AI539 final essay: Information geometry for total-variation denoising of manifold-valued images

Zhanpei Fang

13 June 2025

1 Introduction

I am interested in applying information geometry to remote sensing data, particularly one active remote sensing modality for which it shines: synthetic aperture radar (SAR), created using successive pulses of radio waves which 'illuminate' a target scene where the echo of each pulse is recorded using a single beam-forming antenna. The ordered combination of received signals builds a virtual aperture much longer than the physical antenna width and is not affected by observational conditions such as clouds or nighttime, making it good for, for example, monitoring land changes. SAR images contain both the phase and the amplitude of the backscattered signal for each resolution cell, where the amplitude corresponds to the intensity of the returned radar energy and the absolute phase is related to the sensor-to-target distance (a fraction of a complete wavelength).

A single SAR pixel can be represented as $z = Ae^{i\varphi}$, where A is amplitude and φ is phase, living in $\mathbb{C} \cong \mathbb{R}^2$. Pixel data are therefore complex-valued and better described as points on curved statistical manifolds, such that traditional Euclidean methods fail to capture their intrinsic structure. Other manifold-valued data besides SAR include the chromaticity and brightness of an RGB image ($\mathbb{S}^2 \times \mathbb{R}$), the covariance matrices of EEG brain activity data (and other cases where covariance matrices are adjusted to image pixels), and data with values in the $\mathrm{SO}(n)$ special orthogonal group (such as electron backscatter diffraction) or in $\mathrm{SPD}(n)$ (such as diffusion tensor MRI).

1.1 Interferometric SAR (InSAR) noise model & manifold representation

InSAR in particular measures the signal phase change between at least two SAR images acquired over the same area at different times or points in space. It can be used to measure terrain altitude as well as motion, generating maps of digital elevation and surface deformation respectively. The change in signal phase $\Delta\varphi$ is given by the equation[1]

$$\Delta\varphi = \angle\Gamma = \frac{4\pi}{\lambda}\Delta R + \alpha + t + \text{noise}$$

where λ is the wavelength, ΔR is the target displacement in the line-of-sight, α is the phase shift due to different atmospheric conditions at the times of the two passes, t arises from the topographic distortions from the slightly different viewing angles of the passes and the noise term captures decorrelation effects. Each InSAR pixel represents a phase difference between the outgoing (produced by the satellite) and returning wave which encodes ground deformation along the radar's line of sight, so that the goal is to isolate ΔR even when the measured data may be noisy and/or incomplete.

The phase component $\varphi \in [0, 2\pi)$ on its own is circle-valued and lives on the unit circle $\mathbb{S}^1$, so the manifold is $\mathcal{M} = (\mathbb{S}^1)^{m\times l}$, where $m \times l$ is the number of pixels in the image. Once an interferogram Γ has been produced from two SAR images through complex multiplication $= z_1 \cdot z_2^*$ (* is the conjugate), the consecutive fringes typically have to be unwrapped, which involves interpolating over the 0 to 2π phase jumps

[1] `https://site.tre-altamira.com/insar/`

to produce a continuous deformation field. Phase differences between adjacent pixels also produce random interference effects called 'speckle' (included in the noise term above), which results from the interference of many waves of the same frequency but different phases and amplitudes; it happens because the spatial resolution of the sensor is not sufficient to resolve individual scatterers, so it has dimensions on the order of the resolution. This speckle along with the other noise sources poses an obstacle to phase unwrapping and extracting information from SAR images and is non-Gaussian in general; this denoising step needs to happen before the phase unwrapping [1].

1.2 Variational methods for image processing: from real- to manifold-valued

Processing manifold-valued images has distinct challenges which affect the typical tasks such as denoising, inpainting and segmentation. In the real-valued case, variational methods (such as total variation or Mumford-Shah) are well-adapted to these problems, where we can assume **non-smooth, convex** functionals and so can use convex optimization techniques.

Total variation (TV) denoising assumes that noisy images have high *total variation*–the integral of the image gradient magnitude is quite high–so reducing this should keep the important parts of the image (namely edges) while reducing spurious details. The seminal work in this was done by Rudin, Osher and Fatemi (ROF) in 1992 [17]. Noise in real-valued images is typically assumed to be additive, white, and Gaussian, which is asymptotically justified by the Central Limit Theorem, and a great deal of work has been done on denoising and restoration of such images. Examples include combining the first and second order derivatives in regularization functionals or incorporating anisotropies based on the local structures of the image. Splitting algorithms (splitting the original problem into a sequence of proximal mappings) with primal-dual optimization are used for minimizing the functionals.

The Gaussian assumptions do not hold with manifold-valued images, where the definition of a Gaussian distribution is not canonical on a manifold, though several approaches have been proposed to extend Gaussian distributions in Euclidean space to Riemannian manifolds, generalizing different characteristics of the former. There had historically been few papers combining results on non-smooth optimization with the general manifold-valued setting; Bergmann, Weinmann and collaborators' work generalizes convex models for the restoration of real-valued images to cyclic and manifold-valued images. In this essay I describe the general mathematical structure of the denoising problem and summarize about ten years of work from this group done to apply proposed formulations of manifold-valued regularizers.[2]

1.3 Manifold-valued denoising problem setup

Manifold-valued data live in the space of the tensor product of a manifold with itself (the size of the number of pixels). This yields discrete energy functionals that map from $\mathcal{M}$ into the (extended) real line [4] which can be minimized.

Consider an image as a mapping from the image grid $\mathcal{G} = \{1, \dots, n\} \times \{1, \dots, m\}$ to Riemannian manifold $\mathcal{M}$. A noisy image is $f : \mathcal{V} \to \mathcal{M}$, $\mathcal{V} \subseteq \mathcal{G}$ and we want to reconstruct the original image $u_0 : \mathcal{G} \to \mathcal{M}$. Variational approaches generate a restored image u as the minimizer u^* of some functional of the form

$$\mathcal{J}(u) = \underbrace{\mathcal{D}(u; f)}_{\text{data}} + \underbrace{\alpha\mathcal{R}(u)}_{\text{regularization}}$$

$\mathcal{D}(\cdot; f)$ is a data-fitting term measuring the distance to the given data, $\mathcal{R}$ is a regularization term (prior) containing the assumed properties of an ideal (clean) image, and $\alpha > 0$ is a parameter weighting the influence of the regularizer. In the case that $\mathcal{V} = \mathcal{G}$ this is a typical denoising model in the presence of additive Gaussian noise; if not, missing image values in $\mathcal{G} \backslash \mathcal{V}$ need to be inpainted.

A usual $\mathcal{D}$ for real-valued images with Gaussian noise is the squared Euclidean ℓ^2 distance between f and u, or otherwise ℓ^1 for robustness to outliers. This can be readily replaced by the squared geodesic

[2] Mumford-Shah algorithms have also been proposed for manifold-valued data [20], e.g. as the initial step of a segmentation pipeline, but this is out of the scope of the current essay.

distance $d : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ for manifold-valued images, $\mathcal{D}(u; f) := \frac{1}{2} \sum_{i \in \mathcal{G}} d(u_i, f_i)^2$, or other sort of geodesic distance. Choosing appropriate $\mathcal{R}$ is more involved. The success of methods are dependent on the choice of a suitable regularization functional. In the Euclidean case the first-order total variation (TV) has been shown to perform well for additive white Gaussian noise, as introduced in §1.2. For images $f : \mathcal{V} \to \mathbb{R}$, the total-variation norm proposed by [17] is:

$$\mathcal{R}_{\text{iso}}(u) := \sum_{i,j} |\nabla u| = \sum_{i,j} \sqrt{|u_{i+1,j} - u_{i,j}|^2 + |u_{i,j+1} - u_{i,j}|^2}$$
$$\mathcal{R}_{\text{aniso}}(u) := \sum_{i,j} \left(|u_{i+1,j} - u_{i,j}| + |u_{i,j+1} - u_{i,j}|\right)$$

where $|\nabla u|$ is a discrete gradient. ROF sums up the norms of these gradients at the image points, which preserves sharp discontinuities (edges, unlike classic Tikhonov regularization) and is also convex. The isotropic version of TV is not differentiable and so the anisotropic version is introduced as it may sometimes be easier to minimize. The ROF method [17] uses gradient descent on the Euler-Lagrange equations of the (scalar-valued) TV functional with ℓ^2 norms. Other methods are based on Fenchel duals, ADMM (equivalent to Douglas-Rachford), and split Bregman methods [13].

Our goal is to find $u^* := \arg\min_{u \in \mathcal{G}} \mathcal{J}(u)$. Constraining u to a Riemannian manifold is generally a non-convex constraint which makes optimization much harder. The TV regularization methods are all based on convex duality, which is not available in a manifold setting. The main challenges to developing appropriate analogues of non-smooth regularizers are that

- there is in general no additive group on a Riemannian manifold, so *a priori* addition, scalar multiplication and linear combinations are not well-defined, making it difficult to apply tools from the $\mathcal{M} = \mathbb{R}$ linear case;

- the jumps in the values which are measured by usual TV need to be measured with respect to geodesic distance. Straight lines need to be replaced with geodesics for $\mathcal{R}$ as well as $\mathcal{D}$, the estimation of which may itself be challenging computationally;

- there is a notion of global convexity only for the case of Hadamard manifolds which have nonpositive sectional curvature [3]. There are cases where $\mathcal{J}$ is not convex and has multiple minimizers, such as the sphere.

2 Methodology

2.1 Lifting methods (extrinsic approach) [18, 16, 19]

The simplest idea for how to treat TV regularization is to embed the manifold into Euclidean space and apply Euclidean models, with the constraint that the image values need to lie in $\mathcal{M}$. Lifting methods[19] allow us to transform hard variational problems into convex problems in a suitable higher-dimensional space, via the Whitney embedding theorem. The lifted models can then be solved to a global optimum, which allows us to find approximate global minimizers of the original problem.

[12, 18] proposed such a lifting scheme for circle-valued data. They do a TV approach for $\mathbb{S}^1$-valued data by lifting functions with values in $\mathbb{S}^1$ to functions with values in the universal covering $\mathbb{R}$ of $\mathbb{S}^1$, respecting the periodicity of the data by lifting the involved functionals at the same time. This is a nonconvex problem on real-valued data which can be solved with convex relaxation. Lellman [16] extends the lifting approach of [18] to arbitrary Riemannian manifolds. They reformulate TV as a lifted multilabel optimization problem, where ADMM can be employed for numerically finding the minimizers. However, the components within the extrinsic TV decompositions live in the embedding space which makes their interpretation difficult. This also increases the dimension of the data, and it is beneficial to use the intrinsic metric in many applications in any case.

2.2 ℓ^p-TV regularization (intrinsic) [21]

Building on [16], Weinmann et al [21] use iterative geodesic averaging to implement a cyclic proximal point algorithm (CPPA) to compute a minimizer of the variational model and produce algorithms for ℓ^p-TV for general Riemannian manifold data. CPPA lets you decompose the TV functional into a sum of functionals in order to analytically compute the proximal mappings of the functionals on the Riemannian manifold, given in terms of points on certain geodesics. Their resulting algorithms exclusively perform iterative geodesic averaging. This means that the *only* operations needed are those for computing geodesics. They show the convergence towards a global minimizer for the class of Cartan-Hadamard manifolds (positive definite matrices equipped with the Fisher-Rao metric, containing many symmetric spaces and which are still convex), and show robust but slow performance on sphere- and cylinder-valued data. In their book chapter reviewing these techniques [15] Holler and Weinmann state a theorem that, for data in a (locally compact) Hadamard space, the iterative geodesic averaging algorithms for TV-regularized denoising (based on methods like CPPA) converge towards a minimizer of the ℓ^p-TV functional.

As an aside, Weinmann et al. [21] show their results on InSAR data where TV minimization with CPPA is capable of removing almost all of the noise in the InSAR image, and using ℓ^1 and Huber data terms makes it slightly more robust to outliers than ℓ^2. For InSAR the exponential map, which is needed to compute the geodesic averages, is particularly simple when considering $\mathbb{S}^1$ as a unit circle in the complex plane:

$$\exp_a(\Delta\varphi) = e^{i(\varphi+\Delta\varphi)}$$

where $a = e^{i\varphi}$ and $\Delta\varphi \in]-\pi,\pi[$. For two nonantipodal points a and b the inverse exponential map is $\exp_a^{-1}(b) = \arg(b/a)$, the polar angle of complex number b/a, and the distance between two points on the sphere is $d(a,b) = |\arg(b/a)|$.

Figure 1: Figure reproduced from [15] showing their results using TV denoising on $\mathbb{S}^1$-valued InSAR image (input image on left) using ℓ^2-TV regularization (choosing $\alpha = 0.32$) and ℓ^1-TV regularization (choosing $\alpha = 0.60$). $\mathbb{S}^1$ is represented as an interval where the white and black correspond to the phase endpoints $\varphi = 0$, $\varphi = 2\pi$. TV variation minimization removes noise while preserving image structure.

2.3 Higher-order methods: Total generalized variation (TGV)

Building on [21] and earlier work, non-smooth higher-order models have been derived for the manifold-valued case [9, 2, 6, 7]. These are to address the known drawbacks of TV regularization such as staircasing effects—TV regularization which only involves the first-order derivatives tends to produce piecewise constant results with artificial jump discontinuities.

[10] first proposed the (second order) total generalized variation (TGV) functional which balances between first and second order derivatives. The discrete versions of TV, second order TV, and TGV in the in the

$\mathbb{R}$-valued case are [15]

$$\mathrm{TV}(u) = \|\nabla u\|_1, \mathrm{TV}^2(u) = \|J\nabla u\|_1$$
$$\mathrm{TGV}^2_\alpha(u) = \min_w \alpha_1\|\nabla u - w\|_1 + \alpha_0\|\mathcal{E}w\|_1$$

where $\|\cdot\|_1$ is the ℓ^1 norm with respect to the spatial component, and an implicit ℓ^p norm is taken in the vector components; J is the Jacobian and $\mathcal{E}w = \frac{1}{2}(Jw + Jw^\intercal)$ is a symmetrization of the Jacobian matrix field Jw. $\alpha = (\alpha_1, \alpha_0) \in (0, \infty)^2$ are different weight terms.

[2] generalizes first and second-order nonsmooth variational methods to general symmetric Riemannian manifolds and use CPPA. They use the idea that the second-order difference term in the Euclidean case can be written as $\|x - 2y + z\| = 2\|\frac{1}{2}(x + z) - y\|$ and define a similar second-order spatial difference on manifolds using the geodesic distance from the midpoints of the geodesics connecting x with z to y.

To generalize the above formulation of second-order TGV to a manifold setting, we can consider the above formulation of TGV for vector spaces, which requires us to measure the distance of (discrete) vector fields that are in general defined in different tangent spaces. As we learned in class, we can do this by parallel-transporting vector fields to the same tangent space and measuring the distance there. [6], in addition to defining an extrinsic TGV method and a TGV approach to Lie groups, define a TGV regularizer which works on manifold valued images as

$$\mathrm{TGV}^2_\alpha(u) = \inf_{(\xi_i)_i} \sum_i \alpha\|\log_{u_i}(u_{i+1}) - \xi_i\|_{u_1} + \alpha_0\|\xi_i - P_{u_i}(\xi_{i-1})\|_{u_i}$$

and use gradient descent to compute the minimizers of this intrinsic model. P_{u_i} is the approximation of the parallel transport of ξ_{i-1} to u_i. They approximate parallel transport with the pole ladder (which coincides with the parallel transport for connected, complete, symmetric Riemannian manifolds); one may also approximate parallel transport with Schild's ladder [11], requiring only point operations. They lift the transported tuple again to the tangent space via the logarithmic map.

2.4 Graph-based IRLS TV [8, 4]

Riemannian optimization methods can be used to compute the minimizer of $\mathcal{J}$, which are efficient since they exploit the underlying geometric structure of the manifold. Subsequently, [8, 4] generalize the half-quadratic minimization (iteratively reweighted least squares) introduced by [14] to manifold-valued images using a quasi-Newton method. They treat $\mathcal{G}$ as vertices of a graph with edge set $\mathcal{E} := \{(i, j) : i \in \mathcal{G}, j \in \mathcal{N}(i)\}$ where the regularizing term sums over the edge set $\mathcal{E}$. They focus on the multiplicative half-quadratic minimization method for denoising/inpainting both in the anisotropic and isotropic case, which is to repeatedly compute the exponential map $u^{(k+1)} := \exp_{u^{(R)}} \eta_R$ with gradient descent in the direction $\eta_r \in T_{\tilde{u}^{(r)}}M$ with Newton's method. They show convergence of the algorithm for images with entries in an Hadamard space, for example the manifold of positive definite matrices.

2.5 Recent work

2.5.1 Fenchel duality

In sum, efforts have been made to translate convex optimization techniques to manifolds for the image denoising problem, with certain results shown on Hadamard manifolds (complete, simply-connected Riemanian manifolds of nonpositive sectional curvature), which behave well in certain ways but don't have all of the desired properties for convex analysis. For example, somewhat more recently [5], with a new formulation of the Fenchel conjugate for functions on the manifold, derive a Riemannian primal-dual non-smooth optimization problem and prove its convergence for the case of Hadamard manifolds under appropriate assumptions. They apply this to the anisotropic and isotropic ℓ^2-TV with squared distance $\mathcal{D}$ term, i.e. the ROF variant for Riemannian manifolds, performing favorably against the Douglas-Rachford algorithm on manifolds of nonpositive curvature, and possibly even converging on manifolds of positive curvature, namely the sphere $\mathbb{S}^2$.

2.5.2 Future directions

These manifold-valued image denoising techniques have so far mostly been demonstrated on toy datasets; a potentially rich area of research would be to show how well they can be implemented using these developed methods described in this essay to real datasets. Tools such as MVIRT[3]–which is no longer maintained, but has been succeeded by implementations in Julia, MATLAB and Python[4]–and `manopt`[5] can be used to try out these methods. I would be interested in trying to implement these algorithms for SAR, including Sentinel-1 SAR, TerraSAR-X, and the upcoming NISAR mission, in particular to see whether these algorithms are tractable and scalable for the growing volume of SAR datasets.

References

[1] Roland Akiki et al. “A Brief Evaluation of InSAR Phase Denoising and Coherence Estimation with Phi-Net”. In: *Image Processing On Line* 14 (2024), pp. 205–216.

[2] Miroslav Bacák et al. “A second order nonsmooth variational model for restoring manifold-valued images”. In: *SIAM Journal on Scientific Computing* 38.1 (2016), A567–A597.

[3] Ronny Bergmann, Johannes Persch, and Gabriele Steidl. “A parallel Douglas–Rachford algorithm for minimizing ROF-like functionals on images with values in symmetric Hadamard manifolds”. In: *SIAM Journal on Imaging Sciences* 9.3 (2016), pp. 901–937.

[4] Ronny Bergmann and Daniel Tenbrinck. “A graph framework for manifold-valued data”. In: *SIAM Journal on Imaging Sciences* 11.1 (2018), pp. 325–360.

[5] Ronny Bergmann et al. “Fenchel duality theory and a primal-dual algorithm on Riemannian manifolds”. In: *Foundations of Computational Mathematics* 21.6 (2021), pp. 1465–1504.

[6] Ronny Bergmann et al. “Priors with coupled first and second order differences for manifold-valued image processing”. In: *Journal of mathematical imaging and vision* 60 (2018), pp. 1459–1481.

[7] Ronny Bergmann et al. “Recent advances in denoising of manifold-valued images”. In: *Handbook of Numerical Analysis* 20 (2019), pp. 553–578.

[8] Ronny Bergmann et al. “Restoration of manifold-valued images by half-quadratic minimization”. In: *arXiv preprint arXiv:1505.07029* (2015).

[9] Ronny Bergmann et al. “Second order differences of cyclic data and applications in variational denoising”. In: *SIAM Journal on Imaging Sciences* 7.4 (2014), pp. 2916–2953.

[10] Kristian Bredies, Karl Kunisch, and Thomas Pock. “Total generalized variation”. In: *SIAM Journal on Imaging Sciences* 3.3 (2010), pp. 492–526.

[11] Kristian Bredies et al. “Total generalized variation for manifold-valued data”. In: *SIAM Journal on Imaging Sciences* 11.3 (2018), pp. 1785–1848.

[12] Daniel Cremers and Evgeny Strekalovskiy. “Total cyclic variation and generalizations”. In: *Journal of mathematical imaging and vision* 47.3 (2013), pp. 258–277.

[13] Tom Goldstein and Stanley Osher. “The split Bregman method for L1-regularized problems”. In: *SIAM Journal on Imaging Sciences* 2.2 (2009), pp. 323–343.

[14] Philipp Grohs and Markus Sprecher. “Total variation regularization on Riemannian manifolds by iteratively reweighted minimization”. In: *Information and Inference: A Journal of the IMA* 5.4 (2016), pp. 353–378.

[3] `https://github.com/kellertuer/MVIRT`

[4] `https://www.pymanopt.org/`

[5] `https://github.com/NicolasBoumal/manopt`

[15] Martin Holler and Andreas Weinmann. "Non-smooth variational regularization for processing manifold-valued data". In: *Handbook of Variational Methods for Nonlinear Geometric Data*. Springer, 2020, pp. 51–93.

[16] Jan Lellmann et al. "Total variation regularization for functions with values in a manifold". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2013, pp. 2944–2951.

[17] Leonid I Rudin, Stanley Osher, and Emad Fatemi. "Nonlinear total variation based noise removal algorithms". In: *Physica D: Nonlinear Phenomena* 60.1-4 (1992), pp. 259–268.

[18] Evgeny Strekalovskiy and Daniel Cremers. "Total variation for cyclic structures: Convex relaxation and efficient minimization". In: *CVPR 2011*. IEEE. 2011, pp. 1905–1911.

[19] Thomas Vogt et al. "Lifting methods for manifold-valued variational problems". In: *Handbook of Variational Methods for Nonlinear Geometric Data* (2020), pp. 95–119.

[20] Andreas Weinmann, Laurent Demaret, and Martin Storath. "Mumford–Shah and Potts regularization for manifold-valued data". In: *Journal of Mathematical Imaging and Vision* 55 (2016), pp. 428–445.

[21] Andreas Weinmann, Laurent Demaret, and Martin Storath. "Total variation regularization for manifold-valued data". In: *SIAM Journal on Imaging Sciences* 7.4 (2014), pp. 2226–2257.