

RESEARCH

Open Access

Image denoising by a direct variational minimization

Marko Janev^{1*}, Teodor Atanacković¹, Stevan Pilipović² and Radovan Obradović¹

Abstract

In this article we introduce a novel method for the image de-noising which combines a mathematically well-posedness of the variational modeling with the efficiency of a patch-based approach in the field of image processing. It based on a direct minimization of an energy functional containing a minimal surface regularizer that uses fractional gradient. The minimization is obtained on every predefined patch of the image, independently. By doing so, we avoid the use of an artificial time PDE model with its inherent problems of finding optimal stopping time, as well as the optimal time step. Moreover, we control the level of image smoothing on each patch (and thus on the whole image) by adapting the Lagrange multiplier using the information on the level of discontinuities on a particular patch, which we obtain by pre-processing. In order to reduce the average number of vectors in the approximation generator and still to obtain the minimal degradation, we combine a Ritz variational method for the actual minimization on a patch, and a complementary fractional variational principle. Thus, the proposed method becomes computationally feasible and applicable for practical purposes. We confirm our claims with experimental results, by comparing the proposed method with a couple of PDE-based methods, where we get significantly better denoising results specially on the oscillatory regions.

Keywords: Image denoising, Ritz method, calculus of variations, fractional gradient, anisotropic diffusion, Complementary Principle, saddle point, sparse frame, approximation error bound

1. Introduction

Since the work of Perona and Malik [1], PDE methods have been used for image processing, especially for denoising and stabilizing edges (see [1,2]). They were the first to replace an isotropic diffusion expressed through a linear heat equation with an anisotropic diffusion. Diffusion, in generally, is associated with an energy dissipating process. This process seeks the minima of an energy functional. For example, the well known total variation (TV) minimization model [3,4] is obtained in the case when the energy functional is equal to the TV norm of the image. Although these methods have been demonstrated to be able to achieve a good trade-off between the noise removal and the edge preservation, the resulting image in the presence of the noise often has a “blocky” look.

It is caused by the use of a second-order PDE modeling methods. In order to reduce the “blocky effect”, while

preserving sharp jump discontinuities, many other non-linear filters have been suggested in the literature (see [5-9]). In [5], You and Kaveh proposed a class of fourth-order PDEs that are obtained by the minimization of a functional given as an increasing function of the edge detector Δu . Since the second-order derivatives are zero if the image intensity function is planar, the class of fourth-order PDEs will evolve and settle down to a planar image, if the image support is infinite. This is important, since piecewise planar images look more natural than the step images which are stationary points of the particular nonconvex energy functional [5], whose minimization (after the application of gradient descent) leads to the second-order diffusion. The problem with the use of fourth-order equations is that it tends to leave the image with isolated black and white speckles (so called “speckle effect”) which may be characterized as pixels whose intensity values are either much larger or much smaller than those of the neighboring pixels as it is explained in [5]. Recently, fractional order PDEs have been studied and applied to the problem of image denoising. Bai and

* Correspondence: markojan@uns.ac.rs

¹Faculty of Engineering, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia

Full list of author information is available at the end of the article

Feng in [10] proposed the use of nonlinear anisotropic fractional diffusion equation based on the Euler-Lagrange (EL) equation of a cost functional which is an increasing function of the absolute value of fractional gradient of the image intensity function. They managed to interpolate between second- and fourth-order nonlinear anisotropic diffusion equation to obtain a more natural images.

Nevertheless, only in a very limited number of simple cases, EL PDE that corresponds to the target energy functional can be analytically solved [11,12]. Thus, in all related works, the actual minimization is conducted by the transition from an elliptic EL PDE, to a parabolic PDE with the artificial time. By doing so, a sort of low (in a general case nonlinear) filtering process [13] on a particular image is introduced. This process smoothes the image more and more in time. As a consequence, the problem of obtaining the optimal stopping time of a process emerges since the filtering can easily over-smooth the useful image features (edges, etc.). A similar problem appears with the choice of the optimal time step. Actually, a parabolic PDE model is obtained from the particular EL equation in the limiting process $(\varphi - \varphi_0)/\lambda \rightarrow 0$ as $\lambda \rightarrow 0$, where φ_0 is the noisy image, and λ is a Lagrange multiplier (see [11] or [13]). The role of λ , as the trade of between image smoothness and preservation of image features is lost: it becomes just a time step in the filtering process. The second problem related to the conventional PDE approach is that it is applied on a global image, so that the local image features are not sufficiently taken into account. In recent times, in the fields of Image Analysis, Processing and Synthesis, patch-based techniques emerged and meet with success. Defined as local square neighborhoods of image pixels, patches are very simple objects to work with, but they have the intrinsic ability to catch large-scale structures and textures present in natural images. Some recent image denoising methods are patch-based, such as “Non-Local Means” algorithm [14], and some of its derivatives [15,16].

In this work, we present a novel variational, and at the same time patch-based image smoothing method, which combines a mathematically well-posedness of the variational modeling with the efficiency of a patch-based approach. More-over, the proposed method is based on the direct variational minimization of the appropriate energy functional, which (as in [10]) involves fractional gradient. By doing so, we avoid problems of finding the optimal stopping time and the optimal time step. The role of λ is sustained and the actual minimization is conducted till it converges (with respect to the predefined error bound of the particular optimization method). We note that patch-based approach is also convenient to make the proposed direct variational method computationally feasible and applicable on real images. Actually, if working with the whole image, one

needs a huge approximation bases¹, which is not computationally feasible. According to this, we proceed as follows: The image is divided into relatively small overlapping patches, and the energy functional is minimized on each particular patch independently by using a direct variational minimization. As patches should not be too small, in order to capture enough relevant image features, the computational load would be still unacceptable for any real application if one calculates the minimizer in the whole orthonormal basis of the particular patch. Therefore, we approximate the true minimizer by using the Ritz variational method with a specially chosen trial functions [17]. In the sequel we call the set of those functions: the approximation generator. For that purpose, we derive the complementary fractional variational principle (CFVP) [17] for the corresponding energy functional. The CFVP gives us the explicit upper bound for the L_2 norm of the approximation error. Next, we proceeded with spatial discretization of the continuous model, i.e., we make transition from the continuous image to pixels.

Every discrete patch is analyzed in the chosen discrete over-complete dictionary (same for every patch) that has the sparsity property in the class of discrete images of interest. In this work, we use a simple discrete cosine transform (DCT) over-complete dictionary which possess a sparsity property in the class of images (see [18]). The elements of the actual approximation generator for a particular patch, are chosen to be those with the largest $K \ll N$ projections $\langle \varphi_0, \psi_n \rangle$, where N is size of the orthonormal basis and φ_0 is observed noisy image, so that the upper error bound obtained by a spatially discretized CFVP is below the predefined threshold. Thus, the computational load is additionally rapidly reduced, making the method applicable for practical purposes. Moreover, as we conduct the minimization of the target functional on each patch separately, we use different values for the Lagrange multiplier for each patch. The choice is based on the measure of nonsmoothness of the signal present on that particular patch which is obtained by an appropriate pre-processing. Thus, we obtain additional stronger regularization on the uniform and weaker regularization on the oscillatory patches, which significantly improves resulting image quality. It is an additional adaptive feature of the proposed method which is not applicable to anisotropic diffusion PDE modeling. Actually, for that purpose anisotropic diffusion uses only an appropriate edge stopping function (in our case “mini-mal surface”), which is also included in the proposed model. We also note that we use the functional that contains gradient of a fractional order, in order to gain all benefits of fractional approach (see [10]), in comparison to the classical gradient method or the methods of higher order, as it is previously explained.

The article is organized as follows: In Sect. 2, the basic facts about the fractional order anisotropic [1,10] diffusion, for image smoothing are mentioned. We emphasize the link between the energy functional and the actual parabolic PDE derived from its EL equation. In Sects. 3.1 and 3.2, we derive the CFVP for the energy functional with the fractional “minimal surface” regularizer. In Sect. 3.3, we use the spatially discretized model to obtain denoising on a single image patch, while in Sect. 3.4 we generalize this approach to obtain denoising of the whole image. In Sect. 4, we present experimental results and compare the proposed method with several anisotropic diffusion methods, and show that the proposed method better preserves image features, especially oscillatory regions. The conclusions are given in Sect. 5.

2. Fractional order anisotropic diffusion

In order to give an insight in the PDE approach to anisotropic image smoothing, which we tend to improve, we recall the derivation of the PDE used in [10]. We consider the following energy functional

$$J(\phi) = \int_{\mathbb{R}^2} L(\phi, \nabla_i^\gamma \phi) \, dx dy, \quad (2.1)$$

with the Lagrangian

$$L(\phi, \nabla_i^\gamma \phi) = \frac{1}{2}(\phi - \phi_0)^2 + \lambda \vartheta(|\nabla_i^\gamma \phi|^2), \quad \phi \in H^\gamma(\mathbb{R}^2), \quad (2.2)$$

where ϕ_0 is the initial noisy image. We assume that ϑ is the “minimal surface” edge stopping function $\vartheta(s) = (1 + s^2)^{1/2}$ (see [13]) and λ is the Lagrange multiplier, i.e., a regularization weight. For the sake of simplicity, the image is defined to be the function on the whole space \mathbb{R}^2 (as, for example in [19]). Thus, (2.2) is the functional which we will minimize by a patch-based direct variational minimization (PBDVM).

Denote by $\nabla_i^\gamma \phi = (D_x^\gamma \phi, D_y^\gamma \phi)$ the fractional gradient of order $\gamma > 0$ acting on an admissible scalar field ϕ : The partial fractional derivative operators D_x^γ and D_y^γ of order γ act as

$$D_x^\gamma \phi = \mathcal{F}^{-1}((i\xi_1)^\gamma \hat{\phi}(\xi)), \quad D_y^\gamma \phi = \mathcal{F}^{-1}((i\xi_2)^\gamma \hat{\phi}(\xi)) \quad (2.3)$$

where $\xi = (\xi_1, \xi_2)$, and $\mathcal{F} : L^2(\mathbb{R}^2) \rightarrow L^2(\mathbb{R}^2)$ is the Fourier-Plancherel transform operator. For the set of admissible functions ϕ , we chose $H^\gamma(\mathbb{R}^2)$, $\gamma > 0$. It is defined as

$$H^\gamma(\mathbb{R}^2) = \{\phi \in L^2(\mathbb{R}^2) \mid (1 + |\xi|^2)^{\gamma/2} \hat{\phi} \in L^2(\mathbb{R}^2)\}.$$

Note that $D_x^\gamma \phi$ and $D_y^\gamma \phi$ exist for every $\phi \in H^\gamma(\mathbb{R}^2)$ and belong to $L^2(\mathbb{R}^2)$ due to (2.3) and the fact that $(1 + |\xi|^2)^{\gamma/2} \hat{\phi} \in L^2(\mathbb{R}^2) \Leftrightarrow |\xi|^\gamma \hat{\phi} \in L^2(\mathbb{R}^2)$. We also recall

that the space $H^\gamma(\mathbb{R}^2)$, $\gamma > 0$ is dense in $L^2(\mathbb{R}^2)$. This follows from the fact that $C_0^\infty(\mathbb{R}^2)$ is dense in $L^2(\mathbb{R}^2)$. The density implies the existence of the adjoint operators $(D_x^\gamma)^*$ and $(D_y^\gamma)^*$, for D_x^γ and D_y^γ , defined on $H^\gamma(\mathbb{R}^2)$. This can be easily shown by using the property $\mathcal{F}^{-1} = \mathcal{F}^*$ implying that $(D_x^\gamma)^* = (-1)^\gamma D_x^\gamma$ (and similarly for $(D_y^\gamma)^*$). Consequently, the fractional divergence $\nabla_r^\gamma \mathbf{u} \stackrel{\text{def}}{=} ((D_x^\gamma)^* u_1, (D_y^\gamma)^* u_2)$ is well defined for $\mathbf{u} = (u_1, u_2) \in (H^\gamma(\mathbb{R}^2))^2$.

The Gateaux derivative of a functional J at ϕ in the direction η is defined as $J'(\phi, \eta) = \left. \frac{d}{d\varepsilon} \right|_{\varepsilon=0} J(\phi + \varepsilon\eta)$. Condition $J'(\phi, \eta) = 0$, for admissible functions η , gives the EL equation for (2.1). This is written as

$$\nabla_r^\gamma \left(\frac{\vartheta'(|\nabla_i^\gamma \phi|)}{|\nabla_i^\gamma \phi|} \nabla_i^\gamma \phi \right) = \frac{\phi_0 - \phi}{\lambda}. \quad (2.4)$$

As it is almost always infeasible to solve (2.4) directly [11,12], the problem is solving by the introduction of the artificial time $t \geq 0$. Let $\phi(\cdot, 0) = \phi_0$ and consider the right hand side of (2.4) as the discretization of ϕ_t . Using the time step λ in the limiting process $(\phi - \phi_0)/\lambda \rightarrow 0$ as $\lambda \rightarrow 0$, it follows that (2.4) becomes a fractional order anisotropic diffusion proposed in [10],

$$\phi_t = \nabla_r^\gamma \left(\frac{\vartheta'(|\nabla_i^\gamma \phi|)}{|\nabla_i^\gamma \phi|} \nabla_i^\gamma \phi \right), \quad \phi(\cdot, 0) = \phi_0. \quad (2.5)$$

As it is mentioned in Sect. 1, one is able to smooth the homogenous regions of the image and leave the most of the useful features, due to the anisotropic property of the edge stopping function (see [13]). Moreover, the use of the fractional gradient provides a natural interpolation between the second- and the fourth-order diffusion equations [5,10].

3. Patch-based direct variational minimization

Fractional anisotropic diffusion, as well as all other diffusion PDE models, has a problem of finding the optimal stopping time, as well as the optimal time step (for more details, see for example [11]). Moreover, as it is designed to be applied on a global image, it does not possess the benefits of the patch-based approach. It means that it can not model the local image features such as textures and oscillatory regions of the image such efficiently as the patch-based methods. We propose the novel PBDVM of the energy functional (2.1), which deals with those problems, as it is mentioned in Sect. 1. For that purpose, we divide image into overlapping patches, perform discretization of the continuous model and use a Ritz variational method for finding the minimizer of (2.1) on every particular patch of the image. Ritz variational method (we consider

discretized version) for the particular discretized patch of the image of format $N = M \times M$ is based on representing the minimizer of the functional (2.1) in some fixed basis $\{\psi_i\}_1^N$ of \mathbb{R}^N . Minimizer is represented in the form $\hat{\phi} = \sum_1^N \alpha_i \psi_i$, replaced in (2.1). Thus, one has to find $\alpha^* = \operatorname{argmin}_\alpha J(\hat{\phi})$, $\alpha = (\alpha_1, \dots, \alpha_N)$. Our intention is to find as small as possible approximation generator and still obtain satisfactory approximation. For that purposes, we use the generator $\{\psi_i\}_1^D$ of the space \mathbb{R}^N , with the size D much larger (at least several times) than N ($D \gg N$), which additionally possess the sparsity property in the space of image patches of the particular class of images under consideration. It means that any patch which belongs to an image from that particular class, can be represented in $\{\psi_i\}_1^D$ with a small number of nonzero coefficient α_i^* , i.e., that $\|\alpha^*\|_0 \ll N$, where $\|\cdot\|_0$ stands for l_0 norm (see [18] for more details). In the sequel we call the approximation generator that possess sparsity property “over-complete dictionary” and its elements “atoms”. It is common in the literature (see [18]). As a particular over-complete dictionary in our experiment in Section 4, we use DCT over-complete dictionary which has the sparsity property for a broad class of images (we can say almost all used in real applications). Note that the representation in the over-complete dictionary is not unique. In addition, we deliver the CFVP for the functional (2.1). This gives the L_2 upper bound of the approximation error. In applications, in Sect. 3.3, we use the discretized version of that estimate (which gives the l_2 approximation error bound) in order to find as small as possible approximation generator with sufficiently small approximation error for every patch. The procedure of finding the desired approximation generator, and the usage of CFVP for that purpose are explained in details in Sect. 3.3. Applying the proposed procedure, we significantly reduce the number of parameters for which we search the minimum of the target functional using the Ritz method and thus significantly reduce the computational complexity. The computational complexity is of order $O(PN^3)$ for minimization in the whole orthonormal bases of size N , which is not computationally feasible in real applications. It is reduced to $O(P\bar{K}^3)$ when using the proposed CFVP, where $\bar{K} \ll N$ is the average size of the approximation generator (averaged for all patches) and P is the overall number of overlapped patches. It is the consequence of the fact that Newton-based optimization methods that we use to obtain α^* , uses the inversion of the Hessian of the target function and the order of complexity for the inversion of $n \times n$ matrix is $O(n^3)$.

3.1. Existence of the CFVP for the target energy functional

To formulate CFVP, we proceed by derivation of the canonical Hamiltonian equations. We introduce a new variable (generalized impulse) as

$$\mathbf{u} = \frac{\partial L}{\partial \nabla_i^\gamma \phi} = \frac{\vartheta'(|\nabla_i^\gamma \phi|)}{|\nabla_i^\gamma \phi|} \nabla_i^\gamma \phi. \quad (3.1)$$

Now from the EL equation (2.4), we have the first canonical equation

$$\nabla_r^\gamma \mathbf{u} = \frac{\phi_0 - \phi}{\lambda}. \quad (3.2)$$

As $\vartheta'(s)/s = 1/\sqrt{1+s^2}$, it follows

$$\mathbf{u} = \frac{\nabla_i^\gamma \phi}{\sqrt{1 + |\nabla_i^\gamma \phi|^2}} \quad (3.3)$$

and consequently

$$\mathbf{u} \cdot \mathbf{u} = \frac{|\nabla_i^\gamma \phi|^2}{1 + |\nabla_i^\gamma \phi|^2} \Rightarrow 1 + |\nabla_i^\gamma \phi|^2 = \frac{1}{1 - |\mathbf{u}|^2}. \quad (3.4)$$

From (3.3) and (3.4) now follows the second canonical equation

$$\nabla_i^\gamma \phi = \frac{\mathbf{u}}{\sqrt{1 - |\mathbf{u}|^2}}, \quad (3.5)$$

where $|\mathbf{u}| < 1$ is satisfied by definition (see (3.3)).

We briefly introduce some terms which will be used in the sequel. We denote $T = \nabla_i^\gamma$, and its adjoint operator $T^* = \nabla_r^\gamma$, so that

$$\langle \mathbf{u}, T\phi \rangle_{L^2(\mathbb{R}^2)} = \langle T^*\mathbf{u}, \phi \rangle_{L^2(\mathbb{R}^2)} \quad (3.6)$$

for admissible $\phi \in H^\gamma(\mathbb{R}^2)$ and $\mathbf{u} \in (H^\gamma(\mathbb{R}^2))^2$.

We define the potential $I(\mathbf{u}, \phi)$ as

$$I(\mathbf{u}, \phi) \stackrel{\text{def}}{=} \langle \mathbf{u}, T\phi \rangle_{L^2(\mathbb{R}^2)} - W(\mathbf{u}, \phi) = \langle T^*\mathbf{u}, \phi \rangle_{L^2(\mathbb{R}^2)} - W(\mathbf{u}, \phi), \quad (3.7)$$

where we use (3.6)². The functional W is defined as

$$W(\mathbf{u}, \phi) = \int_{\mathbb{R}^2} H(\mathbf{u}, \phi) dx dy, \quad (3.8)$$

where $H(\mathbf{u}, \phi) = \mathbf{u} \cdot T\phi - L(\phi, T\phi)$ is Hamiltonian.

Now, we have

Proposition 1. *Assuming the existence and the uniqueness of the minimizer for the functional (2.1), there exists a complementary variational principal for (2.1).*

Proof:

In terms of the theory presented in [17] (see [17], p. 94), the canonical equations for the boundary value problem (for which we seek the complementary principle) can be derived using (3.2) and (3.5) as

$$T\phi = W_{\mathbf{u}}(\mathbf{u}, \phi) = f_1(\mathbf{u}), \quad T^*\mathbf{u} = W_{\phi}(\mathbf{u}, \phi) = f_2(\phi), \quad (3.9)$$

where

$$f_1(\mathbf{u}) = \frac{\mathbf{u}}{\sqrt{1 - |\mathbf{u}|^2}}, \quad f_2(\phi) = \frac{\phi_0 - \phi}{\lambda}. \quad (3.10)$$

Theorem 5.3.1, p. 95. [17] states that sufficient conditions for $I(\mathbf{u}, \varphi)$ to be a convex-concave saddle functional (equivalently for a complementary principle) are that W is strictly convex in \mathbf{u} and is concave in φ . Since W is differentiable on its domain, it is equivalent to say that $W_{\mathbf{u}} = f_1(\mathbf{u})$ is strictly monotone and that $W_{\varphi} = f_2(\varphi)$ is anti-monotone. It is clear that f_2 is anti-monotone, as

$$\begin{aligned} \langle \phi_1 - \phi_2, f_2(\phi_1) - f_2(\phi_2) \rangle_{L^2(\mathbb{R}^2)} &= \left\langle \phi_1 - \phi_2, \frac{\phi_0 - \phi_1}{\lambda} - \frac{\phi_0 - \phi_2}{\lambda} \right\rangle_{L^2(\mathbb{R}^2)} \\ &= -\frac{1}{\lambda} \langle \phi_1 - \phi_2, \phi_1 - \phi_2 \rangle_{L^2(\mathbb{R}^2)} \\ &\leq 0. \end{aligned} \quad (3.11)$$

Next, we show that $f_1(\mathbf{u})$ is strictly monotone. By the mean value theorem, there holds

$$F_1 \stackrel{\text{def}}{=} \langle \mathbf{u} - \mathbf{v}, f_1(\mathbf{u}) - f_1(\mathbf{v}) \rangle_{L^2(\mathbb{R}^2)} = \langle \delta \mathbf{u}, f_1'(\mathbf{w}) \delta \mathbf{u} \rangle_{L^2(\mathbb{R}^2)} \quad (3.12)$$

where $\delta \mathbf{u} = \mathbf{u} - \mathbf{v}$, and $\mathbf{w} = \mathbf{v} + \varepsilon \delta \mathbf{u}$, for some $\varepsilon \in (0, 1)$.

Since

$$f_1'(\mathbf{w}) = \frac{1}{(1 - |\mathbf{w}|^2)^{3/2}} > 0, \quad |\mathbf{w}| < 1, \quad (3.13)$$

it follows that

$$\begin{aligned} F_1 &= \left\langle \delta \mathbf{u}, \frac{1}{(1 - |\mathbf{w}|^2)^{3/2}} \delta \mathbf{u} \right\rangle_{L^2(\mathbb{R}^2)} \\ &= \int_{\mathbb{R}^2} \frac{1}{(1 - |\mathbf{w}|^2)^{3/2}} (\delta \mathbf{u} \cdot \delta \mathbf{u}) \, dx dy > 0, \end{aligned} \quad (3.14)$$

for all $|\mathbf{w}| < 1$ and $\delta \mathbf{u} \neq \mathbf{0}$. This implies that $f_1(\mathbf{u})$ is strictly monotone. The proof is completed. \square

3.2. The L^2 bound of the approximation error

From the fact that $W(\mathbf{u}, \varphi) = F_1(\mathbf{u}) + F_2(\varphi)$, where $F_1' = f_1$ and $F_2' = f_2$, and (3.10), it follows that

$$W(\mathbf{u}, \varphi) = - \int_{\mathbb{R}^2} \sqrt{1 - |\mathbf{u}|^2} \, dx dy - \frac{1}{2\lambda} \int_{\mathbb{R}^2} (\varphi_0 - \varphi)^2 \, dx dy \quad (3.15)$$

Now, (3.7) and (3.15) imply that

$$\begin{aligned} I(\mathbf{u}, \varphi) &= \int_{\mathbb{R}^2} \mathbf{u} \cdot \nabla_r^\gamma \varphi \, dx dy + \int_{\mathbb{R}^2} \sqrt{1 - |\mathbf{u}|^2} \, dx dy \\ &\quad + \frac{1}{2\lambda} \int_{\mathbb{R}^2} (\varphi - \varphi_0)^2 \, dx dy \\ &= \int_{\mathbb{R}^2} \varphi \nabla_r^\gamma \mathbf{u} \, dx dy + \int_{\mathbb{R}^2} \sqrt{1 - |\mathbf{u}|^2} \, dx dy \\ &\quad + \frac{1}{2\lambda} \int_{\mathbb{R}^2} (\varphi - \varphi_0)^2 \, dx dy. \end{aligned} \quad (3.16)$$

We denote

$$J(\varphi_1) = I(\mathbf{u}_1, \varphi_1), \quad G(\mathbf{u}_2) = I(\mathbf{u}_2, \varphi_2). \quad (3.17)$$

in which the terms J and G are primal and dual functionals, respectively, and

$$\Omega_1 = \{(\mathbf{u}_1, \varphi_1) : I_{\mathbf{u}} = 0\}, \quad \Omega_2 = \{(\mathbf{u}_2, \varphi_2) : I_{\varphi} = 0\}. \quad (3.18)$$

Note that $I_{\mathbf{u}} = 0$ and $I_{\varphi} = 0$ are equivalent to $T\varphi = W_{\mathbf{u}}(\mathbf{u}, \varphi) = f_1(\mathbf{u})$ and $T^* \mathbf{u} = W_{\varphi}(\mathbf{u}, \varphi) = f_2(\varphi)$, respectively, i.e., (3.5) or (3.2) are satisfied, respectively.

Let $(\mathbf{u}_1, \varphi_1) \in \Omega_1$. Then

$$|\nabla_r^\gamma \varphi_1|^2 = \frac{|\mathbf{u}_1|^2}{1 - |\mathbf{u}_1|^2} \Rightarrow |\mathbf{u}_1|^2 = \frac{|\nabla_r^\gamma \varphi_1|^2}{1 + |\nabla_r^\gamma \varphi_1|^2}. \quad (3.19)$$

From (3.16) and $(\mathbf{u}_1, \varphi_1) \in \Omega_1$ (i.e., (3.5) is satisfied) we have

$$\begin{aligned} I(\mathbf{u}_1, \varphi_1) &= \int_{\mathbb{R}^2} \frac{|\mathbf{u}_1|^2}{\sqrt{1 - |\mathbf{u}_1|^2}} \, dx dy + \int_{\mathbb{R}^2} \sqrt{1 - |\mathbf{u}_1|^2} \, dx dy \\ &\quad + \frac{1}{2\lambda} \int_{\mathbb{R}^2} (\varphi_0 - \varphi_1)^2 \, dx dy. \end{aligned} \quad (3.20)$$

By substituting (3.19) into (3.20), we get

$$J(\varphi_1) = \int_{\mathbb{R}^2} \sqrt{1 + |\nabla_r^\gamma \varphi_1|^2} \, dx dy + \frac{1}{2\lambda} \int_{\mathbb{R}^2} (\varphi_0 - \varphi_1)^2 \, dx dy \quad (3.21)$$

Note that $E(\varphi) = \lambda J(\varphi)$ and that the minimizer φ^* for $J(\varphi)$ is also the minimizer for $E(\varphi)$, so that J is actually a primal functional.

We also derive the relation for the dual functional. As $(\mathbf{u}_2, \varphi_2) \in \Omega_2$ implies (3.2), i.e., $\varphi_2 = \varphi_0 - \lambda \nabla_r^\gamma \mathbf{u}_2$, replacing into (3.16), we get

$$\begin{aligned} G(\mathbf{u}_2) &= \int_{\mathbb{R}^2} \varphi_0 \nabla_r^\gamma \mathbf{u}_2 \, dx dy - \frac{\lambda}{2} \int_{\mathbb{R}^2} (\nabla_r^\gamma \mathbf{u}_2)^2 \, dx dy \\ &\quad + \int_{\mathbb{R}^2} \sqrt{1 - |\mathbf{u}_2|^2} \, dx dy. \end{aligned} \quad (3.22)$$

Next we derive a bound on the L^2 norm of the approximation error of (2.1). We state it as:

Proposition 2. *The L^2 norm of the approximation error $\delta\phi = \hat{\phi} - \phi$, where $\hat{\phi}$ is approximate, and ϕ true minimizer of the functional $J(\varphi)$ given by (3.21), has an upper bound:*

$$\|\delta\phi\|_{L^2}^2 \leq \lambda(J(\varphi_1) - G(\mathbf{u}_2)). \quad (3.23)$$

Proof:

Proposition (1) actually states that $I(\mathbf{u}, \varphi)$ has a saddle point, i.e., that for $(\mathbf{u}_1, \varphi_1) \in \Omega_1$ and $(\mathbf{u}_2, \varphi_2) \in \Omega_2$, so we have

$$G(\mathbf{u}_2) \leq G(\mathbf{u}) = I(\mathbf{u}, \phi) = J(\phi) \leq J(\phi_1). \quad (3.24)$$

Next we determine the second variation for J . The second partial derivatives are

$$\frac{\partial^2 L}{\partial^2 \phi} = \frac{1}{\lambda} \text{ and } \frac{\partial^2 L}{\partial^2 (\nabla_i^\gamma \phi)} = \frac{1}{(1 + |\nabla_i^\gamma \phi|^2)^{3/2}}, \quad (3.25)$$

so it holds that

$$\delta^2 J(\phi) = \int_{\mathbb{R}^2} \left[\frac{1}{\lambda} (\delta\phi)^2 + \frac{|\delta(\nabla_i^\gamma \phi)|^2}{(1 + |\nabla_i^\gamma \phi|^2)^{3/2}} \right] dx dy. \quad (3.26)$$

This implies

$$\delta^2 J(\phi) \geq \frac{1}{\lambda} \|\delta\phi\|_{L_2}^2. \quad (3.27)$$

Now, for $\xi = \phi + t(\phi_1 - \phi)$, for some $t \in (0, 1)$, from (3.24) and (3.27) we have

$$\begin{aligned} J(\phi_1) - G(\mathbf{u}_2) &\geq J(\phi_1) - J(\phi) \\ &= \delta J(\phi) + \delta^2 J(\phi)|_\xi \\ &= \delta^2 J(\phi)|_\xi \\ &\geq \frac{1}{\lambda} \|\delta\phi\|_{L_2}^2. \end{aligned} \quad (3.28)$$

which completes the proof. \square

3.3. Spatial discretization of the continuous model and the single patch denoising procedure

We proceed with the spatial discretization of the continuous model. It involves discretization of the primal and the dual energy functionals given by (3.21) and (3.22), respectively, and the inequality (3.23). We consider the discretized patch φ obtained from the fixed continuous $L \times L$ patch (which is cropped from the original image) by using the uniform $M \times M$ grid as $\varphi(p, l) = \varphi(p\Delta x, l\Delta y)$, for $p, l \in \{0, 1, \dots, M-1\} = I$, where $\Delta x = \Delta y = L/M$. In order to apply DFT transform, we actually impose a periodicity assumption, i.e., the continuous patch is defined on the rectangular domain $\Pi \subset \mathbb{R}^2$, where $\Pi = [a, b]^2$, $b-a = L$, and it is periodically prolonged on the whole \mathbb{R}^2 . For the spatial discretization (and thus approximation) of $D_x^\gamma \phi$ and $D_y^\gamma \phi$, the fractional central differences [10] are used as follows:

$$\widetilde{D}_x^\gamma u = F^{-1} \left[\left(1 - e^{-\frac{-2\pi i \omega_1}{M}} \right)^\gamma e^{\frac{i\pi \gamma \omega_1}{M}} F[\phi(p, l)] \right] \quad (3.29)$$

$$\widetilde{D}_y^\gamma u = F^{-1} \left[\left(1 - e^{-\frac{-2\pi i \omega_2}{M}} \right)^\gamma e^{\frac{i\pi \gamma \omega_2}{M}} F[\phi(p, l)] \right] \quad (3.30)$$

where $\omega_1, \omega_2 \in \{0, 1, \dots, M-1\}$ are DFT frequencies that correspond to the spatial coordinate x and y , respectively, and F is DFT operator.

Operator \widetilde{D}_x^γ has the form $\widetilde{D}_x^\gamma = F^{-1} \circ K_1 \circ F$, where

$$K_1 = \text{diag} \left\{ \left(1 - e^{-\frac{-2\pi i \omega_1}{M}} \right)^\gamma e^{\frac{i\pi \gamma \omega_1}{M}} \right\} \quad (3.31)$$

is a diagonal matrix.

The adjoint operator $(\widetilde{D}_x^\gamma)^*$ of \widetilde{D}_x^γ is expressed [10] by

$$(\widetilde{D}_x^\gamma)^* = (F^{-1} \circ K_1 \circ F)^* = F^{-1} \circ K_1^* \circ F. \quad (3.32)$$

Similarly,

$$\widetilde{D}_y^\gamma = F^{-1} \circ K_2 \circ F, \quad (\widetilde{D}_y^\gamma)^* = F^{-1} \circ K_2^* \circ F, \quad (3.33)$$

where

$$K_2 = \text{diag} \left\{ \left(1 - e^{-\frac{-2\pi i \omega_2}{M}} \right)^\gamma e^{\frac{i\pi \gamma \omega_2}{M}} \right\}. \quad (3.34)$$

The relations (3.29)-(3.34) are used for the actual spatial discretization of the whole problem described by (3.21), (3.22), and (3.23). We thus get the following spatially discretized relations:

$$\begin{aligned} \widetilde{J}(\phi_1) &= \sum_{p,l \in I} \sqrt{1 + (\widetilde{D}_x^\gamma \phi_1(p, l))^2 + (\widetilde{D}_y^\gamma \phi_1(p, l))^2} \\ &\quad + \frac{1}{2\lambda} \sum_{p,l \in I} (\phi_1(p, l) - \phi_0(p, l))^2, \end{aligned} \quad (3.35)$$

$$\begin{aligned} \widetilde{G}(\mathbf{u}_2) &= \sum_{p,l \in I} \phi_0(p, l) ((\widetilde{D}_x^\gamma)^* u_2^{(1)}(p, l) + (\widetilde{D}_y^\gamma)^* u_2^{(2)}(p, l)) \\ &\quad - \frac{\lambda}{2} \sum_{p,l \in I} ((\widetilde{D}_x^\gamma)^* u_2^{(1)}(p, l) + (\widetilde{D}_y^\gamma)^* u_2^{(2)}(p, l))^2 \\ &\quad + \sum_{p,l \in I} \sqrt{1 - u_2^{(1)}(p, l)^2 - u_2^{(2)}(p, l)^2} \end{aligned} \quad (3.36)$$

where $\mathbf{u}_2 = (u_2^{(1)}, u_2^{(2)})$, including the estimate

$$\|\delta\phi\|_2^2 \leq \lambda (\widetilde{J}(\phi_1) - \widetilde{G}(\mathbf{u}_2)) \quad (3.37)$$

which correspond to the continuous relations (3.21), (3.22), and (3.23), respectively.

In the actual denoising procedure, we will also use the discretized version of the relation (3.1):

$$\mathbf{u} = (u^{(1)}, u^{(2)}) = \frac{(\widetilde{D}_x^\gamma \phi(p, l), \widetilde{D}_y^\gamma \phi(p, l))}{\sqrt{1 + (\widetilde{D}_x^\gamma \phi(p, l))^2 + (\widetilde{D}_y^\gamma \phi(p, l))^2}}. \quad (3.38)$$

For a particular discretized $M \times M$ patch ϕ , we proceed with the single patch denoising procedure as follows. We chose the over-complete dictionary $\{\psi_n\}_1^D$ as the fixed approximation generator (see [18]), where $D \gg N$, for $N = M^2$. (We use the DFT over-complete dictionary in our experiments.) We analyze observed noisy patch ϕ_0 in $\{\psi_n\}_1^D$. As noisy patch ϕ_0 and atoms ψ_n are $M \times M$ matrices, when we analyze ϕ_0 by making the projections $\langle \phi_0, \psi_n \rangle$, we use the classical l_2 scalar product in \mathbb{R}^N . Here we present real matrices ϕ_0 and ψ_n as vectors of length N . Then we take the predefined K number ($K \ll N$) of those atoms ψ_i from $\{\psi_n\}_1^D$ with the largest projections $|\langle \phi_0, \psi_i \rangle|$, thus fixing the subspace where the approximation belongs. Next, we proceed with the Ritz method. We minimize (3.35) with respect to $\alpha = (\alpha_1, \dots, \alpha_K)$, where we impose the approximate solution as $\hat{\phi} = \sum_{i=1}^K \alpha_i \psi_i$. Actually, as $\tilde{D}_x^y \hat{\phi} = \sum_{i=1}^K \alpha_i \tilde{D}_x^y \psi_i$ and the terms $\tilde{D}_x^y \psi_i$ can be calculated *a priori* (the same holds for $\tilde{D}_y^x \hat{\phi}$), the functional (3.35) becomes the function $\tilde{J}(\alpha)$ of $\alpha \in \mathbb{R}^K$. In this way we get the unconstrained nonlinear optimization problem. As both, the gradient $\nabla_\alpha \tilde{J}(\alpha)$ and the Hessian $\nabla_\alpha^2 \tilde{J}(\alpha)$ are defined and bounded, we apply classical Newton line search nonlinear optimization method to obtain the numerical solution for the minimizer α^* of $\tilde{J}(\alpha)$. We start from the initial $\alpha_0 = (\langle \phi_0, \psi_1 \rangle, \dots, \langle \phi_0, \psi_K \rangle)$, in order to be initially as close as possible to α^* . This is reasonable since the assumption is that the additive Gaussian noise, superposed with the image, has a not so big variance. When we find $\hat{\phi}$, we calculate \hat{u} using the relation (3.38). Using the estimate (3.37), we obtain the l^2 upper bound for the approximation error induced by the simple shrinkage procedure⁴. If it is above the predefined threshold ε , the approximation error could be greater than ε . So we add next $m \leq K$ vectors $\{\psi_i\}_{K+1}^{K+m}$, with the next m largest values of $|\langle \phi_0, \psi_i \rangle|$ in the approximation generator of that particular patch. Then we set $\hat{\phi}_{\text{new}} = \hat{\phi}_{\text{old}} + \sum_{i=K+1}^{K+m} \alpha_i \psi_i$, where we obtained $\hat{\phi}_{\text{old}}$ in the previous step. We optimize $\tilde{J}(\alpha) = \tilde{J}(\hat{\phi}_{\text{new}})$ with respect to $(\alpha_{K+1}, \dots, \alpha_{K+m})$ and get new, suboptimal, approximate $\hat{\phi}_{\text{new}}$ ⁵. We continue with adding vectors as previously explained, until l^2 upper bound for the approximation error becomes lower than ε . Experimental results presented in the next section show that the introduction of CFVP makes the method computationally feasible as the average number of coefficients \bar{K} (by averaging on all patches) that we have to optimize, for the satisfactory approximation error obtained for every patch, is much (several times) smaller than N .

3.4. Generalization from a single patch to the whole image denoising

In order to effectively apply patch-based approach to the whole image denoising, but also eliminate the border

effects that would emerge, we formulate the overlapped patch approach. We actually consider the generalized minimization task

$$\alpha_{ij}^*, \phi^* = \underset{\alpha_{ij}, \phi}{\operatorname{argmin}} \left\{ \mu \|\phi - \phi_0\|_2^2 + \sum_{ij} \lambda_{ij} \vartheta(|\tilde{\nabla}_i^y \hat{\phi}_{ij}|) + \sum_{ij} \|\hat{\phi}_{ij} - R_{ij} \phi_0\|_2^2 \right\} \quad (3.39)$$

where $\hat{\phi}_{ij} = \Psi \alpha_{ij}$ is reconstructed patch of size $N = M \times M$ with coordinates i and j , with coefficients α_{ij} , and λ_{ij} are Lagrange multipliers for ij th patch. The terms ϕ_0 and ϕ represent the noisy and the target image, respectively, and Ψ is some fixed basis for \mathbb{R}^N , or over-complete dictionary for space of patches of size $M \times M$. We denote $\tilde{\nabla}_i^y = (\tilde{D}_x^y, \tilde{D}_y^y)$. So it holds that $\tilde{\nabla}_i^y \hat{\phi}_{ij} = \sum_{l=1}^K \alpha_{ij}^{(l)} \tilde{\nabla}_i^y \psi_l$, where $\tilde{\nabla}_i^y \psi_l$ are fixed. The operator R_{ij} presents the matrix that crops the ij th patch from the image. The third and the second term in (3.39) correspond to the sum of the patch square errors and the regularization terms, respectively. For fixed ij th patch, $\lambda_{ij} \vartheta(|\tilde{\nabla}_i^y \hat{\phi}_{ij}|) + \|\hat{\phi}_{ij} - R_{ij} \phi_0\|_2^2$ actually corresponds to the discretized version of the primal functional (2.1), and thus to (3.35), defined on the ij th patch. The first term in (3.39) is the log-likelihood global force, that does not allow the target image (i.e., minimizer) to differ substantially from the original noisy image. The level of that proximity can be controlled using $\mu \geq 0$.

Instead of minimizing (3.39) simultaneously for all α_{ij} and ϕ , we perform two step modal approach, thus obtaining the suboptimal solution. First, we optimize α_{ij} for fixed $\phi = \phi_0$, and then optimize ϕ for previously determine α_{ij} . In the first step we obtain

$$\alpha_{ij}^* = \underset{\alpha_{ij}}{\operatorname{argmin}} \lambda_{ij} \vartheta(|\tilde{\nabla}_i^y \hat{\phi}_{ij}|) + \|\hat{\phi}_{ij} - R_{ij} \phi_0\|_2^2 \quad (3.40)$$

for all ij th patches separately. Thus, we proceed with the single patch de-noising procedure, where we fix over-complete dictionary Ψ , obtaining $\alpha_{ij}^* = \operatorname{argmin}_{\alpha_{ij}} \tilde{J}(\alpha_{ij})$ for each ij th patch independently, where \tilde{J} is defined by (3.35), and it is described in details in the previous section. In the second step, by setting $\alpha_{ij} = \alpha_{ij}^*$, we obtain

$$\phi^* = \underset{\phi}{\operatorname{argmin}} \left\{ \mu \|\phi - \phi_0\|_2^2 + \sum_{ij} \|\hat{\phi}_{ij} - R_{ij} \phi\|_2^2 \right\} \quad (3.41)$$

which we solve for ϕ and obtain the close form solution [18]:

$$\phi^* = \left(\mu I + \sum_{ij} R_{ij}^T R_{ij} \right)^{-1} \left(\mu \phi_0 + \sum_{ij} R_{ij}^T \Psi \alpha_{ij}^* \right). \quad (3.42)$$

which is the final denoising result. Our additional goal is to use different values of λ for every patch, based on

information on the level of its smoothness, in order to obtain better edge preservation. We extract that information by performing the synthesis of the auxiliary patch ϕ_{pp} in the subspace of \mathbb{R}^N , generated by $(\psi_1, \dots, \psi_{\tilde{K}})$ obtained by taking first \tilde{K} atoms ψ_i with largest $\langle \phi_0, \psi_j \rangle$, where ϕ_0 is original noisy patch and $\tilde{K} \ll N$ is fixed. If we denote $c_j = \langle \phi_0, \psi_j \rangle$, for $j = 1, \dots, \tilde{K}$, the synthesized auxiliary patch ϕ_{pp} is given using the pseudo inverse of the matrix $\Psi_{\tilde{K}} = [\psi_1, \dots, \psi_{\tilde{K}}]$ containing the elements ψ_i , $i = 1, \dots, \tilde{K}$ as the column vectors:

$$\phi_{pp} = \Psi_{\tilde{K}}^\dagger c = (\Psi_{\tilde{K}}^* \Psi_{\tilde{K}})^{-1} \Psi_{\tilde{K}}^* c. \quad (3.43)$$

This pre-processing can be interpreted as a simple shrinkage of the noisy patch, which “cuts off” some amount of noise, but without significant degradation of important image features. Next we calculate the modulus of gradient $|\nabla \phi_{pp}|$ of the auxiliary patch ϕ_{pp} using (3.29) and (3.30). If it is below the predefined, empirically chosen threshold T_1 , we consider the region predominantly homogeneous, and set $\lambda = \lambda_1$, where predefined λ_1 is large. If it is between T_1 and T_2 , we set $\lambda = \lambda_2$ for some medium λ_2 . Finally, if it is above T_2 , we set $\lambda = \lambda_3$ for some small λ_3 . By doing so, we smooth more intensively homogeneous, and less intensively oscillatory regions of the target image.

4. Experimental results

Experimental results of this section are obtained by applying the proposed method on real images. All test images of size 512×512 , were corrupted with the white Gaussian additive white noise of various standard deviation σ . We compare the proposed method with the several anisotropic diffusion methods. One is the classical Perona-Malik [1] anisotropic diffusion, on which we refer as baseline algorithm 1. The second one is the method proposed by Bai and Feng (also mentioned in Sect. 1) which is reported to give good results in preserving edges. We refer to it as the baseline algorithm 2. The third one is the method proposed by Cao and Yin [20], which belongs to the class of parabolic-hyperbolic equations applied to image smoothing. It is also reported to have very good edge preservation results. We refer to it as the baseline algorithm 3. The fourth one is the recent nonlocal diffusion proposed by Guidotti and Lambers [21]. For the first three baseline algorithms, the edge stopping function was set to $\vartheta(s) = (1 + s^2)^{1/2}$, as it is also used in the proposed model. We use the PSNR as the objective measure of the denoising quality, as it is widespread throughout the image processing community. As PSNR is proved to be inconsistent with the human eye perception, we also use more profound structural similarity index measure (SSIM) [22] to additionally validate the proposed method, in

comparison to the baseline algorithms. The resulting images for the baseline algorithms was chosen so that the best PSNR is reached, as it is usually done in the literature. Results presented show that the proposed method significantly outperform all baseline methods in the means of both, the PSNR and the SSIM measure. We have to mention that in the actual exploitation the referent, i.e., original image is not available. Thus, one has to fix the number of iterations of the used diffusion algorithm. This possibly causes the stronger over-smoothing effect on the target image. So the actual result would be even more in favor of the proposed method.

In Figure 1a,b,c, original test images are presented, while the corresponding noisy images are presented in Figure 2a,b,c, for $\sigma = 20$. We tested the proposed algorithm on “Barbara”, “Fishing boat”, and “Bicycle” test images, with three different values of Gaussian white noise standard deviation σ .

In Figure 3, processing results on the “Barbara” test image are presented for $\sigma = 20$. Results for the baseline algorithms 1, 2, 3, and 4 are presented in Figure 3a,b,c,d, respectively, while the processing result of the proposed method is presented in Figure 3e. The same algorithm order and settings, but for different test images is presented on Figures 4 (“Fishing boat”) and 5 (“Bicycle”). For all the experiments, the order of the fractional gradient in the baseline method 2 and the proposed algorithm was set to be $\gamma = 1.8$, as it is the best reported in [10]. For all the experiments, parameter λ for the baseline algorithm 3 was set to $\lambda = 20$ which is the best reported in [20], while the parameter τ was set to $\tau = 0:1$. Also, for all the experiments, for the baseline method 4, parameters c and “where set to $c = 1.0$, and $\varepsilon = 0.6$, respectively, which are the best reported in [21]. It can be seen visually (Figures 3, 4, and 5), that the proposed method gives much better edge preservation on the oscillatory regions compared to the baseline methods. At the same time, on the homogeneous region, noise is also better removed. The values of PSNR and SSIM are consistent with the visual results, as it can be seen from Tables 1, 2, and 3.

In Table 1, processing results are given for the proposed, and the baseline methods obtained on the “Barbara” test image for three different values of σ . Similar results are given in Tables 2 and 3, for “Fishing boat” and “Bicycle” test images, respectively. It can be seen that the values of PSNR and SSIM are consistent with the visual results, and that the proposed method significantly outperforms all baseline methods in the means of PSNR and specially SSIM measure. In Tables 2 and 3, similar results are given for “Fishing boat” and “Bicycle” test images, respectively. The same can be concluded as for the “Barbara” test image.



Figure 1 Original test images: (a) "Barbara", (b) "Fishing boat", and (c) "Bicycle".



Figure 2 Noisy test images corrupted with the white Gaussian additive noise of $\sigma = 20$: (a) "Barbara", (b) "Fishing boat", and (c) "Bicycle".



Figure 3 Processing results on the “Barbara” test image. Results for the baseline algorithms 1, 2, 3, and 4 are presented in (a), (b), (c), and (d), respectively, while the processing result of the proposed method is presented in (e).

The upper l_2 bound for the patch approximation error for all experiments was set to $\varepsilon = 0.05$. The patch size was fixed to $n \times n = 8 \times 8 = 64$, so it was the size of the orthonormal DCT basis. For all the experiments, we

use 1/4 overlapped patches. We denote by \bar{K} the average size of the approximation generators chosen from the atoms of the overcomplicate DCT dictionary, so that $\|\delta\|_2 < \varepsilon$. As explained in the previous section, for the fix

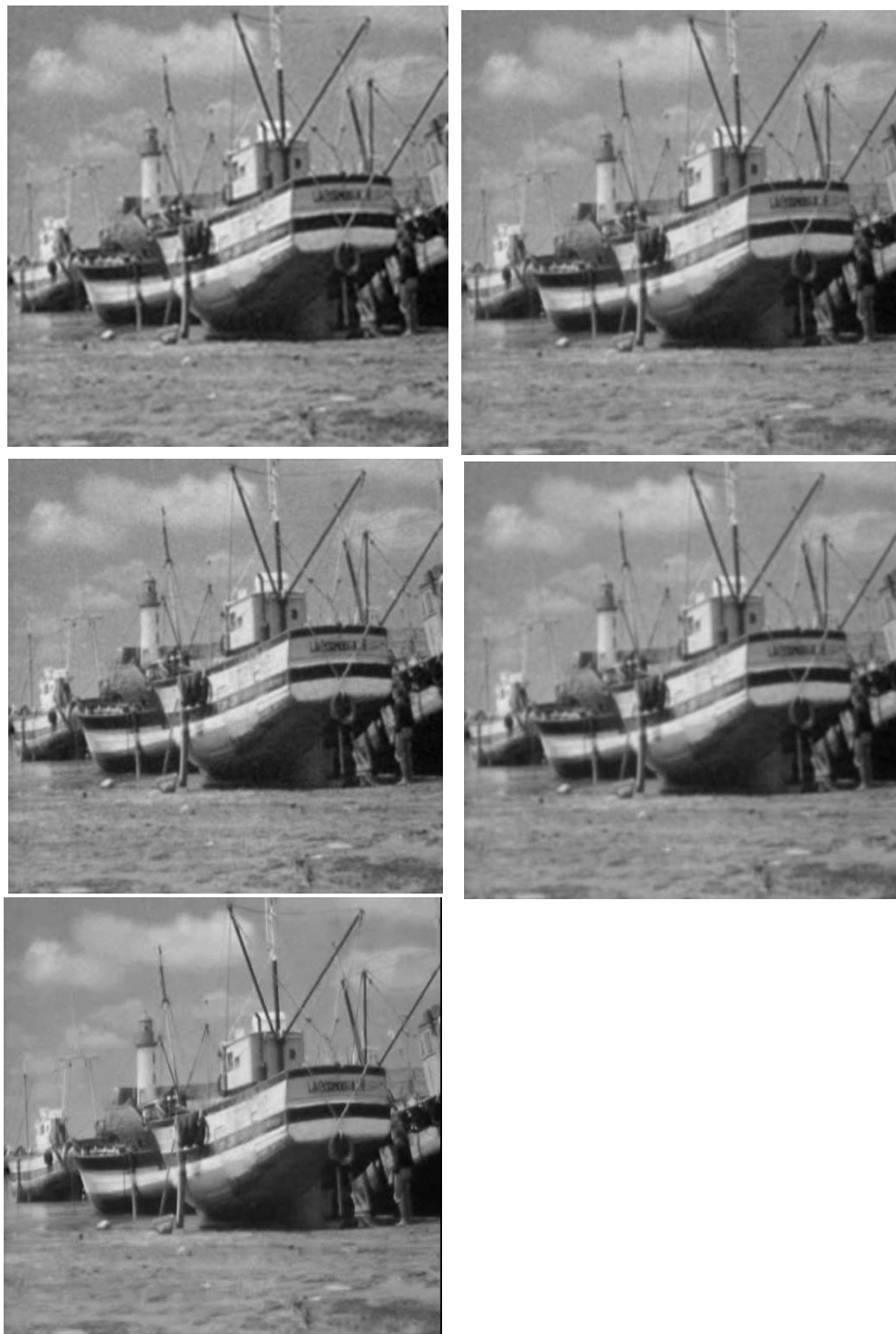


Figure 4 Processing results on the “Fishing boat” test image. Results for the baseline algorithms 1, 2, 3, and 4 are presented in (a), (b), (c), and (d), respectively, while the processing result of the proposed method is presented in (e).

patch, we first take some predefined initial number of atoms K . In all experiments we set $K = 5$. If the upper bound (obtained by (3.37)) for $\|\delta\phi\|_2$ is greater than ε , we add additional $m = 3$ atoms, optimize coefficients,

and continue until we obtain $\|\delta\phi\|_2 < \varepsilon$, as it is explained in the previous section. For all experiments, \bar{K} was obtained several times smaller than the size of the orthonormal DCT basis. For the “Barbara” test image,

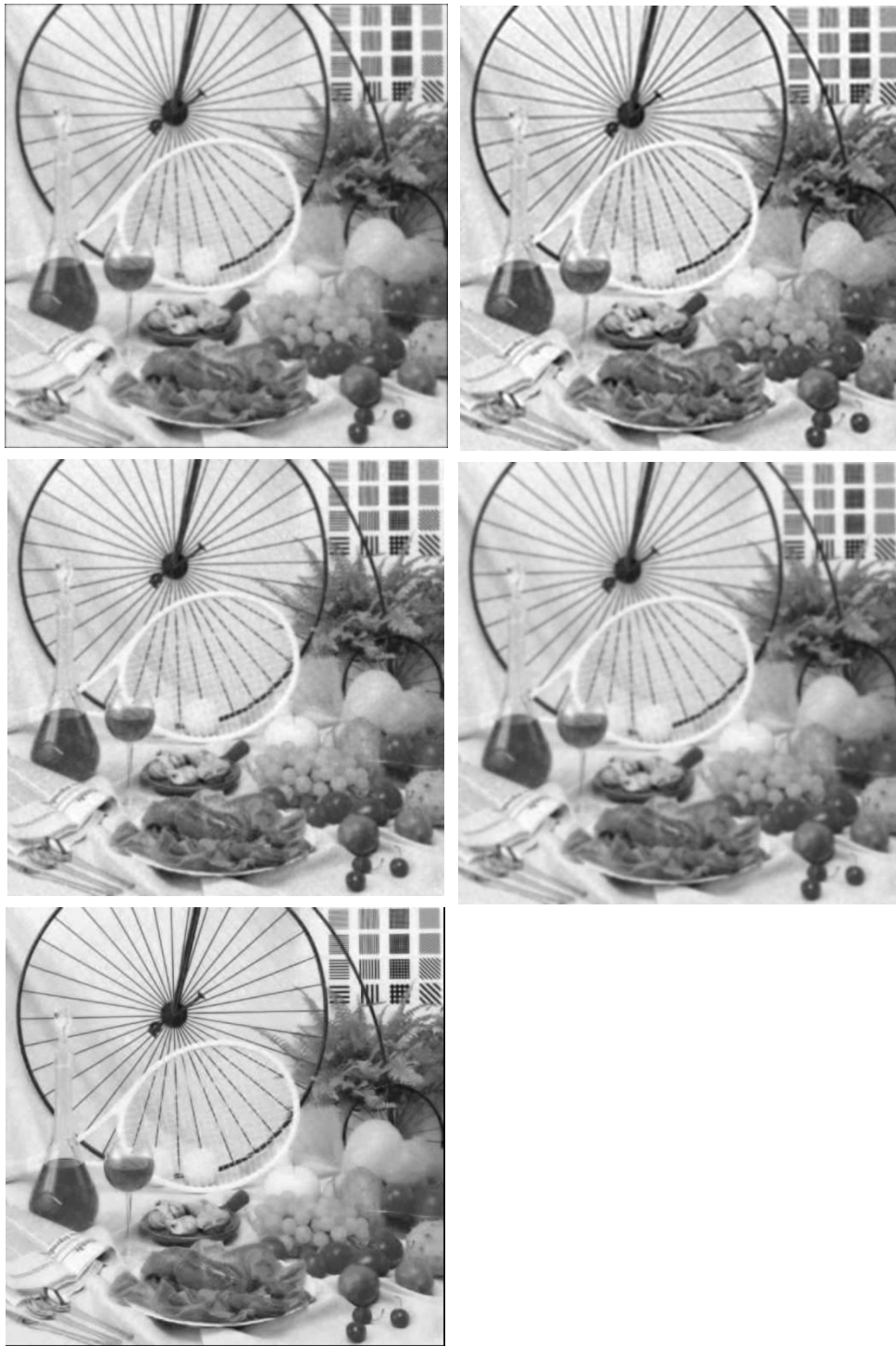


Figure 5 Processing results on the "Bicycle" test image. Results for the baseline algorithms 1, 2, 3, and 4 are presented in (a), (b), (c), and (d), respectively, while the processing result of the proposed method is presented in (e).

Table 1 Processing results for the proposed PBDVM method in comparison to the baseline methods, on “Barbara” test image, given for three different values of the standard deviation σ of the additive Gaussian white noise

Barbara	σ	PSNR	SIMM	σ	PSNR	SIMM	σ	PSNR	SIMM
Baseline 1	15	32.5	0.71	20	30.7	0.67	25	28.8	0.59
Baseline 2	15	32.7	0.69	20	31.2	0.69	25	28.9	0.60
Baseline 3	15	32.9	0.70	20	31.7	0.71	25	30.0	0.62
Baseline 4	15	32.8	0.70	20	31.5	0.70	25	28.9	0.60
PBDVM	15	34.0	0.73	20	33.6	0.79	25	31.3	0.63

we obtain $\lceil \tilde{K} \rceil = 10$. For the “Fishing boat” test image, we obtain $\lceil \tilde{K} \rceil = 8$, and for the “Bicycle” test image, $\lceil \tilde{K} \rceil = 9$. This makes the proposed method considerably faster and thus computationally much more efficient, than it would be the case if one uses the whole orthonormal DCT basis. For all experiments, \tilde{K} used for obtaining the ancillary patch was fixed on $\tilde{K} = 10$. Also, for all experiments, we set the thresholds and the corresponding Lagrange multipliers (explained in the previous section) on the following values: $T_1 = 0.003$ and $T_2 = 0.01$, and $\lambda_1 = 3.7$, $\lambda_2 = 2.3$, and $\lambda_3 = 0.8$, respectively.

In Figure 6, enlargements of the processing results for the “Barbara” test image are presented. The enlargements for the baseline algorithms 1, 2, 3, and 4 are presented in Figure 6a,b,c,d, respectively, while the enlargement for the proposed method is presented in Figure 6e. As it can be seen, the proposed method has saved stripes on Barbaras kerchief, much more efficiently than the baseline algorithms. The similar can be concluded for Figure 7, where the enlargements of the processing results for the “Bicycle” test image are presented in the same order as in Figure 6. Again, it can be seen that the highly oscillatory regions (strips of the various thickness in the right upper corner) are much better preserved in the case of the proposed algorithm, than in the case of baseline algorithms.

Table 2 Processing results for the proposed PBDVM method in comparison to the baseline methods, on “Fishing boat” test image, given for three different values of the standard deviation σ of the additive Gaussian white noise

Fishing boat	σ	PSNR	SIMM	σ	PSNR	SIMM	σ	PSNR	SIMM
Baseline 1	15	33.4	0.69	20	31.6	0.69	25	28.4	0.57
Baseline 2	15	33.7	0.70	20	32.1	0.70	25	28.5	0.58
Baseline 3	15	33.8	0.70	20	32.3	0.72	25	29.0	0.59
Baseline 4	15	33.6	0.69	20	32.1	0.71	25	28.7	0.58
PBDVM	15	36.3	0.73	20	33.2	0.77	25	31.1	0.67

Table 3 Processing results for the proposed PBDVM method in comparison to the baseline methods, on “Bicycle” test image, given for three different values of the standard deviation σ of the additive Gaussian white noise

Bicycle	σ	PSNR	SIMM	σ	PSNR	SIMM	σ	PSNR	SIMM
Baseline 1	15	33.2	0.68	20	31.4	0.66	25	28.2	0.54
Baseline 2	15	33.6	0.69	20	31.7	0.69	25	28.3	0.56
Baseline 3	15	34.2	0.70	20	32.1	0.70	25	28.7	0.59
Baseline 4	15	33.9	0.69	20	32.0	0.68	25	28.5	0.57
PBDVM	15	36.1	0.72	20	33.8	0.78	25	30.2	0.65

5. Conclusion

In this article we introduce a novel patch-based method for image denoising. It is based on a direct minimization of the appropriate energy functional containing a “minimal surface” regularizer with the fractional gradient. We use the Ritz variational method to approximate the true minimizer of the target energy functional on a particular patch. In order to reduce the average number of vectors in the approximation generator, we deliver the CFVP for the target functional. After the spatial discretization, we obtain the l^2 upper bound for the approximation error introduced by a simple shrinkage procedure, where we take the generator vectors from the DCT over-complete dictionary. Thus, we manage to significantly reduce computational load and make the method applicable. Moreover, we control the level of image smoothing on each patch (and thus on the whole image) by adapting the Lagrange multiplier using the information on the level of discontinuities present on a particular patch, which we obtain by pre-processing. We have compared the proposed method with the several anisotropic diffusion methods, and obtained significantly better results in the means of the objective PSNR, SSIM measures, as well as the subjective, visual results. In the future work, as we have manage to minimize the energy functional directly, we will try to introduce a group symmetry approach directly into the variational problem, similarly as it was done in the geometric PDE approach [12].

End notes

1. For example, for 512×512 images it would be of size $512^2 = 262144$.
2. In the framework of theory presented in [17], it holds that $\sigma = 0$, for all values of $\gamma > 0$.
3. In the means of Gateaux derivatives (see [17]).
4. By a simple shrinkage procedure, we consider disposing in synthesis all the other $N - K$ elements of the frame $\{\psi_n\}_1^N$, with lower values for $\langle \varphi_0, \psi_n \rangle$.
5. It is suboptimal, as it is not obtained by simultaneous minimization for all α_i in the representation.



Figure 6 Enlargements for the “Barbara” test image. Enlargements for the baseline algorithms 1, 2, 3, and 4 are presented in (a), (b), (c), and (d), respectively, while the enlargement for the proposed method is presented in (e).

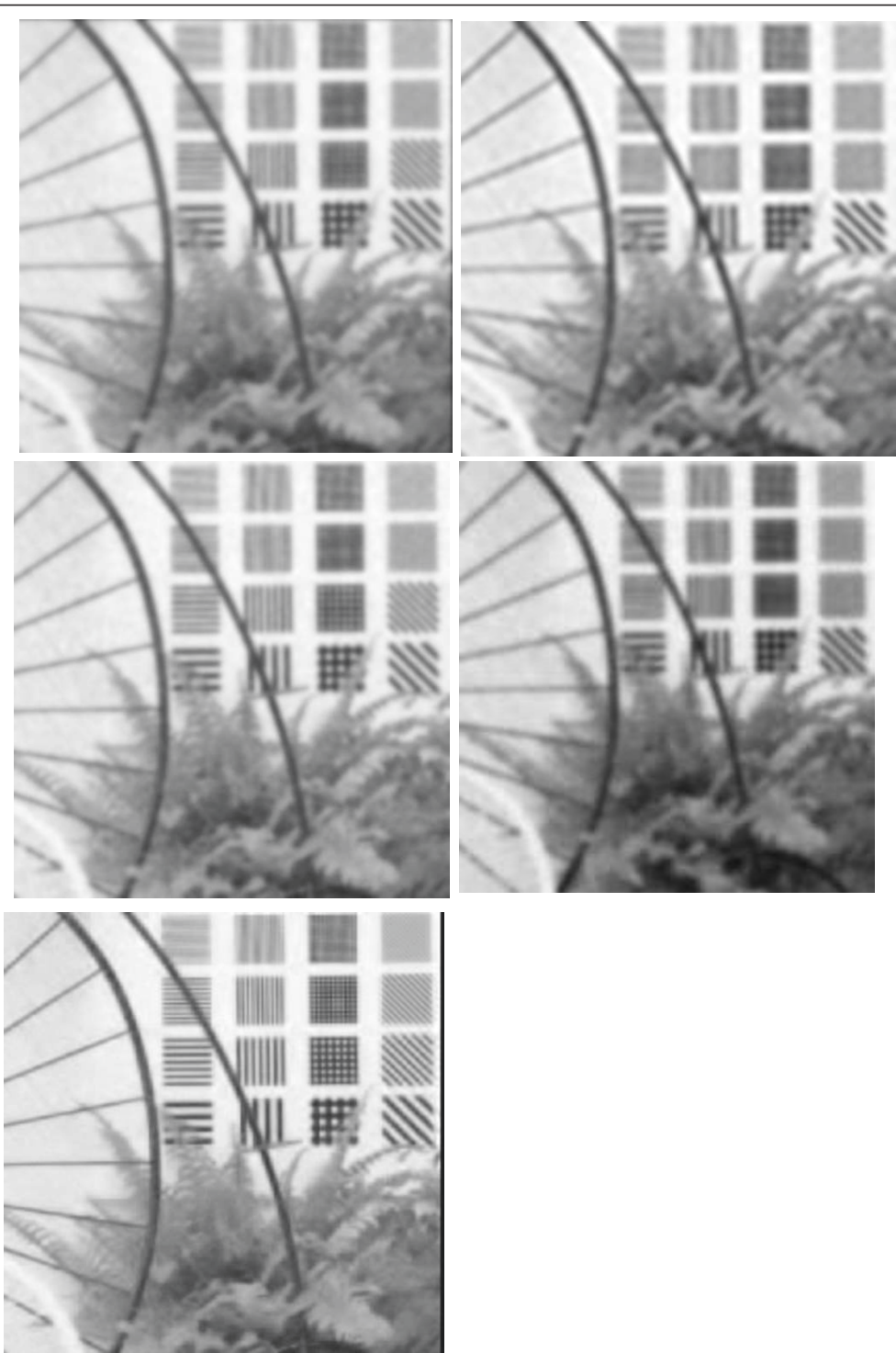


Figure 7 Enlargements for the "Bicycle" test image. Enlargements for the baseline algorithms 1, 2, 3, and 4 are presented in (a), (b), (c), and (d), respectively, while the enlargement for the proposed method is presented in (e).

Abbreviations

CFVP: complementary fractional variational principle; DCT: discrete cosine transform; PBDVM: patch-based direct variational minimization; SSIM: structural similarity index measure.

Acknowledgements

This research work has been supported by the "Serbian Ministry of Science and Technology" and it has been realized as a part of 174005, 174024, III44003, and TR32035 projects.

Author details

¹Faculty of Engineering, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia ²Department of Mathematics and Informatics, University of Novi Sad, Trg Dositeja Obradovića 4, 21000 Novi Sad, Serbia

Competing interests

The authors declare that they have no competing interests.

Received: 9 August 2010 Accepted: 1 June 2011 Published: 1 June 2011

References

1. P Perona, J Malik, Scale-space and edge detection using anisotropic diffusion. *IEEE Trans Pattern Anal Mach Intell.* **12**, 629–639 (1990). doi:10.1109/34.56205
2. F Catte, PL Lions, JM Morel, T Coll, Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J Numer Anal.* **29**, 182–193 (1992). doi:10.1137/0729012
3. LI Rudin, S Osher, E Fatemi, Nonlinear total variation based noise removal algorithms. *Phys D.* **60**, 259–268 (1992). doi:10.1016/0167-2789(92)90242-F
4. C Vogel, M Oman, Iterative methods for total variation denoising. *SIAM J Sci Stat Comput.* **17**, 227–238 (1996). doi:10.1137/0917016
5. YL You, M Kaveh, Fourth-order partial differential equation for noise removal. *IEEE Trans Image Process.* **9**, 1723–1730 (2000). doi:10.1109/83.869184
6. M Lysaker, A Lundervold, XC Tai, Noise removal using fourth-order partial differential equation with application to medical magnetic resonance images in space and time. *IEEE Trans Image Process.* **12**, 1579–1590 (2003). doi:10.1109/TIP.2003.819229
7. A Chambolle, PL Lions, Image recovery via total variation minimization and related problems. *Numer Math.* **76**, 167–188 (1997). doi:10.1007/s002110050258
8. P Blomgren, P Mulet, TF Chan, CK Wong, Total variation image restoration: numerical methods and extensions. *Proc Int Conf Image Process.* **3**, 384–387 (1997)
9. TF Chan, A Marquina, P Mulet, High order total variation-based image restoration. *SIAM J Sci Comput.* **22**, 503–516 (2000). doi:10.1137/S1064827598344169
10. J Bai, X Chu Feng, Fractional order anisotropic diffusion for image denoising. *IEEE Trans Image Process.* **16**, 2492–2502 (2007)
11. S Didas, J Weickert, B Burgeth, Properties of higher order nonlinear diffusion filtering. *J Math Imaging Vis.* **35**, 208–226 (2009). doi:10.1007/s10851-009-0166-x
12. G Sapiro, *Geometric partial differential equations and image analysis.* (Cambridge University Press, Cambridge, 2001)
13. Gilles Aubert, Pierre Kornprobst, *Mathematical Problems in Image Processing, Partial Differential Equations and the Calculus of Variation*, 2nd edn. (Springer, New York, 2006)
14. A Buades, B Coll, JM Morel, A non-local algorithm for image denoising. *IEEE CVPR* 6065. (2005)
15. N Azzabou, N Paragios, F Guichard, F Cao, Variable bandwidth image denoising using image-based noise models. *IEEE CVPR.* (2007)
16. J Boulanger, C Kervrann, P Boutheymy, Space-time adaptation for patch-based image sequence restoration. *IEEE PAMI.* **29**(6):10961102 (2007)
17. AM Arthurs, *Complementary Variational Principles*, 2nd edn. (Clarendon Press, Oxford, 1980)
18. M Elad, M Aharon, Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans Image Process.* **15**(12):3362 (2006)
19. S Didas, B Burgeth, A Imiya, J Weickert, Regularity and Scale Space Properties of Fractional High Order Linear Filtering. (Lecture Notes in Computer Science Springer, Berlin, 2005) **3459**, pp. 13–25
20. Y Cao, J Yin, G Liu, M Li, A class of nonlinear parabolic-hyperbolic equations applied to image restoration. *Nonlinear Anal Real World Appl.* **11**(1):253–261 (2010). doi:10.1016/j.nonrwa.2008.11.004
21. P Guidotti, JV Lambers, Two new nonlinear nonlocal diffusion s for noise reduction. *J Math Imaging Vis.* **33**, 2537 (2009)
22. Z Wang, AC Bovik, HR Sheikh, EP Simoncelli, Image quality assessment: from error to structural similarity. *IEEE Trans Image Process.* **13**(4):600–612 (2004). doi:10.1109/TIP.2003.819861

doi:10.1186/1687-6180-2011-8

Cite this article as: Janev et al.: Image denoising by a direct variational minimization. *EURASIP Journal on Advances in Signal Processing* 2011 2011:8.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com