Journal of Mathematical Imaging and Vision (2021) 63:157–185
https://doi.org/10.1007/s10851-020-00986-1

PDE Evolutions for M-Smoothers in One, Two, and Three Dimensions

Martin Welk[1] · **Joachim Weickert**[2]

Received: 4 November 2019 / Accepted: 30 July 2020 / Published online: 12 August 2020

Abstract
Local M-smoothers are interesting and important signal and image processing techniques with many connections to other methods. In our paper, we derive a family of partial differential equations (PDEs) that result in one, two, and three dimensions as limiting processes from M-smoothers which are based on local order-p means within a ball the radius of which tends to zero. The order p may take any nonzero value > -1, allowing also negative values. In contrast to results from the literature, we show in the space-continuous case that mode filtering does not arise for $p \to 0$, but for $p \to -1$. Extending our filter class to p-values smaller than -1 allows to include, e.g. the classical image sharpening flow of Gabor. The PDEs we derive in 1D, 2D, and 3D show large structural similarities. Since our PDE class is highly anisotropic and may contain backward parabolic operators, designing adequate numerical methods is difficult. We present an L^∞-stable explicit finite difference scheme that satisfies a discrete maximum–minimum principle, offers excellent rotation invariance, and employs a splitting into four fractional steps to allow larger time step sizes. Although it approximates parabolic PDEs, it consequently benefits from stabilisation concepts from the numerics of hyperbolic PDEs. Our 2D experiments show that the PDEs for $p < 1$ are of specific interest: Their backward parabolic term creates favourable sharpening properties, while they appear to maintain the strong shape simplification properties of mean curvature motion.

Keywords M-smoother · Partial differential equation · Mode filter · Mean curvature motion · Shock filter · Backward parabolic operator · Anisotropy · Finite difference method · Operator splitting · Shape analysis

1 Introduction

Partial differential equations (PDEs) constitute a natural framework to model processes in numerous real-world applications, ranging from physics over life sciences to economy. Thus, it is not surprising that they have also contributed substantially to the mathematical foundations of signal and image analysis. For instance, they appear as Euler–Lagrange equations when solving continuous optimisation problems that result from variation models [5,13] or regularisations of ill-posed problems [7]. It has also been shown that they are the natural setting for scale-spaces [1], they are successfully used for image enhancement [55], inpainting [49], and image compression [24]. PDE-based models benefit from many decades of research on their theoretical foundations and efficient numerical algorithms. Since they are continuous concepts, it is also very easy to incorporate useful invariances such as rotation invariance.

One of the most fascinating aspects of PDE-based image analysis is its capability to unify a number of existing methods in image analysis. This has led to deeper structural insights as well as to novel algorithms. For instance, PDE formulations and connections to PDE-based image analysis are known for Gaussian smoothing [31], dilation and erosion [1,3,10,52], morphological amoebas [59], wavelet shrinkage [56], mean and median filtering [29], and mode filtering [27].

Since mean, median and mode filtering are three representatives of M-smoothers based on local order-p means, the question arises if there is a more general PDE formulation that covers the full class of local order-p mean filtering in signal, image and volumetric data processing. This is the topic of

✉ Martin Welk
martin.welk@umit.at

Joachim Weickert
weickert@mia.uni-saarland.de

[1] Institute of Biomedical Image Analysis, UMIT – Private University for Health Sciences, Medical Informatics and Technology, Eduard-Wallnöfer-Zentrum 1, 6060 Hall, Tyrol, Austria

[2] Mathematical Image Analysis Group, Faculty of Mathematics and Computer Science, Campus E1.7, Saarland University, 66041 Saarbrücken, Germany

our paper. Before we go more deeply into our contributions, let us first clarify in more detail the concept of M-estimators and M-smoothers based on order-p means which we will consider in this work.

M-estimators It has been observed long ago by Legendre [38] and Gauß [25] that the mean of a finite multiset $\mathcal{X} = \{a_1, a_2, \ldots, a_n\}$ of real numbers can be described as the minimiser of the sum of squared distances to the given numbers:

$$\operatorname{mean}(\mathcal{X}) = \operatorname*{argmin}_{\mu\in\mathbb{R}} \sum_{i=1}^{n} (\mu - a_i)^2 \,. \tag{1}$$

Likewise it has been proven by Fechner [17] that the median of $\mathcal{X}$ minimises the sum of absolute distances:

$$\operatorname{median}(\mathcal{X}) = \operatorname*{argmin}_{\mu\in\mathbb{R}} \sum_{i=1}^{n} |\mu - a_i| \,. \tag{2}$$

This can be generalised to the notion of *order-p means* given by

$$\operatorname{mean}_p(\mathcal{X}) := \operatorname*{argmin}_{\mu\in\mathbb{R}} \sum_{i=1}^{n} |\mu - a_i|^p \tag{3}$$

for any $p > 0$, with $\operatorname{mean}_2 \equiv \operatorname{mean}$, $\operatorname{mean}_1 \equiv \operatorname{median}$. After more restricted formulations by several predecessors, order-p means for general real-valued $p > 0$ were discussed by Barral Souto [6]. In robust statistics, order-p means belong to the class of *M-estimators* [30].

Including the limiting case of the monomials as

$$|z|^0 = \begin{cases} 0 \,, & z = 0 \,, \\ 1 \,, & z \neq 0 \end{cases} \tag{4}$$

Barral Souto [6] also extends the definition (3) to the case $p = 0$ for which the *mode* of $\mathcal{X}$, i.e. its most frequent value, is obtained. As is also noted in [6], the limit $p \to \infty$ yields what is also called the mid-range value, i.e. the arithmetic mean of the extremal values of $\mathcal{X}$.

Historical remarks In fact, the concept of order-p means has evolved in steps with increasing generality over centuries, which we will briefly mention in the following. The paper [4] was helpful in identifying some of these steps, and provides some further information.

Consolidating a value from observations by minimisation of the sum of absolute differences was proposed by Laplace in 1774 [37]; however, it seems to have been only with Fechner's 1878 proof [17] that the connection to the median of discrete data was clearly established.

Least-squares optimisation was introduced, and put into relation with the arithmetic mean, by Legendre (1805) [38] and Gauss (1809) [25]. However, already Gauss discussed in [25, pp. 221] alternatives to least squares: on one hand, the use of even integer exponents $p > 2$, mentioning even the limit case $p \to \infty$; on the other hand, he also made remarks about Laplace' idea of minimising the sum of absolute differences. Later on, Ellis (1844) [16] pointed out that quite general penalisers $\psi(|\mu - a_i|)$ could be used, thus actually proposing a fairly general class of M-estimators even beyond order-p means.

Fechner (1878) [17] introduced the family of order-p means of discrete data with integer $p \geq 0$, including the case $p = 0$ yielding the mode of a discrete data set. In 1921, Jackson [32] restated the minimising property of the median and introduced order-p means with non-integer $p > 1$, focusing on the use of the limit $p \to 1^+$ as a means to disambiguate the median. Picking up Jackson's notion, Jordan (1927) [33] stated the mode as limit case for $p \to 0^+$. Order-p means with non-integer $p > 0$ in their own right made their appearance in 1938 with Barral Souto's paper [6].

M-estimators, continuous case It is straightforward to rewrite the definition of order-p means for continuous distributions (densities) on $\mathbb{R}$ just by replacing sums with integrals: Let $\gamma : \mathbb{R} \to \mathbb{R}_0^+$ be a density (integrable in a suitable sense), then one defines

$$\operatorname{mean}_p(\gamma) := \operatorname*{argmin}_{\mu\in\mathbb{R}} \int_{-\infty}^{\infty} \gamma(z)|\mu - z|^p \,\mathrm{d}z \,. \tag{5}$$

The general notion of continuous order-p means was investigated in several papers by Fréchet. For $p \geq 1$ it is mentioned in 1946 in [19] where, however, detailed discussion is restricted to $p = 1$ and $p = 2$. A thorough treatment of $p \geq 1$ is provided in 1948 in [20], whereas [21,22], also from 1948, consider the general case $p > 0$ including a thorough discussion of the cases $p \to 0$ and $p \to \infty$. In fact, Fréchet analyses that the mode is obtained in the limit $p \to 0$ for discrete distributions (or such with a discrete component) but not for purely continuous distributions.

The cases of the median and mean had been considered before: In the case $p = 1$, the continuous formulation goes back to Laplace (1774) [37] who, unlike in the discrete case, also identified the minimiser as the median. Fréchet analysed $p = 2$ in [18] from 1943.

M-smoothers In image processing, M-estimators are commonly used to build local filters, see [51] for the median filter (in signal processing) and [50] for order-p means with $p > 0$. In a local filter, one takes at each location the greyvalues from a neighbourhood (selection step) and computes some common value of these (aggregation step) that is assigned to the location in the filtered signal, see e.g. [14,27]. These filters can be iterated to generate a series of progressively processed images.

It has been noticed since long that some of these filters behave similar to certain image filters based on PDEs. Mean filters are a spatial discretisation of linear diffusion. Guichard and Morel [29] have proven that iterated median filtering approximates mean curvature motion [8]. To this end, they consider a space-continuous version of median filtering, in which the selection step is based on a disc-shaped neighbourhood. Sending the radius of the neighbourhood to zero, they show that the effect of the median filtering step becomes asymptotically equal to a time step of an explicit time discretisation of the mean curvature motion PDE.

Griffin [27] proves similar results for three different filters, and puts them in the context of order-p means. In addition to the median ($p = 1$) and the arithmetic mean ($p = 2$) he considers for the first time the mode filter (associating it with $p = 0$). In contrast to [29], the selection step in [27] is based on a Gaussian window, i.e. the input value density of the respective means is made up by the values from the entire image plane but reweighted with a Gaussian function. The limit case is constituted by the standard deviation of the Gaussian window approaching zero. In this framework, the mean curvature motion PDE is re-derived as the limit case of median filtering. For arithmetic mean filtering, the linear diffusion PDE is obtained. For mode filtering, a PDE is derived that combines mean curvature motion (diffusion along level sets) with backward diffusion in gradient flowline direction, see Proposition 2.

Our contributions The goal of our paper is to complete this picture by deriving the PDE limit for arbitrary order-p means and introducing a suitable numerical algorithm. Up to a time rescaling, the PDE limits of all three cases of [27] will be contained in our results in the following. The reason for the time rescaling is that we use for the selection step disc-shaped neighbourhoods such as in [29]. With this choice we aim at modelling the space-continuous filter in an analogy as close as possible to the usual setup of discrete local signal and image filters.

Our paper is based on the conference publication [60]. However, these results are presented in more detail and substantially extended, covering now also the 1D and 3D setting. We also propose a novel splitting-based numerical algorithm with improved efficiency and better rotation invariance.

Starting with the case of planar grey-value images, we derive a family of PDEs associated with M-smoothers based on order-p means with variable p and vanishing disc radius. In contrast to results from the literature, we also permit negative p-values with $p > -1$. Compared to [60], the proof of this approximation result is presented in a more detailed form. Moreover, we discuss the behaviour near critical points (saddle points and extrema) and critical curves. We also analyse the effect of staircasing. Using the calculus of distributions, we can show that the PDEs derived for smooth images remain valid for step functions.

Reconsidering the relation of order-p means and their corresponding PDEs to existing image filters, we show that in the space-continuous setting the mode filter does *not* arise for $p \to 0$, as is commonly assumed [27] (despite the analysis in [21,22]), but for $p \to -1$. Since the common assumption countered by our analysis is derived by analogy from discrete theory, we discuss in this paper also where and why this analogy fails.

In the present work, we extend our results also to 1D signals and 3D grey-value images. Table 1 at the end of Sect. 4 summarises the PDE approximation results obtained in one, two, and three dimensions. The PDEs approximated for $p > -1$ are in full analogy to the 2D case. This is also the case for the mode filter as limiting case for $p \to -1$ in three dimensions. For 1D signals, the limit $p \to -1$ is non-uniform, and the mode filter approximates a shock-filter PDE which was already stated in earlier work [61]. Continuing our PDE family to values $p < -1$ allows to cover also the sharpening Gabor flow [23,39], for which no M-smoothing counterpart is known.

In spite of the fact that our PDE family is anisotropic and may even involve backward parabolic operators, we design an L^∞-stable numerical scheme that enjoys excellent rotation invariance and employs operator splitting to improve its efficiency. Our experiments show that the PDEs for $p < 1$ are particularly attractive since they simultaneously allow image sharpening and shape simplification.

Structure of the paper In Sect. 2, we present our theory that allows us to derive PDE evolutions from M-smoothers, and discuss in detail important aspects of the PDE limit such as its behaviour near critical points and staircasing. Section 3 presents the analogous results for 1D signals, whereas Sect. 4 covers 3D images. In Sect. 5, we discuss the relation between discrete and continuous M-smoothing and explain why the discrete result about approximation of the mode filter for $p \to 0$ cannot be transferred to the continuous situation. Our numerical algorithm is discussed in Sects. 6 and 7 is devoted to an experimental evaluation. The paper is concluded with a summary in Sect. 8. Two appendices provide additional material: The detailed proofs of the results in Sects. 2–4 are collected in Appendix A, whereas Appendix B presents an illustrative example to support the discussion in Sect. 5.

2 M-Smoothers, Mode and Partial Differential Equations for 2D Images

In this section, we derive PDEs for M-smoothers and the mode filter in the case of 2D greyvalue images, and discuss some of their properties.

2.1 Generalised Order-*p* Means

In the following, M-smoothers are based on order-p means with $p > -1$, $p \neq 0$. As this range for p goes beyond the usual $p > 0$, let us first extend the definition of order-p means of continuous-scale distributions accordingly.

Definition 1 Let z be a real random variable with the bounded, piecewise continuous density $\gamma : \mathbb{R} \to \mathbb{R}$. For $p \in (-1, +\infty) \setminus \{0\}$, define the *order-p mean* of γ as

$$\operatorname{mean}_p(\gamma) = \operatorname*{argmin}_{\mu \in \mathbb{R}} \int_{\mathbb{R}} \gamma(z)\ \operatorname{sgn}(p)|\mu - z|^p \,\mathrm{d}z\ . \tag{6}$$

As $|z|^p$ is monotonically increasing on $\mathbb{R}_0^+$ (the set of all nonnegative real numbers) for $p > 0$, but monotonically decreasing on $\mathbb{R}^+$ (the set of all positive real numbers) for $p < 0$, the $\operatorname{sgn}(p)$ factor in (6) ensures that in both cases an increasing penalty function is used.

For $p > 0$ the requirement of continuity of γ in Def. 1 can be relaxed; by modelling a discrete density as a weighted sum of delta peaks, the discrete order-p means as in [6] can be included in this definition.

The continuity is, however, essential for $p < 0$: In this case, the penalty function has a pole at $z = 0$ such that an improper integral is obtained; for $p > -1$ this integral exists provided that γ is continuous, i.e. no delta peaks are allowed. In particular, we cannot define an order-p mean with $-1 < p < 0$ for discrete distributions as considered in [6].

2.2 Infinitesimal Limits of M-Smoothers

We turn now to derive partial differential equations approximated by M-smoothers applied to 2D images. The proofs of the following propositions are given in Appendix A. When speaking of smooth functions we always mean C^∞ functions although a weaker hypothesis could be sufficient for some results. The first proposition contains our first main result.

Proposition 1 (2D PDE limit for $p > -1$) *Let a smooth image $u : \mathbb{R}^2 \to \mathbb{R}$ be given, and let $\boldsymbol{x}_0 = (x_0, y_0)$ be a regular point, $|\nabla u(\boldsymbol{x}_0)| > 0$. One step of order-$p$ mean filtering of u with a disc-shaped window $\mathrm{D}_\varrho(\boldsymbol{x}_0)$ and $p > -1$, $p \neq 0$ approximates for $\varrho \to 0$ a time step of size $\tau = \varrho^2/(2p+4)$ of an explicit time discretisation of the PDE*

$$u_t = u_{\xi\xi} + (p-1)\, u_{\eta\eta} \tag{7}$$

where η and ξ are geometric coordinates referring at each image location to the direction of the positive gradient, and the level-line direction, respectively:

$$\begin{aligned} &\operatorname{mean}_p\{u(x,y) \mid (x,y) \in \mathrm{D}_\varrho(x_0,y_0)\} - u(x_0,y_0) \\ &= \frac{\varrho^2}{2(p+2)}\big(u_{\xi\xi}(x_0,y_0) + (p-1)u_{\eta\eta}(x_0,y_0)\big) \\ &\quad + \mathcal{O}(\varrho^{(\min\{p,0\}+5)/2})\ . \end{aligned} \tag{8}$$

At a local minimum (maximum) of u, i.e. $\boldsymbol{x}_0$ with $|\nabla u(\boldsymbol{x}_0)| = 0$ where the Hessian $\mathbf{D}^2 u(\boldsymbol{x}_0)$ is positive (negative) semidefinite, the same filtering step fulfils for $\varrho \to 0$ the inequality $\operatorname{mean}_p\{u(x,y) \mid (x,y) \in \mathrm{D}_\varrho(x_0,y_0)\} - u(x_0,y_0) \geq 0\ (\leq 0)$, thus approximates an evolution $u_t \geq 0$ ($u_t \leq 0$).

The approximation order in (8) is $\mathcal{O}(\varrho^{1/2})$ for positive p but reduces to $\mathcal{O}(\varrho^{(p+1)/2})$ for negative p.

For $p = 2$ and $p = 1$ the proposition yields the same PDEs as [27] except for a time rescaling which is due to the choice of a Gaussian window in [27].

Under analogous assumptions as in Proposition 1, one can also derive the PDE limit for the mode filter, where the mode is not obtained by a minimisation in the sense of (5) but directly as the maximum of the density of values in $\{u(x,y) \mid (x,y) \in \mathrm{D}_\varrho(x_0,y_0)\}$.

Proposition 2 (2D PDE limit for mode filtering) *Let u and $\boldsymbol{x}_0$ be as in* Proposition 1. *One step of mode filtering of u with a disc-shaped window $\mathrm{D}_\varrho(\boldsymbol{x}_0)$ approximates for $\varrho \to 0$ a time step of size $\tau = \varrho^2/2$ of an explicit time discretisation of the PDE $u_t = u_{\xi\xi} - 2u_{\eta\eta}$ with η, ξ as in* Proposition 1. *At a local minimum (maximum), mode filtering approximates $u_t \geq 0$ ($u_t \leq 0$).*

The PDE for mode filtering coincides with the one given in [27], again up to time rescaling. We see, however, that (8) for $p \to 0$ does not yield the PDE from Proposition 2 but $u_t = u_{\xi\xi} - u_{\eta\eta}$. Instead, the mode filtering PDE is obtained for $p \to -1$. Inserting $p = -2$ into (8) yields $u_t = u_{\xi\xi} - 3u_{\eta\eta}$ which was stated as an image sharpening PDE that has been proposed by Gabor already in 1965 [23,39].

Remarkably, the PDEs for mean ($p = 2$), median ($p = 1$) and mode ($p = -1$) also match the often-stated empirical rule noted first by Pearson [46, p. 376] according to which the median in a large class of skew densities is located at two-thirds the way between mode and mean (which, however, is not a general law).

2.3 Discussion of PDE Evolutions Near Critical Points

Propositions 1 and 2 state PDEs approximated by the respective M-smoothers at regular points, and inequalities that hold at local minima and maxima. Let us briefly discuss how these results determine uniquely the evolutions of the entire image u (including critical points) approximated by the M-smoothers.

2.3.1 Regions of Critical Points

Let us consider first the case of a connected critical region, i.e. a closed set in $\mathbb{R}^2$ consisting entirely of critical points, with nonempty interior. In such a region, the inequalities for minima and maxima together imply $u_t = 0$, which is also consistent with the obvious limit of any M-smoother in all interior points of the region.

2.3.2 Isolated Critical Points

Let us now consider the case of an isolated critical point, i.e. a point $\boldsymbol{x}_0$ with $\nabla u(\boldsymbol{x}_0) = \mathbf{0}$ but $\nabla u(\boldsymbol{x}) \neq \mathbf{0}$ for all other points $\boldsymbol{x}$ within an open neighbourhood of $\boldsymbol{x}_0$.

A direct calculation of the limit for vanishing window size of an M-smoother at $\boldsymbol{x}_0$ would suggest an approximation that differs substantially from that in regular points. We will argue in the following that this naive limit is irrelevant for the time-continuous image evolution approximated by iterated M-smoothing.

On one hand, limit calculations at regular points (see the proofs in Appendix A) require a neighbourhood that contains no critical points at all. Thus, on approaching a critical point of u, the admissible neighbourhood radius ϱ around regular points tends to zero. Therefore, the PDE limit within any open region of the plane that does not contain critical points is not uniform if the boundary of that region contains a critical point. The result of the naive application of the same limit procedure at a critical point can thus not be expected to fit smoothly into the evolution of the regular points around.

On the other hand, for an initial-boundary value problem describing an image evolution, it is in general sufficient for the PDE to be prescribed everywhere except at isolated points. Assuming viscosity solutions as a solution concept, the solution of the initial-boundary value problem will fill in the evolution at the exceptional points.

Revisiting the evolution from Proposition 1, we notice first that the PDE (7) in regular points of u can be rewritten with the Laplacian $\Delta u = u_{xx} + u_{yy} = u_{\xi\xi} + u_{\eta\eta}$ as

$$u_t = (2-p)u_{\xi\xi} + (p-1)\Delta u \;, \tag{9}$$

a linear combination of homogeneous diffusion and curvature motion. Given the smoothness of u, the diffusion term $u_{xx} + u_{yy}$ can obviously be continued smoothly to isolated critical points.

The curvature motion term is more difficult. At an *isotropic critical point*, i.e. $\boldsymbol{x}_0 = (x_0, y_0)$ with $u_x(\boldsymbol{x}_0) = u_y(\boldsymbol{x}_0) = 0, u_{xx}(\boldsymbol{x}_0) = u_{yy}(\boldsymbol{x}_0), u_{xy}(\boldsymbol{x}_0) = 0$, also this term has a unique limit for $(x, y) \to (x_0, y_0)$, namely $u_{xx}(\boldsymbol{x}_0)$. In contrast, when approaching an *anisotropic* critical point (x_0, y_0) (where the Hessian $\mathrm{D}^2 u$ is not a multiple of the unit matrix) from different directions, one obviously obtains different limits such that no unique value can be filled in at this critical point.

To understand the effect of the evolution near a critical point, assume that $\boldsymbol{x}_0 = \mathbf{0}$ is a local minimum of u. For simplicity, we neglect higher order terms of the Taylor expansion and assume at a given time $u(x, y) = \beta x^2 + \delta y^2$ with $\beta \geq \delta > 0$ in a neighbourhood N of $\boldsymbol{x}_0$. For $\boldsymbol{x} = (x, y)^\top \in N$ one has then

$$\nabla u(\boldsymbol{x}) = 2(\beta x, \delta y)^\top \;, \quad \mathrm{D}^2 u(\boldsymbol{x}) = \begin{pmatrix} 2\beta & 0 \\ 0 & 2\delta \end{pmatrix} \;, \tag{10}$$

$$\eta = \frac{(\beta x, \delta y)^\top}{\sqrt{\beta^2 x^2 + \delta^2 y^2}} \;, \quad \xi = \frac{(-\delta x, \beta y)^\top}{\sqrt{\beta^2 x^2 + \delta^2 y^2}} \;, \tag{11}$$

$$u_{\eta\eta} = \frac{2\beta^3 x^2 + 2\delta^3 y^2}{\beta^2 x^2 + \delta^2 y^2} \;, \tag{12}$$

$$u_{\xi\xi} = \frac{2\beta^2 \delta x^2 + 2\beta\delta^2 y^2}{\beta^2 x^2 + \delta^2 y^2} \;. \tag{13}$$

Both $u_{\xi\xi}$ and $u_{\eta\eta}$, and thus also $u_t = u_{\xi\xi} + (p-1)u_{\eta\eta}$, are constant along radial lines through $\mathbf{0}$.

For an isotropic minimum ($\beta = \delta > 0$), the evolution speed is uniform in the neighbourhood N, ensuring that the isotropy of the minimum is preserved during evolution. In particular, for $p \geq 0$, this evolution speed is positive such that the inequality $u_t(\boldsymbol{x}_0) \geq 0$ is automatically preserved.

For an anisotropic minimum ($\beta > \delta > 0$), the evolution speeds u_t along different radial lines differ, with $u_t(z, 0) < u_t(0, z)$, such that the anisotropy is reduced by the evolution. For $p \geq 1 - \delta/\beta$ all evolution speeds are nonnegative, so $u_t(\boldsymbol{x}_0) \geq 0$ is still automatically satisfied, and the anisotropic minimum is converted into an isotropic minimum by the evolution. The position of the minimum can move due to the evolution.

For $1 - \beta/\delta < p < 1 - \delta/\beta$, one has $u_t(z, 0) < 0 < u_t(0, z)$. In this case, the inequality $u_t(\boldsymbol{x}_0) \geq 0$ acts to constrain the evolution near the x axis, implying the immediate formation of a critical line (see Sect. 2.3.3) or plateau (see Sect. 2.3.1) around $\boldsymbol{x}_0$.

If, finally, $p < 1 - \beta/\delta$, the speed u_t is negative throughout N, which implies that a critical plateau is formed immediately.

Analogous considerations apply to local maxima. Finally, a saddle point (with indefinite Hessian) remains a saddle point, and as such is well constrained by the surrounding regular points from above and below. Therefore, the evolution at isolated critical points for $p \geq 1$ is fully determined by filling in the evolution from the surrounding regular points, whereas for $-1 < p < 1$ it is fully determined by filling in combined with the inequality constraints $u_t \geq 0$ at minima and $u_t \leq 0$ at maxima.

2.3.3 Critical Curves

The considerations from Sect. 2.3.2 can be extended to regular curves consisting of critical points. If such a curve is formed by local minima, any point on this curve is a maximally anisotropic local minimum, $\beta > \delta = 0$, yielding evolution speeds (up to higher order terms) $u_t \approx 2(p-1)\beta$ for nearby regular points.

If $p > 1$, one has $u_t > 0$, thus the differential inequality at critical points is automatically satisfied. The critical curve is preserved as a critical curve or may be broken up into isolated critical points.

If $p < 1$, the regular points in N evolve with $u_t > 0$, which leads to an immediate expansion of the critical curve into a plateau.

2.4 Staircasing and Analysis of the PDE for Step Functions

For $p < 1$, the PDE (7) involves a backward parabolic term in gradient flowline direction. In evolutions of this kind staircasing effects are common, i.e. the evolving function turns into a step function which is only piecewise smooth with jumps between the smooth segments. Indeed, staircasing can also observed in numerical experiments with (7). Unfortunately, the staircasing undermines the smoothness assumption underlying the approximation result of Proposition 1. Therefore, we dedicate this section to discuss how our approximation results extend to the situation of step functions. First, we will use the calculus of distributions to generalise the image filtering PDE (7). Afterwards we will discuss order-p mean M-smoothers for step functions. Although we do not possess, at the time being, a full asymptotic analysis of this case, we will consider a simplified case and demonstrate by a combination of analytic and numeric evidence that the behaviour of M-smoothers is still comparable to that of the PDE.

2.4.1 Distributional Analysis

In the following, we will consider piecewise smooth step functions, i.e. functions over $\mathbb{R}^n$ which are smooth except on a set of smooth hypersurfaces that decompose the space $\mathbb{R}^n$ into connected segments $\Omega_1, \dots, \Omega_k$.

A natural way to analyse the effect of the PDE evolution (7) on a step function u is to apply the PDE to a series of smoothed functions that converge to u, and consider the limit of the so obtained evolutions. For example, u could be convolved with Gaussians G_σ of decreasing standard deviation σ, yielding the desired result for $\sigma \to 0$. As G_σ for $\sigma \to 0$ weakly converges to a Dirac delta distribution, the calculus of distributions [54,63] allows to calculate the evolution of interest in a more compact form without explicitly carrying out the limiting procedure. Technically, a distribution, or generalised function, $f \in \mathcal{D}'(\mathbb{R}^n)$ is a functional that acts on smooth basic functions φ by the scalar product of functions $(f, \varphi) := \int_{\mathbb{R}^n} f(\boldsymbol{x})\varphi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$.

Assuming that the jump set of f consists of just one smooth hypersurface S dividing $\mathbb{R}^n$ into domains Ω_1 and Ω_2, we notice that f is differentiable in distributional sense, and we recall the following essential formula from [63, II, §6]:

$$\frac{\partial f}{\partial x_i} = \left\{\frac{\partial f}{\partial x_i}\right\} + [f]_S \cos(\boldsymbol{n}_S, x_i)\delta_S \tag{14}$$

for $i = 1, \dots, n$. Here, $\{\partial f/\partial x_i\}$ denotes the *regular part* of the derivative, i.e. essentially an ordinary function. The vector $\boldsymbol{n}_S$ is the unit outer normal vector of S at a given point, and $[f]_S$ the jump height of u at this point in direction $\boldsymbol{n}_S$. The *single-layer distribution* δ_S is a generalisation of the one-dimensional Dirac delta distribution, behaving like the delta distribution on crossing the hypersurface S in normal direction.

Furthermore, second derivatives of f can be written as [63, II, §6]

$$\begin{aligned}\frac{\partial^2 f}{\partial x_i \partial x_j} = &\left\{\frac{\partial^2 f}{\partial x_i \partial x_j}\right\} + \frac{\partial}{\partial x_j}\left([f]_S \cos(\boldsymbol{n}_S, x_i)\delta_S\right) \\ &+ \left[\left\{\frac{\partial f}{\partial x_i}\right\}\right]_S \cos(\boldsymbol{n}_S, x_j)\delta_S\end{aligned} \tag{15}$$

for $i, j = 1, \dots, n$ where in the second summand a *double-layer distribution* occurs as the derivative of a single-layer distribution; $\partial\delta_S/\partial\boldsymbol{n}_S$ is a generalisation of the derivative of the one-dimensional delta distribution, behaving like δ' on crossing S in normal direction.

To analyse our example, we consider the evolution (7) with a step function u_0 as initial condition. The evolution u will be described by a function u over $\mathbb{R}^2 \times [0, T]$ which is smooth except on a jump set consisting of regular surfaces. Outside the jump set, our previous analysis applies. The case of interest is therefore a point (x, y, t) on a jump surface S. Simplifying further, we assume that the normal vector $\boldsymbol{n}_S$ at (x, y, t) is in the x-t plane and in positive x direction, i.e.

$$\boldsymbol{n}_S = \frac{1}{\sqrt{1+v^2}}(1, 0, -v)^\top \tag{16}$$

where v is the speed at which the jump moves in x direction as the time t progresses. Then (7) becomes $u_t = u_{yy} + (p-1)u_{xx}$. From (14) with $\cos(\boldsymbol{n}_S, t) = -v/\sqrt{1+v^2}$ we calculate

$$u_t(x, y, t) = \{u_t(x, y, t)\}[u]_S(x, y, t)\frac{-v}{\sqrt{1+v^2}}\delta_S\,. \tag{17}$$

Similarly, we obtain from (15) with $\cos(\boldsymbol{n}_S, y) = 0$ and $\partial_y \cos(\boldsymbol{n}_S, y) = -\sin(\boldsymbol{n}_S, y)\cos(\partial_y \boldsymbol{n}_S, y) = -\kappa_S/\sqrt{1+v^2}$

$$\begin{aligned} u_{yy}(x,y,t) &= \{u_{yy}(x,y,t)\} \\ &\quad - \frac{\partial}{\partial y}\big([u]_S(x,y,t)\cos(\boldsymbol{n}_S, y)\delta_S\big) \\ &= \{u_{yy}(x,y,t)\} \\ &\quad + [u]_S(x,y,t)\frac{\kappa_S(x,y,t)}{\sqrt{1+v^2}}\delta_S \end{aligned} \tag{18}$$

where $\kappa_S(x,y,t)$ denotes the curvature of S in the x-y plane at (x,y,t). Finally,

$$\begin{aligned} u_{xx}(x,y,t) &= \{u_{xx}(x,y,t)\} \\ &\quad + \frac{1}{\sqrt{1+v^2}}[u]_S\frac{\partial}{\partial x}\delta_S \\ &\quad + \frac{1}{\sqrt{1+v^2}}[\{u_x\}]_S\delta_S\,. \end{aligned} \tag{19}$$

Inserting (17), (18) and (19) into (7), we have by equating the δ_S contributions

$$v = -\kappa_S - (p-1)\frac{[\{u_x\}]_S}{[u]_S}\,, \tag{20}$$

which describes the speed at which the interface between the two smooth segments of u moves in x direction. The contribution $-\kappa_S$ is in full agreement with the behaviour of the mean curvature motion part of (7) for smooth functions, whereas $-(p-1)[\{u_x\}]_S/[u]_S$ modifies this speed by accelerating or slowing down the inward motion of the interface depending on p and whether the gradient of u is greater or smaller on the outside or inside of the evolving interface. If $p < 1$, the effect is to push the evolution towards increasing contrast at the interface, thus encouraging staircasing. For $p > 1$, the evolution is biased towards reducing contrast at the interface, thus counteracting staircasing.

We notice finally that on the right hand side of (19) also a double-layer term $\partial_x\delta_S$ appears. When equating with the first-order time derivative u_t in (7), this term has no implication for the evolution of u as time integration across S integrates it to zero. However, it indicates that no well-defined regular function value can be assigned to u on the interface S itself.

2.4.2 M-Smoothing a Step Function

For a full theoretical analysis of an M-smoothing step with order-p means for step functions, the limit $\varrho \to 0$ of $\mathrm{mean}_p(u) - u$ at a fixed location of a given step function is of little help, because it just reproduces the result for smooth functions everywhere outside the jump set, and isolated values on the jump set itself are meaningless. To validate (20) in the asymptotic case, it would be necessary instead to determine the displacement of the interface itself by one M-smoothing step for positive ϱ, and consider the asymptotic behaviour of this displacement. This appears substantially more complicated than the proof of Proposition 1, and no analysis of this kind is available at the moment.

In the following, we study instead a simple case of a step function for fixed ϱ by combining analytical with numerical arguments. Let the step function $u : \mathbb{R}^2 \to \mathbb{R}$ be given as

$$u(x,y) = \alpha(x + \delta y^2) + h\mathbb{1}(x + \delta y^2 + \theta > 0) \tag{21}$$

where $\mathbb{1}(x + \delta y^2 + \theta > 0)$ is the Heaviside function of $x + \delta y^2 + \theta$. The jump set of u is the parabola $x + \delta y^2 = -\theta$, and we have $u_x = \alpha$, $u_{xx} = u_{xy} = 0$, $u_{yy} = \frac{1}{2}\alpha\delta$ everywhere outside the jump set.

Let μ be the order-p mean of u within the disc D_ϱ centred at $(0,0)$, for some $p \in (-1, 0)$. We have

$$\mu = \operatorname*{argmin}_{\mu} \operatorname{sgn}(p) \iint_{\mathrm{D}_\varrho} |\mu - u(x,y)|^p \,\mathrm{d}x\,\mathrm{d}y \tag{22}$$

which can be simplified to

$$\begin{aligned} \mu = \operatorname*{argmin}_{\mu} \operatorname{sgn}(p) \Bigg(&\int_{-\varrho}^{-\theta} y^*(x)\,|\mu - \alpha x|^p \,\mathrm{d}x \\ &+ \int_{-\theta}^{\varrho} y^*(x)\,|\mu - \alpha x - h|^p \,\mathrm{d}x \Bigg) \end{aligned} \tag{23}$$

where (for sufficiently small δ)

$$y^*(x) = \sqrt{\varrho^2 - \frac{1}{4\delta^2}\left(1 - \sqrt{1 - 4\delta x + 4\delta^2\varrho^2}\right)^2} \tag{24}$$

assigns to each $x \in [-\varrho, \varrho]$ the y coordinate of the two points $\big(x, \pm y^*(x)\big)$ where the level line of u going through $(x, 0)$ hits the boundary of D_ϱ.

Equation (20) together with the step size $\tau = \varrho^2/(2p+4)$ from Proposition 1 suggests that μ should be in the range of the lower part of u (i.e. $\mu \approx 0$) for $\theta < -\delta/(p+2)$ and in the upper part of u (i.e. $\mu \approx h$) for $\theta > -\delta/(p+2)$.

To check this numerically, we fixed $\varrho = 1$ and evaluated (23) for a set of randomly chosen values of $p \in [-0.99, -0.1]$, $\alpha \in [0.03, 0.15]$, $\delta \in [-0.2, 0.2]$ and for different jump heights $h = 0.01$, $h = 0.1$, $h = 1$.

As is evident from the results visualised in Fig. 1, the values of θ/δ at which μ jumps from the lower to the higher segment of u are close to the predicted ones for small h and deviate somewhat more for larger h, but the overall behaviour of the interface displacement speed is consistent with the analysis of the PDE.

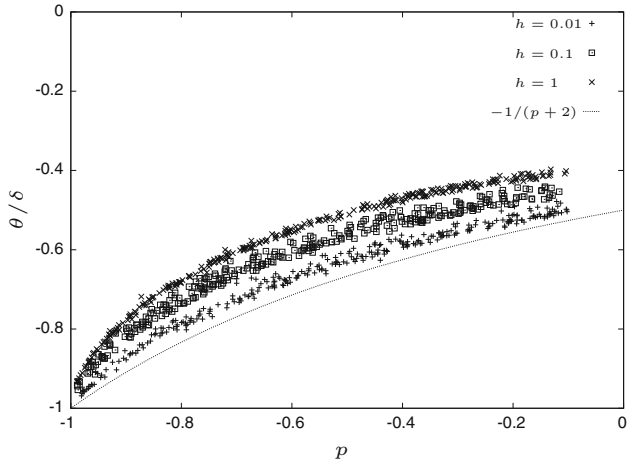

Fig. 1 Jump values of θ/δ for the example function (21) in the range $p \in [-0.99, -0.1]$ for three jump heights h and randomly chosen α, δ. Curve shows the theoretical value $\theta/\delta = -1/(p+2)$ for comparison

3 M-Smoother PDEs for 1D Signals

A similar analysis as in the 2D case can be carried out for 1D signals. The proofs of the next two results are found in Appendix A.

Proposition 3 (1D PDE limit for $p > -1$) *Let a smooth signal $u : \mathbb{R} \to \mathbb{R}$ be given, and let x_0 be a regular point, $u_x \neq 0$. One step of order-p mean filtering of u with a box window $\mathrm{I}_\varrho(x_0) := [x_0 - \varrho, x_0 + \varrho]$ and $p > -1$, $p \neq 0$ approximates for $\varrho \to 0$ a time step of size $\tau = \varrho^2/(2p+2)$ of an explicit time discretisation of the PDE*

$$u_t = (p-1)\, u_{xx} \tag{25}$$

in the sense that

$$\begin{aligned} &\mathrm{mean}_p\{u(x) \mid x \in \mathrm{I}_\varrho(x_0)\} - u(x_0) \\ &= \frac{\varrho^2}{2(p+1)}(p-1)u_{xx}(x_0) + \mathrm{o}(\varrho^2)\,. \end{aligned} \tag{26}$$

At a local minimum (maximum) of u, i.e. x_0 with $u_x = 0$ where u_{xx} is nonnegative (nonpositive), the same filtering step fulfils for $\varrho \to 0$ the inequality $\mathrm{mean}_p\{u(x) \mid x \in \mathrm{I}_\varrho(x_0)\} - u(x_0) \geq 0$ (≤ 0), thus approximates an evolution $u_t \geq 0$ ($u_t \leq 0$).

Proposition 4 (1D PDE limit for mode filtering) *Let u be as in Proposition 3, and x_0 any point in its domain. One step of mode filtering of u with a box window $\mathrm{I}_\varrho(x_0)$ approximates for $\varrho \to 0$ a time step of size $\tau = \varrho$ of an explicit time discretisation of the shock filter PDE* [45] *given by*

$$u_t = -\mathrm{sgn}\,(u_{xx})\,|u_x|\,. \tag{27}$$

This relation between local mode filtering of 1D signals and shock filtering extends a result for discrete signals from [61].

Unlike in the 2D case, the PDE approximated by the mode filter does not fit in as the limit $p \to -1$ of the PDE for order-p means, and the approximation in the mode filter case is with time step size $\mathcal{O}(\varrho)$ instead of $\mathcal{O}(\varrho^2)$. To understand this, notice that in (26) the coefficient in front of u_{xx} goes to infinity for $p \to -1^+$, which means that the approximation holds for ever smaller ϱ as p approaches -1, such that there is no uniform limit of the approximation (26) for any fixed positive radius ϱ.

4 M-Smoother PDEs for 3D Images

In this section, we extend our previous results also to the case of three-dimensional, i.e. volume images. The results are similar to those in two dimensions. However, now the smoothing in level set direction takes place in a surface with two geometric coordinates ξ and χ referring to mutually orthogonal tangential directions of the level set. We give again two propositions referring to $p \in (-1, \infty) \setminus \{0\}$ and the mode case; both proofs are found in Appendix A.

Proposition 5 (3D PDE limit for $p > -1$) *Let a smooth volume image $u : \mathbb{R}^3 \to \mathbb{R}$ be given, and let $\boldsymbol{x}_0 = (x_0, y_0, z_0)$ be a regular point, $|\nabla u(\boldsymbol{x}_0)| > 0$. One step of order-$p$ mean filtering of u with a ball-shaped window $\mathrm{B}_\varrho(\boldsymbol{x}_0)$ and $p > -1$, $p \neq 0$ approximates for $\varrho \to 0$ a time step of size $\tau = \varrho^2/(2p+6)$ of an explicit time discretisation of the PDE*

$$u_t = u_{\xi\xi} + u_{\chi\chi} + (p-1)\, u_{\eta\eta} \tag{28}$$

where η, ξ and χ are geometric coordinates referring at each image location to the direction of the positive gradient, and two mutually orthogonal directions tangent to the level surface, respectively:

$$\begin{aligned} &\mathrm{mean}_p\{u(x,y,z) \mid (x,y,z) \in \mathrm{B}_\varrho(x_0,y_0,z_0)\} \\ &\quad - u(x_0,y_0,z_0) \\ &= \frac{\varrho^2}{2(p+3)}\big(u_{\xi\xi}(x_0,y_0,z_0) + u_{\chi\chi}(x_0,y_0,z_0) \\ &\quad + (p-1)u_{\eta\eta}(x_0,y_0,z_0)\big) \\ &\quad + \mathcal{O}(\varrho^{(\min\{p,0\}+5)/2})\,. \end{aligned} \tag{29}$$

At a local minimum (maximum) of u, i.e. $\boldsymbol{x}_0$ with $|\nabla u(\boldsymbol{x}_0)| = 0$ where the Hessian $\mathbf{D}^2 u(\boldsymbol{x}_0)$ is positive (negative) semidefinite, the same filtering step fulfils for $\varrho \to 0$ the inequality $\mathrm{mean}_p\{u(x,y,z) \mid (x,y,z) \in \mathrm{B}_\varrho(x_0,y_0,z_0)\} - u(x_0,y_0,z_0) \geq 0$ (≤ 0), thus approximates an evolution $u_t \geq 0$ ($u_t \leq 0$).

Table 1 PDE approximation results for order-p mean filters and mode filtering in one to three dimensions

Dimension	PDE for $p > -1$, $p \neq 0$	Time step size τ	PDE for mode filter	Time step size τ
1D	$u_t = (p-1)u_{xx}$	$\frac{\varrho^2}{2p+2}$	$u_t = -\mathrm{sgn}(u_{xx})\lvert u_x\rvert$	ϱ
2D	$u_t = u_{\xi\xi} + (p-1)u_{\eta\eta}$	$\frac{\varrho^2}{2p+4}$	$u_t = u_{\xi\xi} - 2u_{\eta\eta}$	$\frac{\varrho^2}{2}$
3D	$u_t = u_{\xi\xi} + u_{\chi\chi} + (p-1)u_{\eta\eta}$	$\frac{\varrho^2}{2p+6}$	$u_t = u_{\xi\xi} + u_{\chi\chi} - 2u_{\eta\eta}$	$\frac{\varrho^2}{4}$

Like in the 2D case (but in contrast to the 1D case) the mode filter in 3D again fits in as $p \to -1$ into the general case.

Proposition 6 (3D PDE limit for mode filtering) *Let u and $\boldsymbol{x}_0$ be as in* Proposition 5. *One step of mode filtering of u with a ball-shaped window $\mathrm{B}_\varrho(\boldsymbol{x}_0)$ approximates for $\varrho \to 0$ a time step of size $\tau = \varrho^2/4$ of an explicit time discretisation of the PDE $u_t = u_{\xi\xi} + u_{\chi\chi} - 2u_{\eta\eta}$ with η, ξ, χ as in* Proposition 5. *At a local minimum (maximum), mode filtering approximates $u_t \geq 0$ ($u_t \leq 0$).*

We summarise the PDE approximation results from Propositions 1–6 in Table 1. It shows a very systematic behaviour w.r.t. the influence of the dimension, such that it is straightforward to come up with a conjecture for arbitrary dimensions larger than 3: Depending on the order p, we expect a PDE that has forward ($p > 1$) or backward ($p < 1$) parabolic behaviour in the gradient direction $\boldsymbol{\eta}$, combined with forward parabolic smoothing orthogonal to it:

$$\begin{aligned} u_t &= \Delta u - u_{\eta\eta} + (p-1)u_{\eta\eta} \\ &= \Delta u + (p-2)u_{\eta\eta} \ . \end{aligned} \tag{30}$$

5 Discrete Versus Continuous M-Smoothing

The previous results rise the question about the relation between the discrete and continuous situation. For discrete distributions the mode is approximated according to [6] by order-p means for $p \to 0$. Negative orders p in the sense of Sect. 2 cannot be applied at all for discrete distributions. In contrast, in the case of densities over a continuous range the mode is obtained for $p \to -1$. The limit $p \to 0$ for continuous distributions instead results in a mean (that could be called order-0 mean to close the gap of definition) that does in relevant cases not coincide with the mode. As this situation is difficult to grasp intuitively, we present in Appendix B a worked-out example of a simple density function (a cut-off quadratic function) for which the order-p mean and mode can be calculated in closed form, so one can clearly see the discrepancy between order-0 mean and mode as well as the convergence to the mode for $p \to -1$.

Looking at the continuous case first, it is clear that the penaliser function $\mathrm{sgn}\,(p)|z|^p$ converges for $p \to +0$ (from the positive side) to the function (4), with the convergence being non-uniform around $z = 0$. From the negative side, one has non-uniform convergence to the function

$$\lim_{p \to -0} \mathrm{sgn}\,(p)|z|^p = 0 \quad \text{defined for } z \neq 0 \ . \tag{31}$$

As constants are negligible in penalisers for "means" of continuous distributions anyway, (4) and (31) have not only the same effect but they act simply as constants, as the removable discontinuity at 0 is without influence under integration. Thus, they do not give rise to an "order-0 mean" whatsoever.

To model the mode of a continuous distribution, a penaliser is needed that under integration gives positive weight to a single location; thus the penaliser must be a distribution with a (negative, for the mode to arise as minimiser) delta peak at 0. Indeed this is the limit of $\mathrm{sgn}\,(p)|z|^p$ for $p \to -1^+$.

To transfer a continuous penaliser Ψ to the discrete case, the correct way would be to use not sampling,

$$\psi_k := \Psi(kh) \ , \tag{32}$$

(where $h > 0$ is the step width between sampling locations) but a finite-volume discretisation which is essentially the composition of a box-kernel convolution (other low-pass kernels would be possible) with sampling,

$$\psi_k := \frac{1}{h} \int_{(k-\frac{1}{2})h}^{(k+\frac{1}{2})h} \Psi(x)\,\mathrm{d}x = (\Psi * \mathrm{B}_h)(kh) \ , \tag{33}$$

$$\mathrm{B}_h(z) := \frac{1}{h}\chi_{[-h/2,h/2]}(z) \ . \tag{34}$$

The width h of the box kernel B_h takes the role of the bin width of a histogram into which data are aggregated. For originally discrete distributions (finite multisets of data points) one might omit making this step explicit as h may be chosen arbitrarily small such that (33) approximates (32) with arbitrary accuracy (as long as Ψ is integrable in each sampling interval). As soon as h is smaller than the minimal distance between two locations of the discrete distribution, further reduction of h does not increase the number of histogram bins with positive weights.

In contrast, when discretising an originally continuous distribution, the histogram bin width becomes relevant, and it is essential to use (33).

Let us regard now the penaliser $\Psi_p(z) = |z|^p$ for $p > 0$. In order for a finite-volume discretisation of Ψ_p to converge for $p \to +0$ to the naive sampling of (4), i.e.

$$\psi_k = \begin{cases} 0\,, & k = 0\,, \\ 1 & \text{else} \end{cases} \tag{35}$$

the bin width h needs to be sent to zero along with p. For any fixed bin width h, the finite-volume discretisation of Ψ_p converges to the constant unity function, $\psi_k = 1$ for all k, instead.

On the other hand, the finite-volume discretisation of $\Psi_p(z) = -|z|^p$ for $-1 < p < 0$ converges for $p \to -1^+$ exactly to $\psi_0 = -1$ and $\psi_k = 0$ for $k \neq 0$, i.e. (35) up to an irrelevant constant offset. In this sense, the case $p = 0$ as defined in [6] for discrete distributions does indeed correspond to the discretisation of the limit $p \to -1$ of the continuous setting.

As a final remark, we point out that in an image filtering context the discretisation of distributions as discussed in this section is in fact applied to the intensity domain, thus takes the role of quantisation. The mere spatial discretisation of an image leaves the intensity domain continuous, and one should try to approximate the concepts for continuous distributions numerically as accurate as possible. However, the spatial discretisation creates a discrete sample from the continuous distribution, and (at least for $p < 1$) filtering this discrete sample as a finite set will not be a proper approximation of the continuous filter. For example, the mode of the discrete distribution will be meaningless as in generic cases finite samples from continuous distributions consist of distinct values, each with trivial frequency 1. It is therefore necessary to design a numerical process that in the one or other way estimates the continuous distribution from the set of discrete sample values.

6 Numerical Scheme for the 2D PDE Limit

Next, we discuss a numerical algorithm for approximating our two-dimensional PDE limit (7) in an adequate way. The 2D setting is practically most important, and it contains all essential difficulties that also arise in higher dimensions. Our 2D PDE gives rise to two major numerical problems:

- It involves the anisotropic expressions $u_{\xi\xi}$ and $u_{\eta\eta}$. To reproduce their qualitative properties adequately, one has to take care that the discretisation approximates rotationally invariant behaviour well and that it satisfies a discrete maximum–minimum principle which prevents over- and undershoots.
- For $p < 1$, the sign in front of the operator $u_{\eta\eta}$ becomes negative, which results in a backward parabolic operator. Such operators are known to be ill-posed. They require additional stabilisation in the model and the numerics.

These challenges show that great care must be invested in the design of appropriate numerical algorithms. Thus, let us have a deeper look into our efforts along these lines.

Reformulation Using $u_{\eta\eta} = \Delta u - u_{\xi\xi}$ and $u_{\xi\xi} = \operatorname{curv}(u)|\nabla u|$ with the isophote curvature $\operatorname{curv}(u)$ we rewrite (7) in a numerically more convenient form:

$$u_t = (2-p)\operatorname{curv}(u)|\nabla u| + (p-1)\,\Delta u\,. \tag{36}$$

If $p \geq 1$, we apply this equation in all locations, including extrema.

For $p < 1$, the second term describes backward diffusion, which we stabilise by freezing its action in extrema where $|\nabla u|$ vanishes:

$$u_t = (2-p)\operatorname{curv}(u)|\nabla u| + (p-1)\operatorname{sgn}(|\nabla u|)\,\Delta u\,. \tag{37}$$

In practice, our image domain is finite and of rectangular size. This motivates us to equip the equations (36) and (37) with reflecting (i.e. homogeneous Neumann) boundary conditions. Both evolutions (36) and (37) are replaced by finite difference schemes on a regular grid of size h in x- and y-direction and time step size τ. By $u_{i,j}$ we denote an approximation of u in pixel (i, j).

Space discretisation of forward diffusion If $p \geq 1$, we discretise Δu in (36) with a nine-point stencil. It is a weighted average of an approximation aligned along the x- and y-axis with one aligned along the diagonal directions:

$$\frac{1-\nu}{h^2}\begin{array}{|c|c|c|}\hline 0 & 1 & 0 \\ \hline 1 & -4 & 1 \\ \hline 0 & 1 & 0 \\ \hline \end{array} + \frac{\nu}{(\sqrt{2}\,h)^2}\begin{array}{|c|c|c|}\hline 1 & 0 & 1 \\ \hline 0 & -4 & 0 \\ \hline 1 & 0 & 1 \\ \hline \end{array}$$

$$= \frac{1}{2h^2}\begin{array}{|c|c|c|}\hline \nu & 2-2\nu & \nu \\ \hline 2-2\nu & 4\nu-8 & 2-2\nu \\ \hline \nu & 2-2\nu & \nu \\ \hline \end{array}\,, \tag{38}$$

where the weight $\nu \in [0, 1]$ is used to optimise the rotation invariance of the stencil. Since the stencil has an axial size of $3h$ and a diagonal one of $3\sqrt{2}h$, we choose $\nu := \sqrt{2} - 1$. This leads to the weight ratio $(1-\nu) : \nu = \sqrt{2} : 1$, which compensates for the different sizes. Our experiments will show that in this way, rotation invariance is approximated very well.

Space discretisation of backward diffusion For $p < 1$, the term $(p-1)\operatorname{sgn}(|\nabla u|)\,\Delta u$ in (37) creates stabilised

backward diffusion. Here, we base our finite difference approximation on a minmod discretisation of Osher and Rudin [44], but improve its rotation invariance again by a weighted averaging with its diagonally aligned counterpart with weight $\nu = \sqrt{2} - 1$. We denote the forward differences in x-, y-, and the diagonal directions $\boldsymbol{d} = (1, 1)$ and $\boldsymbol{e} = (1, -1)$ by

$$u^x_{i,j} := \frac{u_{i+1,j} - u_{i,j}}{h}, \qquad u^y_{i,j} := \frac{u_{i,j+1} - u_{i,j}}{h}, \tag{39}$$

$$u^d_{i,j} := \frac{u_{i+1,j+1} - u_{i,j}}{\sqrt{2}\,h}, \quad u^e_{i,j} := \frac{u_{i+1,j-1} - u_{i,j}}{\sqrt{2}\,h}. \tag{40}$$

By $\mathrm{M}(a, b, c)$ we abbreviate the minmod function of three arguments which chooses the argument of minimal modulus if the arguments have the same sign, and yields 0 otherwise:

$$\mathrm{M}(a, b, c) := \begin{cases} \operatorname*{argmin}_{z \in \{a,b,c\}} |z| & \text{if } ab \geq 0 \text{ and } ac \geq 0\,, \\ 0 & \text{else}\,. \end{cases} \tag{41}$$

With these notations we approximate $\operatorname{sgn}(|\nabla u|)\,\Delta u$ in pixel (i, j) by

$$\begin{aligned} &\tfrac{1-\nu}{h}\big(\ \mathrm{M}(u^x_{i+1,j}, u^x_{i,j}, u^x_{i-1,j}) \\ &\quad - \mathrm{M}(u^x_{i,j}, u^x_{i-1,j}, u^x_{i-2,j}) \\ &\quad + \mathrm{M}(u^y_{i,j+1}, u^y_{i,j}, u^y_{i,j-1}) \\ &\quad - \mathrm{M}(u^y_{i,j}, u^y_{i,j-1}, u^y_{i,j-2})\ \big) \\ &\quad + \tfrac{\nu}{\sqrt{2}\,h}\big(\ \mathrm{M}(u^d_{i+1,j+1}, u^d_{i,j}, u^d_{i-1,j-1}) \\ &\quad - \mathrm{M}(u^d_{i,j}, u^d_{i-1,j-1}, u^d_{i-2,j-2}) \\ &\quad + \mathrm{M}(u^e_{i+1,j-1}, u^e_{i,j}, u^e_{i-1,j+1}) \\ &\quad - \mathrm{M}(u^e_{i,j}, u^e_{i-1,j+1}, u^e_{i-2,j+2})\ \big)\,. \end{aligned} \tag{42}$$

Space discretisation of mean curvature motion Let us now discuss our approximation of the mean curvature term $(2 - p)\operatorname{curv}(u)|\nabla u|$. The isophote curvature

$$\operatorname{curv}(u) = \frac{u_x^2 u_{yy} - 2u_x u_y u_{xy} + u_y^2 u_{xx}}{(u_x^2 + u_y^2)^{3/2}} \tag{43}$$

can be discretised in a straightforward way with central differences. To avoid a potential singularity in the denominator, we regularise by adding $\epsilon = 10^{-10}$ to $u_x^2 + u_y^2$. Moreover, note that the isophote curvature $\operatorname{curv}(u)$ describes the inverse radius of the osculating circle to the level line. Since a discrete image does not have structures that are smaller than a single pixel, the smallest practically relevant radius is $\frac{h}{2}$. Thus, we impose a curvature limiter that restricts the computed result to the range $\left[-\frac{2}{h}, \frac{2}{h}\right]$.

Depending on the sign of $(2 - p)\operatorname{curv}(u)$, we may interpret $(2-p)\operatorname{curv}(u)|\nabla u|$ either as a dilation term (for positive sign) or an erosion term (for negative sign) with a disc-shaped structuring element of radius $|(2 - p)\operatorname{curv}(u)|$; see e.g. [1]. For a stable discretisation of $|\nabla u|$, we use the Rouy–Tourin upwind scheme [48]. In contrast to our conference paper [60], we again improve its rotation invariance by a weighted averaging of axial and diagonal discretisations with weight $\nu = \sqrt{2} - 1$. In the dilation case, this comes down to

$$\begin{aligned} |\nabla u|_{i,j} \;\approx\; & (1-\nu)\left(\left(\max\left(-u^x_{i-1,j},\, u^x_{i,j},\, 0\right)\right)^2\right. \\ & \left. + \left(\max\left(-u^y_{i,j-1},\, u^y_{i,j},\, 0\right)\right)^2\right)^{1/2} \\ & + \nu\left(\left(\max\left(-u^d_{i-1,j-1},\, u^d_{i,j},\, 0\right)\right)^2\right. \\ & \left. + \left(\max\left(-u^e_{i-1,j+1},\, u^e_{i,j},\, 0\right)\right)^2\right)^{1/2} \end{aligned} \tag{44}$$

and in the erosion case to

$$\begin{aligned} |\nabla u|_{i,j} \;\approx\; & (1-\nu)\left(\left(\max\left(-u^x_{i,j},\, u^x_{i-1,j},\, 0\right)\right)^2\right. \\ & \left. + \left(\max\left(-u^y_{i,j},\, u^y_{i,j-1},\, 0\right)\right)^2\right)^{1/2} \\ & + \nu\left(\left(\max\left(-u^d_{i,j},\, u^d_{i-1,j-1},\, 0\right)\right)^2\right. \\ & \left. + \left(\max\left(-u^e_{i,j},\, u^e_{i-1,j+1},\, 0\right)\right)^2\right)^{1/2}. \end{aligned} \tag{45}$$

Operator splitting The space discretisations we have discussed convert our PDEs (36) and (37) to systems of ordinary differential equations (ODEs). Their general structure is given by

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t} = {} & (1-\nu)\,\boldsymbol{D}_+(\boldsymbol{u}) + \nu\,\boldsymbol{D}_\times(\boldsymbol{u}) \\ & + (1-\nu)\,\boldsymbol{M}_+(\boldsymbol{u}) + \nu\,\boldsymbol{M}_\times(\boldsymbol{u})\,, \end{aligned} \tag{46}$$

where the vector $\boldsymbol{u}$ assembles the function values of u at all grid points in our discretised domain. The expressions $\boldsymbol{D}_+(\boldsymbol{u})$ and $\boldsymbol{D}_\times(\boldsymbol{u})$ stand for axial and diagonal discretisations of the diffusion terms. For $p > 1$ they refer to the forward term $(p-1)\Delta u$, and for $p < 1$ to the stabilised backward term $(p-1)\operatorname{sgn}(|\nabla u|)\Delta u$. Likewise, $\boldsymbol{M}_+(\boldsymbol{u})$ and $\boldsymbol{M}_\times(\boldsymbol{u})$ denote our axial and diagonal discretisations of the mean curvature motion term $(2 - p)\operatorname{curv}(u)|\nabla u|$. All discrete operators take into account the homogeneous Neumann boundary conditions by mirroring one or two layers of boundary pixels. For $t = 0$, the ODE system uses the discretised

original image $\boldsymbol{f}$ as initial condition:

$$\boldsymbol{u}(0) = \boldsymbol{f}\,. \tag{47}$$

For the time discretisation of (46), we proceed in four explicit fractional steps. Denoting the time step size by τ, and $\boldsymbol{u}$ at time level $k\tau$ by $\boldsymbol{u}^k$, our scheme is given by

$$\boldsymbol{u}^{k+1/4} = \boldsymbol{u}^k + \tau\,(1-\nu)\,\boldsymbol{D}_+(\boldsymbol{u}^k)\,, \tag{48}$$
$$\boldsymbol{u}^{k+1/2} = \boldsymbol{u}^{k+1/4} + \tau\,\nu\,\boldsymbol{D}_\times(\boldsymbol{u}^{k+1/4})\,, \tag{49}$$
$$\boldsymbol{u}^{k+3/4} = \boldsymbol{u}^{k+1/2} + \tau\,(1-\nu)\,\boldsymbol{M}_+(\boldsymbol{u}^{k+1/2})\,, \tag{50}$$
$$\boldsymbol{u}^{k+1} = \boldsymbol{u}^{k+3/4} + \tau\,\nu\,\boldsymbol{M}_\times(\boldsymbol{u}^{k+3/4})\,. \tag{51}$$

We will see that compared to an unsplit explicit scheme as was used in our conference paper [60], the split variant allows substantially larger time step sizes. For more information on operator splitting we refer the reader to the classical literature [40,41,64].

Consistency Since our resulting explicit scheme uses various one-sided—and thus first order—finite difference approximations within its upwind and minmod strategies, if follows that its general consistency order outside extrema is $\mathcal{O}(h+\tau)$. For the pure forward diffusion case $p=2$, however, the second-order stencil (38) gives $\mathcal{O}(h^2+\tau)$.

Stability Stability of a numerical algorithm typically refers to the discrete preservation of an essential property of the continuous process. By design, all M-smoothers satisfy a maximum–minimum principle, which states that maxima must not become larger during filtering, and minima not smaller. Since our PDEs of interest have been derived as limits of M-smoothers, it is natural that they should obey a maximum–minimum principle as well. This is also supported by the fact that all our evolutions under consideration satisfy $u_t \le 0$ in maxima and $u_t \ge 0$ in minima. This motivates us to study the stability of our algorithm in terms of a discrete maximum–minimum principle, which obviously also implies L^∞-stability. To this end we show that all fractional steps (48)–(51) in our explicit scheme are designed to satisfy a discrete maximum–minimum principle for suitably chosen time step sizes. Since the details are somewhat cumbersome and do not give more general insights, we sketch only the basic ideas.

For $p>1$ and $\nu<1$, the first fractional step (48) approximates the equation $u_t = (1-\nu)(p-1)\Delta u$ with an explicit axial scheme. This leads to a stencil with weight sum 1. All noncentral weights are nonnegative, and the central stencil weight is given by $1-4(1-\nu)(p-1)\frac{\tau}{h^2}$. It becomes nonnegative for

$$\tau \le \frac{h^2}{4(1-\nu)\,|p-1|} =: \tau_1\,. \tag{52}$$

In this case the scheme computes a convex combination of data from the previous time step, which implies a maximum–minimum principle.

Osher and Rudin [44] report the same stability limit for their minmod scheme for stabilised backward diffusion as one gets for the forward process, and they emphasise that their scheme does neither increase local maxima no does it decrease local minima. Thus, the step size restriction (52) also holds for the stabilised backward PDE $u_t = (1-\nu)(p-1)\mathrm{sgn}(|\nabla u|)\Delta u$ that is approximated by Step (48) for $p<1$. Note that (52) formally becomes singular for $p=1$ or $\nu=1$, when the evolution equation in the first fractional step degenerates to $u_t=0$. In this case the fractional step does nothing at all, such that its stability limit could be seen as $\tau_1=\infty$. The same considerations also apply for the step size limits of the other fractional steps that we discuss below.

The second fractional step (49) approximates the forward diffusion PDE $u_t = \nu(p-1)\Delta u$ for $p>1$, or the stabilised backward process $u_t = \nu(p-1)\mathrm{sgn}(|\nabla u|)\Delta u$ for $p<1$, but in both cases with a diagonal stencil. Hence, we can use the same reasoning as in the first step, if we exchange $1-\nu$ by ν and h by $h\sqrt{2}$. This leads to the stability limit

$$\tau \le \frac{h^2}{2\nu\,|p-1|} =: \tau_2\,. \tag{53}$$

For an axial stencil, the classical Rouy–Tourin scheme for the dilation/erosion evolutions $u_t = \pm|\nabla u|$ is well known to satisfy a maximum–minimum principle if its time step size obeys $\tau \le \frac{h}{2}\sqrt{2}$; see e.g. [9]. Thus, the third fractional step (50), which approximates $u_t = (1-\nu)(2-p)\mathrm{curv}(u)|\nabla u|$ with a curvature limiter interval $\left[-\frac{2}{h},\frac{2}{h}\right]$, must satisfy the step size restriction

$$\tau \le \frac{h^2}{2\sqrt{2}\,(1-\nu)\,|2-p|} =: \tau_3\,. \tag{54}$$

Similar arguments can be used for the fourth fractional step (51). Since it approximates the equation $u_t = \nu(2-p)\mathrm{curv}(u)|\nabla u|$ on a diagonal stencil, we obtain the time step size restriction

$$\tau \le \frac{h^2}{2\,\nu\,|2-p|} =: \tau_4\,. \tag{55}$$

These considerations immediately lead to the following stability result:

Proposition 7 (Numerical Stability) *Let the splitting scheme (48)–(51) be equipped with mirrored boundary layers and initialisation $\boldsymbol{u}^0 = \boldsymbol{f}$. Moreover, let its time step size τ satisfy*

$$\tau \le \min\{\tau_1,\,\tau_2,\,\tau_3,\,\tau_4\} \tag{56}$$

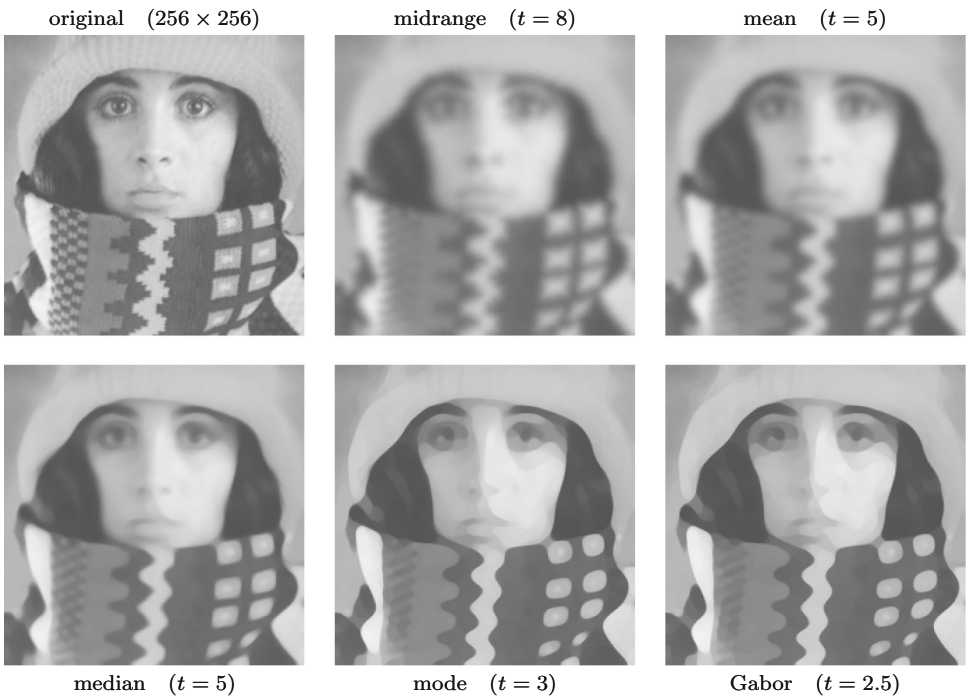

Fig. 2 Smoothing effect of the different evolution equations on the test image *trui*. Recomputed from [60] with our improved algorithm

with $\tau_1, \ldots, \tau_4$ *from* (52)–(55).
Then the scheme is L^∞*-stable,*

$$\|\boldsymbol{u}^k\|_\infty \leq \|\boldsymbol{u}^{k-1}\|_\infty \qquad \forall k \geq 1\,, \tag{57}$$

and respects the discrete maximum–minimum principle

$$\min_{i,j} f_{i,j} \leq u^k_{n,m} \leq \max_{i,j} f_{i,j} \qquad \forall n, m,\ \forall k \geq 1\,. \tag{58}$$

In practice the step size limit (56) is not very restrictive: With $h := 1$ and $\nu = \sqrt{2} - 1$, it comes down to $\tau \leq 0.4267$ for the diffusion evolution ($p = 2$), to $\tau \leq 0.6035$ for mean curvature motion ($p = 1$), to $\tau \leq 0.2011$ for the mode equation ($p = -1$), and to $\tau \leq 0.1422$ for the Gabor flow ($p = -2$). These limits are larger than the ones in our conference paper [60], and they allow efficient numerical approximations of PDE evolutions for M-smoothers.

7 Experiments

In our experiments, we evaluate the PDE (7) with five different settings for p: a temporally rescaled midrange evolution ($p \to \infty$) using $u_t = u_{\eta\eta}$ with $\tau = 0.25$, the mean evolution leading to homogeneous diffusion ($p = 2$, $\tau = 0.25$), the median evolution yielding mean curvature motion ($p = 1$, $\tau = 0.25$), the mode evolution ($p = -1$, $\tau = 0.1$), and the Gabor flow ($p = -2$, $\tau = 0.1$). Unless stated otherwise, we use the diagonal weight $\nu = \sqrt{2} - 1$. The first two experiments recompute results from our conference paper [60] by using our novel algorithm that has been improved w.r.t. rotation invariance and efficiency.

Figure 2 illustrates the effect of these equations on the real-world test image *trui*. The CPU times for computing each of these results on a contemporary laptop are in the order of half a second. We observe that the midrange filter produces fairly jagged results, although it has a clear smoothing effect. Homogeneous diffusion does not suffer from jagged artefacts, but blurs also important structures such as edges. The median evolution is designed to smooth only along isolines which results in a smaller deterioration of edge-like structures. The mode and the Gabor evolutions are very similar. They produce the sharpest results and may even enhance edges due to their backward parabolic term $(p - 1)\, u_{\eta\eta}$.

Figure 3 allows to judge if our numerical algorithm is capable of reproducing the rotationally invariant behaviour of its underlying PDE (7). We observe excellent rotation invariance. Moreover, we see that the mode and Gabor evolutions have comparable shrinkage properties as mean curvature motion. However, they differ from mean curvature motion by their backward term $(p - 1)\, u_{\eta\eta}$, which can compensate dissipative artefacts that are caused by the discretisations of the forward parabolic term $u_{\xi\xi}$.

Figure 4 illustrates the staircasing behaviour of the mode evolution. As already mentioned, staircasing is a common phenomenon for PDEs that enhance images by means of some backward parabolic concepts. It has been observed for the Perona–Malik filter [47], for forward-and-backward (FAB) diffusion [26], and for shock filters [35,45]. Staircasing becomes pronounced if a smoothly varying image

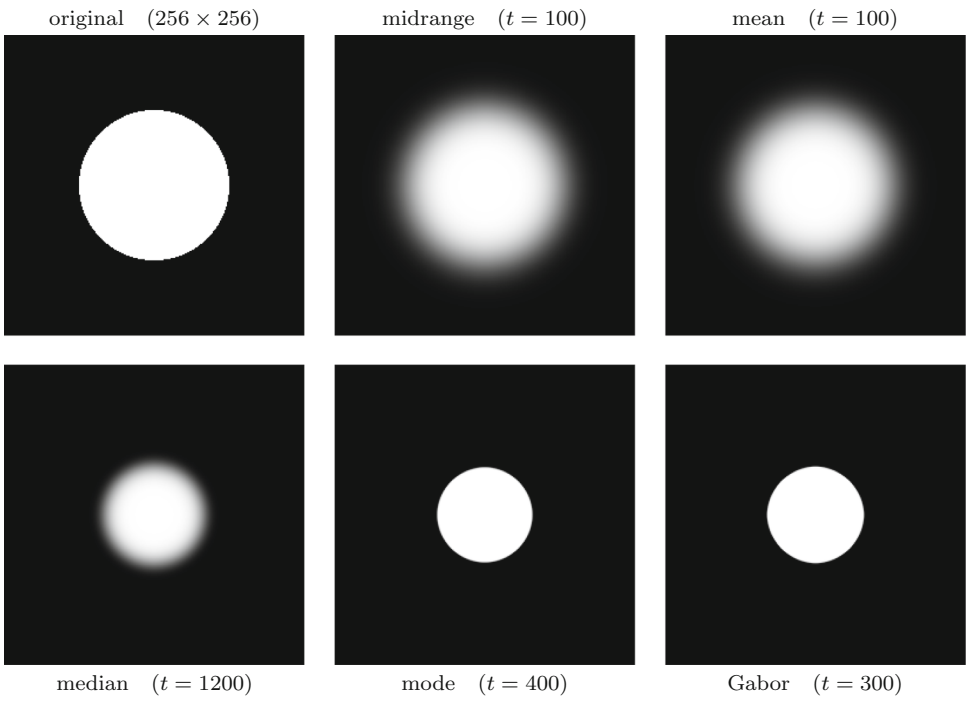

Fig. 3 Effect of the different evolution equations on a disc. Recomputed from [60] with our improved algorithm

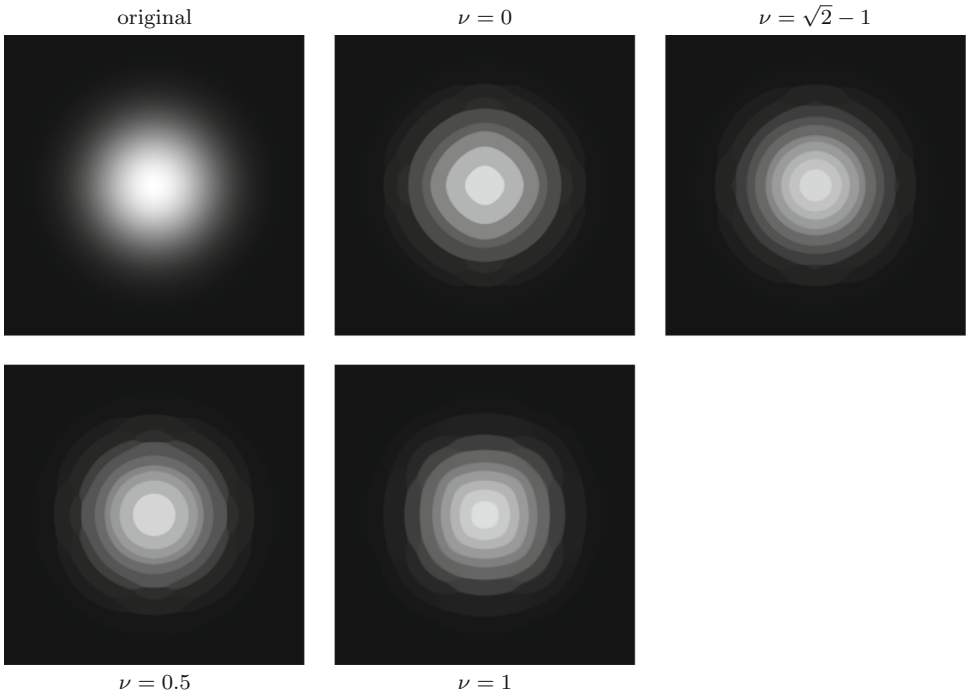

Fig. 4 Staircasing effect of the mode evolution on a Gaussian test image, and influence of the diagonal weight ν on the rotation invariance. Image size: 256×256. Evolution time: $t = 100$

structure is to be enhanced. Therefore, we have chosen a Gaussian-like test image, which also allows to judge the rotation invariance of our algorithm for different values of the diagonal weight ν. We observe that ν also has some impact on the number and size of the evolving stairs: Since backward parabolic processes are very sensitive w.r.t. the data and corresponding algorithms, such a behaviour is not unnatural. The discretisation with $\nu = 0$ produces the coarsest stairs, while the ones for $\nu = \sqrt{2}-1$ are particularly small. Regarding rotation invariance, Fig. 4 shows that a pure axial ($\nu = 0$) or a pure diagonal approximation ($\nu = 1$) perform relatively bad, which is to be expected. We see that the proposed value of $\nu = \sqrt{2}-1 \approx 0.4142$ yields the most favourable result. It also outperforms the result for $\nu = 0.5$. The latter parameter was used in the discretisation of the diffusion term in our conference paper [60].

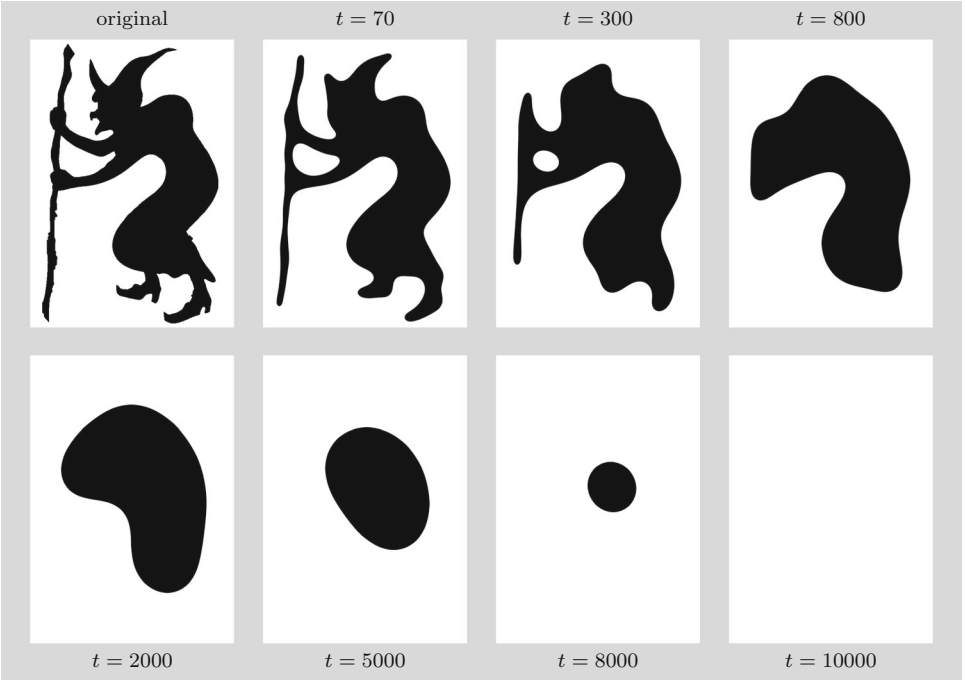

Fig. 5 Shape simplification properties of the mode evolution. Image size: 561 × 792. Source of binarised original image: https://www.kissclipart.com/best-priced-decals-halloween-decor-witch-and-brew-rmm1nq/

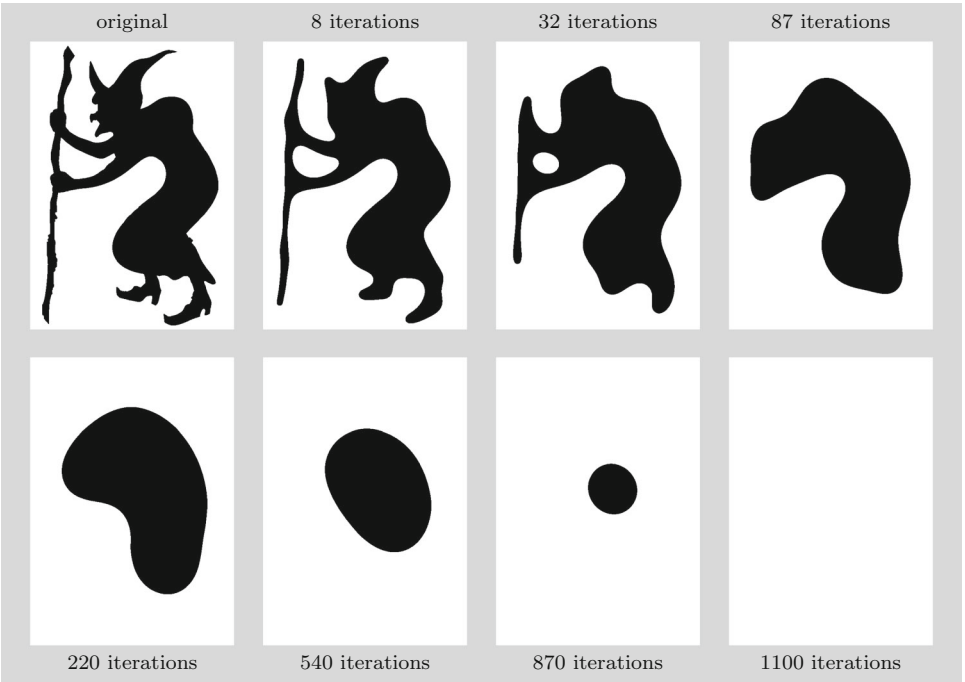

Fig. 6 Shape simplification properties of iterated histogram-based mode filtering with a disc of radius 13. Image size: 561 × 792

In Fig. 5, we study the shape simplification properties of the mode evolution by applying it to the binary image of a witch. We observe that under the mode evolution, connected components remain connected. It shrinks the shape in such a way that highly curved structures evolve faster than less curved ones, resulting in an evolution where nonconvex shapes become convex and vanish in finite time by shrinking to a so-called circular point. Thus, the mode evolution appears to enjoy experimentally the same binary shape simplification qualities as the theory states for mean curvature motion. This may surprise at first glance when looking only at the PDEs: Mean curvature motion is a morphologically invariant geometric PDE in the sense of Alvarez et al. [1], while the mode evolution is not. The M-smoother interpretation can shed some light on this: While the mode evolution is designed to reproduce the qualities of mode filtering, mean curvature motion is related to median filtering. For binary data, we face a specific scenario where median and mode

coincide. Moreover, both median and mode filters preserve the binary nature. Finite difference approximations for mean curvature motion, however, suffer from dissipative artefacts which result in unwanted blurring that destroys the binary nature. Because of its backward parabolic term, the mode evolution does not suffer from these dissipative artefacts. Figure 5 shows that it can preserve the binary nature of the data very well. This property constitutes a distinctive advantage over mean curvature motion and makes the binary mode evolution attractive for shape analysis problems.

The results in Fig. 5 can be juxtaposed to the ones in Fig. 6. The latter one shows the effect of a histogram-based implementation of iterative mode filtering: In every iteration it replaces each pixel by its mode within a disc-shaped neighbourhood of radius 13 pixels. Although our PDE limit has been obtained only for vanishing radii and although its numerical scheme approximates the PDE only with first-order consistency, we observe a large qualitative agreement of Figs. 5 and 6. This confirms the validity of the PDE limit.

8 Summary and Conclusions

We have established a comprehensive analysis that identifies the PDE limit for the full class of iterated M-smoothers with order-p means. Our discussion was not restricted to the two-dimensional case which constitutes the most natural setting in image analysis: We have also derived analog results in the one- and three-dimensional case. This allows to gain deeper structural insights into the general behaviour of this filter class.

In the 2D setting, our analysis does not only reproduce known results for mean and median filtering, but also corrects a common misconception in the literature: We have shown the surprising fact that in the continuous limit, mode filtering does not correspond to $p = 0$, but results from the limit $p \to -1$. Moreover, our filter class $u_t = u_{\xi\xi} + (p-1)\, u_{\eta\eta}$ can also be extended to models that have no interpretation within the setting of M-smoothers, e.g. Gabor's classical method for $p = -2$.

At the time being, our results are restricted to grey-value images. An extension to multivariate data such as colour images or diffusion tensor fields would be interesting but is not straightforward, and has to be left to future research. Available results on multivariate median filters [57,58] indicate that substantial work will be required for such a generalisation.

Since adequate histogram-based implementations of some M-smoothers such as mode filtering can become highly non-trivial when using small local histograms [28,34], we have proposed a novel numerical algorithm in 2D that can handle the PDE evolution for arbitrary values of p. Although these evolutions can be highly anisotropic and may even exhibit backward parabolic behaviour, we managed to come up with an L^∞-stable finite difference scheme that is efficient, satisfies a maximum–minimum principle and shows excellent rotation invariance. This has been partly achieved by employing and adapting powerful stabilisation concepts from the numerics of hyperbolic PDEs, such as upwinding, minmod functions, and curvature limiters.

It should be emphasised that our numerical algorithm is applicable to any stable evolution of type $u_t = a\, u_{\xi\xi} + b\, u_{\eta\eta}$, where a and b may have arbitrary sign. Thus, it is of very general nature and covers also numerous applications beyond M-smoothing, including image interpolation [12], adaptive filter design [2,11], many level set methods [43], as well as other second-order PDEs in gauge coordinates such as p-Laplacian evolutions [15,36].

Our experiments indicate that the PDEs for $p < 1$, such as the mode evolution, are particularly appealing: They combine strong shape simplification properties with pronounced sharpening qualities. They clearly deserve more research.

Connecting the class of M-smoothers to the family of PDE-based methods contributes one more mosaic stone to the mathematical foundations of image analysis. Since M-smoothers themselves are related to many other approaches [42,53,62], including W-smoothers, bilateral filters, mean-shift and robust estimation, our results can help to gain a broader and more coherent view on the entire field.

Acknowledgements This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 741215, ERC Advanced Grant INCOVID). We thank Luis Alvarez (University of Las Palmas de Gran Canaria) for interesting and inspiring discussions on this topic.

A Proofs of PDE Approximation Results

A.1 Proof of Proposition 1

A.1.1 Preliminaries: Some Important Integrals

We start by collecting some definite integrals that will be useful in the following. We define for $\varrho \in (0, 1)$ and $q \in \mathbb{R}$

$$I_q := \int_{\sqrt{\varrho}}^{1} \sqrt{1-\xi^2}\, \xi^q \, \mathrm{d}\xi \,, \tag{59}$$

$$S_q := \int_{\arcsin\sqrt{\varrho}}^{\pi/2} \sin^q \varphi \, \mathrm{d}\varphi \,. \tag{60}$$

With the additional abbreviation

$$R_q := \varrho^{q/2}\sqrt{1-\varrho} \tag{61}$$

we can derive via substituting $\xi = \sin\varphi$ and integration by parts (integrating $\sin^q\varphi\cos\varphi$ and differentiating $\cos\varphi$)

$$\begin{aligned} I_q &= \int_{\arcsin\sqrt{\varrho}}^{\pi/2} \sin^q\varphi\cos^2\varphi\,\mathrm{d}\varphi \\ &= \frac{1}{q+1}\left[\sin^{q+1}\varphi\cos\varphi\right]_{\arcsin\sqrt{\varrho}}^{\pi/2} + \frac{1}{q+1}S_{q+2} \\ &= \frac{-1}{q+1}R_{q+1} + \frac{1}{q+1}S_{q+2} \end{aligned} \tag{62}$$

for $q \neq -1$. Moreover, we have by $1 = \sin^2\varphi + \cos^2\varphi$

$$\begin{aligned} S_q &= \int_{\arcsin\sqrt{\varrho}}^{\pi/2} \sin^q\varphi\cos^2\varphi\,\mathrm{d}\varphi + S_{q+2} \\ &= \frac{-1}{q+1}R_{q+1} + \frac{q+2}{q+1}S_{q+2} \end{aligned} \tag{63}$$

for $q \neq -1$ which allows to transform S_q into S_{q+2} and vice versa.

From (62) we can obtain thereby

$$\begin{aligned} I_{p-4} &= \frac{-1}{p-3}R_{p-3} + \frac{-1}{(p-1)(p-3)}R_{p-1} \\ &\quad + \frac{-p}{(p+1)(p-1)(p-3)}R_{p+1} \\ &\quad + \frac{(p+2)p}{(p+1)(p-1)(p-3)}S_{p+2}\,, \end{aligned} \tag{64}$$

$$\begin{aligned} I_{p-2} &= \frac{-1}{p-1}R_{p-1} + \frac{-1}{(p+1)(p-1)}R_{p+1} \\ &\quad + \frac{p+2}{(p+1)(p-1)}S_{p+2}\,, \end{aligned} \tag{65}$$

$$I_p = \frac{-1}{p+1}R_{p+1} + \frac{1}{p+1}S_{p+2}\,, \tag{66}$$

$$I_{p+2} = \frac{-1}{p+4}R_{p+3} + \frac{1}{p+4}S_{p+2}\,, \tag{67}$$

for real p with exception of some odd integers. Note that also for the exceptional values (where some of the denominators become zero) the integrals exist.

A.1.2 Regular Points: Ansatz via Taylor Expansion

Let the image u and mean order p be given as in the proposition. Assume w.l.o.g. that the regular location $\boldsymbol{x}_0$ is $(0,0)$ with $u(0,0) = 0$, and that the gradient of u at $(0,0)$ is in the positive x direction, i.e. $u_x > 0$, $u_y = 0$. Let a neighbourhood radius $\varrho > 0$ be given, and denote the closed (Euclidean) ϱ-neighbourhood of $(0,0)$ by D_ϱ.

Using Taylor expansion of u up to third order, we can write for $(x,y) \in \mathrm{D}_\varrho$

$$\begin{aligned} u(x,y) = \alpha\big(&x + \beta x^2 + \gamma xy + \delta y^2 + \varepsilon_0 x^3 + \varepsilon_1 x^2 y \\ &+ \varepsilon_2 xy^2 + \varepsilon_3 y^3\big) + \mathcal{O}\big((x+y)^4\big) \end{aligned} \tag{68}$$

where $\alpha = u_x$, $2\beta = u_{xx}/u_x$, $\gamma = u_{xy}/u_x$, $2\delta = u_{yy}/u_x$.

We assume that ϱ is chosen small enough such that u_x is positive everywhere in D_ϱ, each level set of u within the disc D_ϱ is either a smooth line connecting two points at the circular boundary of the disc, or one of two single points on the boundary of D_ϱ where u takes its maximum and minimum on D_ϱ, respectively.

The order-p mean of u within D_ϱ is the minimiser of

$$E_0(\mu) := \operatorname{sgn}(p) \iint_{\mathrm{D}_\varrho} |u(x,y) - \mu|^p \,\mathrm{d}y\,\mathrm{d}x\,. \tag{69}$$

By some rough estimates one can conclude that for $\varrho \to 0$, $\mu \sim \varrho^2$. We substitute therefore

$$x = \varrho\xi\,,\quad y = \varrho\eta\,,\quad \mu = \varrho^2\alpha\kappa\,,\quad u(x,y) = \varrho\alpha\omega(\xi,\eta) \tag{70}$$

and obtain

$$E_0(\mu) = \operatorname{sgn}(p)\varrho^{p+2}\alpha^p E(\kappa)\,, \tag{71}$$

$$E(\kappa) = \iint_{\mathrm{D}_1} |\omega - \kappa\varrho|^p \,\mathrm{d}\eta\,\mathrm{d}\xi\,, \tag{72}$$

$$\begin{aligned} \omega(\xi,\eta) = \xi &+ \beta\xi^2\varrho + \gamma\xi\eta\varrho + \delta\eta^2\varrho + \varepsilon_0\xi^3\varrho^2 + \varepsilon_1\xi^2\eta\varrho^2 \\ &+ \varepsilon_2\xi\eta^2\varrho^2 + \varepsilon_3\eta^3\varrho^2 + \mathcal{O}\big(\varrho^3(\xi+\eta)\big)\,. \end{aligned} \tag{73}$$

In the following we focus therefore on finding the extremum of E (minimum for $p > 0$, maximum for $p < 0$).

A.1.3 Separation of the Integral

The integral E from (72) can be reorganised into a nested integration where the inner integral integrates along a level line of ω going through $(\xi,0)$, and the outer integral then integrates along the ξ axis. We have

$$\begin{aligned} E(\kappa) = &\int_{-1}^{1} \left(\int_{\eta_-^*(\xi)}^{\eta_+^*(\xi)} \frac{1}{\frac{\partial\omega}{\partial\xi}\big(\tilde{\xi}(\eta),\eta\big)}\,\mathrm{d}\eta \right) \\ &\times |\omega(\xi,0) - \kappa\varrho|^p \frac{\partial\omega}{\partial\xi}(\xi,0)\,\mathrm{d}\xi + \mathcal{O}(\varrho^3) \end{aligned} \tag{74}$$

where $\tilde{\xi}$ is a function of η that describes the level line of ω that goes through $(\xi,0)$, and reaches the boundary of D_1 at $\eta_+^* > 0$ and $\eta_-^* < 0$. (Note that the fact that ω_ξ is positive throughout D_1 implies that the level line through $(\xi,0)$ can be described in this way.)

The error term $\mathcal{O}(\varrho^3)$ results from the neglection of those level lines near the maximum and minimum of ω within D_1 that do not reach the ξ axis within D_1.

In (74), the inner integral

$$V(\xi) := \int_{\eta_-^*(\xi)}^{\eta_+^*(\xi)} \frac{1}{\frac{\partial\omega}{\partial\xi}\big(\tilde{\xi}(\eta),\eta\big)}\,\mathrm{d}\eta \tag{75}$$

measures the density of the value $\omega(\xi,0)$ in the overall distribution of ω values within D_1 by integrating along the level line $\tilde{\xi}(\eta)$ with η as integration parameter the inverse density of level lines in ξ direction. It is important here that the inverse density of level lines is measured in a direction perpendicular to that of integration. The density of level lines in ξ direction is exactly the derivative $\partial\omega/\partial\xi$ taken at the point $(\tilde{\xi},\eta)$, i.e. the denominator of the integrand.

Integrating the quantity V multiplied with the penaliser $|\omega-\kappa\varrho|^p$ would directly yield $E(\kappa)$ if the integration were carried out w.r.t. ω. We prefer, however, to keep the integration over ξ in order to avoid plugging in the inverse function of $\omega(\xi) \equiv \omega(\xi,0)$ everywhere in the expressions. This is compensated by the factor $(\partial\omega/\partial\xi)(\xi,0)$ placed at the end of the integrand of (74) that represents just the substitution of ω with ξ (along the ξ axis $\eta=0$) as integration variable.

For ease of evaluation, we combine in the following the substitution factor with the weight $V(\xi)$ in one single expression:

$$W(\xi) := \frac{\partial\omega}{\partial\xi}(\xi,0)\,V(\xi) = \int_{\eta_-^*(\xi)}^{\eta_+^*(\xi)} \frac{\frac{\partial\omega}{\partial\xi}(\xi,0)}{\frac{\partial\omega}{\partial\xi}\big(\tilde{\xi}(\eta),\eta\big)}\,\mathrm{d}\eta \tag{76}$$

A.1.4 Evaluation of the Inner (Weight) Integral

To evaluate (76), we determine first the level line function $\tilde{\xi}(\eta)$ for given $\xi=\tilde{\xi}(0)$ by using the Taylor expansion (73):

$$\omega(\xi,0) = \omega(\tilde{\xi}(\eta),\eta) \tag{77}$$

$$\xi+\beta\xi^2\varrho+\varepsilon_0\xi^3\varrho^2 = \tilde{\xi}+\big(\beta\tilde{\xi}^2+\gamma\tilde{\xi}\eta+\delta\eta^2\big)\varrho + \big(\varepsilon_0\tilde{\xi}^3+\varepsilon_1\tilde{\xi}^2\eta+\varepsilon_2\tilde{\xi}\eta^2+\varepsilon_3\eta^3\big)\varrho^2+\mathcal{O}(\varrho^3) \tag{78}$$

$$\tilde{\xi}(\eta) = \xi-(\gamma\xi+\delta\eta)\eta\varrho + \big((2\beta\xi+\gamma\eta)(\gamma\xi+\delta\eta)-\varepsilon_1\xi^2-\varepsilon_2\xi\eta-\varepsilon_3\eta^2\big)\eta\varrho^2 + \mathcal{O}(\varrho^3)\,. \tag{79}$$

The η coordinates $\eta_\pm^*$ of the end points of the level line are obtained from the condition $\tilde{\xi}^2+\eta^{*2}=1$ as

$$\eta_\pm^* = \pm\sqrt{1-\xi^2}+\big(\gamma\xi^2\pm\delta\xi\sqrt{1-\xi^2}\big)\varrho + \big(\chi(\xi)\pm\psi(\xi)\sqrt{1-\xi^2}\big)\varrho^2+\mathcal{O}(\varrho^3) \tag{80}$$

where

$$\chi(\xi) = \chi_0+\chi_1\xi+\chi_2\xi^2+\chi_3\xi^3+\chi_4\xi^4\,, \tag{81}$$

$$\psi(\xi) = \psi_0+\psi_1\xi+\psi_2\xi^2+\psi_3\xi^3 \tag{82}$$

are polynomials in ξ the exact coefficients of which are not further needed.

Based on the Taylor expansion (73) we obtain

$$\frac{\partial\omega}{\partial\xi}(\xi,\eta) = 1+\big(2\beta\xi+\gamma\eta\big)\varrho + \big(3\varepsilon_0\xi^2+2\varepsilon_1\xi\eta+\varepsilon_2\eta^2\big)\varrho^2+\mathcal{O}(\varrho^3)\,, \tag{83}$$

$$\frac{\partial\omega}{\partial\xi}(\xi,0) = 1+2\beta\xi\varrho+3\varepsilon_0\xi^2\varrho^2+\mathcal{O}(\varrho^3)\,, \tag{84}$$

and with (79)

$$\frac{\partial\omega}{\partial\xi}(\tilde{\xi},\eta) = 1+\big(2\beta\xi+\gamma\eta\big)\varrho+\big(-2\beta\gamma\xi\eta-2\beta\delta\eta + 3\varepsilon_0\xi^2+2\varepsilon_1\xi\eta+\varepsilon_2\eta^2\big)\varrho^2+\mathcal{O}(\varrho^3)\,. \tag{85}$$

Combining (84) and (85) we have

$$\frac{\frac{\partial\omega}{\partial\xi}(\xi,0)}{\frac{\partial\omega}{\partial\xi}(\tilde{\xi},\eta)} = 1-\gamma\eta\varrho+\big(4\beta\gamma\xi\eta+2\beta\delta\eta^2+\gamma^2\eta^2 - 2\varepsilon_1\xi\eta-\varepsilon_2\eta^2\big)\varrho^2+\mathcal{O}(\varrho^3) \tag{86}$$

and therefore

$$\begin{aligned} W(\xi) &= \int_{\eta_-^*}^{\eta_+^*}\mathrm{d}\eta+\big(-\gamma\varrho+4\beta\gamma\xi\varrho^2-2\varepsilon_1\xi\varrho^2\big)\int_{\eta_-^*}^{\eta_+^*}\eta\,\mathrm{d}\eta \\ &\quad+\big(2\beta\delta+\gamma^2-\varepsilon_2\big)\varrho^2\int_{\eta_-^*}^{\eta_+^*}\eta^2\,\mathrm{d}\eta+\mathcal{O}(\varrho^3) \\ &= \big(\eta_+^*-\eta_-^*\big)+\frac{1}{2}\big(-\gamma+4\beta\gamma\xi\varrho-2\varepsilon_1\xi\varrho\big)\varrho\big(\eta_+^{*\,2}-\eta_-^{*\,2}\big) \\ &\quad+\frac{1}{3}\big(2\beta\delta+\gamma^2-\varepsilon_2\big)\varrho^2\big(\eta_+^{*\,3}-\eta_-^{*\,3}\big)+\mathcal{O}(\varrho^3)\,. \end{aligned} \tag{87}$$

From (80) one sees that

$$\eta_+^*-\eta_-^* = 2\sqrt{1-\xi^2}+2\delta\xi\sqrt{1-\xi^2}\varrho + 2\psi(\xi)\sqrt{1-\xi^2}\varrho^2+\mathcal{O}(\varrho^3)\,, \tag{88}$$

$$\eta_+^{*\,2}-\eta_-^{*\,2} = 4\gamma\xi^2\sqrt{1-\xi^2}\varrho+\mathcal{O}(\varrho^2)\,, \tag{89}$$

$$\eta_+^{*\,3}-\eta_-^{*\,3} = 2(1-\xi^2)^{3/2}+\mathcal{O}(\varrho)\,, \tag{90}$$

which allows to continue (87) into

$$\begin{aligned} W(\xi) &= 2\sqrt{1-\xi^2} + 2\delta\xi\sqrt{1-\xi^2}\varrho + 2\psi(\xi)\sqrt{1-\xi^2}\varrho^2 \\ &\quad - 2\gamma^2\xi^2\sqrt{1-\xi^2}\varrho^2 \\ &\quad + \frac{2}{3}\big(2\beta\delta + \gamma^2 - \varepsilon_2\big)(1-\xi^2)^{3/2}\varrho^2 + \mathcal{O}(\varrho^3) \\ &= \left(2 + \left(\frac{4}{3}\beta\delta + \frac{2}{3}\gamma^2 - \frac{2}{3}\varepsilon_2 + 2\psi_0\right)\varrho^2\right)\sqrt{1-\xi^2} \\ &\quad + \left(2\delta + 2\psi_1\varrho\right)\varrho\xi\sqrt{1-\xi^2} \\ &\quad + \left(-2\gamma^2 - \frac{4}{3}\beta\delta - \frac{2}{3}\gamma^2 + \frac{2}{3}\varepsilon_2 + 2\psi_2\right) \\ &\quad \times \varrho^2\xi^2\sqrt{1-\xi^2} \\ &\quad + 2\psi_3\varrho^2\xi^3\sqrt{1-\xi^2} + \mathcal{O}(\varrho^3) \\ &= \Big((w_{0,0} + w_{0,2}\varrho^2) + w_1\varrho\xi + w_2\varrho^2\xi^2 \\ &\quad + w_3\varrho^2\xi^3\Big)\sqrt{1-\xi^2} + \mathcal{O}(\varrho^3) \end{aligned} \tag{91}$$

with

$$\left.\begin{aligned} w_{0,0} &= 2\ , \\ w_{0,2} &= \frac{4}{3}\beta\delta + \frac{2}{3}\gamma^2 - \frac{2}{3}\varepsilon_2 + 2\psi_0\ , \\ w_1 &= 2\delta + 2\psi_1\varrho\ , \\ w_2 &= -2\gamma^2 - \frac{4}{3}\beta\delta - \frac{2}{3}\gamma^2 + \frac{2}{3}\varepsilon_2 + 2\psi_2\ , \\ w_3 &= 2\psi_3\ . \end{aligned}\right\} \tag{92}$$

A.1.5 Domain Splitting of the Outer Integral

The outer integral of (74), i.e. the integration of $W(\xi)$ with the penaliser function $|\omega - \kappa\varrho|^p$, is now split into four parts.

First, we split the integration interval at $\xi = \nu\varrho$ where $\omega(\nu\varrho) = \kappa\varrho$ to reduce $|\omega - \kappa\varrho|$ to either $\omega - \kappa\varrho$ or $-\omega + \kappa\varrho$ in each subinterval. By (73) one has $\nu = \kappa + \mathcal{O}(\varrho^2)$.

Second, the density term $W(\xi)$ contains $\sqrt{1-\xi^2}$ which is not differentiable at ± 1, precluding Taylor expansion of this term near the outer interval boundaries. On the other hand, the p-th power penaliser is for $p \leq 1$ not differentiable at 0 and can therefore not be treated by Taylor expansion at the boundary ν between the two integration intervals. For this reason, we split each of the two intervals again at $|\xi| = \sqrt{\varrho}$. This allows to simplify the integrals in later steps by applying Taylor expansion to either $W(\xi)$ or the penaliser function, safely avoiding the critical regions of each.

As a result, we have

$$E(\kappa) = F_-(\kappa) + G_-(\kappa) + G_+(\kappa) + F_+(\kappa) + \mathcal{O}(\varrho^3)\ , \tag{93}$$

$$\begin{aligned} F_-(\kappa) &= \int_{-1}^{-\sqrt{\varrho}} W(\xi)\left(-\omega(\xi) + \kappa\varrho\right)^p \,\mathrm{d}\xi \\ &= \int_{\sqrt{\varrho}}^{1} W(-\xi)\left(-\omega(-\xi) + \kappa\varrho\right)^p \,\mathrm{d}\xi\ , \end{aligned} \tag{94}$$

$$\begin{aligned} G_-(\kappa) &= \int_{-\sqrt{\varrho}}^{\nu\varrho} W(\xi)\left(-\omega(\xi) + \kappa\varrho\right)^p \,\mathrm{d}\xi \\ &= \int_{0}^{\sqrt{\varrho}+\nu\varrho} W(-(\xi - \nu\varrho)) \\ &\quad \times \left(-\omega(-(\xi - \nu\varrho)) + \kappa\varrho\right)^p \,\mathrm{d}\xi\ , \end{aligned} \tag{95}$$

$$\begin{aligned} G_+(\kappa) &= \int_{\nu\varrho}^{\sqrt{\varrho}} W(\xi)\left(\omega(\xi) - \kappa\varrho\right)^p \,\mathrm{d}\xi \\ &= \int_{0}^{\sqrt{\varrho}-\nu\varrho} W(\xi + \nu\varrho)\left(\omega(\xi + \nu\varrho) - \kappa\varrho\right)^p \,\mathrm{d}\xi\ , \end{aligned} \tag{96}$$

$$F_+(\kappa) = \int_{\sqrt{\varrho}}^{1} W(\xi)\left(\omega(\xi) - \kappa\varrho\right)^p \,\mathrm{d}\xi\ . \tag{97}$$

A.1.6 Evaluation of the Outer Integral I

We start by evaluating the integrals $F_\mp$. In the following the upper signs refer to F_-, the lower ones to F_+. In expanding the power $(1+\ldots)^p$ by a Taylor series, it is important to note that ϱ/ξ is of order $\mathcal{O}\big(\sqrt{\varrho}\big)$ due to the lower integral bound.

$$\begin{aligned} F_\mp &= \int_{\sqrt{\varrho}}^{1} W(\mp\xi)\left(\mp\omega(\mp\xi) \pm \kappa\varrho\right)^p \,\mathrm{d}\xi \\ &= \int_{\sqrt{\varrho}}^{1} W(\mp\xi)\left(\xi \mp \beta\xi^2\varrho \pm \kappa\varrho + \varepsilon_0\xi^3\varrho^2 + \mathcal{O}(\varrho^3\xi)\right)^p \,\mathrm{d}\xi \\ &= \int_{\sqrt{\varrho}}^{1} W(\mp\xi)\xi^p\left(1 \pm \kappa\frac{\varrho}{\xi} \mp \beta\xi\varrho + \varepsilon_0\xi^2\varrho^2 + \mathcal{O}(\varrho^3)\right)^p \,\mathrm{d}\xi \\ &= \int_{\sqrt{\varrho}}^{1} W(\mp\xi)\,\xi^p\left(1 \pm p\kappa\frac{\varrho}{\xi} \mp p\beta\xi\varrho + p\varepsilon_0\xi^2\varrho^2\right. \\ &\quad + \binom{p}{2}\kappa^2\frac{\varrho^2}{\xi^2} - 2\binom{p}{2}\beta\kappa\varrho^2 + \binom{p}{2}\beta^2\xi^2\varrho^2 \\ &\quad \left.\mp\binom{p}{3}\kappa^3\frac{\varrho^3}{\xi^3} + \binom{p}{4}\kappa^4\frac{\varrho^4}{\xi^4} + \mathcal{O}(\varrho^{5/2})\right) \,\mathrm{d}\xi \\ &= \int_{\sqrt{\varrho}}^{1}\Big((w_{0,0} + w_{0,2}\varrho^2) \mp w_1\varrho\xi + w_2\varrho^2\xi^2 \\ &\quad \mp w_3\varrho^2\xi^3 + \mathcal{O}(\varrho^3)\Big)\xi^p\sqrt{1-\xi^2} \\ &\quad \times \left(1 \pm p\kappa\frac{\varrho}{\xi} \mp p\beta\xi\varrho + \binom{p}{2}\kappa^2\frac{\varrho^2}{\xi^2} \mp \binom{p}{3}\kappa^3\frac{\varrho^3}{\xi^3}\right. \end{aligned}$$

$$+ p\varepsilon_0\xi^2\varrho^2 - 2\binom{p}{2}\beta\kappa\varrho^2 + \binom{p}{2}\beta^2\xi^2\varrho^2 + \binom{p}{4}\kappa^4\frac{\varrho^4}{\xi^4} + \mathcal{O}(\varrho^{5/2})\Big)\,\mathrm{d}\xi\,. \tag{98}$$

This gives

$$\begin{aligned}
F_- + F_+ &= 2\int_{\sqrt{\varrho}}^1 w_{0,0}\xi^p\sqrt{1-\xi^2}\Big(1+\binom{p}{2}\kappa^2\frac{\varrho^2}{\xi^2} \\
&\quad + p\varepsilon_0\xi^2\varrho^2 - 2\binom{p}{2}\beta\kappa\varrho^2 + \binom{p}{2}\beta^2\xi^2\varrho^2 \\
&\quad + \binom{p}{4}\kappa^4\frac{\varrho^4}{\xi^4} + \mathcal{O}(\varrho^{5/2})\Big)\,\mathrm{d}\xi \\
&\quad + 2\int_{\sqrt{\varrho}}^1 \Big(w_{0,2} + w_2\xi^2\Big)\varrho^2\xi^p\sqrt{1-\xi^2} \\
&\quad \times \Big(1+\mathcal{O}(\varrho^{1/2})\Big)\,\mathrm{d}\xi \\
&\quad + 2\int_{\sqrt{\varrho}}^1 \Big(w_1\varrho\xi + \mathcal{O}(\varrho^2)\Big)\xi^p\sqrt{1-\xi^2} \\
&\quad \times \Big(-p\kappa\frac{\varrho}{\xi} + p\beta\xi\varrho + \mathcal{O}(\varrho^{3/2})\Big)\,\mathrm{d}\xi \\
&= 2w_{0,0}\binom{p}{4}\kappa^4\varrho^4 I_{p-4} + 2w_{0,0}\binom{p}{2}\kappa^2\varrho^2 I_{p-2} \\
&\quad + 2\Big(w_{0,0} - 2w_{0,0}\binom{p}{2}\beta\kappa\varrho^2 + w_{0,2}\varrho^2 \\
&\quad - w_1 p\kappa\varrho^2\Big) I_p \\
&\quad + 2\Big(w_{0,0}p\varepsilon_0\varrho^2 + w_{0,0}\binom{p}{2}\beta^2\varrho^2 + w_2\varrho^2 \\
&\quad + w_1 p\beta\varrho^2\Big) I_{p+2} + \mathcal{O}(\varrho^{5/2})
\end{aligned} \tag{99}$$

and by (64)–(67) we obtain

$$\begin{aligned}
F_- + F_+ &= 2w_{0,0}\binom{p}{4}\kappa^4\varrho^4\Big(\frac{-R_{p-3}}{p-3} + \frac{-R_{p-1}}{(p-1)(p-3)} \\
&\quad + \frac{-p\,R_{p+1}}{(p+1)(p-1)(p-3)} \\
&\quad + \frac{(p+2)p\,S_{p+2}}{(p+1)(p-1)(p-3)}\Big) \\
&\quad + 2\Big(w_{0,0}\binom{p}{2}\kappa^2\varrho^2\Big)\Big(\frac{-R_{p-1}}{p-1} \\
&\quad + \frac{-R_{p+1}}{(p+1)(p-1)} + \frac{(p+2)\,S_{p+2}}{(p+1)(p-1)}\Big) \\
&\quad + 2\Big(w_{0,0} - 2w_{0,0}\binom{p}{2}\beta\kappa\varrho^2 + w_{0,2}\varrho^2 \\
&\quad - w_1 p\kappa\varrho^2\Big)\Big(\frac{-R_{p+1}}{p+1} + \frac{S_{p+2}}{p+1}\Big) \\
&\quad + 2\Big(w_{0,0}p\varepsilon_0\varrho^2 + w_{0,0}\binom{p}{2}\beta^2\varrho^2 + w_2\varrho^2 \\
&\quad + w_1 p\beta\varrho^2\Big)\Big(\frac{-R_{p+4}}{p+4} + \frac{S_{p+2}}{p+4}\Big) \\
&\quad + \mathcal{O}(\varrho^{5/2})
\end{aligned} \tag{100}$$

$$\begin{aligned}
&= w_{0,0}\frac{2}{p+1}S_{p+2} + \Big(w_{0,0}\frac{(p+2)p}{(p+1)}\kappa^2 \\
&\quad - w_{0,0}\frac{2p(p-1)}{p+1}\beta\kappa + w_{0,2}\frac{1}{p+1} - w_1\frac{p}{p+1}\kappa \\
&\quad + w_{0,0}\frac{2p}{p+4}\varepsilon_0 + w_{0,0}\frac{p(p-1)}{(p+4)}\beta^2 + w_2\frac{2}{p+4} \\
&\quad + w_1\frac{2p}{p+4}\beta\Big)\varrho^2 S_{p+2} \\
&\quad - w_{0,0}\frac{2}{p+1}\varrho^{(p+1)/2}\sqrt{1-\varrho} \\
&\quad - w_{0,0}p\kappa^2\varrho^{(p+3)/2}\sqrt{1-\varrho} \\
&\quad - \Big(w_{0,0}\frac{p(p-1)(p-2)}{12}\kappa^4 + w_{0,0}\frac{p}{(p+1)}\kappa^2 \\
&\quad - w_{0,0}\frac{2p(p-1)}{p+1}\beta\kappa\Big)\varrho^{(p+5)/2}\sqrt{1-\varrho} \\
&\quad + \mathcal{O}(\varrho^{5/2})\,.
\end{aligned} \tag{101}$$

In the intermediate step (100) the factors $p-1$, $p-3$ occur in the denominators of some terms, which would necessitate the exclusion of $p=1$ and $p=3$. However, we see in (99) that the coefficient $\binom{p}{4}$ in front of I_{p-4} vanishes for $p=1$ and $p=3$, and similarly $\binom{p}{2}$ in front of I_{p-2} vanishes for $p=1$, thus sparing the expansion of the respective integrals via (64) and (65). With this consideration, (101) can be obtained also in these cases.

A.1.7 Evaluation of the Outer Integral II

We turn now to evaluating $G_\mp$. After expanding ω in the penaliser function and cancelling terms due to $\nu = \kappa + \mathcal{O}(\varrho^2)$ we substitute $\xi = \sqrt{\varrho}\,\zeta$. Using furthermore the Taylor expansion of ω in ξ direction around $\nu\varrho$,

$$\omega(\nu\varrho+\xi) = \kappa\varrho + (1+2\,\beta\nu\varrho^2)\xi + \beta\varrho\xi^2 + \mathcal{O}(\varrho^3\xi)\,, \tag{102}$$

we obtain

$$\begin{aligned}
G_\mp &= \int_0^{\sqrt{\varrho}\pm\nu\varrho} W(\mp\xi+\nu\varrho)\big(\mp\omega(\mp\xi+\nu\varrho)\pm\kappa\varrho\big)^p\,\mathrm{d}\xi \\
&= \int_0^{\sqrt{\varrho}\pm\nu\varrho} W(\mp\xi+\nu\varrho)\big(\xi\mp\nu\varrho\mp\beta\xi^2\varrho+2\beta\xi\nu\varrho^2 \\
&\quad + \mathcal{O}(\varrho^3\xi)\pm\kappa\varrho\big)^p\,\mathrm{d}\xi \\
&= \int_0^{\sqrt{\varrho}\pm\nu\varrho} W(\mp\xi+\nu\varrho)\big(\xi\mp\beta\xi^2\varrho+2\beta\xi\nu\varrho^2 \\
&\quad + \mathcal{O}(\varrho^3\xi)\big)^p\,\mathrm{d}\xi \\
&= \sqrt{\varrho}\int_0^{1\pm\nu\sqrt{\varrho}} W\big(\mp\zeta\sqrt{\varrho}+\nu\varrho\big)\Big(\zeta\sqrt{\varrho}\mp\beta\zeta^2\varrho^2
\end{aligned}$$

$$
\begin{aligned}
&+2\beta\zeta\nu\varrho^{5/2}+\mathcal{O}(\varrho^{7/2}\zeta)\Big)^p\,\mathrm{d}\zeta\\
&=\varrho^{(p+1)/2}\int_0^{1\pm\nu\sqrt{\varrho}} W\big(\mp\zeta\sqrt{\varrho}+\nu\varrho\big)\\
&\quad\times\Big(1\mp\beta\zeta\varrho^{3/2}+\mathcal{O}(\varrho^2)\Big)^p\,\zeta^p\,\mathrm{d}\zeta\\
&\overset{(91)}{=}\varrho^{(p+1)/2}\int_0^{1\pm\nu\sqrt{\varrho}}\Big(\big(w_{0,0}\mp w_1\varrho^{3/2}\zeta\big)\\
&\quad\times\sqrt{1-(\zeta\sqrt{\varrho}\mp\nu\varrho)^2}+\mathcal{O}(\varrho^2)\Big)\\
&\quad\times\Big(1\mp\beta\zeta\varrho^{3/2}+\mathcal{O}(\varrho^2)\Big)^p\,\zeta^p\,\mathrm{d}\zeta\\
&\overset{(92)}{=}\varrho^{(p+1)/2}\int_0^{1\pm\nu\sqrt{\varrho}}\Big(2\big(1\mp\delta\varrho^{3/2}\zeta\big)\\
&\quad\times\Big(1-\frac{1}{2}\zeta^2\varrho\pm\zeta\nu\varrho^{3/2}\Big)\\
&\quad\times\Big(1\mp p\beta\zeta\varrho^{3/2}\Big)+\mathcal{O}(\varrho^2)\Big)\zeta^p\,\mathrm{d}\zeta\\
&=\varrho^{(p+1)/2}\int_0^{1\pm\nu\sqrt{\varrho}}2\Big(1\mp\delta\varrho^{3/2}\zeta-\frac{1}{2}\zeta^2\varrho\pm\zeta\nu\varrho^{3/2}\\
&\quad\mp p\beta\zeta\varrho^{3/2}\Big)\zeta^p\,\mathrm{d}\zeta+\mathcal{O}(\varrho^{(p+5)/2})\\
&=2\varrho^{(p+1)/2}\Bigg(\int_0^{1\pm\nu\sqrt{\varrho}}\zeta^p\,\mathrm{d}\zeta\\
&\quad\mp(\delta+p\beta)\varrho^{3/2}\int_0^{1\pm\nu\sqrt{\varrho}}\zeta^{p+1}\,\mathrm{d}\zeta\\
&\quad-\frac{1}{2}\varrho\int_0^{1\pm\nu\sqrt{\varrho}}\zeta^{p+2}\,\mathrm{d}\zeta\Bigg)+\mathcal{O}(\varrho^{(p+5)/2})\\
&=2\varrho^{(p+1)/2}\Bigg(\frac{1}{p+1}(1\pm\nu\sqrt{\varrho})^{p+1}\\
&\quad\mp\frac{1}{p+2}(\delta+p\beta)\varrho^{3/2}(1\pm\nu\sqrt{\varrho})^{p+2}\\
&\quad-\frac{1}{2(p+3)}\varrho(1\pm\nu\sqrt{\varrho})^{p+3}\Bigg)+\mathcal{O}(\varrho^{(p+5)/2})\\
&=2\varrho^{(p+1)/2}\Bigg(\frac{1}{p+1}\Big(1\pm(p+1)\nu\sqrt{\varrho}\\
&\quad+\binom{p+1}{2}\nu^2\varrho\pm\binom{p+1}{3}\nu^3\varrho^{3/2}\Big)\\
&\quad\mp\frac{1}{p+2}(\delta+p\beta)\varrho^{3/2}-\frac{1}{2(p+3)}\varrho\\
&\quad\pm(p+3)\nu\varrho^{3/2}\Bigg)+\mathcal{O}(\varrho^{(p+5)/2})
\end{aligned}
\tag{103}
$$

$$
\begin{aligned}
G_-+G_+&=2\varrho^{(p+1)/2}\Big(\frac{2}{p+1}+p\nu^2\varrho-\frac{1}{p+3}\varrho\Big)\\
&\quad+\mathcal{O}(\varrho^{(p+5)/2})\,.
\end{aligned}
\tag{104}
$$

A.1.8 Extremum of the Combined Integral

Combining (93), (101) and (104), applying (92) and $\nu=\kappa+\mathcal{O}(\varrho^2)$ we obtain

$$
\begin{aligned}
E(\kappa)&=\frac{4}{p+1}S_{p+2}+\Big(2\frac{(p+2)p}{p+1}\kappa^2-2\frac{2p(p-1)}{p+1}\beta\kappa\\
&\quad+\Big(\frac{4}{3}\beta\delta+\frac{2}{3}\gamma^2-\frac{2}{3}\varepsilon_2+2\psi_0\Big)\frac{1}{p+1}\\
&\quad-2\delta\frac{p}{p+1}\kappa+2\frac{2p}{p+4}\varepsilon_0+2\frac{p(p-1)}{p+4}\beta^2\\
&\quad+\Big(-2\gamma^2-\frac{4}{3}\beta\delta-\frac{2}{3}\gamma^2+\frac{2}{3}\varepsilon_2+2\psi_2\Big)\frac{2}{p+4}\\
&\quad+2\delta\frac{2p}{p+4}\beta\Big)\varrho^2S_{p+2}\\
&\quad-\frac{4}{p+1}\varrho^{(p+1)/2}\sqrt{1-\varrho}-2p\kappa^2\varrho^{(p+3)/2}\sqrt{1-\varrho}\\
&\quad+2\varrho^{(p+1)/2}\Big(\frac{2}{p+1}+p\kappa^2\varrho-\frac{1}{p+3}\varrho\Big)\\
&\quad+\mathcal{O}(\varrho^{5/2})+\mathcal{O}(\varrho^{(p+5)/2})\\
&=\mathrm{const}(\kappa)\\
&\quad+\Big(-\frac{-4p(p-1)}{p+1}\beta\varrho^2S_{p+2}-\frac{2p}{p+1}\delta\varrho^2S_{p+2}\Big)\kappa\\
&\quad+\Big(\frac{2(p+2)p}{p+1}\varrho^2S_{p+2}-2p\varrho^{(p+3)/2}\sqrt{1-\varrho}\\
&\quad+2p\varrho^{(p+3)/2}\Big)\kappa^2+\mathcal{O}\big(\varrho^{\min\{(p+5)/2,5/2\}}\big)\\
&=\mathrm{const}(\kappa)+\mathcal{O}\big(\varrho^{\min\{(p+5)/2,5/2\}}\big)\\
&\quad+\Big(-\frac{-4p(p-1)}{p+1}\beta\varrho^2S_{p+2}-\frac{2p}{p+1}\delta\varrho^2S_{p+2}\Big)\kappa\\
&\quad+\Big(\frac{2(p+2)p}{p+1}\varrho^2S_{p+2}\Big)\kappa^2\,.
\end{aligned}
\tag{105}
$$

It is worth noting that the $\varrho^{(p+3)/2}$ contributions cancelling in the last step belong to the integration boundaries of $F_\mp$ and $G_\mp$ at $\pm\sqrt{\varrho}$.

For $\varrho\to0$, the last expression (105) is a quadratic function of κ with its apex at

$$
\begin{aligned}
\kappa&=-\frac{-\frac{-4p(p-1)}{p+1}\beta\varrho^2S_{p+2}-\frac{2p}{p+1}\delta\varrho^2S_{p+2}}{2\frac{2(p+2)p}{p+1}\varrho^2S_{p+2}}\\
&\quad+\mathcal{O}(\varrho^{\min\{(p+1)/2,1/2\}})\\
&=\frac{p-1}{p+2}\beta+\frac{1}{p+2}\delta+\mathcal{O}(\varrho^{\min\{(p+1)/2,1/2\}})\,.
\end{aligned}
\tag{106}
$$

Due to the sign of the κ^2 coefficient in (105) the apex is a minimum for $p>0$ and a maximum for $p<0$. The $\mathrm{sgn}\,(p)$ factor in the original energy function E_0 compensates for this such that E_0 is always minimised.

A.1.9 Conclusion for Regular Points

From (106) the claim of the proposition for regular points follows by substituting back $\kappa\alpha\varrho^2 = \mu$, $\alpha\beta = u_{xx}/2$, $\alpha\delta = u_{yy}/2$, and noticing that by our ansatz $u_x > 0$, $u_y = 0$ the coordinates x, y coincide with the geometric coordinates η, ξ as used in the proposition.

A.1.10 Critical Points

The inequalities for local minima (maxima) are obvious consequences of the fact that for any $\varrho > 0$ the mean-p filter value is in the convex hull of values $u(\boldsymbol{x})$, $\boldsymbol{x} \in \mathrm{D}_\varrho(\boldsymbol{x}_0)$. □

A.2 Proof of Proposition 2

With the same substitutions as in the previous proof, the mode of ω is given by the maximiser of $V(\xi)$. By a slight modification of the calculations of the previous proof one finds

$$V(\xi) = 2\,(1 + \delta\xi\varrho - 2\beta\xi\varrho)\sqrt{1-\xi^2} + \mathcal{O}(\xi^2\varrho^2)\,. \tag{107}$$

Equating $V'(\xi)$ to zero yields $\omega(\xi) = (\delta - 2\beta)\varrho + \mathcal{O}(\varrho^2)$ for the mode. For local minima (maxima), the same reasoning as in the previous proof applies. □

A.3 Proof of Proposition 3

We proceed largely analogous to the proof of Proposition 1 in Appendix A.1. However, the integral decomposition gets simpler since no infinite ascents of the weighting at the integral boundaries ± 1 need to be controlled.

A.3.1 Regular Points: Ansatz via Taylor Expansion

Let the signal u and mean order p be given as in the proposition. Assume w.l.o.g. that the regular location x_0 is 0 with $u(0) = 0$, and that the derivative of u at 0 is positive, $u_x(0) > 0$. Let a neighbourhood radius ϱ be given.

Using Taylor expansion of u up to third order, we obtain for $-\varrho \le x \le \varrho$ the following expression:

$$u(x) = \alpha(x + \beta x^2 + \varepsilon x^3) + \mathcal{O}(x^4) \tag{108}$$

where $\alpha = u_x$, $2\beta = u_{xx}/u_x$.

We assume that ϱ is chosen small enough so u_x is positive throughout $[-\varrho, \varrho]$, i.e. u is strictly monotonic within this interval. The order-p mean of u within $[-\varrho, \varrho]$ is the minimiser of

$$E_0(\mu) := \operatorname{sgn}(p) \int_{\varrho}^{\varrho} |u(x) - \mu|^p \,\mathrm{d}x\,. \tag{109}$$

By rough estimates one can again conclude that for $\varrho \to 0$, $\mu \sim \varrho^2$. We substitute therefore

$$x = \varrho\xi\,, \quad \mu = \varrho^2\alpha\kappa\,, \quad u(x) = \varrho\alpha\omega(\xi) \tag{110}$$

and obtain

$$E_0(\mu) = \operatorname{sgn}(p)\varrho^{p+1}\alpha^p E(\kappa)\,, \tag{111}$$

$$E(\kappa) = \int_{\varrho}^{\varrho} |\omega - \kappa\varrho|^p \,\mathrm{d}\xi\,, \tag{112}$$

$$\omega(\xi) = \xi + \beta\xi^2\varrho + \varepsilon\xi^3\varrho^2 + \mathcal{O}(\varrho^3\xi)\,. \tag{113}$$

In the following we focus therefore on finding the extremum of E (minimum for $p > 0$, maximum for $p < 0$).

A.3.2 Domain Splitting of the Integral

We split the integral (112) into two parts, using again the location $\xi = \nu\varrho$ where $\omega(\nu\varrho) = \kappa\varrho$ as splitting point. By (113), one has $\nu = \kappa + \mathcal{O}(\varrho^2)$. We have then

$$E(\kappa) = F_-(\kappa) + F_+(\kappa)\,, \tag{114}$$

$$F_-(\kappa) = \int_{-1}^{\nu\varrho} \big(\kappa\varrho - \omega(\xi)\big)^p \,\mathrm{d}\xi\,, \tag{115}$$

$$F_+(\kappa) = \int_{\nu\varrho}^{1} \big(\omega(\xi) - \kappa\varrho\big)^p \,\mathrm{d}\xi\,. \tag{116}$$

By substituting the integration variables, one obtains

$$F_\mp = \int_0^{1\pm\nu\varrho} \big(\mp\omega(\mp\xi + \nu\varrho) \pm \kappa\varrho\big)^p \,\mathrm{d}\xi \tag{117}$$

where again the upper and lower signs refer to F_- and F_+, respectively.

A.3.3 Evaluation of the Integrals

The Taylor expansion for ω around $\nu\varrho$ is identical with (102). Inserting this into (117), we have further

$$\begin{aligned}
F_\mp &= \int_0^{1\pm\nu\varrho} \big((1 + 2\,\beta\nu\varrho^2)\xi \mp \beta\varrho\xi^2 + \mathcal{O}(\varrho^3\xi)\big)^p \,\mathrm{d}\xi \\
&= \int_0^{1\pm\nu\varrho} \xi^p\big(1 + 2\,\beta\nu\varrho^2 \mp \beta\varrho\xi + \mathcal{O}(\varrho^3)\big)^p \,\mathrm{d}\xi \\
&= \int_0^{1\pm\nu\varrho} \xi^p\left(1 + 2\,p\beta\nu\varrho^2 \mp p\beta\varrho\xi + \frac{p(p-1)}{2}\beta^2\varrho^2\xi^2 \right. \\
&\qquad \left. + \mathcal{O}(\varrho^3)\right) \mathrm{d}\xi \\
&= \big(1 + 2\,p\beta\nu\varrho^2 + \mathcal{O}(\varrho^3)\big) \int_0^{1\pm\nu\varrho} \xi^p \,\mathrm{d}\xi
\end{aligned}$$

$$\mp p\beta\varrho \int_0^{1\pm\nu\varrho} \xi^{p+1}\,\mathrm{d}\xi + \frac{p(p-1)}{2}\beta^2\varrho^2 \int_0^{1\pm\nu\varrho} \xi^{p+2}\,\mathrm{d}\xi\ , \tag{118}$$

from which by evaluating the standard integrals, adding F_- and F_+ and inserting $\nu = \kappa + \mathcal{O}(\varrho^2)$ we reach

$$E(\kappa) = F_-(\kappa) + F_+(\kappa) = \mathrm{const}(\kappa) + p\varrho^2\left(\kappa^2 - 2\frac{p-1}{p+1}\beta\kappa\right) + \mathcal{O}(\varrho^3)\ . \tag{119}$$

The extremum of E is again found as the apex of the quadratic function on the r.h.s., from which the claim for regular points follows.

For critical points, the reasoning from Appendix A.1 applies. □

A.4 Proof of Proposition 4

Assuming again that the regular location for the signal u is $x_0 = 0$, and u is strictly monotonically increasing and Lipschitz within $[-\varrho, \varrho]$, the density of each value $u(x)$ for $-\varrho \le x \le \varrho$ is proportional to $1/u'(x)$. The maximum of these values is reached at $u(-\varrho)$ if u is convex, or $u(\varrho)$ if u is concave. This proves the claim for regular points. If x_0 is a local extremum, the density has a pole at $u(x_0)$ and is finite for all other values, making $u(x_0)$ the mode. □

A.5 Proof of Proposition 5

A.5.1 Regular Points: Ansatz via Taylor Expansion

Let the volume image u and mean order p be given as in the proposition. Assume w.l.o.g. that the regular location $\boldsymbol{x}_0$ is $(0,0,0)$ with $u(0,0,0) = 0$, and that the gradient of u at $(0,0,0)$ is in the positive x direction, i.e. $u_x > 0, u_y = u_z = 0$. Let a neighbourhood radius $\varrho > 0$ be given, and denote the closed (Euclidean) ϱ-neighbourhood of $(0,0,0)$ by B_ϱ.

Using Taylor expansion of u up to third order, we can write for $(x,y,z) \in \mathrm{B}_\varrho$ the ansatz

$$\begin{aligned} u(x,y,z) = \alpha\big(&x + \beta x^2 + \gamma_0 xy + \gamma_1 xz + \delta_0 y^2 + \delta_1 yz + \delta_2 z^2 \\ &+ \varepsilon_0 x^3 + \varepsilon_{10} x^2 y + \varepsilon_{01} x^2 z + \varepsilon_{20} xy^2 + \varepsilon_{11} xyz \\ &+ \varepsilon_{02} xz^2 + \varepsilon_{30} y^3 + \varepsilon_{21} y^2 z + \varepsilon_{12} yz^2 + \varepsilon_{03} z^3\big) \\ &+ \mathcal{O}\big((x+y+z)^4\big)\ . \end{aligned} \tag{120}$$

We assume that ϱ is chosen small enough such that u_x is positive everywhere in B_ϱ, each level set of u within the ball B_ϱ is either a smooth surface patch bounded by a closed regular curve on the boundary of the ball, or one of two single points on the boundary of B_ϱ where u takes its maximum and minimum on B_ϱ, respectively.

The order-p mean of u within B_ϱ is the minimiser of

$$E_0(\mu) := \mathrm{sgn}\,(p) \iiint_{\mathrm{B}_\varrho} |u(x,y,z) - \mu|^p\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x\ . \tag{121}$$

Rough estimates again ensure $\mu \sim \varrho^2$ for $\varrho \to 0$. Combining an appropriate rescaling with a transition to cylindrical coordinates with the axis in gradient (x) direction, we substitute

$$x = \varrho\xi\ ,\quad y = \varrho\eta\cos\varphi\ ,\quad z = \varrho\eta\sin\varphi\ ,\quad \mu = \varrho^2\alpha\kappa\ , \tag{122}$$

$$u(x,y,z) = \varrho\alpha\omega(\xi,\eta,\varphi) \tag{123}$$

and obtain

$$E_0(\mu) = \mathrm{sgn}\,(p)\varrho^{p+3}\alpha^p E(\kappa)\ , \tag{124}$$

$$E(\kappa) = \iint_{\mathrm{D}_1}\int_0^{2\pi} |\omega(\xi,\eta,\varphi) - \kappa\varrho|^p\,\eta\,\mathrm{d}\varphi\,\mathrm{d}\eta\,\mathrm{d}\xi\ , \tag{125}$$

where the integration in cylindrical coordinates has been written using the disc D_1 for the ξ,η coordinates. The Taylor expansion of u transfers to

$$\begin{aligned} \omega(\xi,\eta,\varphi) = \xi &+ \beta\xi^2\varrho + \gamma(\varphi)\xi\eta\varrho + \delta(\varphi)\eta^2\varrho \\ &+ \varepsilon_0\xi^3\varrho^2 + \varepsilon_1(\varphi)\xi^2\eta\varrho^2 + \varepsilon_2(\varphi)\xi\eta^2\varrho^2 \\ &+ \varepsilon_3(\varphi)\eta^3\varrho^2 + \mathcal{O}\big(\varrho^3(\xi+\eta)\big)\ , \end{aligned} \tag{126}$$

$$\gamma(\varphi) := \gamma_0\cos\varphi + \gamma_1\sin\varphi\ , \tag{127}$$

$$\delta(\varphi) := \delta_0\cos^2\varphi + \delta_1\cos\varphi\sin\varphi + \delta_2\sin^2\varphi\ , \tag{128}$$

$$\varepsilon_1(\varphi) := \varepsilon_{10}\cos\varphi + \varepsilon_{01}\sin\varphi\ , \tag{129}$$

$$\varepsilon_2(\varphi) := \varepsilon_{20}\cos^2\varphi + \varepsilon_{11}\cos\varphi\sin\varphi + \varepsilon_{02}\sin^2\varphi\ , \tag{130}$$

$$\begin{aligned} \varepsilon_3(\varphi) := \varepsilon_{30}&\cos^3\varphi + \varepsilon_{21}\cos^2\varphi\sin\varphi \\ &+ \varepsilon_{12}\cos\varphi\sin^2\varphi + \varepsilon_{03}\sin^3\varphi\ . \end{aligned} \tag{131}$$

We aim again at finding the extremum of E.

A.5.2 Separation of the Integral

Similar to Appendix A.1.3, the integral E from (125) can be reorganised into a nested integration where the inner double integral (in polar coordinates) integrates over a level surface of ω going through $(\xi,0,0)$, and the outer integral then integrates along the ξ axis. We have

$$E(\kappa) = \int_{-1}^{1}\left(\int_0^{2\pi}\int_0^{\eta^*(\varphi)} \frac{\eta}{\frac{\partial\omega}{\partial\xi}\big(\tilde{\xi}(\eta),\eta,\varphi\big)}\,\mathrm{d}\eta\,\mathrm{d}\varphi\right)$$

$$\times\, |\omega(\xi,0,0) - \kappa\varrho|^p \frac{\partial\omega}{\partial\xi}(\xi,0,0)\,\mathrm{d}\xi + \mathcal{O}(\varrho^3) \tag{132}$$

where $\tilde{\xi}$ is a function of η, φ that describes the level set of ω which goes through $(\xi,0,0)$, and reaches the boundary of B_1 at $(\eta^*(\varphi),\varphi)$. (Note that our initial assumptions on u ensure that the level set can be described in this way.)

Analogously to Appendix A.1.3 we rewrite (132) as

$$E(\kappa) = \int_{-1}^{1} W(\xi)\,\mathrm{d}\xi\ , \tag{133}$$

$$W(\xi) := \int_0^{2\pi}\int_0^{\eta^*(\varphi)} \eta\, \frac{\frac{\partial\omega}{\partial\xi}(\xi,0,0)}{\frac{\partial\omega}{\partial\xi}\big(\tilde{\xi}(\eta),\eta,\varphi\big)}\,\mathrm{d}\eta\,\mathrm{d}\varphi\ . \tag{134}$$

A.5.3 Evaluation of the Weight Integral

Within any axial plane ($\varphi = \text{const}$), (126) is exactly (73). We can therefore transfer verbatim the analysis from Appendix A.1.4, which leads to the expression (79) for $\tilde{\xi}$, the expression for η_+^* from (80) for $\eta^*(\varphi)$, and (86) for $\frac{\partial\omega}{\partial\xi}(\xi,0,0)/\frac{\partial\omega}{\partial\xi}(\tilde{\xi},\eta,\varphi)$.

Inserting (86) into the inner integral of (134) leads to

$$\begin{aligned}
W(\xi,\varphi) &:= \int_0^{\eta^*(\varphi)} \eta\, \frac{\frac{\partial\omega}{\partial\xi}(\xi,0,0)}{\frac{\partial\omega}{\partial\xi}\big(\tilde{\xi}(\eta),\eta,\varphi\big)}\,\mathrm{d}\eta \\
&= \int_0^{\eta^*(\varphi)} \eta\,\mathrm{d}\eta \\
&\quad + \big(-\gamma(\varphi)\varrho + 2\beta\gamma(\varphi)\xi\varrho^2 - 2\varepsilon_1(\varphi)\xi\varrho^2\big)\int_0^{\eta^*} \eta^2\,\mathrm{d}\eta \\
&\quad + \big(2\beta\delta(\varphi) + \gamma(\varphi)^2 - \varepsilon_2(\varphi)\big)\varrho^2\int_0^{\eta^*} \eta^3\,\mathrm{d}\eta \\
&\quad + \mathcal{O}(\varrho^3) \\
&= \frac{1}{2}\eta^*(\varphi)^2 \\
&\quad + \frac{1}{6}(-\gamma_0 + 2\beta\gamma_0\xi\varrho - 2\varepsilon_{10}\xi\varrho)\varrho\eta^*(\varphi)^3\cos\varphi \\
&\quad + \frac{1}{6}(-\gamma_1 + 2\beta\gamma_1\xi\varrho - 2\varepsilon_{01}\xi\varrho)\varrho\eta^*(\varphi)^3\sin\varphi \\
&\quad + \frac{1}{4}(2\beta\delta_0 + \gamma_0^2 - \varepsilon_{20})\varrho^2\eta^*(\varphi)^4\cos^2\varphi \\
&\quad + \frac{1}{2}(2\beta\delta_1 + \gamma_0\gamma_1 - \varepsilon_{11})\varrho^2\eta^*(\varphi)^4\cos\varphi\sin\varphi \\
&\quad + \frac{1}{4}(2\beta\delta_2 + \gamma_1^2 - \varepsilon_{02})\varrho^2\eta^*(\varphi)^4\sin^2\varphi \\
&\quad + \mathcal{O}(\varrho^3)\ .
\end{aligned} \tag{135}$$

To finally obtain $W(\xi)$, the latter expression needs to be integrated over φ. From (80) one obtains by lengthy but straightforward calculation

$$\begin{aligned}
&\int_0^{2\pi} \eta^*(\varphi)^2\,\mathrm{d}\varphi \\
&= \pi\Big(2(1-\xi^2) + 2(\delta_0+\delta_2)(1-\xi^2)\xi\varrho + (\gamma_0^2+\gamma_1^2)\xi^4\varrho^2 \\
&\quad + \tfrac{1}{4}(3\delta_0^2 + \delta_1^2 + 3\delta_2^2 + 2\delta_0\delta_2) \\
&\quad + 2\Psi(\xi)(1-\xi^2)\varrho^2\Big) + \mathcal{O}(\varrho^3)\ ,
\end{aligned} \tag{136}$$

where Ψ is a third-order polynomial in ξ obtained by integrating (82) w.r.t. φ,

$$\Psi(\xi) = \frac{1}{\pi}\int_0^{2\pi} \psi(\xi,\varphi)\,\mathrm{d}\varphi = \Psi_0 + \Psi_1\xi + \Psi_2\xi^2 + \Psi_3\xi^3\ . \tag{137}$$

Analogously one obtains

$$\int_0^{2\pi} \eta^*(\varphi)^3\cos\varphi\,\mathrm{d}\varphi = 3\pi\gamma_0(1-\xi^2)\xi^2\varrho + \mathcal{O}(\varrho^2) \tag{138}$$

$$\int_0^{2\pi} \eta^*(\varphi)^3\sin\varphi\,\mathrm{d}\varphi = 3\pi\gamma_1(1-\xi^2)\xi^2\varrho + \mathcal{O}(\varrho^2)\ , \tag{139}$$

$$\int_0^{2\pi} \eta^*(\varphi)^4\cos^2\varphi\,\mathrm{d}\varphi = \pi(1-\xi^2)^2 + \mathcal{O}(\varrho)\ , \tag{140}$$

$$\int_0^{2\pi} \eta^*(\varphi)^4\cos\varphi\sin\varphi\,\mathrm{d}\varphi = 0 + \mathcal{O}(\varrho)\ , \tag{141}$$

$$\int_0^{2\pi} \eta^*(\varphi)^4\sin^2\varphi\,\mathrm{d}\varphi = \pi(1-\xi^2)^2 + \mathcal{O}(\varrho)\ . \tag{142}$$

Inserting (135)–(142) into (134) yields after sorting terms, similarly to (91), (92),

$$\begin{aligned}
W(\xi) = \Big((w_{0,0} + w_{0,2}\varrho^2) + (w_{1,1} + w_{1,2}\varrho)\varrho\xi \\
+ (w_{2,0} + w_{2,2}\varrho^2)\xi^2 + (w_{3,1} + w_{3,2}\varrho)\varrho\xi^3 \\
+ w_{4,2}\varrho^2\xi^4 + w_{5,2}\varrho^2\xi^5\Big)\pi + \mathcal{O}(\varrho^3)
\end{aligned} \tag{143}$$

with

$$\left.\begin{aligned}
w_{0,0} &= 1 \,, \\
w_{0,2} &= \tfrac{1}{2}\beta(\delta_0+\delta_2) + \tfrac{1}{4}(\gamma_0^2+\gamma_1^2) \\
&\quad - \tfrac{1}{4}(\varepsilon_{20}+\varepsilon_{02}) + \Psi_0 \,, \\
w_{1,1} &= \delta_0+\delta_2 \,, \\
w_{1,2} &= \Psi_1 \,, \\
w_{2,0} &= -1 \,, \\
w_{2,2} &= -\beta(\delta_0+\delta_2) - \tfrac{3}{2}(\gamma_0^2+\gamma_1^2) + \tfrac{1}{2}(\varepsilon_{20}+\varepsilon_{02}) \\
&\quad + \tfrac{1}{8}(3\delta_0^2+\delta_1^2+3\delta_2^2+2\delta_0\delta_2) + \Psi_2 - \Psi_0 \,, \\
w_{3,1} &= -(\delta_0+\delta_2) \,, \\
w_{3,2} &= \Psi_3 - \Psi_1 \,, \\
w_{4,2} &= \tfrac{1}{2}\beta(\delta_0+\delta_2) + \tfrac{9}{4}(\gamma_0^2+\gamma_1^2) - \tfrac{1}{4}(\varepsilon_{20}+\varepsilon_{02}) \\
&\quad - \tfrac{1}{8}(3\delta_0^2+\delta_1^2+3\delta_2^2+2\delta_0\delta_2) - \Psi_2 \,, \\
w_{5,2} &= -\Psi_3 \,.
\end{aligned}\right\} \tag{144}$$

A.5.4 Domain Splitting of the Outer Integral

As the outer integral of (132) has the same structure as in the 2D case, we use the same domain splitting (93).

A.5.5 Evaluation of the Outer Integral I

The first steps in evaluating the integrals $F_\mp$ are as in the 2D case. In (98), the longer expansion (143) has to be used for $W(\mp)$, which then leads to

$$\begin{aligned}
F_\mp = \int_{\sqrt{\varrho}}^1 &\Big((w_{0,0}+w_{0,2}\varrho^2) + (w_{1,1}+w_{1,2}\varrho)\varrho\xi \\
&+ (w_{2,0}+w_{2,2}\varrho^2)\xi^2 + (w_{3,1}+w_{3,2}\varrho)\varrho\xi^3 \\
&+ w_{4,2}\varrho^2\xi^4 + w_{5,2}\varrho^2\xi^5 + \mathcal{O}(\varrho^3)\Big)\pi\xi^p \\
&\times \left(1 \pm p\kappa\frac{\varrho}{\xi} \mp p\beta\xi\varrho + \binom{p}{2}\kappa^2\frac{\varrho^2}{\xi^2} \mp \binom{p}{3}\kappa^3\frac{\varrho^3}{\xi^3} \right. \\
&+ p\varepsilon_0\xi^2\varrho^2 - 2\binom{p}{2}\beta\kappa\varrho^2 + \binom{p}{2}\beta^2\xi^2\varrho^2 \\
&\left. + \binom{p}{4}\kappa^4\frac{\varrho^4}{\xi^4} + \mathcal{O}(\varrho^{5/2})\right) \mathrm{d}\xi \,,
\end{aligned} \tag{145}$$

which yields

$$\begin{aligned}
F_- + F_+ = 2\pi &\int_{\sqrt{\varrho}}^1 (w_{0,0}+w_{2,0}\xi^2)\xi^p \left(1 + \binom{p}{2}\kappa^2\frac{\varrho^2}{\xi^2} \right. \\
&+ p\varepsilon_0\xi^2\varrho^2 - 2\binom{p}{2}\beta\kappa\varrho^2 + \binom{p}{2}\beta^2\xi^2\varrho^2 \\
&\left. + \binom{p}{4}\kappa^4\frac{\varrho^4}{\xi^4} + \mathcal{O}(\varrho^{5/2})\right) \mathrm{d}\xi \\
&+ 2\pi \int_{\sqrt{\varrho}}^1 (w_{0,2}+w_{2,2}\xi^2+w_{4,2}\xi^4)\varrho^2\xi^p \\
&\times \left(1 + \mathcal{O}(\varrho^{1/2})\right) \mathrm{d}\xi \\
&+ 2\pi \int_{\sqrt{\varrho}}^1 (w_{1,1}\xi + w_{3,1}\xi^3)\varrho\xi^p \\
&\times \left(-p\kappa\frac{\varrho}{\xi} + p\beta\xi\varrho + \mathcal{O}(\varrho^{3/2})\right) \mathrm{d}\xi \\
&+ 2\pi \int_{\sqrt{\varrho}}^1 (w_{1,2}\xi + w_{3,2}\xi^3 + w_{5,2}\xi^5)\varrho^2\xi^p \\
&\times \left(-p\kappa\frac{\varrho}{\xi} + \mathcal{O}(\varrho^{1/2})\right) \mathrm{d}\xi \,.
\end{aligned} \tag{146}$$

Using the abbreviation

$$J_q := \int_{\sqrt{\varrho}}^1 \xi^q \,\mathrm{d}\xi \,, \tag{147}$$

we can sort this into

$$\begin{aligned}
F_- + F_+ = 2\pi &\Big(w_{0,0}\binom{p}{2}\kappa^2\varrho^2 J_{p-2} \\
&+ \Big(w_{0,0} + \Big(-2w_{0,0}\binom{p}{2}\beta\kappa + w_{2,0}\binom{p}{2}\kappa^2 \\
&+ w_{0,2} - w_{1,1}p\kappa\Big)\varrho^2\Big) J_p \\
&+ \Big(w_{2,0} + \Big(w_{0,0}p\varepsilon_0 + w_{0,0}\binom{p}{2}\beta^2 \\
&- 2w_{2,0}\binom{p}{2}\beta\kappa + w_{2,2} + w_{1,1}p\beta \\
&- w_{3,1}p\kappa\Big)\varrho^2\Big) J_{p+2} \\
&+ \Big(w_{2,0}p\varepsilon_0 + w_{2,0}\binom{p}{2}\beta^2 + w_{4,2} \\
&+ w_{3,1}p\beta\Big)\varrho^2 J_{p+4}\Big) + \mathcal{O}(\varrho^{5/2}) \,,
\end{aligned} \tag{148}$$

which by $J_q = \frac{1}{q+1}(1-\varrho^{(q+1)/2})$ for $q \neq -1$ (the special case J_{-1} only occurs as J_{p-2} for $p=1$ and has then a vanishing coefficient) yields

$$\begin{aligned}
F_- + F_+ = 2\pi &\Big(w_{0,0}\frac{1}{p+1} + w_{2,0}\frac{1}{p+3} \\
&+ \Big(w_{0,0}\frac{p}{2}\kappa^2 - 2w_{0,0}\frac{p(p-1)}{2(p+1)}\beta\kappa \\
&+ w_{2,0}\frac{p(p-1)}{2(p+1)}\kappa^2 + w_{0,2} - w_{1,1}\frac{p}{p+1}\kappa \\
&+ w_{0,0}\frac{p}{p+3}\varepsilon_0 + w_{0,0}\frac{p(p-1)}{2(p+3)}\beta^2
\end{aligned}$$

$$
\begin{aligned}
&- 2w_{2,0}\frac{p(p-1)}{2(p+3)}\beta\kappa + w_{2,2}\frac{1}{p+3}\\
&+ w_{1,1}\frac{p}{p+3}\beta - w_{3,1}\frac{p}{p+3}\kappa\\
&+ w_{2,0}\frac{p}{p+5}\varepsilon_0 + w_{2,0}\frac{p(p-1)}{2(p+5)}\beta^2\\
&+ w_{4,2}\frac{1}{p+5} + w_{3,1}\frac{p}{p+5}\beta\Big)\varrho^2\\
&- w_{0,0}\frac{1}{p+1}\varrho^{(p+1)/2}\\
&- \left(w_{0,0}\frac{p}{2}\kappa^2 + w_{2,0}\frac{1}{p+3}\right)\varrho^{(p+3)/2}\Big)\\
&+ \mathcal{O}(\varrho^{5/2}) + \mathcal{O}(\varrho^{(p+5)/2})\,. \qquad (149)
\end{aligned}
$$

A.5.6 Evaluation of the Outer Integral II

Starting with the same substitution $\xi = \sqrt{\varrho}\,\zeta$ and Taylor expansion of ω in ξ direction as in Appendix A.1.7, we evaluate

$$
\begin{aligned}
G_\mp &\overset{(143)}{=} \pi\varrho^{(p+1)/2}\int_0^{1\pm\nu\sqrt{\varrho}} \big(w_{0,0} + w_{2,0}\varrho\zeta^2 \mp w_{1,1}\varrho^{3/2}\zeta\\
&\quad \mp 2w_{2,0}\nu\varrho^{3/2}\zeta + \mathcal{O}(\varrho^2)\big)\\
&\quad \times \big(1 \mp \beta\zeta\varrho^{3/2} + \mathcal{O}(\varrho^2)\big)^p\,\zeta^p\,\mathrm{d}\zeta\\
&= \pi\varrho^{(p+1)/2}w_{0,0}\int_0^{1\pm\nu\sqrt{\varrho}}\zeta^p\,\mathrm{d}\zeta\\
&\quad \mp \pi\varrho^{(p+4)/2}(w_{1,1} + 2w_{2,0}\nu + w_{0,0}p\beta)\int_0^{1\pm\nu\sqrt{\varrho}}\zeta^{p+1}\,\mathrm{d}\zeta\\
&\quad + \pi\varrho^{(p+3)/2}w_{2,0}\int_0^{1\pm\nu\sqrt{\varrho}}\zeta^{p+2}\,\mathrm{d}\zeta + \mathcal{O}(\varrho^{(p+5)/2})\\
&= \frac{\pi\varrho^{(p+1)/2}}{p+1}w_{0,0}(1\pm\nu\sqrt{\varrho})^{p+1}\\
&\quad \mp \frac{\pi\varrho^{(p+4)/2}}{p+2}(w_{1,1} + 2w_{2,0}\nu + w_{0,0}p\beta)(1\pm\nu\sqrt{\varrho})^{p+2}\\
&\quad + \frac{\pi\varrho^{(p+3)/2}}{p+3}w_{2,0}(1\pm\nu\sqrt{\varrho})^{p+3} + \mathcal{O}(\varrho^{(p+5)/2})\\
&= \frac{\pi\varrho^{(p+1)/2}}{p+1}w_{0,0}(1+\nu^2\varrho)\\
&\quad \mp \frac{\pi\varrho^{(p+4)/2}}{p+2}(w_{1,1} + 2w_{2,0}\nu + w_{0,0}p\beta)\\
&\quad + \frac{\pi\varrho^{(p+3)/2}}{p+3}w_{2,0} + \mathcal{O}(\varrho^{(p+5)/2})\,, \qquad (150)
\end{aligned}
$$

$$
\begin{aligned}
G_- + G_+ &= 2\pi\varrho^{(p+1)/2}\left(\frac{1}{p+1}w_{0,0} + \frac{p}{2}\nu^2\varrho w_{0,0}\right.\\
&\quad \left. + \frac{1}{p+3}\varrho w_{2,0}\right) + \mathcal{O}(\varrho^{(p+5)/2})\,. \qquad (151)
\end{aligned}
$$

A.5.7 Extremum of the Combined Integral

When we finally combine (93), (149) and (151) and apply (144) and $\nu = \kappa + \mathcal{O}(\varrho^2)$, we observe as in the 2D case than all terms originating from $G_- + G_+$ (151) cancel, and it remains

$$
\begin{aligned}
E(\kappa) &= \mathrm{const}(\kappa) + \mathcal{O}\big(\varrho^{\min\{(p+5)/2,5/2\}}\big)\\
&\quad + \left(p - \frac{p(p-1)}{p+1}\right)\pi\varrho^2\kappa^2\\
&\quad + \left(-\frac{2p(p-1)}{(p+1)}\beta + \frac{2p(p-1)}{(p+3)}\beta\right.\\
&\quad \left. - \frac{2p}{p+1}(\delta_0+\delta_2) + \frac{2p}{p+3}(\delta_0+\delta_2)\right)\pi\varrho^2\kappa\\
&= \mathrm{const}(\kappa) + \mathcal{O}\big(\varrho^{\min\{(p+5)/2,5/2\}}\big)\\
&\quad + \frac{2p}{p+1}\pi\varrho^2\kappa^2 - \left(\frac{4p(p-1)}{(p+1)(p+3)}\beta\right.\\
&\quad \left. + \frac{4p}{(p+1)(p+3)}(\delta_0+\delta_2)\right)\pi\varrho^2\kappa\,. \qquad (152)
\end{aligned}
$$

For $\varrho \to 0$, the extremum of $E(\kappa)$ can again be found as the apex of the quadratic function in (152), which yields

$$
\begin{aligned}
\kappa &= \frac{\frac{4p(p-1)}{(p+1)(p+3)}\beta + \frac{4p}{(p+1)(p+3)}(\delta_0+\delta_2)}{\frac{4p}{p+1}\pi\varrho^2}\\
&\quad + \mathcal{O}(\varrho^{\min\{(p+1)/2,1/2\}})\\
&= \frac{p-1}{p+3}\beta + \frac{1}{p+3}(\delta_0+\delta_2)\\
&\quad + \mathcal{O}(\varrho^{\min\{(p+1)/2,1/2\}})\,. \qquad (153)
\end{aligned}
$$

A.5.8 Conclusion of the Proof

From (153) the claim of the proposition for regular points follows by substituting back $\kappa\alpha\varrho^2 = \mu$, $\alpha\beta = u_{xx}/2$, $\alpha\delta_0 = u_{yy}/2$, $\alpha\delta_2 = u_{zz}/2$ and noticing that by our ansatz $u_x > 0$, $u_y = u_z = 0$ the coordinates x, y, z coincide with the geometric coordinates η, ξ, χ as used in the proposition.

For critical points, the reasoning from Appendix A.1 applies. □

A.6 Proof of Proposition 6

Analogous to Appendix A.2, we calculate

$$
V(\xi) = \pi(1 + (\delta_0+\delta_2)\xi\varrho - 2\beta\xi\varrho)(1-\xi^2) + \mathcal{O}(\xi^2\varrho^2)\,. \qquad (154)
$$

The relevant solution of $V'(\xi) = 0$ yields up to higher order terms $\omega(\xi) = \frac{1}{2}(\delta_0 + \delta_2 - 2\beta)\varrho$. □

B Continuous Order-*p* Means and Mode: A Toy Example

To understand the behaviour of order-p mean filters for $p > -1$, $p \neq 0$ and their relation to the mode of a continuous density, we consider the following simple example. Let z be a real random variable with (non-normalised) density

$$\gamma(z) = \begin{cases} 1 - \lambda(z-m)^2\,, & -1 \le z \le 1\,, \\ 0 & \text{otherwise.} \end{cases} \tag{155}$$

Here, $0 < m << 1$ is a fixed parameter, and $0 < \lambda \le (1+m)^{-2}$ to ensure that $\gamma(z) \ge 0$ for all z. Obviously, the mode of z is the maximum of γ, i.e. m.

For any $p > -1$, $p \neq 0$ the order-p mean of z is given by the minimiser of $E(\mu)$ where

$$\begin{aligned}
\operatorname{sgn}(p) E(\mu) &= \int_{-1}^{1} \gamma(z)\,|z-\mu|^p \,\mathrm{d}z \\
&= \int_{-1}^{\mu} \big(1-\lambda(z-m)^2\big)(-z+\mu)^p \,\mathrm{d}z \\
&\quad + \int_{\mu}^{1} \big(1-\lambda(z-m)^2\big)(z-\mu)^p \,\mathrm{d}z \\
&= \int_{0}^{1+\mu} \big(1-\lambda(z-\mu+m)^2\big)\, z^p \,\mathrm{d}z \\
&\quad + \int_{0}^{1-\mu} \big(1-\lambda(z+\mu-m)^2\big)\, z^p \,\mathrm{d}z \\
&= \big(1-\lambda(\mu-m)^2\big)\left(\int_0^{1+\mu} z^p \,\mathrm{d}z + \int_0^{1-\mu} z^p \,\mathrm{d}z\right) \\
&\quad + 2\lambda(\mu-m)\left(\int_0^{1+\mu} z^{p+1} \,\mathrm{d}z - \int_0^{1-\mu} z^{p+1} \,\mathrm{d}z\right) \\
&\quad - \lambda\left(\int_0^{1+\mu} z^{p+2} \,\mathrm{d}z + \int_0^{1-\mu} z^{p+2} \,\mathrm{d}z\right) \\
&= \frac{1}{p+1}\big(1-\lambda(\mu-m)^2\big)\Big((1+\mu)^{p+1} + (1-\mu)^{p+1}\Big) \\
&\quad + \frac{2\lambda}{p+2}(\mu-m)\Big((1+\mu)^{p+2} - (1-\mu)^{p+2}\Big) \\
&\quad - \frac{\lambda}{p+3}\Big((1+\mu)^{p+3} + (1-\mu)^{p+3}\Big) \\
&= \frac{2}{p+1}\big(1-\lambda(\mu-m)^2\big)\left(1 + \binom{p+1}{2}\mu^2 + \mathcal{O}(\mu^4)\right) \\
&\quad + \frac{4\lambda}{p+2}(\mu-m)\Big((p+2)\mu + \mathcal{O}(\mu^3)\Big) \\
&\quad - \frac{2\lambda}{p+3}\left(1 + \binom{p+3}{2}\mu^2 + \mathcal{O}(\mu^4)\right) \\
&= \frac{2}{p+1} + p\mu^2 - \frac{2\lambda}{p+1}\mu^2 + \frac{4\lambda}{p+1}m\mu - \frac{2\lambda}{p+1}m^2 \\
&\quad - p\lambda m^2\mu^2 + 4\lambda\mu^2 - 4\lambda m\mu - \frac{2\lambda}{p+3} - (p+2)\lambda\mu^2 \\
&\quad + \mathcal{O}(\mu^3) \\
&= \operatorname{const}(\mu) + \left(\frac{4\lambda}{p+1} - 4\lambda\right) m\mu \\
&\quad + \left(p - \frac{2\lambda}{p+1} - p\lambda m^2 + 4\lambda - (p+2)\lambda\right)\mu^2 + \mathcal{O}(\mu^3) \\
&= \operatorname{const}(\mu) + \frac{-4p\lambda}{p+1} m\mu \\
&\quad + \left(p - \frac{p(p-1)}{p+1}\lambda - p\lambda m^2\right)\mu^2 + \mathcal{O}(\mu^3)
\end{aligned} \tag{156}$$

from which the minimiser μ^* of $E(\mu)$ can be read off as the apex of the quadratic function of μ as

$$\begin{aligned}
\mu^* &= -\frac{-2p\lambda}{p+1}\, m \Big/ \left(p - \frac{p(p-1)}{p+1}\lambda - p\lambda m^2\right) \\
&= \frac{2\lambda}{(p+1) - (p-1)\lambda}\, m + \mathcal{O}(m^3)\,.
\end{aligned} \tag{157}$$

For any fixed $\lambda \in \big(0, (1+m)^{-2}\big)$, the minimiser μ^* goes to m for $p \to -1$, but approaches $2\lambda m/(1+\lambda) < m - m^2/3$ for $p \to 0$. Moreover, if we send λ to 0, making the density more and more uniform, for any $p > -1$ we have $\mu^* \to 0$ which comes as no surprise as for a flattening out density, any penalisation where the penaliser increases with distance will end up in the symmetry centre of the support interval $[-1, 1]$.

References

1. Alvarez, L., Guichard, F., Lions, P.L., Morel, J.M.: Axioms and fundamental equations in image processing. Arch. Ration. Mech. Anal. **123**, 199–257 (1993)
2. Alvarez, L., Lions, P.L., Morel, J.M.: Image selective smoothing and edge detection by nonlinear diffusion, II. SIAM J. Numer. Anal. **29**, 845–866 (1992)
3. Arehart, A.B., Vincent, L., Kimia, B.B.: Mathematical morphology: the Hamilton–Jacobi connection. In: Proceedings of Fourth International Conference on Computer Vision, pp. 215–219. IEEE Computer Society Press, Berlin (1993)
4. Armatte, M.: Fréchet et la médiane: un moment dans une histoire de la robustesse. J. Soc. Fr. Stat. **147**(2), 23–37 (2006)
5. Aubert, G., Kornprobst, P.: Mathematical problems in image processing: partial differential equations and the calculus of variations, Applied Mathematical Sciences, vol. 147, second edn. Springer, New York (2006)
6. Barral Souto, J.: El modo y otras medias, casos particulares de una misma expresión matemática. Report No. 3, Cuadernos de Trabajo, Instituto de Biometria, Universidad Nacional de Buenos Aires, Argentina (1938)
7. Bertero, M., Poggio, T.A., Torre, V.: Ill-posed problems in early vision. Proc. IEEE **76**(8), 869–889 (1988)
8. Brakke, K.A.: The Motion of a Surface by Its Mean Curvature. Princeton University Press, Princeton (1978)
9. Breuß, M., Weickert, J.: A shock-capturing algorithm for the differential equations of dilation and erosion. J. Math. Imaging Vis. **25**(2), 187–201 (2006)

10. Brockett, R.W., Maragos, P.: Evolution equations for continuous-scale morphological filtering. IEEE Trans. Signal Process. **42**, 3377–3386 (1994)
11. Carmona, R., Zhong, S.: Adaptive smoothing respecting feature directions. IEEE Trans. Image Process. **7**(3), 353–358 (1998)
12. Caselles, V., Morel, J.M., Sbert, C.: An axiomatic approach to image interpolation. IEEE Trans. Image Process. **7**(3), 376–386 (1998)
13. Chan, T.F., Shen, J.: Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods. SIAM, Philadelphia (2005)
14. Chu, C.K., Glad, I., Godtliebsen, F., Marron, J.S.: Edge-preserving smoothers for image processing. J. Am. Stat. Assoc. **93**(442), 526–556 (1998)
15. Cohen, I., Falik, A., Gilboa, G.: Stable explicit p-Laplacian flows based on nonlinear eigenvalue anaysis. In: Lellmann, J., Burger, M., Modersitzki, J. (eds.) Scale Space and Variational Methods in Computer Vision. Lecture Notes in Computer Science, vol. 11603, pp. 315–327. Springer, Cham (2019)
16. Ellis, R.L.: On the foundations of the theories of probablities. Trans. Cambridge Philos. Soc. **8**(1), 204–219 (1844)
17. Fechner, G.T.: Ueber den Ausgangswerth der kleinsten Abweichungssumme, dessen Bestimmung, Verwendung und Verallgemeinerung. Abhandlungen der mathematisch-physischen Classe der Königlich Sächsischen Gesellschaft der Wissenschaften **11**, 1–76 (1878)
18. Fréchet, M.: Sur l'extension de certaines evaluations statistiques au cas de petits echantillons. Revue de l'Institut International de Statistique **11**(3), 182–205 (1943)
19. Fréchet, M.: Nouvelles définitions de la valeur moyenne et des valeurs equiprobables d'un nombre aléatoire. Sciences, pp. 5–26. Annales de l'Université de Lyon, Lyon (1946)
20. Fréchet, M.: Les élements aléatoires de nature quelconque dans un espace distancié. Annales de l'Institut Henri Poincaré **10**, 215–310 (1948)
21. Fréchet, M.: Les valeurs typiques d'ordre nul ou infini d'un nombre aléatoire. Revue de l'Institut International de Statistique **16**(1), 1–22 (1948)
22. Fréchet, M.: Positions typiques d'un élément aléatoire de nature quelconque. Annales scientifiques de l'É.N.S., 3e série **65**, 211–237 (1948)
23. Gabor, D.: Information theory in electron microscopy. Lab. Investig. **14**, 801–807 (1965)
24. Galić, I., Weickert, J., Welk, M., Bruhn, A., Belyaev, A., Seidel, H.P.: Towards PDE-based image compression. In: Paragios, N., Faugeras, O., Chan, T., Schnörr, C. (eds.) Variational Geometric and Level-Set Methods in Computer Vision. Lecture Notes in Computer Science, vol. 3752, pp. 37–48. Springer, Berlin (2005)
25. Gauss, C.F.: Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium. Perthes & Besser, Hamburg (1809)
26. Gilboa, G., Sochen, N.A., Zeevi, Y.Y.: Forward-and-backward diffusion processes for adaptive image enhancement and denoising. IEEE Trans. Image Process. **11**(7), 689–703 (2002)
27. Griffin, L.D.: Mean, median and mode filtering of images. Proc. R. Soc. Lond. Ser. A **456**(2004), 2995–3004 (2000)
28. Griffin, L.D., Lillholm, M.: Mode estimation using pessimistic scale space tracking. In: Griffin, L.D., Lillholm, M. (eds.) Scale-Space Methods in Computer Vision. Lecture Notes in Computer Science, vol. 2695, pp. 266–280. Springer, Berlin (2003)
29. Guichard, F., Morel, J.M.: Partial differential equations and image iterative filtering. In: Duff, I.S., Watson, G.A. (eds.) The State of the Art in Numerical Analysis, no. 63 in IMA Conference Series (New Series), pp. 525–562. Clarendon Press, Oxford (1997)
30. Huber, P.J.: Robust Statistics. Wiley, New York (1981)
31. Iijima, T.: Basic equation of figure and observational transformation. Syst. Comput. Controls **2**(4), 70–77 (1971)
32. Jackson, D.: Note on the median of a set of numbers. Bull. Am. Math. Soc. **27**, 160–164 (1921)
33. Jordan, C.: Statistique Mathématique. Gauthier-Villars, Paris (1927)
34. Kass, M., Solomon, J.: Smoothed local histogram filters. ACM Trans. Graph. **29**(4), 100 (2010)
35. Kramer, H.P., Bruckner, J.B.: Iterations of a non-linear transformation for enhancement of digital images. Pattern Recogn. **7**, 53–58 (1975)
36. Kuijper, A.: Geometrical PDEs based on second-order derivatives of gauge coordinates in image processing. Image Vis. Comput. **29**, 1023–1034 (2009)
37. Laplace, P.S.: Mémoire sur la probabilité des causes par les événements. Mémoires de l'Académie royale des Sciences de Paris (Savants étrangers) **VI**, 621pp. (1774). In: Oeuvres complètes de Laplace, vol. 8, 25–65, Gauthier-Villars, Paris (1891)
38. Legendre, A.M.: Nouvelles Méthodes pour la détermination des Orbites des Comètes. Firmin Didot, Paris (1805)
39. Lindenbaum, M., Fischer, M., Bruckstein, A.: On Gabor's contribution to image enhancement. Pattern Recogn. **27**, 1–8 (1994)
40. Marchuk, G.I.: Splitting and alternating direction methods. In: Ciarlet, P.G., Lions, J.L. (eds.) Handbook of Numerical Analysis, vol. I, pp. 197–462. North Holland, Amsterdam (1990)
41. Mitchell, A.R., Griffiths, D.F.: The Finite Difference Method in Partial Differential Equations. Wiley, Chichester (1980)
42. Mrázek, P., Weickert, J., Bruhn, A.: On robust estimation and smoothing with spatial and tonal kernels. In: Klette, R., Kozera, R., Noakes, L., Weickert, J. (eds.) Geometric Properties from Incomplete Data, Computational Imaging and Vision, vol. 31, pp. 335–352. Springer, Dordrecht (2006)
43. Osher, S., Paragios, N. (eds.): Geometric Level Set Methods in Imaging Vision and Graphics. Springer, New York (2003)
44. Osher, S., Rudin, L.: Shocks and other nonlinear filtering applied to image processing. In: Tescher, A.G. (ed.) Applications of Digital Image Processing XIV. Proceedings of SPIE, vol. 1567, pp. 414–431. SPIE Press, Bellingham (1991)
45. Osher, S., Rudin, L.I.: Feature-oriented image enhancement using shock filters. SIAM J. Numer. Anal. **27**, 919–940 (1990)
46. Pearson, K.: Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. Philos. Trans. R. Soc. A **186**, 343–414 (1895)
47. Perona, P., Malik, J.: Scale space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Anal. Mach. Intell. **12**, 629–639 (1990)
48. Rouy, E., Tourin, A.: A viscosity solutions approach to shape-from-shading. SIAM J. Numer. Anal. **29**, 867–884 (1992)
49. Schönlieb, C.B.: Partial Differential Equation Methods for Image Inpainting. Cambridge University Press, New York (2015)
50. Torroba, P.L., Cap, N.L., Rabal, H.J., Furlan, W.D.: Fractional order mean in image processing. Opt. Eng. **33**(2), 528–534 (1994)
51. Tukey, J.W.: Exploratory Data Analysis. Addison-Wesley, Menlo Park (1971)
52. van den Boomgaard, R., Smeulders, A.: The morphological structure of images: the differential equations of morphological scale-space. IEEE Trans. Pattern Anal. Mach. Intell. **16**, 1101–1113 (1994)
53. van den Boomgaard, R., van de Weijer, J.: On the equivalence of local-mode finding, robust estimation and mean-shift analysis as used in early vision tasks. In: Proceedings of 16th International Conference on Pattern Recognition, vol. 3, pp. 927–930. Quebec City, Canada (2002)
54. Vladimirov, V.S.: Generalized Functions in Mathematical Physics. Mir, Moscow (1979)
55. Weickert, J.: Anisotropic Diffusion in Image Processing. Teubner, Stuttgart (1998)

56. Weickert, J., Steidl, G., Mrázek, P., Welk, M., Brox, T.: Diffusion filters and wavelets: What can they learn from each other? In: Paragios, N., Chen, Y., Faugeras, O. (eds.) Handbook of Mathematical Models in Computer Vision, pp. 3–16. Springer, New York (2006)
57. Welk, M.: Multivariate median filters and partial differential equations. J. Math. Imaging Vis. **56**, 320–351 (2016)
58. Welk, M.: Multivariate medians for image and shape analysis. arXiv:1911.00143 (2019)
59. Welk, M., Breuß, M., Vogel, O.: Morphological amoebas are self-snakes. J. Math. Imaging Vis. **39**(2), 87–99 (2011)
60. Welk, M., Weickert, J.: PDE evolutions for M-smoothers: from common myths to robust numerics. In: Burger, M., Lellmann, J., Modersitzki, J. (eds.) Scale Space and Variational Methods in Computer Vision. Lecture Notes in Computer Science, vol. 11603, pp. 248–263. Springer, Cham (2019)
61. Welk, M., Weickert, J., Galić, I.: Theoretical foundations for spatially discrete 1-D shock filtering. Image Vis. Comput. **25**(4), 455–463 (2007)
62. Winkler, G., Aurich, V., Hahn, K., Martin, A.: Noise reduction in images: some recent edge-preserving methods. Pattern Recogn. Image Anal. **9**(4), 749–766 (1999)
63. Wladimirow, W.S.: Gleichungen der mathematischen Physik. Deutscher Verlag der Wissenschaften, Berlin (1972). English edition: V. S. Vladimirov, Equations of Mathematical Physics, Marcel Dekker, New York (1971)
64. Yanenko, N.N.: The Method of Fractional Steps: the Solution of Problems of Mathematical Physics in Several Variables. Springer, New York (1971)

Martin Welk received his Ph.D. in mathematics in 1998 from the University of Leipzig. From 2002 till 2010, he worked with the Mathematical Image Analysis group at Saarland University, Saarbrücken, Germany. In 2010, he moved to UMIT – Private University for Health Sciences, Medical Informatics and Technology, Hall/Tyrol, where he currently holds an Associate Professor position. His research interests are centered around mathematical principles of image analysis, especially PDE and variational methods.

Joachim Weickert is professor of Mathematics and Computer Science at Saarland University (Saarbrücken, Germany), where he heads the Mathematical Image Analysis Group. He graduated and obtained his Ph.D. from the University of Kaiserslautern (Germany) in 1991 and 1996. He worked as post-doctoral researcher at the University Hospital of Utrecht (The Netherlands) and the University of Copenhagen (Denmark), and as assistant professor at the University of Mannheim (Germany). Joachim Weickert has developed many models and efficient algorithms for image processing and computer vision using partial differential equations and variational methods. His scientific work covers more than 320 refereed publications. He has served in the editorial boards of ten international journals or book series and is Editor-in-Chief of the Journal of Mathematical Imaging and Vision. In 2010, he has received a Gottfried Wilhelm Leibniz Prize and in 2017 an ERC Advanced Grant.