

A Primal-Dual Projected Gradient Algorithm for Efficient Beltrami Regularization[☆]

Dominique Zosso^{a,*}, Aurélien Bustin^{a,b,c}

^aDepartment of Mathematics, University of California, Los Angeles (UCLA), 520 Portola Plaza, Box 951555, Los Angeles, CA 90095-1555, U.S.A.

^bTechnische Universität München, IMETUM, Boltzmannstr. 11, 85748 Garching, Germany

^cGE Global Research, Freisinger Landstr. 50, 85748 Garching, Germany

Abstract

Inverse problems in imaging and computer vision are typically addressed as data-fidelity optimization problems, where data-regularizers such as H^1 or TV (total variation) are included to render the problem well-posed. However, while H^1 regularization is known to produce overly smooth reconstructions, the TV (or ROF) model is feature-preserving but introduces staircasing artifacts. The geometrically derived Beltrami framework, introduced by Sochen, Kimmel and Malladi (1998) offers an ideal compromise between feature preservation and avoidance of staircasing artifacts. Until now, one of the main limiting factors of the Beltrami regularizers was the lack of really efficient optimization schemes. Here, we start from one of the most efficient TV-optimization methods, primal-dual projected gradients, and apply it to the Beltrami functional. Doing so, we achieve better performance than ROF denoising for the basic grey-scale denoising problem, then extend the method to more involved problems such as inpainting, deconvolution, and the color case, all in a straightforward fashion. With the proposed primal-dual projected gradients optimization algorithm, the benefits of the geometric Beltrami regularizer become available at no extra computational cost, compared to state-of-the-art TV/ROF regularizers.

Keywords: Beltrami Energy, Image Denoising, Image Deblurring, Image Inpainting, Primal-Dual, Regularization, Gradient Projection, Constrained Optimization

2010 MSC: 68U10, 65K10, 49M29

1. Introduction

Modern day imaging is mainly challenged by so-called inverse problems, namely to find an underlying image I , given derived measurements I_0 through the system K , affected by noise n , e.g.:

$$\text{find } I : I_0 = KI + n. \quad (1)$$

Typical tasks are image restoration, reconstruction, segmentation, registration, classification, and many others. These problems are ill-posed, i.e., not all of the following conditions for well-posedness are true: (1) a solution exists, (2) the solution is unique, (3) the solution depends smoothly on the data [1, 2]. One classical resolution scheme consists in minimizing a quadratic energy E in terms of the forward problem, while adding a quadratic regularity penalty to render the problem well-posed [3, 4]. Formally:

$$\min_u \left\{ E = \|KI - I_0\|_2^2 + \alpha \|\Gamma I\|_2^2 \right\} \quad (2)$$

where typically $\Gamma = \nabla$, requiring the solution to be smooth.

Although widely used today thanks to its simplicity, this resolution scheme has some important shortcomings. Despite



Figure 1: An image with values in $[0, 256]$ with Gaussian noise $\sigma = 10$ denoised with three different regularity priors. One can see the oversmoothing associated to the H^1 regularization, and the staircasing effect of TV, while Beltrami combines smooth gradients with feature preservation.

their popularity due to computational ease (fast analytical global solution), it is well-known, that these H^1 regularizers are not feature-preserving and that the obtained solutions are typically overly smooth. Different anisotropic, feature-preserving regularization schemes have therefore been researched, e.g., [5, 6, 7, 8, 9].

The Rudin-Osher-Fatemi (ROF) or Total Variation (TV) model [10] has particularly gained in importance, where the following minimization problem using the TV-norm is proposed. Formally:

$$\min_I \left\{ \int_{\Omega} |\nabla I| + \frac{\lambda}{2} \int_{\Omega} (I - I_0)^2 \right\}, \quad (3)$$

for a given balancing parameter $\lambda > 0$. Here, $\int_{\Omega} (I - I_0)^2$ is a

[☆]This work was supported by the Swiss National Science Foundation (SNF) under grants PBELP2-137727 and P300P2-147778.

*Corresponding author

Email addresses: zosso@math.ucla.edu (Dominique Zosso), aurelien.bustin@ge.com (Aurélien Bustin)

fidelity term wrt. the observed, noise image I_0 , and $\int_{\Omega} |\nabla I|$ is a total variation regularizing term, where we suppose $I \in BV(\Omega)$,¹⁵ the space of functions of bounded variation, see also [11]. Soon after its introduction, the TV norm was locally weighted, in order to make the diffusion even more feature-aware and spatially adaptive [12]. The weighted TV was later rediscovered in the context of global solutions to active contours [13]. One main problem with TV-regularization is the appearance of staircasing artifacts as shown in figure 1.²⁰

Another, edge-aware feature-preserving diffusion flow stems from the Beltrami embedding [14, 15]. Most simply:²⁵

$$\min_I \left\{ \int_{\Omega} \sqrt{1 + \beta^2 |\nabla I|^2} + \frac{\lambda}{2} \int_{\Omega} (I - I_0)^2 \right\}. \quad (4)$$

A particular choice of gradient-descent preconditioner allows for a geometric Laplace-Beltrami flow, which is parametrization invariant, and which successfully reduces diffusion close to features. More generally, this model is based on the Polyakov model [16] introduced in string theory for physics. The Polyakov model represents strings as harmonic maps in high-dimensional and curved spaces defined by Riemannian manifolds. Adopting this pure geometric point of view amounts to seeing objects such as images, shapes, or vector fields as geodesics or harmonic maps, much like a 2D topographic map corresponds to a three-dimensional surface in the real world.³⁰

The potential of this geometric framework lies in the general definition of the space-feature manifold and the choice of its metric. In particular, the metric can be chosen such that the Polyakov energy corresponds to an interpolation between quadratic or total variation gradient penalty. The features are not restricted to scalar values but include vector features encountered in color, texture or multispectral image analysis [17].³⁵ Similarly, the embedding is not limited to 2-dimensional image surfaces and generalizes naturally to n -dimensional manifolds associated to volumetric or time varying images or videos. Moreover, the choice of the metric enables the study of complex geometries inherent to scale-space methods [18] and non-flat images generated, e.g., by catadioptric or omnidirectional cameras [19]. Local, pixel-based Beltrami-regularization was already successfully used for image denoising [20, 21, 22].⁴⁵

The direct way of implementing the Beltrami flow is obtained by using an Euler explicit forward scheme of gradient descent [23]. This scheme, however, is heavily limited by the CFL condition [24] and the time step has to be chosen carefully. An upper bound that ensures stability of such a scheme is provided in [25]. In order to make the scheme more robust, semi-implicit schemes have been used instead, e.g., [26, 27, 28, 29], largely based on work by [30]. In practice, many authors reduce the multidimensional diffusion problem into a series (LOD, [31])¹⁰⁵ or average (AOS, [32]) of several independent 1D diffusion processes, e.g., [28]. Another speedup of the PDE can be achieved by using vector extrapolation techniques [33, 34, 35]. An entirely different approach makes use of decoupling schemes that have been known in computer vision for quite some time [36, 20, 21, 22]. Equivalent splitting schemes have previously been applied to TV-regularized image denoising with great success¹¹⁰ [37].

Another technique that has successfully been used with TV regularization problems is the use of *primal-dual optimization* [38, 39, 40]. These algorithms are simple and easy to implement and are potentially faster than primal gradient-descent methods. While Chambolle's primal-dual TV algorithm has clearly led the way, more recent, very similar projected-gradient type methods were shown to be even more efficient [41, 42].

Here, we propose to use such a *primal-dual projected gradient* approach to design an efficient recovery algorithm for grey-scale images by employing the Beltrami energy as regularity prior. We present a family of simple yet efficient primal-dual projected gradient algorithms for Beltrami minimization problems with applications in image denoising, and straightforward extensions to different image processing applications such as inpainting/superresolution, deblurring/deconvolution, and compressed sensing. Beyond, we present a convex simplification of the color-version of the Beltrami energy, and thereby render our algorithm applicable to multichannel images. We show that the proposed algorithms can be easily adapted to all these applications. Numerical results are presented on different images to show the performance of the primal-dual projected gradient algorithm using the Beltrami framework. We compare our results against the similarly implemented ROF model. We can show that the proposed Beltrami algorithms converge faster, and slightly improve the quality of results wrt. TV: Beltrami regularization maintains TV's feature preservation but reduces the staircasing effect in restored images.

The rest of this paper is organized as follows. First, we provide a short introduction to the Beltrami framework in section 2. Section 3 presents the primal-dual model, which is then related to a projection type algorithm and variational inequalities. In section 4 we combine results from sections 2 and 3 to describe the primal-dual projected gradient algorithm as used for Beltrami regularization, and apply our method to denoising problems. Denoising results and a short comparison to the related TV/ROF implementation are shown in section 5. In sections 6-8 we apply our proposed model to different imaging problems. A simplified generalization to color images is presented in section 9. Then we conclude the paper in section 10.

2. Beltrami Framework

The two main ingredients of the Beltrami framework come from differential geometry: The embedding of images as a Riemannian manifold, and the metric tensor used to measure image regularity. Here, we provide a very brief review of some essential concepts and definitions from differential geometry, and refer the reader, e.g., to [43, 44, 45] for more details. Readers who are happy with the Beltrami denoising problem as stated in (4) may skip this section altogether.

2.1. Definitions from differential geometry

First, we give a definition of a differentiable manifold.

Definition 1 (Differentiable manifold). A topological space \mathcal{M} is called an n -dimensional differentiable manifold if:

- i. \mathcal{M} is provided with a family of pairs $\{(\mathcal{M}_\alpha, \phi_\alpha)\}$,
ii. $\{\mathcal{M}_\alpha\}$ is a family of open sets covering \mathcal{M} :

$$\bigcup_\alpha \mathcal{M}_\alpha = \mathcal{M},$$

- iii. ϕ_α are homeomorphisms to an open subset O_α :

$$O_\alpha \subset \mathbb{R}^n : \quad \phi_\alpha : \mathcal{M}_\alpha \rightarrow O_\alpha$$

- iv. For any pair of overlapping patches \mathcal{M}_α and \mathcal{M}_β , such that $\mathcal{M}_\alpha \cap \mathcal{M}_\beta \neq \emptyset$, the map

$$\phi_\beta \circ \phi_\alpha^{-1} : \phi_\alpha(\mathcal{M}_\alpha \cap \mathcal{M}_\beta) \subset \mathbb{R}^n \rightarrow \phi_\beta(\mathcal{M}_\alpha \cap \mathcal{M}_\beta) \subset \mathbb{R}^n \quad ^{135}$$

is C^∞ .

Note: each pair $(\mathcal{M}_\alpha, \phi_\alpha)$ is called a chart, and the entire family $\{(\mathcal{M}_\alpha, \phi_\alpha)\}$ forms an atlas. Items (i)-(iii) provide that the manifold \mathcal{M} is locally Euclidean, while (iv) ensures smooth parametrization.
¹²⁰

Now, let us introduce a Riemannian metric on such a differentiable manifold \mathcal{M} :

Definition 2 (Riemannian metric). A Riemannian metric on $\mathcal{M} \subset \mathbb{R}^m$ is a family of inner products $g_p^\mathcal{M}$ of elements of the tangent space $T_p \mathcal{M} \subset \mathbb{R}^m$,

$$\begin{aligned} g_p^\mathcal{M} : T_p \mathcal{M} \times T_p \mathcal{M} &\rightarrow \mathbb{R} \\ (u, v) &\mapsto g_p^\mathcal{M}(u, v) \end{aligned} \quad p \in \mathcal{M} \subset \mathbb{R}^m \quad (5)$$

such that for any two differentiable vector fields X, Y on \mathcal{M} , the map

$$\begin{aligned} \mathcal{M} &\rightarrow \mathbb{R} \\ p &\mapsto g_p(X(p), Y(p)) \end{aligned} \quad ^{140} \quad (6)$$

defines a smooth function.

Now, we may define a Riemannian manifold as follows:

Definition 3 (Riemannian manifold). A Riemannian manifold, $(\mathcal{M}, g_p^\mathcal{M})$, is a differentiable manifold \mathcal{M} equipped with an associated Riemannian metric $g_p^\mathcal{M}$.
¹⁴⁵

An important relation is given by the pullback metric:

Definition 4 (Pullback metric). Let $f : \mathcal{M} \rightarrow \mathcal{N}$ be a differentiable map, and $(\mathcal{N}, g^\mathcal{N})$ be a Riemannian manifold. Now, the pullback of $g^\mathcal{N}$ along f is a quadratic form on the tangent space of \mathcal{M} :

$$\begin{aligned} f^* g^\mathcal{N} : T_p \mathcal{M} \times T_p \mathcal{M} &\rightarrow \mathbb{R} \\ (f^* g^\mathcal{N})(u, v) &\mapsto g^\mathcal{N}(df(u), df(v)), \end{aligned} \quad (7)$$

where $df(u)$ is the pushforward of u by f . In particular, if f is ¹⁴⁵ diffeomorphic, then this defines the induced, or pullback metric on \mathcal{M} .

Definition 5 (Isometric maps). Two Riemannian manifolds $(\mathcal{M}, g^\mathcal{M})$ and $(\mathcal{N}, g^\mathcal{N})$ are said to be isometric under $f : \mathcal{M} \rightarrow \mathcal{N}$, if their metric tensors are related by the pullback relation:
¹⁵⁰

$$g^\mathcal{M} = f^* g^\mathcal{N}. \quad (8)$$

In practical terms, the pullback metric is easily obtained as follows. Let (x_1, \dots, x_m) be the local coordinates of the manifold \mathcal{M} , and the coordinates of the manifold \mathcal{N} be given as $(y_1, \dots, y_n) = f(x_1, \dots, x_m)$. The components of the metric tensor $g^\mathcal{M} = f^* g^\mathcal{N}$ are calculated as:

$$g_{ij}^\mathcal{M} = \sum_{\mu\nu} \frac{\partial y_\mu}{\partial x_i} \frac{\partial y_\nu}{\partial x_j} g_{\mu\nu}^\mathcal{N}. \quad (9)$$

Finally, let us introduce one practical use of metric tensors to calculate the area of a bounded domain on a Riemannian manifold:

Definition 6 (Area on a manifold). Let $\mathcal{R} \subset \mathcal{M}$ be a bounded region of an n -dimensional Riemannian manifold $(\mathcal{M}, g^\mathcal{M})$. Its area (or hyper-surface, for $n > 2$) is defined as

$$A(\mathcal{R} \subset \mathcal{M}) := \int_{\phi(\mathcal{R}) \subset \mathbb{R}^n} \sqrt{\det g^\mathcal{M}}, \quad (10)$$

where $\det g^\mathcal{M}$ is the determinant of the metric tensor expressed in the coordinate chart (\mathcal{M}, ϕ) .

2.2. Embedding

Let us consider a real-valued (grayscale) image function $I(x) \in W^{2,2} : \mathbb{R}^n \rightarrow \mathbb{R}$, where $W^{2,2}$ denotes a Sobolev space, i.e. the L^2 norm of both the function and its first and second order derivatives are finite:

$$\|I(x)\|_2 + \|\nabla I(x)\|_2 + \|\Delta I\|_2 < \infty \quad (11)$$

In simple terms, the function I is required to be finite and twice differentiable. This is easily guaranteed if we consider the image to be the result of an adequate (smooth) interpolation from a finite, discrete quantization.

In this general form, $x \in \Omega \subset \mathbb{R}^n$ denotes the coordinates in n -dimensional space, and $I(x) \in \mathbb{R}$ are the associated features. We thus typically see an image as a mapping from space to features:

$$\begin{aligned} I : \Omega \subset \mathbb{R}^n &\rightarrow \mathbb{R} \\ (x_1, \dots, x_n) &\mapsto I(x_1, \dots, x_n). \end{aligned} \quad (12)$$

In contrast, the Beltrami embedding associates the spatial coordinates along with the features:

Definition 7 (Beltrami embedding). The Beltrami embedding defines a diffeomorphic map as follows:

$$\begin{aligned} X : \Omega \subset \mathbb{R}^n &\rightarrow \Omega \times \mathbb{R} \\ (x_1, \dots, x_n) &\mapsto (x_1, \dots, x_n, I(x_1, \dots, x_n)). \end{aligned} \quad (13)$$

This amounts to seeing the function I as a non-flat surface embedded in a higher dimensional space, much like a topographic map corresponds to a three-dimensional surface in the real world.

Definition 8 (Image domain). In terms of differential geometry, we have the following situation: We have a first Riemannian manifold defined by the image domain, $(\Omega \subset \mathbb{R}^n, g_{\mu\nu}^\Omega)$, where the metric $g_{\mu\nu}^\Omega$ is unspecified for now. A second Riemannian manifold, $(\mathcal{M} \subset \Omega \times \mathbb{R}, h_{ij}^\mathcal{M})$,

is formed by the joint space-feature manifold. These two manifolds are related by the embedding, diffeomorphic map $X : \Omega \rightarrow \mathcal{M}$.

In the case of a 2D gray scale image, $I(x, y) \in \mathbb{R}$, the manifold \mathcal{M} is the space-feature manifold embedded in 3-dimensional space, and Ω is a 2D surface with (x, y) denoting coordinates on it. The embedding map is given by three functions of two variables:

$$\begin{aligned} X : \Omega \subset \mathbb{R}^2 &\rightarrow \mathcal{M} \subset \mathbb{R}^3 \\ (x, y) &\mapsto (x, y, I(x, y)). \end{aligned} \quad (14)$$

2.3. Regularizing functional

To measure the “weight” of the embedding $X : \Omega \rightarrow M$, Sochen et al. use the Polyakov action from high energy physics [15, 14, 16]. This action does not need the two metrics be specifically related in the first place. However, it was shown that the Polyakov action is minimized (*ceteris paribus*) if the two manifolds are chosen to be isometric, *i.e.*, the metric tensor on Ω is obtained by pullback of the metric tensor on \mathcal{M} according to (9).

Definition 8 (Beltrami functional). In this case, the Polyakov action simplifies to the following functional:

$$E_{Bel}(X^i, g_{\mu\nu}, h_{ij}) := \int_{\Omega} \sqrt{g} dx, \quad (15)$$

which simply measures the surface of the embedded manifold \mathcal{M} . We will use this area as a measure of regularity of the underlying image I .

2.4. Choosing the metric tensors

The embedding space of the manifold $\mathcal{M} \subset \Omega \times \mathbb{R}$ has dimensions of different nature, namely spatial components versus feature components. The relative scaling between these components is arbitrary in most applications, and it is therefore natural to choose a metric tensor $h_{ij}^{\mathcal{M}}$ that incorporates tuning of the aspect ratio [14, 15, 17], by a factor β :

$$\{h_{ij}^{\mathcal{M}}\} := \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 \\ 0 & \cdots & 0 & \beta^2 \end{bmatrix} = \text{diag}(1, \dots, 1, \beta^2). \quad (16)$$

Having chosen the metric tensor of the embedding space, one can now easily compute the corresponding pullback metric tensor for $g_{\mu\nu}^{\Omega}$, to be used on the original image domain manifold Ω , according to (9).

For the particular case at hand, this yields

$$\{g_{\mu\nu}^{\Omega}\} := \begin{bmatrix} 1 + \beta^2 I_1^2 & \beta^2 I_1 I_2 & \cdots & \beta^2 I_1 I_n \\ \beta^2 I_2 I_1 & 1 + \beta^2 I_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \beta^2 I_{n-1} I_n \\ \beta^2 I_n I_1 & \cdots & \beta^2 I_n I_{n-1} & 1 + \beta^2 I_n^2 \end{bmatrix}, \quad (17)$$

where $I_i := \partial I / \partial x_i$ is shorthand for the spatial derivatives of the image. Beyond, the determinant of this metric tensor is calculated as simply

$$g = \det g_{\mu\nu}^{\Omega} = 1 + \beta^2 |\nabla I|^2, \quad (18)$$

and the energy functional now simply reads

$$E_{Bel}(I, \beta^2) := \int_{\Omega} \sqrt{1 + \beta^2 |\nabla I|^2} dx. \quad (19)$$

In the context of image processing, this energy is often referred to as the *Beltrami energy*, and this is the form we have used in (4). The standard method of minimizing this energy is to use gradient descent (nonlinear Euler-Lagrange equations):

$$I^{k+1} = I^k + \tau \frac{\beta^2}{\sqrt{g}} \operatorname{div} \left(\frac{\nabla I}{\sqrt{g}} \right) \quad (20)$$

where τ is the time-step of the PDE integration. There are, however, serious numerical issues. Indeed, this explicit scheme is heavily restricted by the CFL condition and the time step τ has to be chosen carefully.

Remark 1. This Beltrami energy looks surprisingly similar to the regularized TV functional, which integrates $\sqrt{\epsilon + |\nabla I|^2}$. However, the role of the addition under the square root is a different one. In regularized TV, the goal is to make the energy differentiable by smoothing around zero; thus ϵ is typically chosen very small. In the present case, however, the aspect ratio β allows us to consider different norms. For $\beta \rightarrow \infty$, the 1 becomes negligible and the energy truly approaches the TV semi-norm. On the other hand, if $\beta \rightarrow 0$, the minimizing flow approaches isotropic heat diffusion.

2.5. Discretized Beltrami energy

For practical purposes, we consider here a discretized grey-scale image modeled as a function $I : \Omega \rightarrow \mathbb{R}$ where $\Omega \subseteq \mathbb{R}^2$ is a rectangle and $I(x, y)$ is the intensity of the pixel (x, y) . In the discrete domain an image is usually a function $I : \{1, \dots, m\} \times \{1, \dots, n\} \rightarrow \mathbb{R}$ where $m \times n$ is the image size.

Definition 9 (Discretized differential operators). By choosing Neumann boundary conditions, the discrete gradient ∇I can be discretized with a forward difference scheme,

$$(\nabla_x I)(i, j) := \begin{cases} I(i+1, j) - I(i, j) & \text{if } 1 \leq i < m \\ 0 & \text{if } i = m \end{cases} \quad (21)$$

$$(\nabla_y I)(i, j) := \begin{cases} I(i, j+1) - I(i, j) & \text{if } 1 \leq j < n \\ 0 & \text{if } j = n. \end{cases} \quad (22)$$

In order to preserve the adjointness of gradient and divergence, $\langle \nabla I, p \rangle = -\langle I, \operatorname{div} p \rangle$, the divergence of a vector field $p = (p_x, p_y)$

is then defined as

$$(\operatorname{div} p)(i, j) := \begin{cases} p_x(i, j) - p_x(i-1, j) & \text{if } 1 < i < m \\ p_x(i, j) & \text{if } i = 1 \\ -p_x(i-1, j) & \text{if } i = m \end{cases} + \begin{cases} p_y(i, j) - p_y(i, j-1) & \text{if } 1 < j < n \\ p_y(i, j) & \text{if } j = 1 \\ -p_y(i, j-1) & \text{if } j = n. \end{cases} \quad (23)$$

From now, and until the end of the paper, we consider Ω as a discrete set of points, and we will limit our attention to this discrete, finite case. For example, the discrete equivalent of (4) is

$$\min_I \left\{ \sum_{\Omega} \sqrt{1 + \beta^2 |\nabla I|^2} + \frac{\lambda}{2} (I - I_0)^2 \right\}. \quad (24)$$

3. Primal-Dual Projected Gradients

In this section, we entirely switch gears and provide a short description of primal-dual projected gradient optimization. We recall the main results existing in the literature to find the primal-dual model of an initial convex problem, and see that it can be transformed into a variational problem. Then we give a brief overview on how to solve this new model by using a gradient projection type method. Here, we only recall some core principal results, and refer the interested reader to, e.g., [46, 41, 42] for further details, analysis and improvements of this general method.

We start with a definition of the convex conjugate (a.k.a. Legendre-Fenchel transform) of a function:

Definition 10 (Legendre-Fenchel transform). The convex conjugate of a function f is the function f^* defined by

$$f^*(s) = \sup_{x \in \mathbb{R}^n} \left\{ \langle s, x \rangle - f(x) \right\}, \quad s \in \mathbb{R}^n. \quad (25)$$

The convex conjugate of a closed convex (= convex and lower semi-continuous) function is again a closed convex function. Also, the biconjugate, $f^{**} := (f^*)^*$, is the largest closed convex function with $f^{**} \leq f$. As a result, $f = f^{**}$ iff f is closed convex (Fenchel-Moreau theorem).

The Legendre-Fenchel transform is used in the following primal-dual equivalence:

Theorem 1. Let $F : W \rightarrow \mathbb{R}$ be a closed and convex functional on the set W , G a closed and convex functional on the set V and let $K : V \rightarrow W$ be a continuous linear operator. Then we have the following equivalence:

$$\underbrace{\min_{x \in V} \left\{ F(Kx) + G(x) \right\}}_{\text{Primal}} = \underbrace{\min_{x \in V} \max_{\varphi \in W^*} \left\{ \langle Kx, \varphi \rangle - F^*(\varphi) + G(x) \right\}}_{\text{Primal-Dual}} \quad (26)$$

where x and φ are the primal and dual variables, respectively, F^* is the convex conjugate of F , and W^* is the dual space of W .

For a proof, see [46].

3.1. Variational problem

We will now see how the primal-dual problem of Theorem 1 can be written as a variational problem. First, let us define sub- and supergradients as follows:

Definition 11 (Subgradient). The subderivative of a function $f : U \rightarrow \mathbb{R}$ at a point x_0 in the convex open set U is the set $\partial^- f(x_0)$ of all elements $c \in U^*$ such that

$$f(x) - f(x_0) \geq \langle c, x - x_0 \rangle. \quad (27)$$

For a convex function f , the set $\partial^- f(x_0)$ is always closed and non-empty. If in addition, f is differentiable at x_0 , then $\partial^- f(x_0)$ contains exactly one element.

Definition 12 (Supergradient). The superderivative of a function $f : U \rightarrow \mathbb{R}$ at a point x_0 in the convex open set U is the set $\partial^+ f(x_0)$ of all elements $c \in U^*$ such that

$$f(x) - f(x_0) \leq \langle c, x - x_0 \rangle. \quad (28)$$

For a concave function f , the set $\partial^+ f(x_0)$ is always closed and non-empty. If in addition, f is differentiable at x_0 , then $\partial^+ f(x_0)$ contains exactly one element.

Sub- and supergradient are related by the equivalence

$$\partial^- f = \{c \mid -c \in \partial^+(-f)\} =: -\partial^+(-f). \quad (29)$$

Further, we require partial sub- and superderivatives as follows:

Definition 13 (Partial sub- and superderivatives). Let $f : U_1 \times \dots \times U_k \rightarrow \mathbb{R}$ be a multivariate function. We define the i -th partial subderivative of f at (x_1, \dots, x_k) as

$$\partial_{x_i}^- f(x_1, \dots, x_k) := \partial^- g_i(0), \quad (30)$$

where $g_i : U_i \rightarrow \mathbb{R}$, $g_i(t) := f(x_0, \dots, x_i + t, \dots, x_k)$. Partial superderivatives are defined analogously.

We now consider the general saddle point problem:

$$\min_{u \in U} \max_{v \in V} \left\{ f(u, v) \right\} \quad (31)$$

where U and V are closed convex, $f(\cdot, v)$ is convex in the first argument for any $v \in V$ and $f(u, \cdot)$ is concave in the second argument for any $u \in U$.

The point $(u^*, v^*) \in U \times V$ is a solution to the problem (31) iff

$$\forall (u, v) \in U \times V : \quad f(u^*, v) \leq f(u^*, v^*) \leq f(u, v^*). \quad (32)$$

Such a $(u^*, v^*) \in U \times V$ is called a saddle point of f .

Further, for any $(u, v) \in U \times V$, and $c_1 \in \partial_u^- f(u, v)$ and $c_2 \in \partial_v^+ f(u, v)$, we have

$$\begin{cases} f(u^*, v) \geq f(u, v) + \langle c_1, u^* - u \rangle \\ f(u, v^*) \leq f(u, v) + \langle c_2, v^* - v \rangle, \end{cases} \quad (33)$$

and from (32) and (33) we can conclude that

$$\langle c_1, u - u^* \rangle + \langle -c_2, v - v^* \rangle \geq 0 \quad (34)$$

We can thus rewrite (31) as the following variational inequality problem: find $x^* \in X := U \times V$ s.t.

$$\langle x - x^*, H(x^*) \rangle \geq 0 \quad \forall x \in X, \quad (35)$$

where

$$x = \begin{pmatrix} u \\ v \end{pmatrix} \quad \text{and} \quad H(x) = \begin{pmatrix} \partial_u f(u, v) \\ -\partial_v f(u, v) \end{pmatrix}. \quad (36)$$

3.2. A projection method

In this section, we make use of a gradient projection method for solving the general variational inequality (35). Indeed, given a vector x^* and the functional H , we have the important well-known results that follow (see [47]).

Definition 14 (Projection). We define $P_X(\cdot)$ as the orthogonal projection, under the Euclidean metric, onto the convex set X :

$$P_X[x] = \operatorname{argmin}_{z \in X} |z - x|^2 \quad (37)$$

Proposition 2. Let r be a positive parameter and X a convex set. An element x^* is solution of (35) if and only if

$$x^* = P_X(x^* - rH(x^*)). \quad (38)$$

To prove the proposition we first need the following results:

Lemma 3. Given $x \in \mathbb{R}^n$, a vector $z \in X$ is equal to $P_X[x]$ if and only if

$$\langle y - z, x - z \rangle \leq 0 \quad \forall y \in X. \quad (39)$$

PROOF. $P_X[x]$ is the minimizer of $F(z) = |z - x|^2$ over all $z \in X$. Moreover, if a vector $x \in X$ minimizes $F: \mathbb{R}^n \rightarrow \mathbb{R}$, then $\langle y - x, \nabla F(x) \rangle \geq 0, \forall y \in X$. We have

$$\nabla F(z) = 2(z - x), \quad (40)$$

and thus

$$\langle y - z, x - z \rangle \leq 0 \quad \forall y \in X. \quad (41)$$

□

PROOF (PROPOSITION 2). Suppose that $x^* = P_X[x^* - rH(x^*)]$, then by lemma 3:

$$\langle x - x^*, -rH(x^*) \rangle \leq 0 \quad \forall x \in X, \quad (42)$$

and since r is positive, it follows that x^* solves (35). Conversely, suppose that x^* solves inequality (35), then

$$\langle x - x^*, H(x^*) \rangle \geq 0 \quad \forall x \in X, \quad (43)$$

which can be rewritten

$$\langle x - x^*, x^* - (x^* - rH(x^*)) \rangle \geq 0 \quad \forall x \in X; \quad (44)$$

and using lemma 3 we get

$$x^* = P_X[x^* - rH(x^*)]. \quad (45)$$

□

This equation can be solved using a simple fixed-point method: Given $x_0 \in X$ compute the solution at step $n + 1$ by iterating the scheme

$$x_{n+1} = P_X(x_n - rH(x_n)), \quad (46)$$

until convergence. Different strategies for choosing r are discussed in [42]. Here, for simplicity we'll pick $r = 0.2$ as a constant.

4. Primal-Dual Beltrami Regularization

Let us now return to the Beltrami regularization problem, and apply the primal-dual projected gradient method for its optimization. We first consider the simple case of denoising, as given by the discrete problem stated in (24). Once the primal-dual projected gradient method has been described for this starting case, it will be easily adapted to more complicated image restoration problems.

4.1. Primal-Dual approach

Computationally, the Beltrami denoising model (24) is usually solved by its formal Euler-Lagrange equations. This would be viewed as the primal approach to solving the Beltrami denoising problem, and I is the primal variable. Based on the results of the previous section, we can now formulate the primal-dual problem. By comparison of our Beltrami-denoising optimization problem with the primal-dual model in theorem 1, we get the following notation:

$$F(KI) := \sum_{ij} \sqrt{1 + \beta^2 |(\nabla I)(i, j)|^2} \quad \text{and} \quad (47)$$

$$G(I) := \sum_{ij} \frac{\lambda}{2} (I(i, j) - I_0(i, j))^2 \quad (48)$$

with the reflexive spaces $V = \mathbb{R}^{m \times n}$ and $W = \mathbb{R}^{(m \times n) \times 2}$. Now, we define the functionals:

$$\begin{aligned} K: V &\rightarrow W \\ I &\mapsto \nabla I \end{aligned} \quad \text{and} \quad \begin{aligned} F: W &\rightarrow \mathbb{R} \\ x &\mapsto \sum_{ij} f(x(i, j)), \end{aligned} \quad (49)$$

with $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(s) = \sqrt{1 + \beta^2 |s|^2}$. The Legendre-Fenchel transform f^* of the function f is found as:

$$\begin{aligned} f^*: \mathbb{R}^2 &\rightarrow \mathbb{R} \\ s &\mapsto \begin{cases} -\frac{\sqrt{\beta^2 - |s|^2}}{\beta} & \text{if } \beta^2 \geq |s|^2 \\ +\infty & \text{otherwise,} \end{cases} \end{aligned} \quad (50)$$

and therefore the convex conjugate F^* of F is

$$\begin{aligned} F^*: W &\rightarrow \mathbb{R} \\ \varphi &\mapsto \sum_{ij} f^*(\varphi(i, j)). \end{aligned} \quad (51)$$

It is easy to prove that F is convex. Thus, using theorem 1, we now get the equivalent primal-dual problem:

$$\min_I \max_{\varphi \in X} \left\{ \sum_{ij} \langle \nabla I, \varphi \rangle + \frac{\sqrt{\beta^2 - |\varphi|^2}}{\beta} + \frac{\lambda}{2} (I - I_0)^2 \right\}, \quad (52)$$

where $X = W^* = \left\{ \varphi \in \mathbb{R}^{(m \times n) \times 2} \mid |\varphi(i, j)|^2 \leq \beta^2 \right\}$. Further, using $\nabla^* = -\operatorname{div}$ by construction of the discretized differential operators, (52) can be rewritten:

$$\min_I \max_{\varphi \in X} \left\{ \sum_{ij} -I \operatorname{div} \varphi + \frac{\sqrt{\beta^2 - |\varphi|^2}}{\beta} + \frac{\lambda}{2} (I - I_0)^2 \right\}. \quad (53)$$

4.2. Algorithm

Having reformulated the Beltrami denoising problem in a primal-dual form, we can now make use of the projected gradient method for optimization.

Using the notation in (35) we get for each (i, j)

$$H(I, \varphi) = \begin{pmatrix} -\operatorname{div} \varphi + \lambda(I - I_0) \\ -\beta \nabla I / \sqrt{\beta^2 - |\varphi|^2} + \varphi \end{pmatrix} \quad (54)$$

In order to avoid a division by zero, we have multiplied the second term of H by $\beta \sqrt{\beta^2 - |\varphi|^2}$. Given a current estimate (I^k, φ^k) at iteration step k , the projection algorithm (46) then yields the following update formula (for each (i, j)):

- 1) $\varphi^{k+1} = P_X \left(\varphi^k - r_1 \left(-\beta \nabla I^k / \sqrt{\beta^2 - |\varphi^k|^2} + \varphi^k \right) \right)$
- 2) $I^{k+1} = I^k - r_2 \left(-\operatorname{div} \varphi^{k+1} + \lambda (I^k - I_0) \right)$

where r_1, r_2 are positive constants and the projection on the $X = W^*$ is simply given by

$$P_X(\varphi)(i, j) = \frac{\beta \varphi(i, j)}{\max(|\varphi(i, j)|, \beta)}. \quad (55)$$

Indeed,

$$|P_X(\varphi)(i, j)|^2 = \begin{cases} |\varphi(i, j)|^2 & \text{if } \beta \geq |\varphi(i, j)| \\ \beta^2 & \text{otherwise} \end{cases} \quad (56)$$

as required. This primal-dual projected gradient algorithm for Beltrami-regularized image denoising is summarized in algorithm 1.

Algorithm 1 Beltrami primal-dual denoising

Initialize $I^0 = I_0, \varphi^0 = 0$.

repeat

$$\begin{cases} \bar{\varphi}^{k+1} & \leftarrow (1 - r_1) \varphi^k + \beta r_1 \nabla I^k / \sqrt{\beta^2 - |\varphi^k|^2} \\ \varphi^{k+1} & \leftarrow \frac{\beta \bar{\varphi}^{k+1}}{\max(|\bar{\varphi}^{k+1}|, \beta)} \\ I^{k+1} & \leftarrow (1 - \lambda r_2) I^k + r_2 (\operatorname{div} \varphi^{k+1} + \lambda I_0) \end{cases} \quad (280)$$

until convergence

5. Denoising results

We have implemented the proposed primal-dual projected gradient Beltrami-denoising algorithm. Here, we want to demonstrate its use, and compare our Beltrami algorithm with a simple primal-dual formulation of the ROF model [41, 42], known to be efficient. The methods are very comparable in structure. Indeed, the only difference between the proposed Beltrami denoising and the primal-dual ROF model is in the projection and the update of the dual variable. Here, the ROF algorithm (see Algorithm 2) uses

$$\varphi^{k+1} = P_X(\varphi^k - r_1 \nabla I^k) \text{ with } P_X(\varphi) = \frac{\varphi}{\max(|\varphi|, 1)}. \quad (57)$$

It has been shown that this algorithm converges very fast and gives a better convergence rate, e.g., compared with Chambolle's projection method [39].

Algorithm 2 ROF primal-dual denoising

Initialize $I^0 = I_0, \varphi^0 = 0$.

repeat

$$\begin{cases} \bar{\varphi}^{k+1} & \leftarrow \varphi^k - r_1 \nabla I^k \\ \varphi^{k+1} & \leftarrow \frac{\bar{\varphi}^{k+1}}{\max(|\bar{\varphi}^{k+1}|, 1)} \\ I^{k+1} & \leftarrow (1 - \lambda r_2) I^k + r_2 (\operatorname{div} \varphi^{k+1} + \lambda I_0) \end{cases}$$

until convergence

In figure 2 we illustrate results for denoising a series of test images using $\lambda = 0.075$ and $\beta = 1$. The images have been degraded by an additive Gaussian noise of standard deviation $\sigma = 10$ (compared to an image intensity range of $[0, 256]$). In all examples we use the same descent parameter $r = 0.2$. We choose to terminate the algorithm when the primal relative error improvement, $(P(u^{k+1}) - P(u^k)) / P(u^0)$, reaches a desired precision, here $\varepsilon = 2 \cdot 10^{-5}$. Both methods perform almost equally; only subtle differences are visible in image locations with smoothly varying intensities, such as Lenna's cheek.

Indeed, this is the major difference between the two methods. While the ROF model is feature preserving, it also has a tendency to produce artificial intensity jumps (staircasing) in the output images. The Beltrami method is also feature preserving, but has greatly reduced staircasing. This difference is best appreciated in Figure 3.

To measure more quantitatively the similarity between the recovered image and the original (supposedly noise-free) image, we use both the signal-to-noise ratio, and the structural similarity index (SSIM), the latter being well known to better reflect perceived visual quality [48]. Comparative numbers are reported for both methods and over a series of 16 test images, in Table 1. It can be seen that the quality difference between ROF and Beltrami is not extreme, but almost consistently in favor of the proposed Beltrami model. It is also to note that given the



Figure 2: Denoising results: Beltrami and TV denoising applied to 7 artificially degraded test images. Noise level $\sigma = 10$ for images in $[0, 256]$. Parameters have been chosen to denoise somewhat excessively, in order to make differences between Beltrami and TV more visible. Beltrami denoising has smoother gradients compared to the staircasing seen with TV.

Image	Beltrami				TV			
	SNR	SSIM	Iter.	CPU [s]	SNR	SSIM	Iter.	CPU [s]
cameraman*	25.632	0.888	33	0.624	25.574	0.886	74	1.248
circuit*	24.562	0.877	34	0.764	24.401	0.870	59	1.108
coins*	25.968	0.930	37	0.733	25.980	0.932	77	1.451
concordorthophoto	23.605	0.828	35	0.546	23.531	0.823	55	0.655
moon	28.662	0.897	46	2.683	28.659	0.908	86	4.696
liftingbody	32.106	0.951	47	4.228	31.981	0.948	92	6.505
mri	23.027	0.910	27	0.109	23.090	0.941	63	0.328
rice*	23.838	0.811	31	0.546	23.728	0.801	59	0.764
snowflakes	15.660	0.921	41	0.328	15.518	0.917	80	0.530
text	24.574	0.901	33	0.515	24.913	0.950	70	0.967
tire	22.880	0.916	41	0.406	22.803	0.914	67	0.593
bag	21.984	0.912	31	0.296	21.938	0.910	49	0.484
boat*	26.291	0.942	37	3.089	26.174	0.939	69	5.132
lenna*	25.084	0.948	40	3.448	24.940	0.945	78	5.710
lincoln*	28.519	0.845	46	2.012	28.306	0.834	78	2.839
t1brain	24.514	0.926	36	0.546	24.461	0.934	68	0.874

Table 1: Beltrami and TV denoising applied to a larger set of test images. Asterisk denotes images shown in figure 2. Signal-to-noise ratio and structural similarity Index are in favor of the Beltrami model (**bold**). Also, the Beltrami model converges in a little more than half the iterations required for the TV model. Each iteration is only slightly more computationally complex than the TV algorithm, since both implementations are based on a very similar primal-dual projected gradient method and thus structurally identical. Convergence criterion is based on the decrease of the primal problem energy.

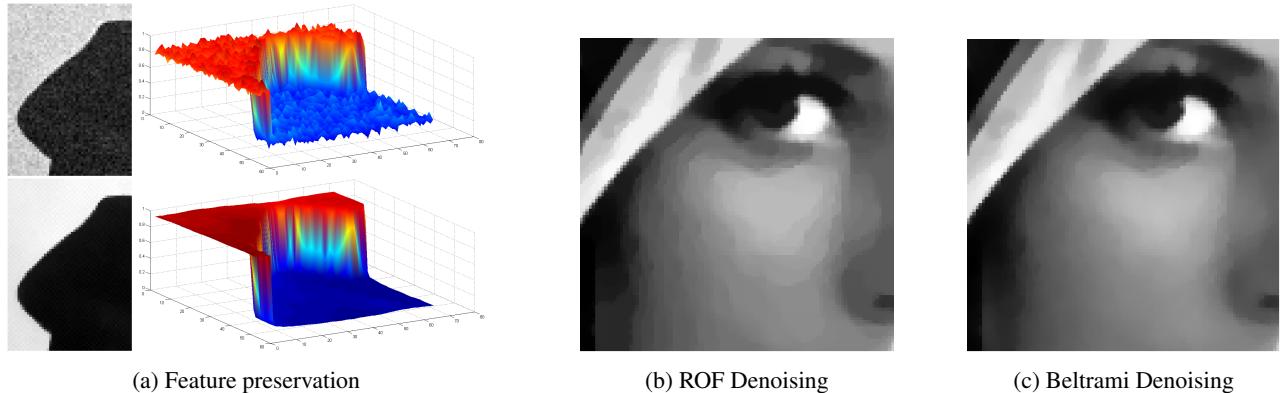


Figure 3: Feature preservation versus staircasing. (a) A surface plot of a section of the cameraman image before and after Beltrami denoising illustrates the feature preserving nature of the denoising model: The steep edge is preserved (feature), while the little oscillations (noise) are removed. (b)–(c) The TV model is also feature preserving, but has a strong tendency to introduce piecewise constant patches separated by artificial steps (staircasing). This effect is much less prominent in the Beltrami model, where soft gradients remain smooth.

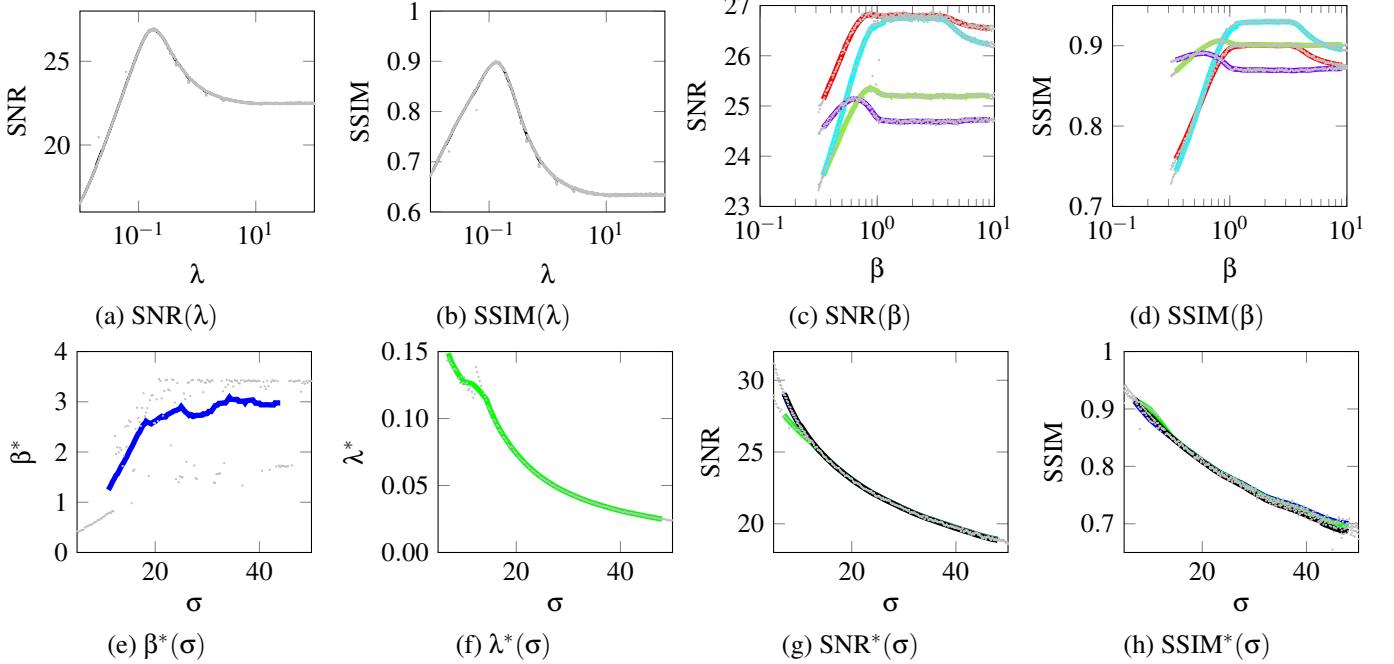


Figure 4: Parameter choice and sensitivity. (a)–(b) For a given image and a fixed noise level, the balancing parameter λ has a prominent peak in the resulting signal-to-noise ratio (SNR) and the structural similarity index (SSIM). (c)–(d) The optimal aspect ratio β is image dependent, but has also exhibits a flatter plateau of near-optimality. (e)–(f) the locations of the optimal parameters β^* and λ^* as determined by numerical optimization. For significant noise levels ($\sigma > 10$), the optimal β is almost noise independent, and exhibits large variation. Due to the flat plateau, almost any value between 1 and 3.5 is “good”. In contrast, the optimal balancing parameter λ is well-defined by an inverse proportional relationship with the noise-level. (g)–(h) SNR/SSIM at SNR-optimal parameters β^* and λ^* . (blue: optimize both β and λ , green: keep $\beta = 2$ and only optimize λ , black: TV with optimal λ). Optimally tuned Beltrami model slightly outperforms TV denoising.

primal-energy decrease criterion, the Beltrami model converges in roughly half as many iterations as the TV model.

While the ROF model has only one balancing parameter, which is essentially governed by the noise level, the Beltrami model involves the aspect ratio β , in addition. We study the influence of parameter selection (β and λ) in terms of SNR and SSIM. The curves are shown in Figure 4(a)–(d). In addition, we have performed numerical optimization to search for optimal β^* and λ^* over a broad range of noise levels, σ , see Figure 4(e)–(f). The variation on β^* is large, and above a certain noise level the optimal choice seems largely independent of σ . In contrast, the optimal balancing parameter λ is clearly anti-proportional to the noise-level. In conclusion, as with the ROF model, the balancing parameter λ is governed by the noise level, is largely image independent and exhibits a prominent peak. Conversely, the aspect ratio β is mostly image dependent, and typically has a broader plateau. In a typical application, the balancing parameter λ can be inferred from the estimated noise level, while the optimal β should be learned from a set of representative training images. In our case, the curves suggest optimal values in a typical range of $1 - 3$ (and less in low-noise regimes). Finally, we have traced the SNR and SSIM achieved at optimal parameter choices, for both the Beltrami and the ROF model, see Figure 4(g)–(h). The curves suggest that the Beltrami model typically outperforms ROF in both SNR and SSIM, albeit only marginally so.

6. Application to Inpainting and Super-resolution

Inpainting techniques aim to reconstruct or filling a lost part of an image. In this section, we apply our proposed primal-dual projected gradient Beltrami method to image inpainting and we show on numerical results that our method is well adapted to this application.

6.1. Problem presentation

Suppose we have an image I defined on a domain $\Omega \subset \mathbb{R}^2$ and a subset $D \subset \Omega$ thereof, where data of the image has been lost. We now want to find

$$u = \Phi^{-1}(f) \quad (58)$$

where Φ is a damaging operator. In fact, this is a projection operator on $\Omega \setminus D$ which can be written:

$$\Phi(u)(x) = \begin{cases} u(x) & \text{if } x \notin D, \\ 0 & \text{if } x \in D. \end{cases} \quad (59)$$

We suppose that the set of lost data D is finite. The inpainting model using the Beltrami energy is then defined as

$$\min_I \left\{ \sum_{\Omega} \sqrt{1 + \beta^2 |\nabla I|^2} + \frac{\lambda}{2} \sum_{\Omega \setminus D} (I - I_0)^2 \right\}. \quad (60)$$

Using the same approach as in the denoising minimization problem, we reach the following primal-dual formulation:

$$\min_{I \in \mathbb{R}^{m \times n}} \max_{\varphi \in X} \left\{ \sum_{\Omega} \left(-I \operatorname{div} \varphi + \frac{\sqrt{\beta^2 - |\varphi|^2}}{\beta} \right) + \frac{\lambda}{2} \sum_{\Omega \setminus D} (I - I_0)^2 \right\}, \quad (61)$$

where $X = \left\{ \varphi \in \mathbb{R}^{(m \times n) \times 2}, |\varphi(i, j)|^2 \leq \beta^2 \right\}$. This problem can be solved analogously to the algorithm of the previous section. The dual update is exactly the same, while we have the following update of I :

$$I^{k+1} = (1 - \lambda_e r_2) I^k + r_2 \left(\operatorname{div} \varphi^{k+1} + \lambda_e I_0 \right) \quad (62)$$

where the balancing weight λ_e is simply given by $\lambda_e = \lambda(1 - \chi_D)$, and χ_D represents the characteristic function on the inpainting domain D . This means that the fidelity term is only active where image data is available, while on D only the Beltrami regularity prior is acting.

When the original image is considered as clean, this would correspond to picking $\lambda \rightarrow \infty$; however, doing so greatly affects the conditioning of the system. Instead, we enforce the data-fidelity constraint on $\Omega \setminus D$ directly. Here, the image update equation is replaced by

$$I^{k+1}(x) = \begin{cases} I^k(x) + r_2 \operatorname{div} \varphi^{k+1}(x) & \text{if } x \in D \\ I_0(x) & \text{otherwise} \end{cases} \quad (63)$$

The two inpainting algorithms, corresponding to noisy and clean images, are summarized below, as algorithms 3 and 4.

Algorithm 3 Beltrami primal-dual inpainting (noisy image)

Initialize $I^0 = I_0, \varphi^0 = 0$.

repeat

$$\begin{cases} \bar{\varphi}^{k+1} \leftarrow (1 - r_1) \varphi^k + \beta r_1 \nabla I^k \sqrt{\beta^2 - |\varphi^k|^2} \\ \varphi^{k+1} \leftarrow \frac{\beta \bar{\varphi}^{k+1}}{\max(|\varphi^{k+1}|, \beta)} \\ I^{k+1} \leftarrow (1 - \lambda_e r_2) I^k + r_2 (\operatorname{div} \varphi^{k+1} + \lambda_e I_0) \end{cases}$$

until convergence

6.2. Results

We demonstrate the proposed versions of Beltrami inpainting algorithms on two test images, as shown in Figure 5. The two test images have been artificially corrupted by discarding $p = \{15, 50, 85\}\%$ of the image pixels, randomly. Our examples illustrate the subtle difference between the two inpainting versions, *noisy* and *clean*: While the *clean* algorithm will reconstruct an image that exactly matches the input on $\Omega \setminus D$, including potentially noisy pixels and artifacts, the *noisy* algorithm incorporates some slack in this data-fidelity and we achieve a certain amount of denoising in the process.

Algorithm 4 Beltrami primal-dual inpainting (clean image)

Initialize $I^0 = I_0, \varphi^0 = 0$.

repeat

$$\begin{cases} \bar{\varphi}^{k+1} \leftarrow (1 - r_1) \varphi^k + \beta r_1 \nabla I^k \sqrt{\beta^2 - |\varphi^k|^2} \\ \varphi^{k+1} \leftarrow \frac{\beta \bar{\varphi}^{k+1}}{\max(|\varphi^{k+1}|, \beta)} \\ I^{k+1} \leftarrow \begin{cases} I^k(x) + r_2 \operatorname{div} \varphi^{k+1}(x) & \text{if } x \in D \\ I_0(x) & \text{otherwise} \end{cases} \end{cases}$$

until convergence

Although not shown here, the TV-based results compare analogously to the denoising problem.

Also note that this Beltrami inpainting model can be easily extended to image zooming, where simply the missing pixels will have a regular geometric structure. Indeed, we can think of image zooming as filling a lost part of an image, the lost part being now between each pixels.

7. Application to non-blind deconvolution

In this section we apply the primal-dual projected gradient approach to the Beltrami deconvolution (deblurring) problem.

7.1. Problem statement

We assume here that the observed image I_0 is a blurred and noisy version of an unobserved image I :

$$I_0 = h * I + n, \quad (64)$$

where h is the blurring kernel (point spread function), and $*$ denotes convolution. We can formulate the Beltrami-regularized inverse problem as follows:

$$\min_{I \in \mathbb{R}^{m \times n}} \max_{\varphi \in X} \left\{ \sum_{\Omega} -I \operatorname{div} \varphi + \frac{\sqrt{\beta^2 - |\varphi|^2}}{\beta} + \frac{\lambda}{2} (h * I - I_0)^2 \right\}. \quad (65)$$

We can notice that when $h = \delta$ (δ being the Dirac impulse, i.e., the identity element under convolution), the problem becomes the initial Beltrami denoising model.

The only difference between the denoising and the deconvolution algorithm appears in the update of the primal variable through gradient descent. Here, as a little extra, we propose an implicit gradient descent:

$$I^{k+1} = I^k + r_2 \left(\frac{\operatorname{div} \varphi^{k+1}}{\lambda} - h^s * (h * I^{k+1} - I_0) \right), \quad (66)$$

and solve it spectrally as follows:

$$I^{k+1} = \mathcal{F}^{-1} \left\{ \frac{\mathcal{F} \{ I^k + \frac{r_2}{\lambda} \operatorname{div} \varphi^{k+1} \} + r_2 H \mathcal{F} \{ I_0 \}}{1 + r_2 |H|^2} \right\}, \quad (67)$$

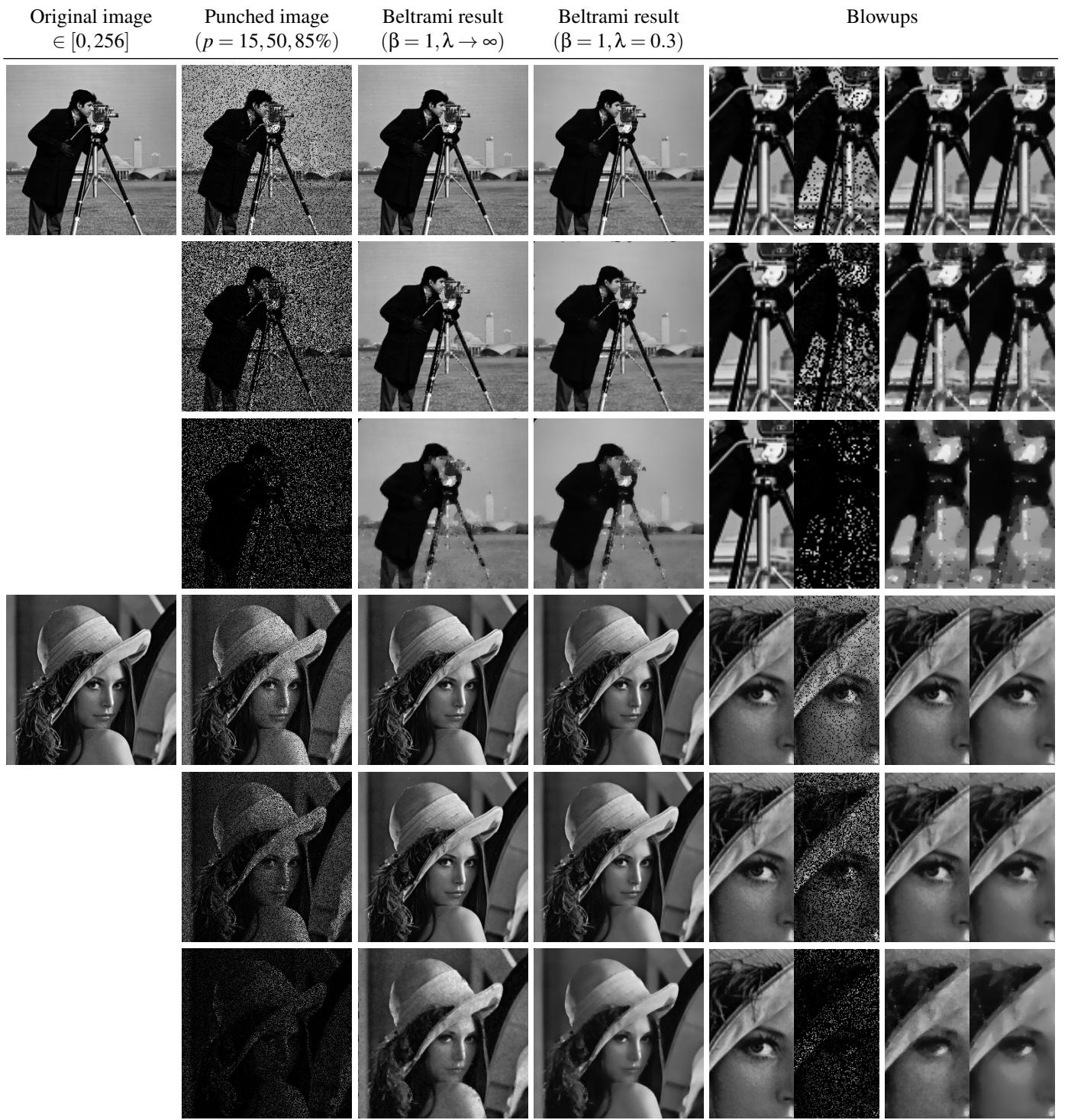


Figure 5: Beltrami regularized inpainting results. Two test images are artificially degraded (removal of 15-85% of the pixels). Both “clean” ($\lambda \rightarrow \infty$) and “noisy” ($\lambda = 0.3$) algorithms are applied.

where \mathcal{F} denotes the Fourier transform, $H = \mathcal{F}\{h\}$, and \bar{H} denotes the complex conjugate. The spectral approach is computationally interesting for larger convolution kernels, where FFT and spectral multiplication are more efficient than convolution in time domain. For the complete non-blind deconvolution algorithm, see algorithm 5.

Algorithm 5 Beltrami primal-dual non-blind deconvolution

Initialize $I^0 = I_0, \phi^0 = 0$.
repeat

$$\left\{ \begin{array}{l} \bar{\phi}^{k+1} \leftarrow (1 - r_1)\phi^k + \beta r_1 \nabla I^k \sqrt{\beta^2 - |\phi^k|^2} \\ \phi^{k+1} \leftarrow \frac{\beta \bar{\phi}^{k+1}}{\max(|\bar{\phi}^{k+1}|, \beta)} \\ I^{k+1} \leftarrow \mathcal{F}^{-1} \left\{ \frac{\mathcal{F}\{I^k + \frac{r_2}{\lambda} \operatorname{div} \phi^{k+1}\} + r_2 \bar{H} \mathcal{F}\{I_0\}}{1 + r_2 |H|^2} \right\} \end{array} \right.$$

until convergence

7.2. Results

We showcase the proposed Beltrami-regularized deconvolution algorithm in Figure 6. Six test images were artificially degraded by a linear motion-blur kernel (120° angle, 12 pixels long), and affected by a low level of noise ($\sigma = 5$). The deconvolution is generally good and relates similarly to the TV-based results (not shown) as in the denoising case; for the black-and-white text image, however, the Beltrami regularizer clearly is not the ideal prior.

8. Application in Compressed Sensing

In this section, we consider the problem of reconstructing an object from partial frequency information. For instance, MRI measurements can be interpreted as a selection of a few Fourier frequencies, because densely sampling the entire k-space would take too long.

8.1. Problem statement

The problem is to recover f from incomplete Fourier samples. This problem can be seen as a non-blind deconvolution (see section 7, above). Indeed, here, compressed sensing refers to applying a binary frequency mask H to the spectrum of the image,

$$\mathcal{F}\{I_0\} = H \cdot \mathcal{F}\{I\}, \quad (68)$$

where \cdot denotes point-wise multiplication. Due to the convolution theorem, this can be interpreted as convolution by the filter $h = \mathcal{F}^{-1}\{H\}$,

$$(h * I)(x) = (\mathcal{F}^{-1}\{H\} * I)(x). \quad (69)$$

Eventually, we are thus dealing with a non-blind deconvolution problem, and the image is recovered using the same algorithm as in section 7. Note that in this case, the filter h will not be

sparsely supported in time domain, but has a much nicer definition in spectral domain by construction, and therefore the spectral update proposed in the deconvolution algorithm is particularly interesting.

8.2. Results

To show successful image reconstruction from sparse Fourier samples using the proposed Beltrami model, we test our algorithm on the same series of 6 test images. For each of these images, we keep frequencies on a set of 25, 50, and 100 radial lines only. This corresponds to between 5 and 35 % of spectral coefficients, depending on image dimensions. The results of our algorithm are shown in Figure 7.

9. Extension to multichannel and color images

So far, we have dealt grey-scale image reconstruction problems, only. However, this is just a particular case, and more generally images have higher co-dimension. Such multi-channel images occur for example with color, texture, multimodal or hyperspectral features. In particular, RGB-color images have co-dimension 3.

Definition 15 (Multichannel image). We consider a d -channel image as a function \mathbf{u} defined as

$$\begin{aligned} \mathbf{u}: \Omega \subset \mathbb{R}^2 &\rightarrow \mathbb{R}^d \\ x &\mapsto \mathbf{u}(x) := (u_1(x), \dots, u_d(x)). \end{aligned} \quad (70)$$

Thus, d -channel images are simply a collection of d grey-scale images u_1, \dots, u_d .

9.1. Simplified multichannel Beltrami functional

The Beltrami functional derived in differential geometry terms now leads to particular gradient coupling terms in the determinant [15, 17]. However, these gradient terms render the functional non-convex, and are thus a computational obstacle in the present context.

Definition 16 (Simplified multichannel Beltrami energy). Here, we thus formally define a simpler d -channel Beltrami energy as

$$E(\mathbf{u}) := \int_{\Omega} \sqrt{1 + \beta^2 \sum_{s=1}^d |\nabla u_s|^2}. \quad (71)$$

With this simplified d -channel Beltrami regularizer discretized as before, we rewrite the initial denoising model using Beltrami regularization and L_2 fidelity term wrt. to given data \mathbf{u}_0 :

$$\min_{\mathbf{u}} \left\{ \sum_{\Omega} \sqrt{1 + \beta^2 \sum_{s=1}^d |\nabla u_s|^2} + \frac{\lambda}{2} \sum_{s=1}^d (u_s - u_{0,s})^2 \right\} \quad (72)$$

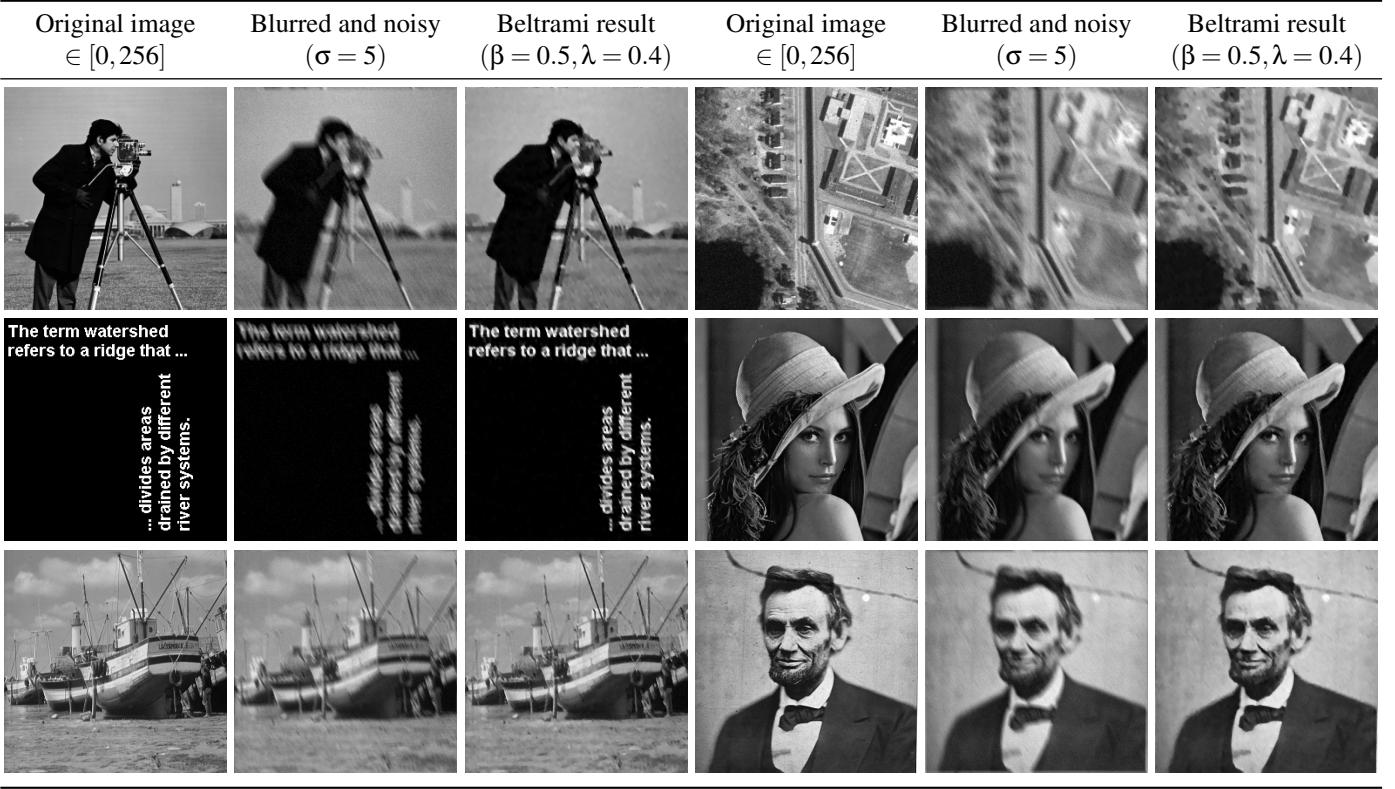


Figure 6: Beltrami-regularized deconvolution results. 6 test images are artificially degraded by a known motion blur filter (orientation 120° , length 12 pixels), and a low amount of noise is added to the blurred image ($\sigma = 5$). Estimates of the original images are recovered using fixed Beltrami parameters ($\beta = 0.5, \lambda = 0.4$). The Beltrami prior seems appropriate with natural images, but clearly underperforms on the black-and-white text image.

Using the same method as in the scalar case, we can proceed⁴¹⁵ to the definition of the following equivalent primal-dual variational model:

$$\min_{\mathbf{u}} \max_{\sum_{s=1}^d |\varphi_s|^2 \leq \beta^2} \left\{ \sum_{\Omega} - \sum_{s=1}^d u_s \operatorname{div} \varphi_s + \frac{\sqrt{\beta^2 - \sum_{s=1}^d |\varphi_s|^2}}{\beta} \right. \\ \left. + \frac{\lambda}{2} \sum_{s=1}^d (u_s - u_{0,s})^2 \right\} \quad (73)$$

The modified variational inequality problem can again be solved using a projection type method:⁴²⁵

$$u_s^{k+1} = u_s^k - r_1 \left(-\operatorname{div} \varphi_s^{k+1} + \lambda (u_s^k - u_{s,0}) \right) \quad (74)$$

and

$$\varphi_s^{k+1} = P_X \left(\varphi_s^k - r_2 \left(-\beta \nabla u_s^k \sqrt{\beta^2 - \sum_{l=1}^d |\varphi_l|^2} + \varphi_s^k \right) \right). \quad (75)$$

Here, the projection on the ball $X = \{x \mid |x|^2 \leq \beta^2\}$ is simply given by⁴³⁵

$$P_X(p_s) = \frac{\beta p_s}{\max(\beta, \sqrt{\sum_{l=1}^d |p_l|^2})}. \quad (76)$$

9.2. Results

We demonstrate the denoising power of the proposed simplified color-Beltrami model on 4 different color images. Images and denoising results are shown in Figure 8. We have chosen parameters such that the denoising is rather excessive, in order to better highlight the nature of the denoising. The simplified color-Beltrami model can be seen to effectively remove color granularity (noise, texture), while preserving features such as lines and edges. Color gradients are smooth and do not exhibit artificial staircasing.

10. Conclusions

In this paper, we have proposed a primal-dual projected gradient method to efficiently solve the Beltrami regularization based image denoising. This algorithm is inspired by very successful schemes used with the ROF model, proposed by Zhu and Chan in [41, 42]. Beltrami regularization offers an interesting compromise between the smooth regularization of H^1 priors, and the feature preservation of TV regularization. Until now, this compromise came at a cost in the form of increased computational complexity and was thus rarely used in practice. Here, we present a generic Beltrami-based denoising algorithm and compare its performance to the corresponding TV-based algorithm. We can show that increase denoising quality can be achieved at comparable computational complexity.



Figure 7: Beltrami-regularized compressed sensing results. 6 test images are artificially degraded by removing large parts of their Fourier spectrum. For each image, only Fourier coefficients on 25, 50, and 100 radial lines are conserved. For cameraman, orthophoto and text, the three masks correspond to $\{9, 18, 35\}$ % of Fourier coefficients observed, for Lenna and boat to $\{5, 9, 19\}$ %, and for Lincoln to $\{7, 13, 25\}$ %, respectively, due to larger image dimensions.

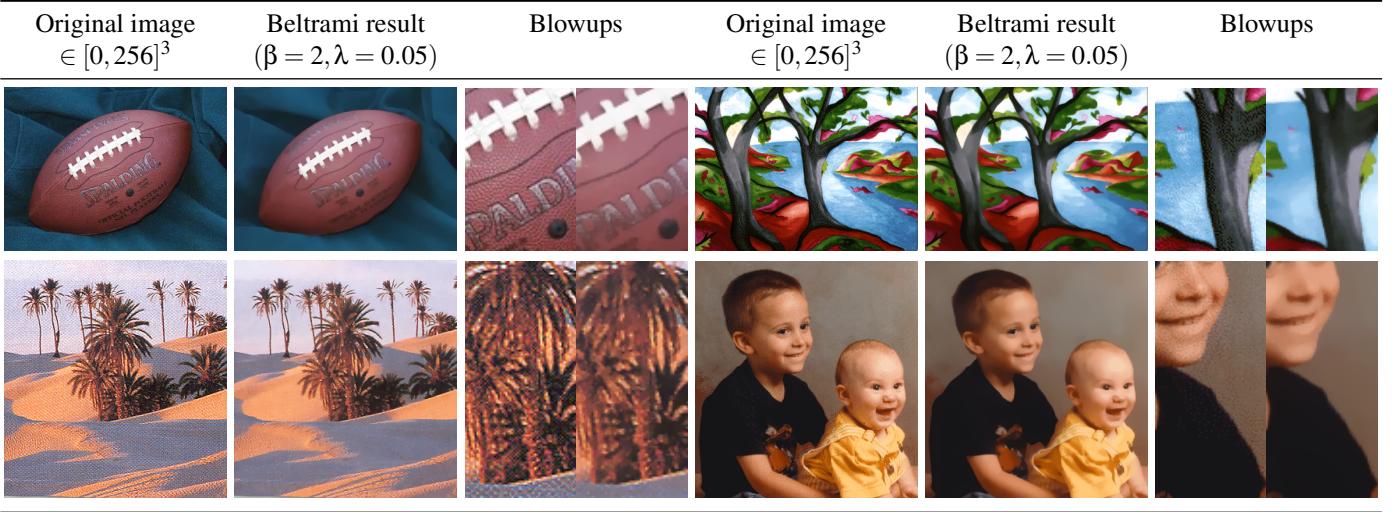


Figure 8: Simplified Color-Beltrami denoising results. 4 color test images with strong inherent noise (object texture, half toning).

The fundamental idea is to formulate an equivalent optimization problem involving a dual variable to the primal (image) variable, and where the Beltrami-regularizer is replaced by the convex conjugate of this dual variable. The resulting saddle-point problem is solved through a variational inequality.⁴⁷⁵ The primal variable is optimized by gradient descent, and the dual variable are optimized by gradient ascent followed by projection on the appropriate convex set. In this work we use a fixed steps size in the gradient descent/ascent equations, however, more elaborate schemes could be employed in order to further increase the computational efficiency of the algorithms, as has been done for the ROF model [41, 42].

Beyond the simple denoising problem, we extend the primal-dual projected gradient Beltrami regularization algorithm to a larger family of inverse image processing problems, such as inpainting, deconvolution and compressed sensing. All these cases can be tackled in a very similar way. Eventually, we also propose a simplified color-extension of the Beltrami regularization model, that lends itself to optimization in a very similar fashion. Here, simplification with respect to the “true” color-Beltrami model [17] is necessary in order to maintain a convex functional. Future work will focus on convex relaxations of the true multichannel-Beltrami model to not have to make such simplifications.

A MATLAB implementation of the proposed algorithms⁵⁰⁵ will be made available at <http://www.math.ucla.edu/~zosso/code.html>.

References

- [1] J. Hadamard, Sur les problèmes aux dérivées partielles et leur signification physique, Princeton University Bulletin 13 (1902) 49–52.
- [2] M. Bertero, T. A. Poggio, V. Torre, Ill-Posed Problems in Early Vision, in: Proceedings of the IEEE, Vol. 76, 1988, pp. 869–889.
- [3] A. N. Tikhonov, Solution of incorrectly formulated problems and the regularization method, Soviet Mathematics 4 (1963) 1035–1038.
- [4] T. Poggio, V. Torre, C. Koch, Computational vision and regularization theory, Nature 317 (6035) (1985) 314–319. doi:10.1038/317314a0.
- [5] N. K. Nordström, Biased anisotropic diffusion: a unified regularization and diffusion approach to edge detection, Image and Vision Computing 8 (4) (1990) 318–327. doi:10.1016/0262-8856(90)80008-H.
- [6] L. Alvarez, P.-L. Lions, J.-M. Morel, Image Selective Smoothing and Edge Detection by Nonlinear Diffusion. II, SIAM Journal on Numerical Analysis 29 (3) (1992) 845–866. doi:10.1137/0729052.
- [7] Y.-L. You, M. Kaveh, Blind Image Restoration by Anisotropic Regularization, IEEE Transactions on Image Processing 8 (3) (1999) 396–407.
- [8] D. Forsberg, M. Andersson, H. Knutsson, Adaptive anisotropic regularization of deformation fields for non-rigid registration using the morphon framework, no. x, IEEE, 2010. doi:10.1109/ICASSP.2010.5495704.
- [9] H. Hong, I. K. Park, Single Image Motion Deblurring Using Anisotropic Regularization, IEEE International Conference on Image Processing ICIP 2010 (2010) 1149–1152.
- [10] L. Rudin, S. Osher, E. Fatemi, Nonlinear Total Variation based Noise Removal Algorithms, Physica D 60 (1-4) (1992) 259–268. doi:10.1016/0167-2789(92)90242-F.
- [11] L. Vese, A Study in the BV Space of a Denoising–Deblurring Variational Problem, Applied Mathematics and Optimization 44 (2) (2001) 131–161. doi:10.1007/s00245-001-0017-7.
- [12] D. M. Strong, P. Blomgren, T. F. Chan, Spatially adaptive local-feature-driven total variation minimizing image restoration, in: Proceedings of SPIE 3167, Vol. 3167, SPIE, 1997, pp. 222–233. doi:10.1117/12.279642.
- [13] X. Bresson, S. Esedoglu, P. Vandergheynst, J.-P. Thiran, S. Osher, Fast Global Minimization of the Active Contour/Snake Model, Journal of Mathematical Imaging and Vision 28 (2) (2007) 151–167. doi:10.1007/s10851-007-0002-0.
- [14] R. Kimmel, N. Sochen, R. Malladi, From high energy physics to low level vision, Scale-Space Theory in Computer Vision 1252 (1997) 236–247.
- [15] N. Sochen, R. Kimmel, R. Malladi, A general framework for low level vision, IEEE Transactions on Image Processing 7 (3) (1998) 310–318.
- [16] A. M. Polyakov, Quantum geometry of bosonic strings, Phys. Lett. B 103 (3) (1981) 207–210.
- [17] R. Kimmel, R. Malladi, N. Sochen, Images as Embedded Maps and Minimal Surfaces: Movies, Color, Texture, and Volumetric Medical Images, Int. J. Comput. Vis. 39 (2) (2000) 111–129.
- [18] X. Bresson, P. Vandergheynst, J.-P. Thiran, Multiscale Active Contours, Int. J. Comput. Vis. 70 (3) (2006) 197–211.
- [19] I. Bogdanova, X. Bresson, J.-P. Thiran, P. Vandergheynst, Scale space analysis and active contours for omnidirectional images., IEEE Transactions on Image Processing 16 (7) (2007) 1888–1901.
- [20] G. Rosman, X.-C. Tai, L. Dascal, R. Kimmel, Augmented Lagrangian for Polyakov Action Minimization in Color Images, in: ICNAAM 2010: International Conference of Numerical Analysis and Applied Mathematics 2010, Vol. 1281, 2010, pp. 1018–1021. doi:10.1063/1.3497793.
- [21] G. Rosman, X.-C. Tai, L. Dascal, R. Kimmel, Polyakov Action Mini-

- mization for Efficient Color Image Processing, Tech. Rep. 10-46, CAM, Department of Mathematics, UCLA (2010).
- [22] G. Rosman, X.-C. Tai, L. Dascal, R. Kimmel, Polyakov Action for Efficient Color Image Processing, Tech. Rep. 623, Technion CIS-2010-04 (2010).
- [23] R. Ben-Ari, N. Sochen, A geometric framework and a new criterion in optical flow modeling, *J. Math. Imaging Vis.* 33 (2) (2009) 178–194.
- [24] R. Courant, K. Friedrichs, H. Lewy, Über die partiellen Differenzengleichungen der mathematischen Physik, *Mathematische Annalen* 100 (1) (1928) 32–74.
- [25] L. Dascal, N. A. Sochen, A Maximum Principle for Beltrami Color Flow, *SIAM J. Appl. Math.* 65 (5) (2005) 1615–1632.
- [26] R. Malladi, I. Ravve, Fast Difference Schemes for Edge Enhancing Beltrami Flow, in: *ECCV 2002*, Vol. 2350 of *Lecture Notes in Computer Science*, 2002, pp. 343–357. doi:[10.1007/3-540-47969-4_23](https://doi.org/10.1007/3-540-47969-4_23).
- [27] D. Barash, T. Schlick, M. Israeli, R. Kimmel, Multiplicative Operator Splittings in Nonlinear Diffusion: From Spatial Splitting to Multiple Timesteps, *J. Math. Imaging Vis.* 19 (1) (2003) 33–48. doi:[10.1023/A:1024484920022](https://doi.org/10.1023/A:1024484920022).
- [28] L. Dascal, G. Rosman, X.-C. Tai, R. Kimmel, On Semi-implicit Splitting Schemes for the Beltrami Color Flow, in: *Scale Space and Variational Methods in Computer Vision*, Vol. 5567 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, 2009, pp. 259–270. doi:[10.1007/978-3-642-02256-2_22](https://doi.org/10.1007/978-3-642-02256-2_22).
- [29] D. Zosso, J.-P. Thiran, A scale-space of cortical feature maps, *IEEE Signal Processing Letters* 16 (10) (2009) 873–876.
- [30] J. Weickert, B. M. ter Haar Romeny, M. A. Viergever, Efficient and reliable schemes for nonlinear diffusion filtering., *IEEE Transactions on Image Processing* 7 (3) (1998) 398–410. doi:[10.1109/83.661190](https://doi.org/10.1109/83.661190).
- [31] N. Janenko, *The method of fractional steps*, Springer, 1971.
- [32] T. Lu, P. Neittaanmaki, X.-C. Tai, A parallel splitting up method for partial differential equations and its application to Navier-Stokes equations, *RAIRO Mathematical Modelling and Numerical Analysis* 26 (6) (1992) 673–708.
- [33] K. Jbilou, H. Sadok, Vector extrapolation methods. Applications and numerical comparison, *Journal of Computational and Applied Mathematics* 122 (2000) 149–165.
- [34] L. Dascal, G. Rosman, R. Kimmel, Efficient Beltrami filtering of color images via vector extrapolation, in: *SSVM'07: Proceedings of the 1st international conference on Scale space and variational methods in computer vision*, Springer-Verlag, 2007, pp. 92–103.
- [35] G. Rosman, L. Dascal, A. Sidi, R. Kimmel, Efficient Beltrami Image Filtering via Vector Extrapolation Methods, *SIAM J. Imaging Sci.* 2 (3) (2009) 858–878. doi:[10.1137/080728391](https://doi.org/10.1137/080728391).
- [36] L. D. Cohen, Auxiliary Variables and Two-Step Iterative Algorithms in Computer Vision Problems, *J. Math. Imaging Vis.* 6 (1) (1996) 59–83. doi:[10.1007/BF00127375](https://doi.org/10.1007/BF00127375).
- [37] X.-C. Tai, C. Wu, Augmented Lagrangian Method, Dual Methods and Split Bregman Iteration for ROF Model, in: *Scale Space and Variational Methods in Computer Vision*, 2009, pp. 502–513.
- [38] T. F. Chan, G. H. Golub, P. Mulet, A Nonlinear Primal-Dual Method for Total Variation-Based Image Restoration, *SIAM Journal on Scientific Computing* 20 (6) (1999) 1964–1977. doi:[10.1137/S1064827596299767](https://doi.org/10.1137/S1064827596299767).
- [39] A. Chambolle, An Algorithm for Total Variation Minimization and Applications, *Journal of Mathematical Imaging and Vision* 20 (2004) 89–97.
- [40] X. Bresson, T. Chan, Fast dual minimization of the vectorial total variation norm and applications to color image processing, *Inverse Problems and Imaging* 2 (4) (2008) 455–484. doi:[10.3934/ipi.2008.2.455](https://doi.org/10.3934/ipi.2008.2.455).
- [41] M. Zhu, T. Chan, An efficient primal-dual hybrid gradient algorithm for total variation image restoration, Tech. rep., UCLA CAM Report 08-34 (2008).
- [42] M. Zhu, S. J. Wright, T. F. Chan, Duality-based algorithms for total-variation-regularized image restoration, *Computational Optimization and Applications* 47 (3) (2010) 377–400. doi:[10.1007/s10589-008-9225-2](https://doi.org/10.1007/s10589-008-9225-2).
- [43] M. P. do Carmo, *Differential geometry of curves and surfaces*, Prentice Hall Inc., 1976.
- [44] M. Spivak, *A comprehensive introduction to differential geometry*, Vol. 1, Publish or Perish, Berkeley, 1979.
- [45] E. Kreyszig, *Introduction to differential geometry and Riemannian geometry*, University of Toronto Press, 1968.
- [46] I. Ekeland, R. Témam, *Convex Analysis and Variational Problems*, Society for Industrial and Applied Mathematics, 1999. doi:[10.1137/1.9781611971088](https://doi.org/10.1137/1.9781611971088).
- [47] D. P. Bertsekas, J. N. Tsitsiklis, *Parallel and distributed computation*, Prentice Hall Inc., Old Tappan, NJ (USA), 1989.
- [48] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image Quality Assessment: From Error Visibility to Structural Similarity, *IEEE Transactions on Image Processing* 13 (4) (2004) 600–612. doi:[10.1109/TIP.2003.819861](https://doi.org/10.1109/TIP.2003.819861).