14786435.2020.1790053.
- [15] D. P. BOURNE, M. PEARCE, AND S. M. ROPER, *Geometric modelling of polycrystalline materials: Laguerre tessellations and periodic semi-discrete optimal transport*, Mechanics Research Communications, 127 (2023), p. 104023, doi:10.1016/j.mechrescom.2022.104023.
- [16] D. P. BOURNE, M. PEARCE, AND S. M. ROPER, *Inverting Laguerre tessellations: Recovering tessellations from the volumes and centroids of their cells using optimal transport*, ESAIM: Mathematical Modelling and Numerical Analysis, 59 (2025), pp. 841–871, doi:10.1051/m2an/2025004.
- [17] S. BOYD AND L. VANDENBERGHE, *Convex Optimization*, Cambridge University Press, 2004.
- [18] M. BUZE, J. FEYDY, S. M. ROPER, K. SEDIGHIANI, AND D. P. BOURNE, *Anisotropic power diagrams for polycrystal modelling: Efficient generation of curved grains via optimal transport*, Computational Materials Science, 245 (2024), p. 113317, doi:10.1016/j.commatsci.2024.113317.
- [19] B. CHARLIER, J. FEYDY, J. A. GLAUNÈS, F.-D. COLLIN, AND G. DURIF, *Kernel operations on the GPU, with autodiff, without memory overflows*, Journal of Machine Learning Research, 22 (2021), pp. 1–6, <http://jmlr.org/papers/v22/20-275.html>.
- [20] W.-T. CHIU, P. WANG, AND P. SHAFTO, *Discrete probabilistic inverse optimal transport*, in International Conference on Machine Learning, PMLR, 2022, pp. 3925–3946.
- [21] D. DEPRIESTER AND R. KUBLER, *Radical Voronoï tessellation from random pack of polydisperse spheres: Prediction of the cells’ size distribution*, Computer-Aided Design, 107 (2019), pp. 37–49, doi:10.1016/j.cad.2018.09.001.
- [22] Z. FAN, Y. WU, X. ZHAO, AND Y. LU, *Simulation of polycrystalline structure with Voronoï diagram in Laguerre geometry based on random closed packing of spheres*, Computational Materials Science, 29 (2004), pp. 301–308, doi:10.1016/j.commatsci.2003.10.006.
- [23] W. GAUTSCHI, *How (un) stable are vandermonde systems?*, in Asymptotic and computational analysis, CRC Press, 2020, pp. 193–210.
- [24] E. O. HALL, *The deformation and ageing of mild steel: III discussion of results*, Proceedings of the Physical Society. Section B, 64 (1951), pp. 747–753, doi:10.1088/0370-1301/64/9/303.

- [25] T. HASTIE, R. TIBSHIRANI, AND J. FRIEDMAN, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2nd ed., 2009.
- [26] C. JUNG AND C. REDENBACH, *An analytical representation of the 2D generalized balanced power diagram*, Computational Geometry, 121 (2024), p. 102101, doi:10.1016/j.comgeo.2024.102101.
- [27] M. KRAUSE, T. BÖHLKE, AND M. SCHNEIDER, *Generating high-fidelity microstructures of polycrystalline materials with prescribed higher-order texture tensors*, Computer Methods in Applied Mechanics and Engineering, 452 (2026), p. 118690, doi:10.1016/j.cma.2025.118690.
- [28] J. KUHN, M. SCHNEIDER, P. SONNWEBER-RIBIC, AND T. BÖHLKE, *Fast methods for computing centroidal Laguerre tessellations for prescribed volume fractions with applications to microstructure generation of polycrystalline materials*, Computer Methods in Applied Mechanics and Engineering, 369 (2020), p. 113175, doi:10.1016/j.cma.2020.113175.
- [29] M. KÜHN AND M. O. STEINHAUSER, *Modeling and simulation of microstructures using power diagrams: Proof of the concept*, Applied Physics Letters, 93 (2008), p. 034102, doi:10.1063/1.2959733.
- [30] D.-T. LEE, *On k -nearest neighbor Voronoi diagrams in the plane*, IEEE Transactions on Computers, C-31 (1982), pp. 478–487.
- [31] R. LI, X. YE, H. ZHOU, AND H. ZHA, *Learning to match via inverse optimal transport*, Journal of Machine Learning Research, 20 (2019), pp. 1–37.
- [32] D. C. LIU AND J. NOCEDAL, *On the limited memory BFGS method for large scale optimization*, Mathematical Programming, 45 (1989), pp. 503–528, doi:10.1007/BF01589116.
- [33] A. LYCKEGAARD, E. M. LAURIDSEN, W. LUDWIG, R. W. FONDA, AND H. F. POULSEN, *On the use of Laguerre tessellations for representations of 3D grain structures*, Advanced Engineering Materials, 13 (2011), pp. 165–170, doi:10.1002/adem.201000258.
- [34] H. MECKING AND U. F. KOCKS, *Kinetics of flow and strain-hardening*, Acta Metallurgica, 29 (1981), pp. 1865–1875, doi:10.1016/0001-6160(81)90112-7.
- [35] J. MEYRON, *Initialization procedures for discrete and semi-discrete optimal transport*, Computer-Aided Design, 115 (2019), pp. 13–22, doi:10.1016/j.cad.2019.05.037.
- [36] N. MOELANS, B. BLANPAIN, AND P. WOLLANTS, *An introduction to phase-field modeling of microstructure evolution*, Calphad, 32 (2008), pp. 268–294, doi:10.1016/j.calphad.2007.11.003.
- [37] K. P. MURPHY, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.
- [38] E. PAPADOPOULOU AND M. ZAVERSHYNSKYI, *The higher-order Voronoi diagram of line segments*, Algorithmica, 74 (2016), pp. 415–439, doi:10.1007/s00453-014-9950-0.

- [39] L. PETRICH, O. FURAT, M. WANG, C. E. KRILL III, AND V. SCHMIDT, *Efficient fitting of 3D tessellations to curved polycrystalline grain boundaries*, *Frontiers in Materials*, 8 (2021), p. 760602, doi:10.3389/fmats.2021.760602.
- [40] L. PETRICH, J. STANĚK, M. WANG, D. WESTHOFF, L. HELLER, P. ŠITTNER, C. E. KRILL III, V. BENEŠ, AND V. SCHMIDT, *Reconstruction of grains in polycrystalline materials from incomplete data using Laguerre tessellations*, *Microscopy and Microanalysis*, 25 (2019), pp. 743–752, doi:10.1017/S1431927619000485.
- [41] C. PLATEAU-HOLLEVILLE, B. STAMM, V. NIVOLIERS, M. MARIA, AND S. MÉRILLOU, *In search of empty spheres: 3D Apollonius diagrams on GPU*, *ACM Transactions on Graphics (TOG)*, 44 (2025), pp. 1–15, doi:10.1145/3730868.
- [42] C. POON, *Some brief notes on inverse optimal transport*. https://cmhsp2.github.io/files/chemnitz2025/iOT_notes.pdf, 2025. Lecture notes for the 2025 Chemnitz Summer School on Applied Mathematics.
- [43] R. QUEY AND L. RENVERSADE, *Optimal polyhedral description of 3D polycrystals: Method and application to statistical and synchrotron X-ray diffraction data*, *Computer Methods in Applied Mechanics and Engineering*, 330 (2018), pp. 308–333, doi:10.1016/j.cma.2017.10.029.
- [44] R. RIFKIN AND A. KLAUTAU, *In defense of one-vs-all classification*, *Journal of Machine Learning Research*, 5 (2004), pp. 101–141, <https://jmlr.org/papers/volume5/rifkin04a/rifkin04a.pdf>.
- [45] F. ROTERS, P. EISENLOHR, L. HANTCHERLI, D. D. TIAHJANTO, T. R. BIELER, AND D. RAABE, *Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: Theory, experiments, applications*, *Acta Materialia*, 58 (2010), pp. 1152–1211, doi:10.1016/j.actamat.2009.10.058.
- [46] T. J. SANTNER AND D. E. DUFFY, *A note on A. Albert and J.A. Anderson’s conditions for the existence of maximum likelihood estimates in logistic regression models*, *Biometrika*, 73 (1986), pp. 755–758.
- [47] O. ŠEDIVÝ, T. BRERETON, D. WESTHOFF, L. POLÍVKA, V. BENEŠ, V. SCHMIDT, AND A. JÄGER, *3D reconstruction of grains in polycrystalline materials using a tessellation model with curved grain boundaries*, *Philosophical Magazine*, 96 (2016), pp. 1926–1949, doi:10.1080/14786435.2016.1183829.
- [48] O. ŠEDIVÝ, D. WESTHOFF, J. KOPEČEK, C. E. KRILL III, AND V. SCHMIDT, *Data-driven selection of tessellation models describing polycrystalline microstructures*, *Journal of Statistical Physics*, 172 (2018), pp. 1223–1246, doi:10.1007/s10955-018-2096-8.
- [49] A. SPETTL, T. BRERETON, Q. DUAN, T. WERZ, C. E. KRILL III, D. P. KROESE, AND V. SCHMIDT, *Fitting Laguerre tessellation approximations to tomographic image data*, *Philosophical Magazine*, 96 (2016), pp. 166–189, doi:10.1080/14786435.2015.1125540.
- [50] K. TEFERRA AND D. J. ROWENHORST, *Direct parameter estimation for generalised balanced power diagrams*, *Philosophical Magazine Letters*, 98 (2018), pp. 79–87, doi:10.1080/09500839.2018.1472399.

- [51] L. N. TREFETHEN, *Approximation theory and approximation practice, extended edition*, SIAM, 2019.
- [52] P. J. M. VAN LAARHOVEN AND E. H. L. AARTS, *Simulated Annealing: Theory and Applications*, Springer, 1987.

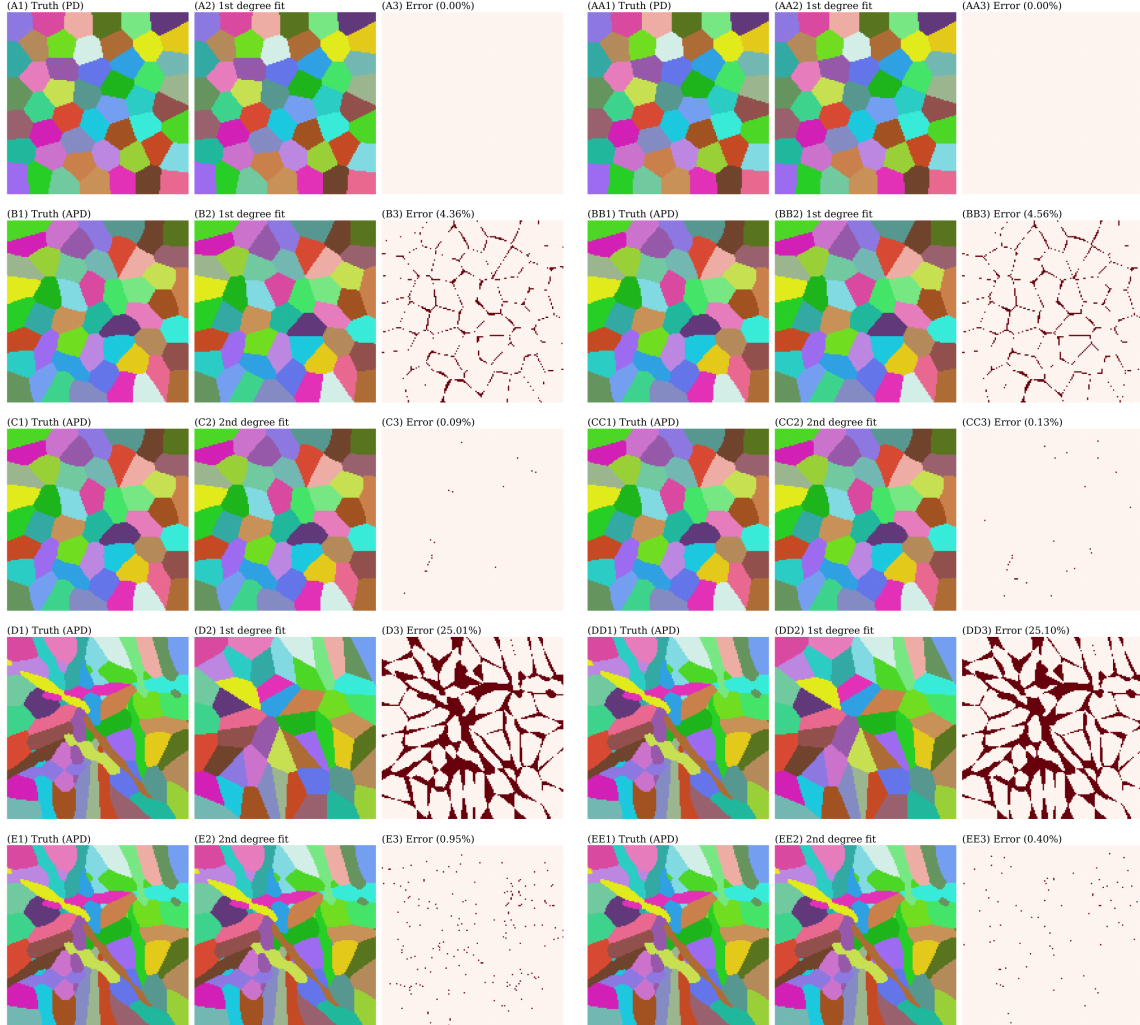


Figure 1: The reconstruction of (anisotropic) power diagrams as described in Section 5.4. Column 1: Ground truth. The given grain map G is generated by a power diagram (PD) or anisotropic power diagram (APD), where $G(x) = i$ if pixel x is in cell i . The colours correspond to the assignment of pixels. Note that (B1)=(C1) is an APD with low anisotropy and (D1)=(E1) is an APD with high anisotropy. Column 2: A power diagram (1st degree fit) or anisotropic power diagram (2nd degree fit) is fitted to the grain map G by approximately maximising Φ_ε^G with initial guess $\theta^{(0)} = \mathbf{0}$. Column 3: The misassignment error. The light pink pixels are correctly assigned, the dark pixels are incorrectly assigned. Columns 4–6: The same as columns 1–3 but with the initial guess $\theta^{(0)} = \theta'$ from Section 5.6.2. In each case $\varepsilon = 0.01$, there are $N = 50$ cells and $|\Omega| = 19,600$ pixels, and the algorithm was run for 2000 iterations and took around 10–41 seconds.

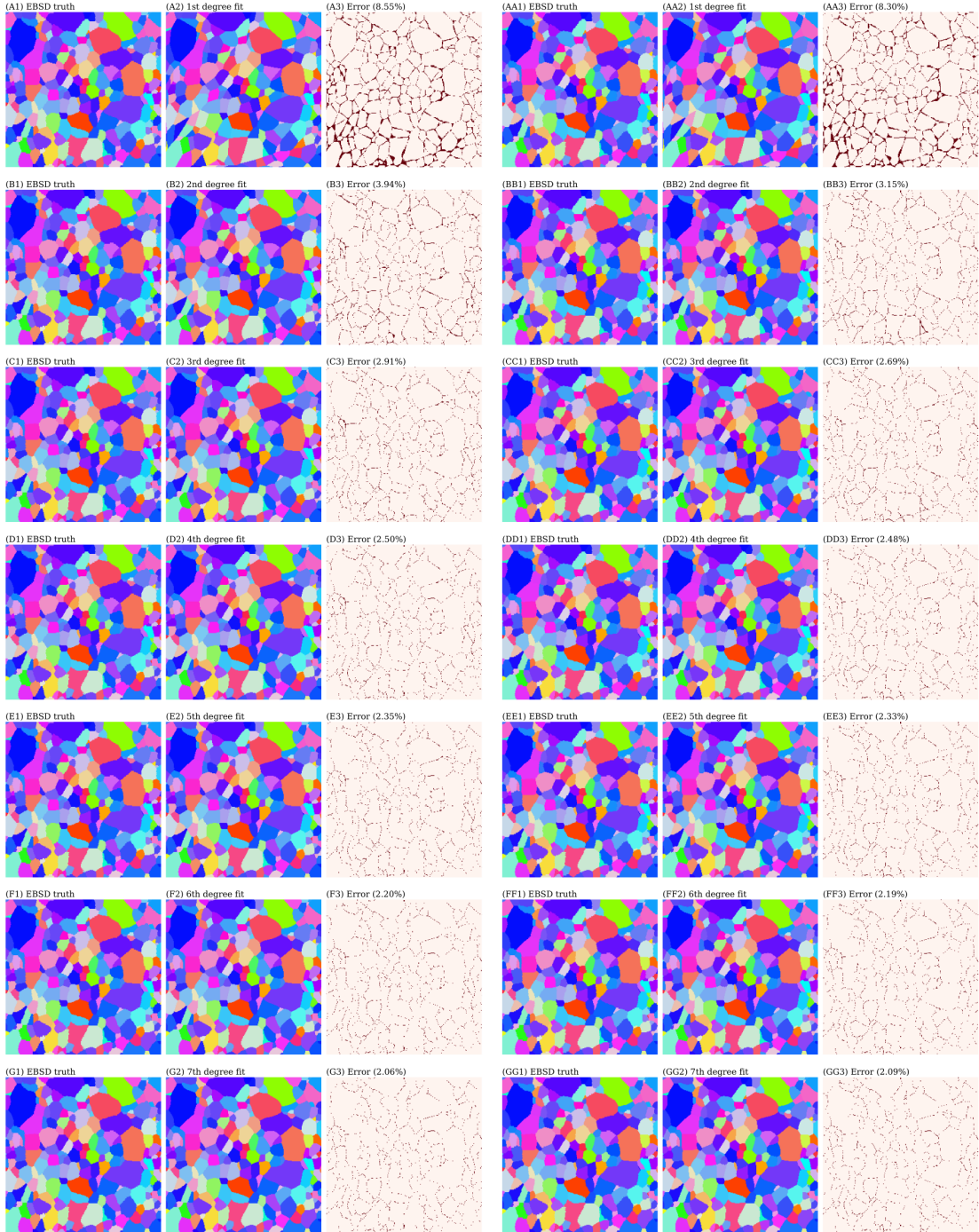


Figure 2: The fitting of polynomial diagrams to the small EBSD dataset as described in Section 5.5. Column 1: Ground truth. The given grain map G comes from an EBSD image of steel, where $G(x) = i$ if pixel x is in grain i . The colours correspond to the assignment of pixels. Column 2, row d : A polynomial diagram of degree d , for $d \in [7]$, is fitted to the grain map G by approximately maximising Φ_ε^G with initial guess $\theta^{(0)} = \mathbf{0}$. Column 3: The misassignment error. The light pink pixels are correctly assigned, the dark pixels are incorrectly assigned. Columns 4–6: The same as columns 1–3 but with the initial guess $\theta^{(0)} = \theta'$ from Section 5.6.2. In each case $\varepsilon = 0.01$, there are $N = 245$ grains and $|\Omega| = 63,252$ pixels, and the algorithm was run for 1000 iterations and took around 25–64 seconds (with run time increasing with d).

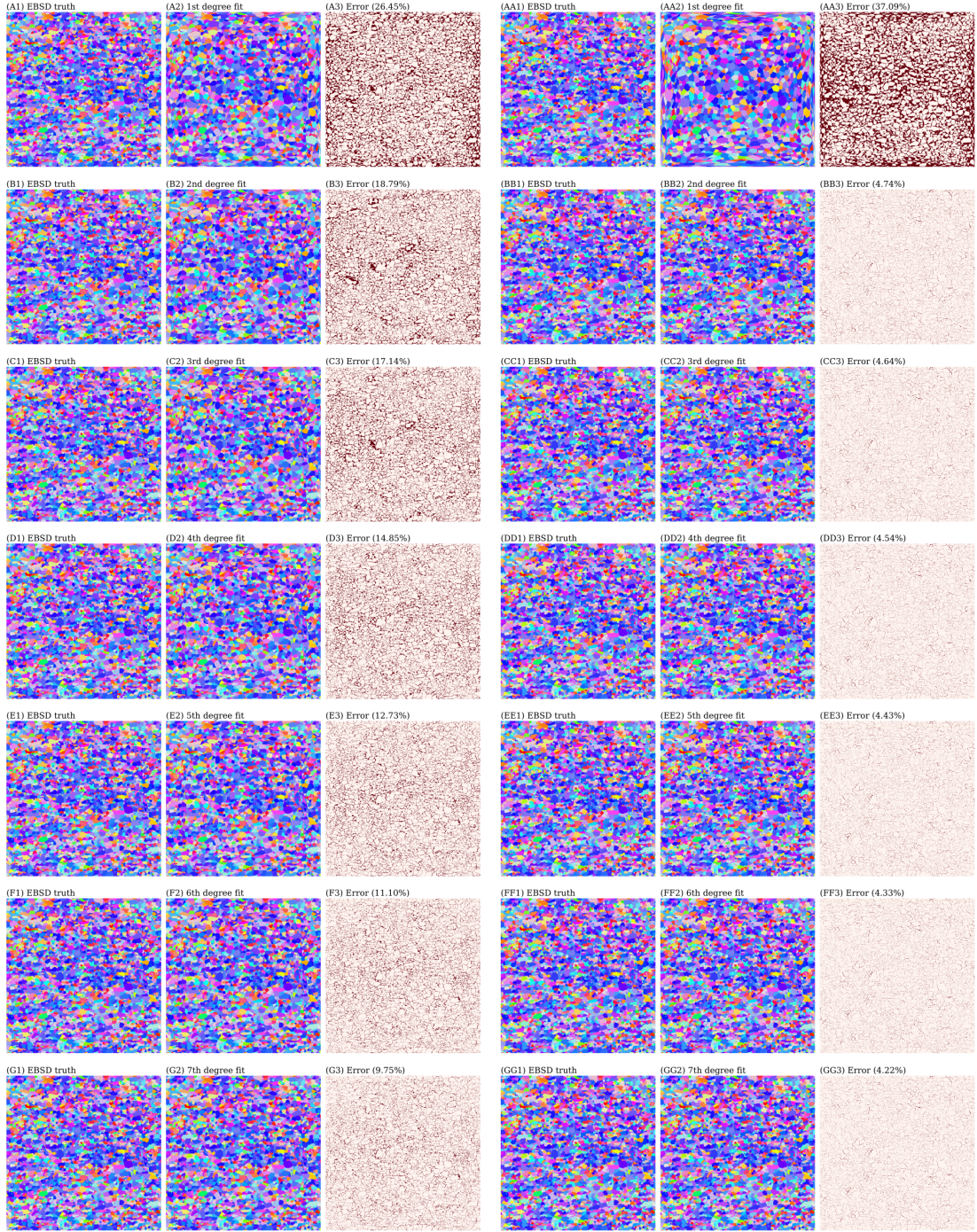


Figure 3: The fitting of polynomial diagrams to the big EBSD dataset as described in Section 5.5. Column 1: Ground truth. The given grain map G comes from an EBSD image of steel, where $G(x) = i$ if pixel x is in grain i . The colours correspond to the assignment of pixels. Column 2, row d : A polynomial diagram of degree d , for $d \in [7]$, is fitted to the grain map G by approximately maximising Φ_ε^G with initial guess $\theta^{(0)} = \mathbf{0}$. Column 3: The misassignment error. The light pink pixels are correctly assigned, the dark pixels are incorrectly assigned. Columns 4–6: The same as columns 1–3 but with the initial guess $\theta^{(0)} = \theta'$ from Section 5.6.2. In each case $\varepsilon = 0.01$, there are $N = 4686$ grains and $|\Omega| = 1,033,376$ pixels, and the algorithm was run for 1000 iterations and took around 4.5-16 minutes (with run time increasing with d).