Diffusion, Pre-Smoothing and Gradient Descent

Martin Welk

Institute of Biomedical Image Analysis
UMIT TIROL – Private University of Health Sciences, Medical Informatics and Technology
Eduard-Wallnöfer-Zentrum 1, 6060 Hall/Tyrol, Austria
`martin.welk@umit.at`

Abstract. Nonlinear diffusion of images, both isotropic and anisotropic, has become a well-established and well-understood denoising tool during the last three decades. Moreover, it is a component of partial differential equation methods for various further tasks in image analysis. For the analysis of such methods, their understanding as gradient descents of energy functionals often plays an important role. Often the diffusivity or diffusion tensor field for nonlinear diffusion is computed from pre-smoothed image gradients. What was not clear so far was whether nonlinear diffusion with this pre-smoothing step still is the gradient descent for some energy functional. This question is answered to the negative in the present paper. Triggered by this result, possible modifications of the pre-smoothing step to retain the gradient descent property of diffusion are discussed.

1 Introduction

Nonlinear diffusion processes are well-established image denoising methods. They also form indispensable building blocks in numerous image analysis methods involving partial differential equations (PDEs).

The starting point for the modern development of nonlinear diffusion methods in image processing was Perona and Malik's paper [13]. It proposes the isotropic nonlinear diffusion model which is nowadays mostly stated as

$$u_t = \mathrm{div}\left(g(|\boldsymbol{\nabla}u|^2)\,\boldsymbol{\nabla}u\right),\tag{1}$$

with $g : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ being a decreasing diffusivity function. From the two candidates for g proposed in [13], the function

$$g(s^2) = \frac{1}{1 + s^2/\lambda^2}\tag{2}$$

prevails as a widespread standard choice. The parameter $\lambda > 0$ can be interpreted intuitively as a threshold. It separates small gradients $|\boldsymbol{\nabla}u| < \lambda$ that are smoothed out as noise from large ones $|\boldsymbol{\nabla}u| > \lambda$ presumed to represent valuable image structures that should be preserved.

Other diffusivities have been proposed, with total variation diffusion [1, 2, 6] being the most prominent example besides (2).

In semidiscrete (space-discrete, time-continuous) and fully discrete settings well-posedness of nonlinear isotropic diffusion processes has been proved, see [4, 3, 15].

In the fully continuous setting, Perona-Malik diffusion in its original form is not generally well-posed. Solutions of its initial-boundary value problem in different function spaces have been investigated [8, 7, 9, 10, 23] with mixed results: Although classical solutions exist in certain conditions, either locally or even globally, this is not generally the case. Weak solutions can be highly non-unique. Stability results are largely limited to extremum principles (L^∞-stability).

A pivotal role in the stability issues of Perona-Malik diffusion is played by the staircasing phenomenon, by which even smooth initial data develop discontinuities within finite time. This is essentially caused by the local appearance of inverse diffusion in regions with large gradients where the flux $g(|\boldsymbol{\nabla} u|^2)|\boldsymbol{\nabla} u|$ decreases with $|\boldsymbol{\nabla} u|$. With the diffusivity (2), for example, this is the case for $|\boldsymbol{\nabla} u| > \lambda$, compare [15].

Regularised nonlinear diffusion. An explanation for the discrepancy between the space-continuous and discrete behaviour of Perona-Malik diffusion is given in [17] where it is pointed out that discretisation by itself introduces a regularising effect. However, relying on regularisation by discretisation means that important features of the actual image enhancement process are not part of the space-continuous model. Therefore it is desirable to have an explicit regularisation already in the space-continuous model.

Indeed, explicit regularisations in the continuous setting have been proposed as early as 1992 in [5, 11]. The pre-smoothing by Gaussian smoothing of the gradient within the diffusivity expression as introduced by [5] enjoys most popularity till today,

$$u_t = \operatorname{div}\left(g(|G_\sigma * \boldsymbol{\nabla} u|^2)\,\boldsymbol{\nabla} u\right). \tag{3}$$

Well-posedness of (3) has been proven in [5].

Introducing an additional directional dependency of the diffusivity, one arrives at anisotropic diffusion processes. One such process, *edge-enhancing anisotropic diffusion (EED)*, can be stated as [14, 15]

$$u_t = \operatorname{div}\left(g(\boldsymbol{J}(u))\,\boldsymbol{\nabla} u\right). \tag{4}$$

Herein, the decreasing diffusivity function g acts on the rank-one symmetric outer-product matrices

$$\boldsymbol{J}(u) = (G_\sigma * \boldsymbol{\nabla} u)(G_\sigma * \boldsymbol{\nabla} u)^{\mathrm{T}} \tag{5}$$

to yield at each image location a diffusion tensor $g(\boldsymbol{J})$. The diffusion tensor possesses one eigenvalue $g(|G_\sigma * \boldsymbol{\nabla} u|^2)$ with an eigenvector parallel to the pre-smoothed gradient $G_\sigma * \boldsymbol{\nabla} u$, and one eigenvalue $g(0) = 1$ for an eigenvector orthogonal to the gradient. Like isotropic nonlinear diffusion, EED is capable of preserving and even enhancing sharp edges because it suppresses, or even reverts, diffusion across edges; in contrast to its isotropic predecessor, however, it allows for undiminished diffusion flow along edges, thus achieving better denoising near edges.

It is worth noting that for grey-value images, the pre-smoothing of the gradient is essential for true anisotropy because the application of g to the rank-one matrices $\boldsymbol{\nabla} u \boldsymbol{\nabla} u^{\mathrm{T}}$ would yield just the exact same flux field as Perona-Malik diffusion, and

thereby effectively reproduce isotropic diffusion. For multi-channel (e.g., colour) images, anisotropy can arise even without pre-smoothing if the gradient directions of the colour channels do not coincide.

Diffusion as gradient descent. PDE-based methods in image processing, be it for denoising or for other purposes, are often derived from variational approaches, where an image processing task is formulated as the minimisation of some energy functional dependent on the sought image. Variational calculus then allows to derive PDEs either as Euler-Lagrange equations or as gradient descent of the functional. Indeed, the energy functional

$$E[u] = \frac{1}{2} \int_{\Omega} \Psi\big(|\boldsymbol{\nabla} u|^2\big) \, \mathrm{d}\boldsymbol{x} \tag{6}$$

immediately leads to the gradient descent

$$u_t = \mathrm{div}\big(\Psi'(|\boldsymbol{\nabla} u|^2) \, \boldsymbol{\nabla} u\big) \, , \tag{7}$$

which is exactly the original Perona-Malik equation without pre-smoothing, with diffusivity $g \equiv \Psi'$; compare [12]. For example, the diffusivity (2) arises from $\Psi(s^2) = \lambda^2 \ln(1 + s^2/\lambda^2)$. More general variational models often contain summands of the type (6) which then yield diffusion components in their gradient descent equations.

Similarly, the EED equation without pre-smoothing in the case of multi-channel images can be obtained as gradient descent of an energy functional in which a penalty Ψ as before is applied to the sum of outer products $\boldsymbol{\nabla} u_c \boldsymbol{\nabla} u_c^{\mathrm{T}}$ of the colour channels u_c.

For PDEs arising from gradient descent, it can be attractive to design finite-difference discretisations in a way that they preserve this connection. To this end, the energy functional is discretised, which yields a function of a large finite number of real variables (one per pixel). Gradient descent for this function takes the form of a system of ordinary differential equations which is a discretisation of the PDE, see [19] for an example.

Derivation as a gradient descent from an energy functional can serve as a strong theoretical justification that singles out a particular PDE evolution from similar ones by an optimality criterion. The analysis of energy functionals often provides powerful tools to derive important properties of the models such as convergence to a unique steady state (e.g. for convex energy functionals).

Thus, the gradient descent property makes a strong case for the original Perona-Malik diffusion (1). In practice, however, it is very often the pre-smoothed version (3) which is used for denoising, or as building block within some other PDE-based image processing method. This raises naturally the question whether (3) and similar diffusion processes are gradient descents for suitable energy functionals, too.

Whereas researchers in the field have pondered about this question time and again, it seems that it has attracted little real effort over the years. Researchers noted that there is no known energy functional yielding this evolution as gradient descent but could not agree whether this would be principally impossible, or whether they had not been inventive enough to write down the proper energy functional, or whether maybe such an energy functional existed but would not admit being stated in a closed form. For example, [19, p. 199, footnote 3] states that "For $\sigma \neq 0$, no energy functional is known that has [isotropic nonlinear diffusion with pre-smoothing] as gradient descent."

However, given the theoretical advantages of PDE methods being derived from variational models, we believe that this question is worth settling, and will do so in this paper. Unfortunately, the answer is negative. We will therefore add some – explorative – discussion of alternatives.

Structure of the paper. In Section 2, we prove that neither 1D nonlinear diffusion, nor, as a consequence, 2D isotropic nonlinear diffusion, nor EED can be stated as gradient descents. In Section 3, we discuss how to design regularised diffusion processes that are exact gradient descents. Experimental demonstration of one such process in comparison to Perona-Malik diffusion is provided in Section 4. A short summary in Section 5 concludes the paper.

2   Integrability Analysis of Diffusion with Pre-Smoothing

In order to investigate whether diffusion with pre-smoothing can be stated as a gradient descent, let us recall first the situation in classical vector field analysis.

Classical integrability conditions. Consider a continuously differentiable vector field $\boldsymbol{v} : \mathbb{R}^d \to \mathbb{R}^d$, where $\boldsymbol{v}(\boldsymbol{x}) = (v_1(x_1, \ldots, x_d), \ldots, v_d(x_1, \ldots, x_d))^{\mathrm{T}}$. A necessary criterion for $\boldsymbol{v}$ to be the gradient field of a potential V is then given by the *integrability condition*

$$\frac{\partial v_i}{\partial x_j} = \frac{\partial v_j}{\partial x_i} \qquad \text{for all } i, j = 1, \ldots, d,\, i \neq j. \tag{8}$$

In particular, for $d = 2$ or $d = 3$ this boils down to the well-known condition $\operatorname{rot} \boldsymbol{v} = 0$ where $\operatorname{rot} \boldsymbol{v} = \boldsymbol{\nabla} \wedge \boldsymbol{v} = \partial_1 v_2 - \partial_2 v_1$ (scalar-valued) in two, and $\operatorname{rot} \boldsymbol{v} = \boldsymbol{\nabla} \times \boldsymbol{v}$ (vector-valued) in three dimensions.

Coordinate-free integrability conditions. Whereas it is usually assumed in (8) that coordinates x_i, v_i are taken w.r.t. some orthonormal basis, it is obvious that this is not necessary: Note that (8) is trivially fulfilled also for $i = j$. Thus it can easily be extended to linear combinations (with $\boldsymbol{y} = (y_1, \ldots, y_d)^{\mathrm{T}}$ and $\boldsymbol{z} = (z_1, \ldots, z_d)^{\mathrm{T}}$ being unit vectors), yielding

$$\sum_{i=1}^{d} \sum_{j=1}^{d} \alpha_i \beta_j \frac{\partial v_i}{\partial x_j} = \sum_{i=1}^{d} \sum_{j=1}^{d} \beta_j \alpha_i \frac{\partial v_j}{\partial x_i} \tag{9}$$

which by $\langle \boldsymbol{v}, \boldsymbol{y} \rangle = \sum_{i=1}^{d} y_i v_i$, $\partial/\partial \boldsymbol{z} = \sum_{j=1}^{d} z_j\, \partial/\partial x_j$ means that the set of integrability conditions can be re-stated as

$$\frac{\partial \langle \boldsymbol{v}, \boldsymbol{y} \rangle}{\partial \boldsymbol{z}} = \frac{\partial \langle \boldsymbol{v}, \boldsymbol{z} \rangle}{\partial \boldsymbol{y}} \tag{10}$$

for arbitrary unit vectors $\boldsymbol{y}$, $\boldsymbol{z}$, i.e. the component of $\boldsymbol{v}$ in direction of $\boldsymbol{y}$ has a directional derivative in direction of $\boldsymbol{z}$ equal to that of the $\boldsymbol{z}$ component in $\boldsymbol{y}$ direction. The virtue of (10) is that it is coordinate-free, i.e. it does not depend on any choice of orthonormal basis.

Moreover, each integrability condition does in fact involve only the projection of $\boldsymbol{v}$ onto a two-dimensional subspace. This is natural since the restriction and projection

of a gradient descent to a subspace is again a gradient descent in that subspace. This argument does obviously hold not only in finite, but also in infinite dimensions, which means that the set of necessary conditions (10) remains valid even if $\boldsymbol{v} : \mathcal{V} \to \mathcal{V}$ with any Hilbert space $\mathcal{V}$.

Transfer to function spaces. Assume that $\mathcal{V} = \mathcal{V}(\Omega)$ is a Hilbert space of sufficiently smooth functions over some domain Ω. We consider a time-dependent partial differential equation

$$u_t = F[u] \, , \qquad F[u](\boldsymbol{x}) = f\big(u(\boldsymbol{x}), \partial_{\alpha_1} u(\boldsymbol{x}), \partial_{\alpha_2} u(\boldsymbol{x}), \dots\big) \qquad (11)$$

where $\partial_{\alpha_i} u$ are partial derivatives of u w.r.t. spatial coordinates in Ω, and f some sufficiently smooth function combining these. The flux $F[u]$ defined by f can then be considered as a mapping from the space $\mathcal{V}$ of functions to the space of perturbations of functions in $\mathcal{V}$, which can be identified with $\mathcal{V}$. Thus, F is a vector field on $\mathcal{V}$.

If (11) is to be the gradient descent of some functional $E[u]$ over $\mathcal{V}$, F needs to be the gradient field of $E[u]$. Translating the integrability conditions (10) to this situation, we obtain as the set of necessary conditions

$$\frac{\partial \langle u_t, v \rangle}{\partial \langle u, w \rangle} = \frac{\partial \langle u_t, w \rangle}{\partial \langle u, v \rangle} \qquad (12)$$

where $v, w \in \mathcal{V}$ are arbitrary perturbation functions. Using $\partial(\,\cdot\,)/\partial\langle u, v \rangle = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}(\,\cdot\,)|_{\varepsilon=0}$, this set of conditions can be translated further into

$$\left.\frac{\mathrm{d}}{\mathrm{d}\varepsilon} \langle F[u + \varepsilon v], w \rangle\right|_{\varepsilon=0} = \left.\frac{\mathrm{d}}{\mathrm{d}\varepsilon} \langle F[u + \varepsilon w], v \rangle\right|_{\varepsilon=0} \qquad (13)$$

for arbitrary perturbation functions v, w.

Analysis of pre-smoothed nonlinear diffusion in 1D. We turn now to apply (13) to analyse the 1D nonlinear diffusion process with pre-smoothing given by

$$u_t = \partial_x \big(g((G_\sigma * u_x)^2) u_x\big) \, . \qquad (14)$$

Let us assume that u and its perturbation functions come from a suitable Hilbert space of functions over a domain $\Omega \subseteq \mathbb{R}$ with the standard scalar product

$$\langle u, v \rangle = \int_\Omega u(x) v(x) \, \mathrm{d}x \, . \qquad (15)$$

(Note that if Ω is not the entire $\mathbb{R}$, the boundary treatment for the convolution must be specified appropriately.) Assume further that the perturbation functions v and w vanish on the boundary of Ω if any (this technical condition could be relaxed but it simplifies the expressions arising from integration by parts later on). Setting for abbreviation $\tilde{u} := G_\sigma * u$, we can calculate

$$F[u] = \partial_x \big(g(\tilde{u}_x^2) u_x\big) = g'(\tilde{u}_x^2) \, 2 \, \tilde{u}_x \, \tilde{u}_{xx} \, u_x + g(\tilde{u}_x^2) \, u_{xx} \qquad (16)$$

and thus

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\langle F[u+\varepsilon v], w\rangle\Big|_{\varepsilon=0} &= \int_{\Omega} \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big(2\, g'((\tilde{u}_x + \varepsilon\tilde{v}_x)^2)\,(\tilde{u}_x + \varepsilon\tilde{v}_x)\,(\tilde{u}_{xx} + \varepsilon\tilde{v}_{xx})\cdot \\
&\qquad (u_x + \varepsilon v_x) + g((\tilde{u}_x + \varepsilon\tilde{v}_x)^2)(u_{xx} + \varepsilon v_{xx})\Big)\Big|_{\varepsilon=0} w\,\mathrm{d}x \\
&= \int_{\Omega}\Big(2\, g''(\tilde{u}_x^2)\cdot 2\,\tilde{u}_x\,\tilde{v}_x\,\tilde{u}_x\,\tilde{u}_{xx}\,u_x + 2\, g'(\tilde{u}_x^2)\,\tilde{v}_x\,\tilde{u}_{xx}\,u_x \\
&\qquad + 2\, g'(\tilde{u}_x^2)\,\tilde{u}_x\,\tilde{v}_{xx}\,u_x + 2\, g'(\tilde{u}_x^2)\,\tilde{u}_x\,\tilde{u}_{xx}\,v_x \\
&\qquad + g'(\tilde{u}_x^2)\cdot 2\,\tilde{u}_x\,\tilde{v}_x\,u_{xx} + g(\tilde{u}_x^2)\,v_{xx}\Big)w\,\mathrm{d}x \\
&= \int_{\Omega} 4\, g''(\tilde{u}_x^2)\,\tilde{u}_x^2\,u_x\,\tilde{u}_{xx}\,\tilde{v}_x\,w + 2\, g'(\tilde{u}_x^2)\,u_x\,\tilde{u}_{xx}\,\tilde{v}_x\,w \\
&\qquad + 2\, g'(\tilde{u}_x^2)\,\tilde{u}_x\,u_x\,\tilde{v}_{xx}\,w + 2\, g'(\tilde{u}_x^2)\,\tilde{u}_x\,\tilde{u}_{xx}\,v_x\,w \\
&\qquad + 2\, g'(\tilde{u}_x^2)\,\tilde{u}_x\,u_{xx}\,\tilde{v}_x\,w + g(\tilde{u}_x^2)\,v_{xx}\,w\,\mathrm{d}x \qquad (17)
\end{aligned}
$$

Using integration by parts for the summands involving second derivatives of perturbation functions, most summands cancel, leaving

$$
\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\langle F[u+\varepsilon v], w\rangle\Big|_{\varepsilon=0} = \int_{\Omega} -g(\tilde{u}_x^2)\,v_x\,w_x - 2\, g'(\tilde{u}_x^2)\tilde{u}_x\,u_x\,\tilde{v}_x\,w_x\,\mathrm{d}x . \qquad (18)
$$

We combine this expression with its counterpart for $\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\langle F[u+\varepsilon v], w\rangle\big|_{\varepsilon=0}$ to obtain, with cancellation of $g(\tilde{u}_x^2)\,v_x\,w_x$,

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\langle F[u+\varepsilon v], w\rangle\Big|_{\varepsilon=0} - \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\langle F[u+\varepsilon v], w\rangle\Big|_{\varepsilon=0} \\
&\qquad = 2\int_{\Omega} g'(\tilde{u}_x^2)\tilde{u}_x\,u_x\,(v_x\,\tilde{w}_x - \tilde{v}_x\,w_x)\,\mathrm{d}x . \qquad (19)
\end{aligned}
$$

Unfortunately, this expression does not vanish identically for non-constant diffusivity g and arbitrary functions u, v, w. We have therefore proven the following statement.

Proposition 1. *The regularised nonlinear 1D diffusion process* (14) *with a non-constant, twice continuously differentiable diffusivity function g is not the gradient descent for any energy functional on functions $u : \Omega \to \mathbb{R}$, $\Omega \subseteq \mathbb{R}$, w.r.t. the standard metric in function space induced by the scalar product* (15).

Implications for 2D diffusion processes. We turn now to consider the nonlinear isotropic diffusion process from [13] with pre-smoothing [5] in 2D (or higher dimension) as given in (3).

We assume that the domain $\Omega \subseteq \mathbb{R}^d$ is of the form $\Omega = \Omega_1 \times \Omega_2$ where $\Omega_1 \subseteq \mathbb{R}$ and $\Omega_2 \subseteq \mathbb{R}^{d-1}$; this is obviously the case e.g. for rectangular images. We assume

further that the scalar product of functions on Ω (and thus the Hilbert function space $\mathcal{V}(\Omega)$) is chosen in such a way that the functions that are constant along all but the first coordinate direction and are given by $u(x_1, x_2, \ldots, x_d) = u_1(x_1)$ with $u_1 \in \mathcal{V}(\Omega_1)$ belong to $\mathcal{V}(\Omega)$. This can be ensured, e.g., by taking Ω_2 as an interval or Cartesian product of intervals with periodic boundary conditions, or by equipping $\Omega_2 = \mathbb{R}^{d-1}$ with a weighted scalar product in which the local weights decay quickly enough to ensure finiteness of $\langle 1, 1 \rangle$, e.g.

$$\langle u, v \rangle = \int_{\Omega_2} G_\sigma(\boldsymbol{x}) u(\boldsymbol{x}) v(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \; . \tag{20}$$

By restriction to functions as described above that depend on the first coordinate only, the process (3) reduces verbatim to (14). As Proposition 1 shows, it cannot be represented as gradient descent in this particular case, and therefore neither in general. Using a suitable limit argument where necessary, the result can be transferred to the function space $\mathcal{V}(\Omega)$ with standard scalar product as stated in the following corollary.

Corollary 1. *The regularised nonlinear isotropic diffusion process* (3) *with a non-constant, twice continuously differentiable diffusivity function g is not the gradient descent for any energy functional on functions $u : \Omega \to \mathbb{R}$, $\Omega \subseteq \mathbb{R}^d$, w.r.t. the standard metric in function space.*

Similar arguments apply to EED, yielding the following statement.

Corollary 2. *The regularised nonlinear anisotropic diffusion process* (4) *with a non-constant, twice continuously differentiable diffusivity function g is not the gradient descent for any energy functional on functions $u : \Omega \to \mathbb{R}$, $\Omega \subseteq \mathbb{R}^d$, w.r.t. the standard metric in function space.*

3 Alternatives

In this section we discuss possible alternatives to the established pre-smoothing in diffusion methods which could be compatible with the gradient descent framework that exists for nonlinear diffusion without pre-smoothing, given that this framework often also inspires applications of the pre-smoothed variants.

We remark first that using the traditional pre-smoothed Perona-Malik diffusion (3) for edge-enhancing denoising, the smoothing of the gradient has a two-fold role. On one hand, it yields an overall smoother diffusivity field, thus supporting edge enhancement in creating a more regular set of edges. On the other hand, it boosts the removal of small-scale structures such as single noise pixels which would otherwise be stabilised longer by their surrounding high gradients and thus low diffusivities.

We also notice that "regularisation by discretisation", as undesirable an intertwining of model and numerics it involves, has the advantage to retain the gradient descent property if an appropriate discretisation is used. However, it does not provide a means to steer the degree of regularisation.

Modified energy functional. In order to find diffusion equations with adjustable regularisation parameters that are consistent with gradient descent in the space-continuous setting, we consider modifications of the energy functional (6).

In coherence-enhancing anisotropic diffusion (CED) [16], another anisotropic diffusion process introduced by Weickert that is designed to denoise and enhance line-like structures rather than providing general-purpose denoising such as EED, a smoothed structure tensor is used that involves, besides the smoothing of the gradients $\boldsymbol{\nabla} u$, a second Gaussian convolution that applies to the outer product matrices, leading to

$$\boldsymbol{J}_\varrho(u) = G_\varrho * \big((G_\sigma * \boldsymbol{\nabla} u)(G_\sigma * \boldsymbol{\nabla} u)^{\mathrm{T}}\big) . \tag{21}$$

Moreover, following [18], the energy functional (6) can be rewritten as

$$E[u] = \frac{1}{2} \int_\Omega \Psi\big(\mathrm{trace}(\boldsymbol{\nabla} u \boldsymbol{\nabla} u^{\mathrm{T}})\big) \, \mathrm{d}\boldsymbol{x} . \tag{22}$$

Inspired by this observation, we consider smoothing of $|\boldsymbol{\nabla} u|^2 = \mathrm{trace}(\boldsymbol{\nabla} u \boldsymbol{\nabla} u^{\mathrm{T}})$. Generalising Gaussian convolutions to linear operators L_1, L_2, we write down the ansatz

$$E[u] = \frac{1}{2} \int_\Omega \Psi\big(L_1(|L_2(\boldsymbol{\nabla} u)|^2)\big) \, \mathrm{d}\boldsymbol{x} . \tag{23}$$

For the following, we denote by L^* the adjoint operator of a linear operator L, i.e. the linear operator that satisfies

$$\int_\Omega L^*(f) \cdot g \, \mathrm{d}\boldsymbol{x} = \int_\Omega f \cdot L(g) \, \mathrm{d}\boldsymbol{x} \tag{24}$$

for all f, g. Gaussian convolution on $\mathbb{R}^n$ is self-adjoint, i.e. $(G_\sigma *)^* = G_\sigma *$.

Gradient descent. To determine the gradient descent of (26), we calculate, for some perturbation function v that vanishes on the boundary of Ω,

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} E[u + \varepsilon v]\Big|_{\varepsilon=0} = \int_\Omega \Psi'\big(L_1(|L_2(\boldsymbol{\nabla} u)|^2)\big) \cdot L_1\big(\langle L_2(\boldsymbol{\nabla} u), L_2(\boldsymbol{\nabla} v)\rangle\big) \, \mathrm{d}\boldsymbol{x}$$

$$= \int_\Omega L_1^*\Big(\Psi'\big(L_1(|L_2(\boldsymbol{\nabla} u)|^2)\big)\Big) \cdot \langle L_2(\boldsymbol{\nabla} u), L_2(\boldsymbol{\nabla} v) \, \mathrm{d}\boldsymbol{x}$$

$$= \int_\Omega \Big\langle L_2^*\Big(L_1^*\big(\Psi'\big(L_1(|L_2(\boldsymbol{\nabla} u)|^2)\big)\big) \cdot L_2(\boldsymbol{\nabla} u)\Big), \boldsymbol{\nabla} v \Big\rangle \, \mathrm{d}\boldsymbol{x}$$

$$= -\int_\Omega \mathrm{div}\Big(L_2^*\Big(L_1^*\big(\Psi'\big(L_1(|L_2(\boldsymbol{\nabla} u)|^2)\big)\big) \cdot L_2(\boldsymbol{\nabla} u)\Big)\Big) \cdot v \, \mathrm{d}\boldsymbol{x} \tag{25}$$

from which we read off the desired gradient descent as

$$u_t = \mathrm{div}\Big(L_2^*\Big(L_1^*\big(\Psi'\big(L_1(|L_2(\boldsymbol{\nabla} u)|^2)\big)\big) \cdot L_2(\boldsymbol{\nabla} u)\Big)\Big) . \tag{26}$$

Note that in this diffusion-like process the flux is the smoothed version of a vector field which is at each location a scalar multiple of $L_2(\boldsymbol{\nabla} u)$. If L_2 is not the identical operator id, the flux direction can, and will at most locations with non-trivial structure, deviate from that of $\boldsymbol{\nabla} u$. Thus, (26) is an anisotropic process. However, there is an important difference to established anisotropic diffusion processes like EED or CED: In these, the flux is always the product of a positive semidefinite diffusion tensor with $\boldsymbol{\nabla} u$. Therefore, the projection of the flux onto the gradient direction always points in positive gradient direction, ensuring a forward diffusion component in that direction. In contrast, the flux in (26) can even have a negative projection onto $\boldsymbol{\nabla} u$, thus performing inverse diffusion with negative diffusivity. Inverse diffusion is a prototype of an unstable evolution. Although diffusion processes that involve local inverse diffusion can still be stable for entire images, compare [20, 22], it is not obvious whether this is true here. To decide this requires a more detailed stability analysis which is beyond the scope of this paper. At any rate, to devise stable numerical schemes for such a process would be challenging [20–22]. The gradient descent (26) with $L_2 \neq$ id does therefore not lend itself as a promising candidate to replace (3).

With $L_2 \equiv$ id, instead, (26) simplifies into

$$u_t = \operatorname{div}\left(L_1^*\Big(\Psi'\big(L_1(|\boldsymbol{\nabla} u|^2)\big)\Big) \cdot \boldsymbol{\nabla} u \right). \tag{27}$$

Specifically for L_1 being Gaussian convolution, we have

$$u_t = \operatorname{div}\Big(\big(G_\sigma * \Psi'(G_\sigma * |\boldsymbol{\nabla} u|^2)\big)\boldsymbol{\nabla} u \Big). \tag{28}$$

We will demonstrate the effect of the evolution (28) compared with traditional Perona-Malik diffusion with pre-smoothing (3) by an experiment in the next section. Before we do so, let us shortly discuss what effect can be expected from the modified pre-smoothing in (28). Unlike in (3), where pre-smoothing amounted to a local averaging of (oriented) gradient directions, the Gaussian convolution in (28) locally averages the (non-oriented) gradient flow-line directions (or, equivalently, level line directions).

Regarding the twofold effect of traditional pre-smoothing discussed at the begin of this section, this means that the first effect, creating a more regular set of edges, will still happen in a similar way. The second effect, fast removal of small-scale structures, cannot be expected to the same extent because opposing gradients do no longer cancel, thus leaving a higher average gradient magnitude to be estimated around small-scale structures.

4 Experiments

We use a test image with substantial additive Gaussian noise, Fig. 1a, to compare the gradient-descent-based regularised isotropic nonlinear diffusion (28) with traditional pre-smoothed Perona-Malik diffusion (3). Both PDEs are discretised by essentially the same explicit Euler forward scheme with central spatial differences, see e.g. [22, eq. (10)]. In all experiments, we use the diffusivity (2) with the same threshold, and the same standard deviation for pre-smoothing by Gaussian convolution.
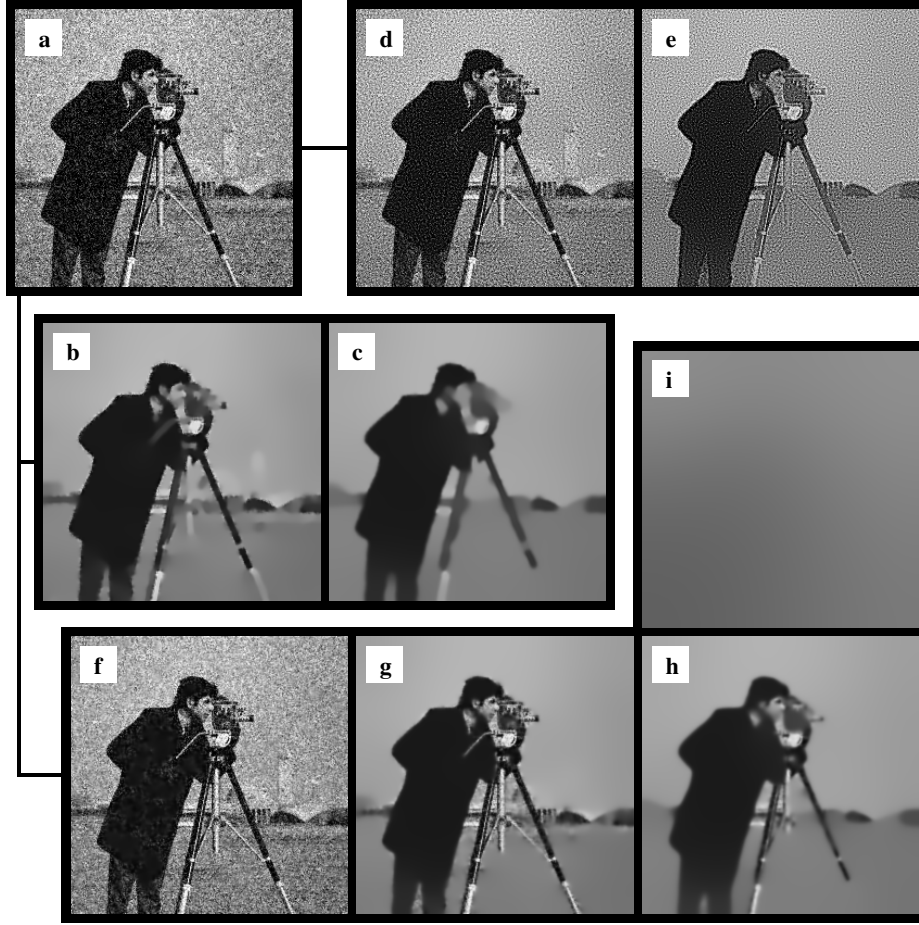
Fig. 1. Comparison of regularised isotropic nonlinear diffusion evolutions. In all experiments, an explicit forward scheme with central spatial differences and time step size $\tau = 0.25$ was employed, and the diffusivity function was fixed to (2) with $\lambda = 1$. All Gaussian convolutions used $\sigma = 2$. **Top left, a** Test image *Cameraman* (256×256 pixels) with Gaussian noise of standard deviation 40. – **Middle horizontal strip (reference):** Pre-smoothed Perona-Malik diffusion (3). **b** diffusion time $T = 125$. – **c** $T = 500$. – **Top right strip:** Unstable evolution (29). **d** Diffusion time $T = 125$ (same as b). – **e** $T = 1000$. – **Lower strip:** Evolution (28). **f** Diffusion time $T = 125$ (same as b). – **g** $T = 500$ (same as c). – **h** $T = 1000$ (visual effect comparable to c). – **i** $T = 10000$.

To start with, Fig. 1b, c show the result of (3). As expected, noise is removed quickly, and progressive simplification of image structures, and smoothing of edges takes place.

The next two frames, Fig. 1d, e show a failed evolution: In order to give an indication of the difficulties arising from anisotropy with inverse diffusion that occur in (26) with non-trivial L_2, we show this process with L_2 being Gaussian convolution, and $L_1 \equiv \mathrm{id}$,

i.e.

$$u_t = \mathrm{div}\Big(G_\sigma * \big(\Psi'(|G_\sigma * \boldsymbol{\nabla} u|^2)(G_\sigma * \boldsymbol{\nabla} u)\big)\Big) \,. \tag{29}$$

Using the same evolution times as in b, one observes small-scale oscillatory artifacts or ripples, Fig. 1d, that persist even at an evolution time when many meaningful image structures have already been removed, see frame e. Although it remains open whether this process can be stabilised by more advanced numerics, the experiment supports that (29) is not a convincing replacement for (3).

The remaining frames, Fig. 1f–i, demonstrate the evolution (28). Frames f and g show the same evolution times as b and c. As expected, small-scale noise takes longer to be removed but is eventually eliminated. In frame g, still many more small-scale details are preserved than in c although the larger-scale smoothing effect is not that much behind that in c. With doubling the evolution time, frame h, the overall image smoothing is visually comparable to that in c, but still with a stronger tendency to preserve small features. In the long run, of course, (28) converges to a flat homogeneous image, as the final frame Fig. 1i illustrates.

5   Summary

In this paper we have closed a long-standing – minor, but, in our opinion, relevant – gap in the theoretical framework of diffusion methods in image processing. We have proven that nonlinear isotropic and anisotropic diffusion with the commonly used pre-smoothing of image gradients is not the gradient descent of any energy functional, despite the fact that applications of these components in image processing applications are often justified with energy minimisation arguments.

This result raised the question whether the concepts of diffusion as a gradient descent, and pre-smoothing as an important ingredient for stability in diffusion methods can be reconciled. By analysing a generalised ansatz for an energy functional with pre-smoothing, we could single out pre-smoothing of the squared gradients as a candidate regularisation procedure for nonlinear diffusion that retains the gradient descent property.

References

1. Andreu-Vaillo, F., Caselles, V., Mazon, J.M.: Parabolic Quasilinear Equations Minimizing Linear Growth Functionals, Progress in Mathematics, vol. 223. Birkhäuser, Basel (2004)
2. Bellettini, G., Caselles, V., Novaga, M.: The total variation flow in R^N. Journal of Differential Equations **184**(2), 475–525 (2002)
3. Bellettini, G., Novaga, M., Paolini, M.: Convergence for long-times of a semidiscrete Perona-Malik equation in one dimension. Mathematical Models and Methods in Applied Sciences **21**(2), 241–265 (2011)
4. Bellettini, G., Novaga, M., Paolini, M., Tornese, C.: Convergence of discrete schemes for the Perona-Malik equation. Journal of Differential Equations **245**, 892–924 (2008)
5. Catté, F., Lions, P.L., Morel, J.M., Coll, T.: Image selective smoothing and edge detection by nonlinear diffusion. SIAM Journal on Numerical Analysis **32**, 1895–1909 (1992)

6. Dibos, F., Koepfler, G.: Global total variation minimization. SIAM Journal on Numerical Analysis **37**(2), 646–664 (2000)
7. Ghisi, M., Gobbino, M.: A class of local classical solutions for the one-dimensional Perona-Malik equation. Transactions of the American Mathematical Society **361**(12), 6429–6446 (2009)
8. Ghisi, M., Gobbino, M.: An example of global classical solution for the Perona-Malik equation. Communications in Partial Differential Equations **36**(8), 1318–1352 (2011)
9. Guidotti, P.: Anisotropic diffusions of image processing from Perona-Malik on. In: Ambrosio, L., Giga, Y., Rybka, P., Tonegawa, Y. (eds.) Variational Methods for Evolving Objects, Advanced Studies in Pure Mathematics, vol. 67, pp. 131–156. Mathematical Society of Japan, Tokyo (2015)
10. Kawohl, B., Kutev, N.: Maximum and comparison principle for one-dimensional anisotropic diffusion. Mathematische Annalen **311**, 107–123 (1998)
11. Nitzberg, M., Shiota, T.: Nonlinear image filtering with edge and corner enhancement. IEEE Transactions on Pattern Analysis and Machine Intelligence **14**, 826–833 (1992)
12. Nordström, N.: Biased anisotropic diffusion – a unified regularization and diffusion approach to edge detection. Image and Vision Computing **8**, 318–327 (1990)
13. Perona, P., Malik, J.: Scale space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence **12**, 629–639 (1990)
14. Weickert, J.: Theoretical foundations of anisotropic diffusion in image processing. Computing Supplement **11**, 221–236 (1996)
15. Weickert, J.: Anisotropic Diffusion in Image Processing. Teubner, Stuttgart (1998)
16. Weickert, J.: Coherence-enhancing diffusion filtering. International Journal of Computer Vision **31**(2/3), 111–127 (1999)
17. Weickert, J., Benhamouda, B.: A semidiscrete nonlinear scale-space theory and its relation to the Perona–Malik paradox. In: Solina, F., Kropatsch, W.G., Klette, R., Bajcsy, R. (eds.) Advances in Computer Vision, pp. 1–10. Springer, Wien (1997)
18. Weickert, J., Schnörr, C.: A theoretical framework for convex regularizers in PDE-based computation of image motion. International Journal of Computer Vision **45**(3), 245–264 (2001)
19. Welk, M., Steidl, G., Weickert, J.: Locally analytic schemes: A link between diffusion filtering and wavelet shrinkage. Applied and Computational Harmonic Analysis **24**, 195–224 (2008)
20. 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 (SSVM) 2019, Lecture Notes in Computer Science, vol. 11603, pp. 236–248. Springer, Cham (2019)
21. Welk, M., Weickert, J.: PDE evolutions for M-smoothers in one, two, and three dimensions. Journal of Mathematical Imaging and Vision **63**(2), 157–185 (2021)
22. Welk, M., Weickert, J., Gilboa, G.: A discrete theory and efficient algorithms for Forward-and-Backward diffusion filtering. Journal of Mathematical Imaging and Vision **60**(9), 1399–1426 (2018)
23. Zhang, K.: Existence of infinitely many solutions for the one-dimensional Perona-Malik model. Calculus of Variations and Partial Differential Equations **26**(2), 126–171 (2006)