On Regularized Reconstruction of Vector Fields

Pouya Dehghani Tafti*, *Member, IEEE,* and Michael Unser, *Fellow, IEEE*

***Abstract*—In this paper we give a general characterization of regularization functionals for vector field reconstruction, on the basis of requiring that the said functionals satisfy certain geometric invariance properties with respect to transformations of the coordinate system. In preparation for our general result, we also address some commonalities of invariant regularization in scalar and vector settings—and give a complete account of invariant regularization for scalar fields—before focusing on their main points of difference, which lead to an entirely different class of regularization operators in the vector case. Finally, as an illustration of potential, we formulate and compare quadratic (L_2) and total-variation-type (L_1) regularized denoising of vector fields in the proposed framework.**

***Index Terms*—Regularization, vector fields, rotation-invariance, scale-invariance, vector L_p spaces, total variation, fractional Laplacian, curl and divergence in higher dimensions, fractional vector calculus.**

I. Introduction

Our aim in the present paper is to derive, in a principled manner, formulae for regularization functionals suitable for reconstructing *vector fields,* with a view to applications such as denoising, deconvolution, and reconstruction from incomplete (that is, scalar) measurements [1, 2], among others. Our motivation in approaching the question of vector field reconstruction derives from the increasing prevalence of imaging modalities that produce measurements of vector quantities, and the need to design algorithms for treating such data [3]. Such algorithms can also be applicable in other contexts where vector fields appear, such as estimating optical flow and image registration [4–7].

Throughout the paper, we take *invariance under coordinate transformations* as our guiding principle. The importance of invariance in reconstruction was already apparent to Duchon [8], who considered the problem of interpolating or approximating scalar fields in $\mathbb{R}^d$; however, the mathematical formulation of invariance laws is in general different for scalars and vectors, as we shall see briefly in §II and in more detail in §IV and §V. The appeal of the notion of invariance partly lies in the fact that invariant regularizers do not impose a preferential choice of coordinate system on the model. We give a rather complete characterization of invariant vector regularization operators in §IV and §V, after initially reviewing the related scalar theory in §III.

Regularized reconstruction of vector fields has been previously considered, notably by Suter and Chen [9], who proposed quadratic (L_2) regularization with mixed-order differentials of the vector field. Arigovindan *et al.* [2, 10] studied quadratic regularization with fractional-order differential operators and paid particular attention to the invariance properties of the regularization term with respect to vector rotation, translation, and change of scale, characterizing the complete family of quadratic regularization functionals with the required invariances, which essentially extend Duchon's thin-plate splines [8] to the vector setting. Specialized examples of such functionals, involving curl and divergence regularization, had been considered earlier by Dodu and Rabut [11] and (for the problem of interpolation) by Amodei and Benbourhim [12] before them.

All of the previous schemes fall under the general heading of smoothing spline and spline interpolation methods. They thus exhibit similar advantages (efficient resolution by linear methods and connection with splines) and limitations (most notably, over-smoothing of discontinuities and edges which, e.g., occur naturally at fluid interfaces in fluid dynamical systems and at object boundaries in optical flow). In this connection, it has been observed in the scalar setting that schemes using L_1 regularization—in particular total variation (TV) type methods—do a better job of preserving edges and discontinuities than their quadratic (L_2) counterparts [13, 14]. The framework we have adopted in this paper allows us to find natural vector equivalents of these non-quadratic methods. (On the algorithmic side, the non-quadratic problems we formulate here can be solved using techniques similar to those employed in the scalar case (see for instance Figueiredo *et al.* [15]) as we show by way of examples in §VI.)

On the theoretical side, another common property of quadratic schemes is that, due to the association of quadratic functionals with inner products, they can all be reduced to regularization with *self-adjoint* differential operators (essentially fractional Laplacians and their extensions; see §IV). This is in contrast to the general non-quadratic case considered here, where the factorization of these self-adjoint operators into skew-symmetric ones becomes relevant (see §V).

Finally, we wish to point out that unlike at least some of the previous works which have been concerned exclusively with 2D and/or 3D vector fields, the approach we have adopted in the present paper makes it possible to consider vector fields in any number number of dimensions on the same footing. This is particularly apparent in our dimensionless formulation of fractional Laplacians in §III, §IV, and of curl-

 The authors are with the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne (EPFL), Station 17, CH-1015 Lausanne, Switzerland (phone: +41 21 693 11 85; fax: +41 21 693 68 10; e-mail: pouya.tafti@a3.epfl.ch, michael.unser@epfl.ch)
DOI: 10.1109/TIP.2011.2159230.

and divergence-like operators in §V.

Regularized reconstruction

The standard scenario for regularized reconstruction is as follows. We are given a vector y of measurements or observations that are assumed to depend, in a known probabilistic fashion (or deterministically but with some measurement and/or modelling error), on the unknown entity f_{true} which we wish to reconstruct. f_{true} will in our case be a function defined on some finite or infinite domain. We then define the regularized reconstruction of f_{true} as the (hopefully unique) minimizer over f of a cost functional

$$\mathfrak{J}_\lambda(f\,;y) := \mathfrak{d}(f\,;y) + \lambda\mathfrak{R}(f) \tag{1}$$

composed of a *fidelity criterion* $\mathfrak{d}$ quantifying the proximity of the observed measurements (y) to hypothetical measurements made from some possible reconstruction f, and a *regularization functional* $\mathfrak{R}$ that measures the undesirability of f based on our (deterministic or probabilistic) prior information or assumptions about the solution. The above formulation can be arrived at in different ways, some of which we shall now mention in passing for the sake of motivation, while reminding the reader that our primary purpose here is to derive some specific families of regularization functionals, and not to justify the regularized variational framework for reconstruction in general (for comparable classifications see Poggio *et al.* [4] and Figueiredo *et al.* [16]).

(I.1) In finite sample/parameter dimensions, that is, when both f and y are finite vectors, it is often possible to view the minimization of (1) as a case of maximum a posteriori (MAP) estimation. In this interpretation, $\mathfrak{d}$ and $\lambda\mathfrak{R}$ essentially play the respective roles of the negative log-likelihood and the negative log of the prior, usually up to some normalization (and possibly also discarding some terms that do not modify the solution). To come up with a prior, one might seek an operator R that *whitens* the vector f (i.e. renders its components independent); the log of the joint probability of elements of Rf then becomes additive due to independence. This fits nicely with the most common form of regularization functionals used in practice, i.e. sums of the form

$$\sum_i \Phi([Rf]_i) \tag{2}$$

where R is a *regularization operator,* $[Rf]_i$ is the i-th element of Rf, and Φ is a convex function such as $|\cdot|^2$ for ℓ_2 regularization or the absolute value for ℓ_1 regularization.
Although MAP estimation is not the only purely probabilistic interpretation of (1), it is by far the most common one, and hence the only one we shall mention here.

(I.2) The form given in (1) can also be justified from a hybrid probabilistic-deterministic standpoint, where $\mathfrak{d}$ again represents a negative log-likelihood, while $\mathfrak{R}$ now corresponds to the constraint

$$\mathfrak{R}(f) \leq a \tag{3}$$

on the solution, put in Lagrange form with λ serving as the Lagrange multiplier. Such schemes are known under the names of constrained or penalized likelihood. In addition, a connection can often be made with Grenander's method of sieves [17] (where one considers a limiting sequence of minimizers of the cost functional with varying λ).

(I.3) Finally, a purely deterministic interpretation is also possible, where $\lambda\mathfrak{R}$ is again the Lagrange relaxation of the constraint $\mathfrak{R}(f) \leq a$, while $\mathfrak{d}$ is a deterministic measure of data fidelity such as the Euclidean distance between y and samples of f. However, we remark that in many practical situations, the constraint bound a on which λ depends is not known (or the constraint is not really a hard one); consequently, λ can also be seen as a tuning parameter of the reconstruction algorithm.

Among the above justifications for the regularized reconstruction framework, the MAP interpretation does not trivially generalize to the case where an infinite number of values need to be estimated, which occurs for example when the domain of f is an infinite set such as $\mathbb{R}^d$ (rather than a finitely countable set); for one thing, it is generally not possible to associate a probability distribution function, in its finite-dimensional sense, with probabilities on the function space to which f belongs, due to the fact that the Lebesgue measure does not admit of an infinite-dimensional generalization.

It is therefore constructive, in what follows, to imagine that the term $\lambda\mathfrak{R}(f)$ is derived from an inequality constraint as in the second and third interpretations. Moreover, we shall consider all algorithms based on the same $\mathfrak{R}$ at the same time, and consider λ (or equivalently, the constraint bound), as a tuning parameter of the algorithm. Probabilistic considerations then become secondary to geometrical/analytical ones, for which reason they shall not be emphasized in the remainder of the paper.

Even so, we still draw inspiration from the observation made at the end of paragraph (I.1) to define our regularization functionals as integrals of the form

$$\mathfrak{R}(f) = \int_{\mathrm{Dom} f} \Phi(Rf(x))\,\mathrm{d}x \tag{4}$$

where R (previously the whitening operator) is now referred to as the *regularization operator.* Formally, the above integral—which replaces the sum in (2)—can be thought of as the normalized aggregate contribution of individual independent point-wise innovations; i.e., the values of $Rf(x)$ as a function of x (even though, strictly speaking, without proper normalization such a contribution should be infinite from the probabilistic point of view).

After this brief introduction to regularized reconstruction, let us now describe the direction and contents of the paper. Our primary focus in the present work is vector field regularization. Thus, assuming the general form given in (4) for the reason we just described, our task is then to specify the linear operator R as well as the function Φ. We shall derive the general form of admissible Rs and Φs—for scalars as well as for vectors—by imposing invariances under certain geometric transformations, namely rotation (and reflection), translation, and scaling.

The motivation behind using invariances is that, in many physical systems, there exists no obvious preferential choice of direction, position, or scale, at least within a reasonably wide range relevant in many applications. We therefore seek reconstruction algorithms that lead to a consistent solution under such transformations, possibly by appropriately adjusting a single parameter (λ). This requires the regularization functional $\mathfrak{R}$ to be invariant under such transformations (possibly up to a computable multiplicative factor).

Although our main goal here is to formulate regularization functionals for vector fields, we begin the exposition by general considerations that apply equally to scalars and vectors (§II) and, for completeness, proceed to include a detailed account of invariant scalar regularization in §III, where we derive the general form of R and Φ for the scalar case under suitable assumptions. Next, in §IV, we turn our attention to vector fields and invariances relevant for them. This is followed by some extensions of the framework in §V, where we additionally consider regularization operators that map vector fields to scalars and tensors. It will become clear by the end of §V that, with a high level of generality, the functional $\mathfrak{R}$ takes the form $\|Rf\|_p^q$ where:

- In the scalar setting, $\|\cdot\|_p$ is the standard Lebesgue L_p norm and R is either a fractional scalar Laplacian (defined in §III) or a fractional gradient (defined in §V).
- In the vector setting, $\|\cdot\|_p$ is a suitable generalization of the scalar L_p norm to vector- or matrix-valued functions (introduced, respectively, in §II and Appendix A), and R is either a generalized fractional vector Laplacian (introduced in §IV) which incorporates a Helmholtz decomposition into curl- and divergence-free components, or else, it is a fractional curl or a fractional divergence (both introduced in §V).

We then illustrate the proposed construction in §VI, where we consider the problem of vector field denoising in 2D and 3D, and compare two solutions (quadratic and TV-like) that fall within our framework. Some remarks in §VII conclude the paper.

Symbols and other notation are defined when first used and summarized in Table I for reference.

Table I: Notation

symbol	description
d	number of spatial dimensions
$\mathbb{R}_+$	set of positive reals $= \{a \in \mathbb{R} \mid a > 0\}$
$x = (x_1, \ldots, x_d)$	spatial coordinates
$\xi = (\xi_1, \ldots, \xi_d)$	Fourier coordinates (dual to x)
$f = f(x)$	field of scalars, vectors, or bivectors (usu. denoting a possible reconstruction)
$\hat{f} = \hat{f}(\xi)$	Fourier transform of f
$\Omega = \{\omega\}$	group of orthogonal matrices $\omega \in \mathbb{R}^{d\times d}$
$T = \{\tau\}$	group of displacement vectors $\tau \in \mathbb{R}^d$
$A = \{\alpha\}$	group of spatial scale factors $\alpha \in \mathbb{R}_+$
S	placeholder for ω, τ, or α
$[S]f$	transformation of f by S; $[S]$ can be understood as the *operator* that transforms f
$[S]_s f$	same as above, additionally indicating that f is scalar-valued
$[S]_v f$	same as above, additionally indicating that f is vector-valued
$[S]_b f$	same as above, additionally indicating that f is bivector-valued
$\lvert a\rvert$	for $a \in \mathbb{R}$, the absolute value of a
	for $a \in \mathbb{R}^d$, the Euclidean length of a
$\|f\|_p$	for scalar-valued f, the standard Lebesgue L_p norm of f
	for vector-valued f, the L_p norm defined in Corollary 1
	for matrix-valued f, the L_p norm defined in Appendix A
$\langle a, b\rangle$	the scalar product of vectors $a, b \in \mathbb{R}^d$
$\langle f, g\rangle$	the scalar product of functions f, g $(= \int_{\mathbb{R}^d} \overline{f^{\mathrm{T}}} g)$
R	regularization operator
div	divergence operator (vector to scalar)
grad	gradient operator (scalar to vector)
curl	curl operator (vector to bivector)
curl*	adjoint curl operator (bivector to vector)
$\star$	star operation (bivector to pseudo-vector, see (21))

II. Generalities regarding regularization and invariance

As noted in the introduction, in identifying suitable families of regularization functionals we are guided by the principle of invariance under specific geometric transformations. With any such transformation is associated a symbol S that can be a scalar $\sigma > 0$ (the scale) for changes of scale, a vector $\tau \in \mathbb{R}^d$ (the displacement vector) in the case of translations, an orthogonal transformation matrix $\omega \in \mathbb{R}^{d\times d}$ when considering rotations and reflections, or, once again, a scalar $\alpha > 0$ (the gain) when multiplication by positive reals (change of units) is considered.

Since, in general, the *same* transformation group can act differently on scalars and vectors (this is particularly true for rotations, as we shall see in §IV), the same symbol S can describe different laws of transformation depending on whether it is acting on scalars or vectors or other entities. For this reason, we introduce the notation $[S]$ to denote the *operator* associated with the symbol S, and distinguish between scalar and vector operators by using subscripts as per $[S]_s$ and $[S]_v$ where necessary.

Definition 1: In mathematical terms, we assume that S belongs to one of several *transformation groups* $T = \{\tau \in \mathbb{R}^d\}$ (the *translation* group), $\Omega = \{\omega \in \mathbb{R}^{d\times d} : \omega \text{ orthogonal}\}$ (the *orthogonal* group), $\Sigma = \{\sigma \in \mathbb{R}_+\}$ (the *scaling* group), or $A = \{\alpha \in \mathbb{R}_+\}$ (the *gain* group), and consider maps (isomorphisms) $S \mapsto [S]_C$ between transformation groups and groups of operators (*actions* or *transformation laws*) acting on objects of some class C ($C = s$ for scalar fields, $= v$ for vector fields, $= b$ for bivector fields, *etc.*).

We then define $[\tau]f(\cdot) = f(\cdot - \tau)$ for $\tau \in T$ (translation); $[\alpha]f(\cdot) = \alpha f(\cdot)$ for $\alpha \in A$ (gain); $[\sigma]f(\cdot) = f(\sigma^{-1}\cdot)$ for $\sigma \in \Sigma$ (scaling); and $[\omega]_s f(\cdot) = f(\omega^{\mathrm{T}}\cdot)$ for $\omega \in \Omega$ (orthogonal transformation *of scalars*). Note that the first three identities are valid for scalars as well as for vector fields, while the last one only applies to scalars. Vector rotation follows a different rule: $[\omega]_v f(\cdot) = \omega f(\omega^{\mathrm{T}}\cdot)$. The reason is that the coordinates $f_1, f_2, \dots, f_d$ of a vector field $f(x)$, $x \in \mathbb{R}^d$, are specified in the same coordinate system as that of its argument x, which means that if the coordinate system of the argument is rotated by ω^{T}, the coordinates $f_1, \dots, f_d$ have to be transformed by the inverse (ω) in order to keep the direction of the vectors fixed. □

We recall (cf. (4)) that we shall be seeking invariant regularization functionals of the form

$$\mathfrak{R}(f) = \int_{\mathbb{R}^d} \Phi(Rf(x))\,\mathrm{d}x$$

where the scalar-valued function Φ and the operator R are to be determined.

Requiring that the regularization be *S-invariant* up to some re-adjustment of the parameter λ amounts to demanding that

$$\mathfrak{R}([S]f) = c_{S,\mathfrak{R}}\mathfrak{R}(f) \tag{5}$$

for all f under consideration, where $c_{S,\mathfrak{R}}$ is a constant. In order to have more flexibility in constructing regularization functionals, we wish to find families of functions Φ and operators R that we can then pick and combine independently. In particular, since we shall always include *identity* in our family of regularization operators, we require Φ to satisfy

$$\int_{\mathbb{R}^d} \Phi([S]f(x))\,\mathrm{d}x = c_S \int_{\mathbb{R}^d} \Phi(f(x))\,\mathrm{d}x \tag{6}$$

for all f and all S and for some constant $c_S > 0$ that depends on S.

From (6) immediately follows

Proposition 1: Let Φ satisfy (6) and be continuous on some open neighbourhood. Φ is then equivalent to a homogeneous function; that is, $\Phi(a) = c|a|^p$ (almost everywhere) for some $c > 0$ and $p \in \mathbb{R}$ ($|a|$ denotes the absolute value or the modulus of a as appropriate).

Conversely, (6) holds for any such Φ as long as the integrals are well-defined. □

Proof: From (6) we have $\int[\Phi([S]f(\cdot)) - c_S\Phi(f(\cdot))] = 0$ for all f and therefore

$$\Phi([S]f(x)) = c_S\Phi(f(x)) \qquad \text{for almost all } x.$$

We shall first consider the case of scalar f, where Φ is a function of the reals. Let $S = \alpha$ belong to the gain group A, and let $\phi(\alpha) := c_\alpha > 0$. We then have, for arbitrary $\alpha = -1$ and arbitrary $a = f(x)$, $\Phi(a) = \Phi(-(-a)) = [\phi(-1)]^2\Phi(a)$ whence $\phi(-1) = 1$ and $\Phi(a) = \Phi(|a|)$. Next, for arbitrary α and a, we may write

$$\phi(|\alpha|)\Phi(|a|) = \Phi(\alpha a) = \Phi(a\alpha) = \phi(|a|)\Phi(|\alpha|).$$

Fixing either α or a then proves that $\phi = c'\Phi$ for some constant c'. Therefore, for all $\alpha, a \in \mathbb{R}_+$,

$$\Phi(\alpha a) = c'\Phi(\alpha)\Phi(a).$$

This shows that Φ is an exponential function and can therefore be written as $\Phi(\cdot) = c(\cdot)^p$ for some constants c, p, as claimed.

When f is vector-valued, rotation invariance implies that Φ is in fact only a function of the modulus $|\alpha|$ of α; we may then repeat the argument of the previous paragraph to once again deduce that $\Phi(\alpha) = c|\alpha|^p$ for some c, p.

To prove the converse one can directly inspect each of the groups of transformations involved by a simple change of variables in the integrals, whence it is observed that the desired result follows immediately from the invariances of the Lebesgue measure. ■

The following corollary is immediate.

Corollary 1: The vector norms

$$\|f\|_p := \begin{cases} \left(\int_{\mathbb{R}^d} |f(x)|^p\,\mathrm{d}x\right)^{1/p}, & 1 \le p < \infty, \\ \operatorname{ess\,sup}|f(x)|, & p = \infty, \end{cases} \tag{7}$$

are S-invariant in the sense that $\|[S]f\|_p = c_{S,p}\|f\|_p$ for all (vector-valued) f. Conversely, any convex S-invariant integral functional (as defined in (6)) that satisfies the requirements of Proposition 1 is of the form $\|f\|_p^p$ for some $p \ge 1$. □

It is then sufficient, in order to have the desired independence between the choice of Φ and R, to require that R *commute* with coordinate transformations up to a multiplicative constant $k_S \neq 0$, in the sense that

$$R\,[S]f = k_S[S]Rf \tag{8}$$

for all f. This, we note, is the quintessence of invariance, as it means that applying the coordinate transformation before or after the application of R yields the same result (up to normalization).

Consequently, the regularization functional given in (4) can be written as (the p-th power of) the L^p norm of Rf (we absorb all the constants in λ; p is required to be ≥ 1 for the sake of convexity). We may also include the ∞-norm $\mathfrak{R}(f) = \|Rf\|_\infty$ for completeness since, even though it is not strictly derived from an integral, it nevertheless satisfies the required invariances.

It is worth noting that, following the Lagrangian interpretation given in the introduction (cf. (3)), we may in practice replace $\mathfrak{R}(f)$ by $\Psi(\mathfrak{R}(f))$ where Ψ is an arbitrary continuous strictly increasing function on $\mathbb{R}_+$, since all such functions define equivalent inequality constraints in the Lagrangian formulation, for $\mathfrak{R}(f) \le a \Leftrightarrow \Psi(\mathfrak{R}(f)) \le \Psi(a)$. Such a function Ψ can therefore be introduced as convenient. However, if it is desired to have (5) hold with a constant $c_{s,\mathfrak{R}}$ not depending on f, one can then show that Ψ needs to be a multiple of the homogeneous function $|a|^q$ for some q (cf. the proof of Proposition 1). Putting all this together, we get

Proposition 2: Let R be S-invariant in the sense of (8). Then, given $p \in [1,\infty]$ and any $q \in \mathbb{R}_+$, the regularization functionals

$$\mathfrak{R}_p^q(f) := \|Rf\|_p^q \quad (9)$$

are S-invariant up to a multiplicative factor; that is, we have,

$$\mathfrak{R}_p^q([S]f) = c_{S,p,q}\mathfrak{R}_p^q(f)$$

for some $c_{S,p,q} \neq 0$. □

Proof: This is an immediate consequence of Corollary 1 and (8). ∎

Excepting the case of $p = \infty$ where one normally takes $q = 1$, the preferred choice of q in practice is $q = p$, which simplifies the formulae by getting rid of the p-th algebraic root hidden in the definition of the L_p norm.

As a reminder, in (8) (reproduced below for convenience) we required that the operator $R : X \to Y$ commute with the transformation associated with S, where S is taken from one of the transformation groups Σ, A, T, Ω (cf. Definition 1):

$$R\,[S]_X = k_S [S]_Y\, R. \quad (10)$$

Note that, in general, when R maps objects of type X to those of a different type Y (such as vectors to scalars or vice versa), the operator associated with S will be different on the two sides of (10); we have emphasized this in the above equation by subscripting the operator with X and Y as appropriate.

We say that R is $\{S\}$-invariant if it satisfies (10) for all S in some understood transformation group(s) (*strictly* $\{S\}$-invariant if in addition $k_S = 1$). For instance, we shall talk about Ω-invariant (T-invariant, *etc.*) operators, by which we mean operators that satisfy (10) for $S \in \Omega$ ($S \in T$, *etc.*). One notes that for an $\{S\}$-invariant operator the map

$$S \mapsto k_S \quad (11)$$

is a group homomorphism from any of the transformation groups under consideration (typically Σ, A, T, and Ω) onto its image under R.

In the sequel, we shall limit ourselves to *linear* regularization operators R, while reminding the reader that in general, the reconstruction problem remains non-linear due to the L_p norms involved. We shall also assume that R is stable under shifts in the sense defined below.

Definition 2: An operator R is said to be *minimally T-stable* in L_p if there exists a subset E of L_p, not entirely inside the kernel of R, that is invariant under the action of T and on which R has a bounded operator norm; that is, if the following conditions are simultaneously satisfied.

$$[\tau]f \in E \qquad \text{for all } f \in E \text{ and all } \tau \in T;$$
$$\frac{\|Rf\|_p}{\|f\|_p} < C \qquad \text{for some } C < \infty \text{ and all } f \in E; \quad (12)$$
$$0 < \|Rf\|_p \qquad \text{for some } f \in E.$$
□

In some problems of practical interest, one may wish to consider a combination of, say, N regularization terms rather than a single one of them. These different regularizers may, for instance, measure the regularity of the projections of f onto different subspaces with special physical significance (we shall see some examples of these in §IV where we consider curl- and divergence-free subspaces). In this case, the cost functional to be minimized takes the form

$$\mathfrak{J}_{(\lambda_1:\lambda_N)}(f;y) = \mathfrak{d}(f;y) + \sum_{1\le i\le N} \lambda_i \|R_i f\|_{p_i}^{q_i},$$

which can also be interpreted as the Lagrange relaxation of a constrained optimization problem with several inequality constraints (i.e. $\|R_i f\|_{p_i}^{q_i} < a_i$, $1 \le i \le N$). Since, per Proposition 2, each of the regularization terms is invariant under the desired geometric transformations, their weighted sum will also have this property, up to a suitable independent adjustment of the λ_is for each given geometric transformation. As such, all that was or will be said here in connection with the interplay of invariance and regularization, will be understood to generalize in the sense just described to linear combinations of regularization terms.

Having established the general form of regularization functionals in terms of L_p norms of Rf, where R is the regularization operator with invariance properties dictated by (10), we shall now take up the task of identifying such operators. This will require us to consider scalar and vector cases separately, primarily due to the difference in the law of rotation in the two settings.

III. Regularization operators: Scalar case

In this section we shall derive the general form of linear regularization operators that possess specific invariance properties in the sense of (10). Our main result here is stated in Theorem 1, which shows that these operators take the form of fractional Laplacians.

We refer the reader to Definition 1 for a list of invariances that are of interest to us. Some peculiarities of the translation group (T) and the orthogonal group (Ω), together with the stability assumption described in Definition 2, allow us to show that for transformations in these two groups the constant k_S in (10) is always 1:

Lemma 1: A minimally T-stable operator R (cf. Definition 2) that is invariant under the action of Ω and T in the sense of (10) is *strictly* invariant under Ω and T, that is, it has

$$k_S = 1$$

for all $S \in \Omega \cup T$. □

Proof: First, note that for those elements of Ω that are of some finite order m, i.e. for any orthogonal matrix ω such that $\omega^m = \mathrm{Id}$, by the $\omega \mapsto k_\omega$ homomorphism (cf. (11)) we have $1 = k_{\mathrm{Id}} = k_{\omega^m} = k_\omega^m \Rightarrow k_\omega = 1$ (knowing that $k_\omega \in \mathbb{R}_+$).

Furthermore, any element ω of Ω, including those of infinite order, can be written as a product of at most d reflections ω_i, where d is the dimension (this is the Cartan-Dieudonné theorem). Reflections are of order 2, and hence have coefficient $k_{\omega_i} = 1$ by the previous paragraph. We therefore have, for arbitrary $\omega \in \Omega$, $k_\omega = \prod_i k_{\omega_i} = 1$. This proves the Ω part of the lemma.

We shall prove the second part by contradiction. To this end, assume that there exists $\tau \in T$ with $k_\tau \neq 1$. Without loss of generality we may assume $k_\tau > 1$ (simply replace τ by $-\tau$ in the other case). Then, for some $f \in E$ not in the kernel of R, with E defined in Definition 2,

$$\lim_{m\to\infty} \frac{\|R[m\tau]f\|}{\|[m\tau]f\|} = \lim_{m\to\infty} k_\tau^m \frac{\|Rf\|}{\|f\|} \to \infty,$$

which contradicts (12).

Finally, note that if we had restricted ourselves to rotation matrices instead of general orthogonal transformations in the first part of the lemma, we could still have proved $k_\omega = 1$ with the aid of an additional minimal Ω-stability assumption, arguing as we did for T. ∎

We also have

Lemma 2: The factor k_σ corresponding to scaling with $\sigma > 0$ (cf. (10)) is homogeneous in σ, that is, it can be written as

$$k_\sigma = \sigma^{-\gamma}$$

for some $\gamma \in \mathbb{R}$. □

The proof is very similar to that of Proposition 1, hence we omit it.

The stage is now set for the following result. It essentially goes back to Duchon, although here we derive it from somewhat different premises (such as minimal T-stability).

Theorem 1: Let R be a real, minimally T-stable, Fourier integral operator, initially defined from the Schwartz space $\mathscr{S}$ to L_p for some $p \geq 1$, that is invariant under the action of T, Ω, and Σ in the sense of (10). R is then characterized by a Fourier multiplier of the form

$$\hat{R}(\xi) = c|\xi|^\gamma \tag{13}$$

where $\gamma > d - d/p$ is the exponent identified in Lemma 2.

Conversely, Fourier operators with symbols given by (13) are strictly invariant under the action of T and Ω, and invariant under the action of Σ with the same coefficient k_σ as in Lemma 2. □

Proof: First observe that, by Lemma 1, R is *strictly* T- and Ω-invariant and, by Lemma 2, its Σ-invariance coefficient k_σ is a homogeneous function $\sigma^{-\gamma}$ of $\sigma \in \mathbb{R}_+$. Since R is a linear and translation-invariant Fourier operator, it is associated with an integral as per

$$\begin{aligned} Rf(x) &= (2\pi)^{-d} \int_{\mathbb{R}^d} e^{i\langle x,\xi\rangle} \hat{R}(\xi) \hat{f}(\xi)\, d\xi \\ &= (2\pi)^{-d} \langle \hat{R}, e^{i\langle x,\cdot\rangle} \hat{f} \rangle, \end{aligned}$$

where $\hat{R}$ is the Fourier multiplier corresponding to R.

One can then directly verify that in order for R to commute with rotations and scalings (the latter up to a homogeneous multiplicative factor of $k_\sigma = \sigma^{-\gamma}$), its Fourier expression $\hat{R}$ must be rotationally symmetric and homogeneous of degree γ. It is known [18, 19] that, subject to L_p boundedness, all such distributions can be represented in the form:

$$\hat{R}(\xi) = c\ \ |\xi|^\gamma \qquad \text{with } \gamma > d - d/p.$$

The same proof goes through when restricting ourselves to rotations instead of general orthogonal transformations if we make the additional assumption of minimal Ω-stability.

The converse is easily verified by simple changes of variables in the Fourier domain. ∎

Note that $|\xi|^\gamma$ is the Fourier symbol of the $\frac{1}{2}\gamma$-th (fractional) power of the negative Laplacian $(-\Delta)$. We can therefore write the reconstruction cost functional as

$$\mathfrak{J}_\lambda(f;y) = \mathfrak{d}(f;y) + \lambda\|(-\Delta)^{\gamma/2} f\|_p^q.$$

Moreover, by the argument given at the end of the previous section, we may additionally consider multiple additive regularization terms, as in

$$\mathfrak{J}_{(\lambda_1:\lambda_N)}(f;y) = \mathfrak{d}(f;y) + \sum_{1\leq i\leq N} \lambda_i \|(-\Delta)^{\gamma_i/2} f\|_{p_i}^{q_i}.$$

Two of the most important regularization functionals traditionally used in image processing are the total variation of f and its L_2 relaxation: $\|\operatorname{grad} f\|_1$ and $\|\operatorname{grad} f\|_2^2$ (using the vector L_p norms of (7)), both of which satisfy the required invariances. Note, however, that these regularizers, as such, fall outside the scope of this section for the reason that they incorporate an operator (grad) that maps scalars to vectors, while all operators considered so far map to scalars and not vectors. Nevertheless, due to a peculiar property of the L_2 norm (namely, that it is a Hilbert space and has an inner product structure), in the L_2 case one can write

$$\begin{aligned} \|Rf\|_2^2 &= \langle Rf, Rf\rangle = \langle R^*Rf, f\rangle \\ &= \langle (R^*R)^{\frac{1}{2}} f, (R^*R)^{\frac{1}{2}} f\rangle = \|(R^*R)^{\frac{1}{2}} f\|_2^2, \end{aligned} \tag{14}$$

where R^* is the adjoint of R, and the self-adjoint operator $(R^*R)^{\frac{1}{2}}$ maps scalars to scalars, and is therefore included in our framework. Hence, in particular, for the L_2 grad regularizer we have $\|\operatorname{grad} f\|_2^2 = \|(-\Delta)^{\frac{1}{2}} f\|_2^2$, which belongs to the family we derived above; something that cannot be said about the L_1 total variation.

Partly in order to overcome the latter limitation, later, in §V, we shall also develop the theory of scalar-to-vector regularization operators and introduce fractional gradients $\operatorname{grad}^\gamma$. From there it then follows that, more generally, invariant scalar cost functionals can be of the form

$$\begin{aligned} \mathfrak{J}_{(\lambda_1:\lambda_N)}(f;y) = \mathfrak{d}(f;y) &+ \sum_{1\leq i\leq N} \lambda_i \|(-\Delta)^{\gamma_i/2} f\|_{p_i}^{q_i} \\ &+ \sum_{1\leq i\leq N'} \lambda_i' \|\operatorname{grad}^{\gamma_i'} f\|_{p_i'}^{q_i'}. \end{aligned}$$

IV. Regularization operators: Vector case

Translations $\tau \in T$ and scalings $\sigma \in \Sigma$ act in the same way on vector fields as they do on scalars. On the other hand, we shall need to redefine the action of the orthogonal group Ω in the vector setting. Since a vector field is specified in the same coordinate system in which its argument is given, when transforming the domain one has to recompute the coordinates of the vector field accordingly. More precisely, the formula for transforming a vector field $f=(f_1,\dots,f_d)^{\mathrm{T}}$ by an orthogonal matrix ω is

$$[\omega]_v f=\omega f(\omega^{\mathrm{T}}\cdot); \tag{15}$$

that is, the coordinates of the vector are transformed by the inverse of the domain transformation matrix. On occasion, we shall refer to invariance as in (15) as *contra-variance* (recall that we distinguish between the scalar and vector operators associated with ω by subscripting $[\omega]$ by s and v respectively).

The following result, proved indirectly for $d=2,3$ in Arigovindan [10], is the vector counterpart of Theorem 1. In the Appendix, we give a different and more general proof of this theorem, valid in any number of dimensions.

Theorem 2: Let R be a real, minimally T-stable, Fourier operator initially defined $\mathscr{S}^d\to \mathrm{L}_p^d$ and mapping vector fields to vector fields, which is invariant under the action of T, Ω, and Σ in the sense of (10). R is then characterized by a (matrix-valued) Fourier multiplier of the form

$$\hat{R}(\xi)=|\xi|^\gamma\left[c_1\frac{\xi\xi^{\mathrm{T}}}{|\xi|^2}+c_2\Big(I-\frac{\xi\xi^{\mathrm{T}}}{|\xi|^2}\Big)\right] \tag{16}$$

where $c_1,c_2\in\mathbb{R}$ are Helmholtz coefficients (see below) and $\gamma>d-d/p$ is the exponent identified in Lemma 2.

Conversely, operators with Fourier multipliers as above satisfy all of the required invariances. □

Sketch of the proof: The complete proof appears in the Appendix. Here is an introduction to it.

The $k_\tau=1$ part of Lemma 1, and Lemma 2 (which says that $k_\sigma=\sigma^{-\gamma}$ for some γ) apply directly and without modification in the vector setting. Also, following the same line of argument as in the proof of Lemma 1, one can prove that once again $k_\omega=1$ for all $\omega\in\Omega$, as was the case for scalars. Since R is linear and translation-invariant, it admits a Fourier-domain representation as

$$Rf(x)=(2\pi)^{-d}\int_{\mathbb{R}^d}\mathrm{e}^{\mathrm{i}\langle x,\xi\rangle}\hat{R}(\xi)\hat{f}(\xi)\,\mathrm{d}\xi,$$

where $\hat{R}$ is now a *matrix-valued* Fourier kernel. The scale-invariance of R with coefficient $k_\sigma=\sigma^{-\gamma}$ and its strict orthogonal contra-variance translate to the following Fourier-domain identities:

$$\hat{R}(\sigma\cdot)=\sigma^\gamma\hat{R}(\cdot) \qquad \text{for all } \sigma>0; \tag{17}$$

$$\hat{R}(\omega\cdot)=\omega\hat{R}(\cdot)\omega^{\mathrm{T}} \qquad \text{for all } \omega\in\Omega. \tag{18}$$

In the Appendix, we prove the forward direction in two steps, first showing that orthogonal contra-variance implies that $\hat{R}(\xi)$ at any $\xi\neq0$ has the eigen-decomposition

$$\hat{R}(\xi)=\mu_1(\xi)\frac{\xi\xi^{\mathrm{T}}}{|\xi|^2}+\mu_2(\xi)\Big(I-\frac{\xi\xi^{\mathrm{T}}}{|\xi|^2}\Big), \tag{19}$$

and then noting that, by Theorem 1, μ_1 and μ_2 must be of the form $c_i|\xi|^\gamma$, $i=1,2$.

The converse of the theorem can be verified easily by Fourier-domain changes of variables. ∎

With regard to the parameters c_1,c_2, three cases are of particular interest; namely, those of $c_1=c_2$, $c_1=0$, and $c_2=0$.

For $c_1=c_2=c$, the operator defined in Theorem 2 has the Fourier expression $c\|\xi\|^\gamma$ and therefore corresponds, up to normalization, to the fractional vector Laplacian $(-\Delta)^{\gamma/2}$, that is, the scalar Laplacian applied coordinate-wise. For this reason we shall refer to the family of operators identified by (19) as *generalized vector Laplacians,* with the notation $(-\Delta)^{\gamma/2}_{(c_1,c_2)}$ [20].

To better understand the behaviour of the operator when either c_1 or c_2 is zero, note that $(-\Delta)^{\gamma/2}_{(c_1,c_2)}$ can be decomposed as

$$\left[c_1(\mathrm{Id}-\mathbb{P})+c_2\mathbb{P}\right](-\Delta)^{\gamma/2} \tag{20}$$

where the operator $\mathbb{P}$ is defined by its Fourier multiplier $\hat{\mathbb{P}}=\xi\xi^{\mathrm{T}}/|\xi|^2$. It is straightforward to see that $\mathbb{P}$ and its complement $\mathrm{Id}-\mathbb{P}$ are projections and that they in fact project their argument onto its curl-free and divergence-free components respectively; in other words, taken together they provide a Helmholtz decomposition of their argument.

To summarize, the operators identified in Theorem 2 effectively combine a fractional vector Laplacian with a re-weighting of Helmholtz components. Moreover, one has

$$(-\Delta)^{\gamma/2}_{(c_1,c_2)}(-\Delta)^{\gamma'/2}_{(c_1',c_2')}=(-\Delta)^{(\gamma+\gamma')/2}_{(c_1c_1',c_2c_2')}.$$

We can now give the general form of our cost functional for vector fields, as we did for scalar fields in §II. Once again, we may consider linear combinations of some N regularization terms, which retain the same invariances as the individual terms, up to re-adjustment of $\lambda_1,\dots,\lambda_N$:

$$\mathfrak{J}_{(\lambda_1:\lambda_N)}(f;y)=\mathfrak{d}(f;y)+\sum_{1\le i\le N}\lambda_i\|(-\Delta)^{\gamma_i/2}_{(c_{1,i},c_{2,i})}f\|_{p_i}^{q_i}.$$

However, the above family is still not complete, for reasons similar to those given at the end of §III. This, in fact, will be the subject of the next section.

V. More on L_p regularization of vector fields

A. Motivation

In our discussion in the preceding sections we implicitly assumed that R mapped scalar or vector fields to similar objects and in the same number of dimensions. In other words, we considered the operator associated with S in (10) to be the same on the left and right sides. In this way, we overlooked some important possibilities for vector regularization operators, such as the divergence operator (mapping vector fields to fields of scalars) or the curl (mapping vector fields to pseudo-vector fields in 3D; see below). In this section, we shall remedy this by studying operators that generalize divergences and curls (and their adjoints), in the same way that the operators of the preceding sections generalized scalar and vector Laplacians.

The generalization to d dimensions of the divergence raises no difficulty. Indeed, the divergence of a vector field is defined in any number of dimensions by means of the Fourier multiplier $\mathrm{i}\xi^{\mathrm{T}}$ (given in Cartesian coordinates). The divergence maps vector fields to scalar fields. Its adjoint is the negative gradient with Fourier multiplier $-\mathrm{i}\xi$, which maps scalar fields to vector fields.

It is less obvious how the usual three-dimensional definition of the curl can be generalized to d dimensions. This difficulty is essentially rooted in the fact that the curl of a vector in 3D is not a *true* vector: per the right-hand rule of physics, the curl of a vector field transforms as an ordinary vector field under proper rotations, but it flips sign under *improper* rotations (those with determinant -1). For this reason, curl fields in 3D are usually referred to as *pseudo-vector* fields.

It is in fact this notion of pseudo-vector that does not generalize directly to arbitrary d. For this reason, in higher dimensions, it is constructive to consider the curl operator as a map from vector fields to *bivector* fields ($d \times d$ matrix fields with specific transformation laws). We may identify bivectors with fields of $d \times d$ anti-symmetric tensors [21]. These have $d(d-1)/2$ independent components, corresponding to the upper-diagonal elements of the matrix (only in 3D is $d = d(d-1)/2$, hence the difficulty in generalizing the customary definition of curl and pseudo-vectors to $d > 3$).

In three dimensions, identification between pseudo-vectors and anti-symmetric matrices (bivectors) can be made by the $\star$-map that we introduce below.

$$\frac{1}{\sqrt{2}}\begin{bmatrix} 0 & -f_3 & f_2 \\ f_3 & 0 & -f_1 \\ -f_2 & f_1 & 0 \end{bmatrix} \mapsto \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}. \tag{21}$$

The d-dimensional generalization of the curl, as a map from vector fields to bivector fields, is then given by the Fourier expression

$$(\operatorname{curl} f)^\wedge = \tfrac{1}{\sqrt{2}}\big(\mathrm{i}\xi\, \hat{f}^{\mathrm{T}} - \hat{f}\, \mathrm{i}\xi^{\mathrm{T}}\big).$$

Defining the vector-gradient $\operatorname{grad} f$ of the vector field f as the matrix

$$\operatorname{grad} f = \begin{bmatrix} \partial_1 f_1 & \partial_1 f_2 & \partial_1 f_3 \\ \partial_2 f_1 & \partial_2 f_2 & \partial_2 f_3 \\ \partial_3 f_1 & \partial_3 f_2 & \partial_3 f_3 \end{bmatrix},$$

we may write the curl of f as

$$\operatorname{curl} f = \tfrac{1}{\sqrt{2}}\left(\operatorname{grad} f - (\operatorname{grad} f)^{\mathrm{T}}\right).$$

In combination with (21), the above relation yields the usual definition of curl in 3D.

The adjoint of the curl, which maps bivector fields to vector fields, is given by the expression

$$(\operatorname{curl}^* f)^\wedge = \tfrac{1}{\sqrt{2}}(\hat{f} - \hat{f}^{\mathrm{T}})\mathrm{i}\xi$$

(note that in the former equation f is a vector field, whereas in the latter it denotes a tensor).

Finally, we note that in 3D, under an orthogonal transformation of the domain by $\omega \in \Omega$, the bivector and pseudo-vector representations (the sides of (21)) transform respectively as

$$\tfrac{1}{\sqrt{2}}\omega^{\mathrm{T}} \begin{bmatrix} 0 & -f_3(\omega^{\mathrm{T}}\cdot) & f_2(\omega^{\mathrm{T}}\cdot) \\ f_3(\omega^{\mathrm{T}}\cdot) & 0 & -f_1(\omega^{\mathrm{T}}\cdot) \\ -f_2(\omega^{\mathrm{T}}\cdot) & f_1(\omega^{\mathrm{T}}\cdot) & 0 \end{bmatrix} \omega$$

and

$$(\det \omega)\, \omega \begin{bmatrix} f_1(\omega^{\mathrm{T}}\cdot) \\ f_2(\omega^{\mathrm{T}}\cdot) \\ f_3(\omega^{\mathrm{T}}\cdot) \end{bmatrix}$$

(the determinant captures the sign flip of pseudo-vectors under parity transformations). The first of these two defines the general law of action of Ω on bivectors in d dimensions:

$$[\omega]_b f = \omega^{\mathrm{T}} f(\omega^{\mathrm{T}}\cdot)\, \omega.$$

B. Curl-like and divergence-like operators and their adjoints

In this subsection we shall give a categorical definition of d-dimensional *curl-like* and *divergence-like* families of operators, and make the connection between these operators and the scalar and vector Laplacians of previous sections. But before this, let us first briefly recall in a single place the law of action of transformation groups on scalars, vectors, and bivectors. For orthogonal transformation with $\omega \in \Omega$ we have

$$[\omega]_s f = f(\omega\cdot); \tag{22}$$
$$[\omega]_v f = \omega^{\mathrm{T}} f(\omega\cdot); \tag{23}$$
$$[\omega]_b f = \omega\, f(\omega\cdot)\, \omega^{\mathrm{T}}; \tag{24}$$

for all $\omega \in \Omega$. Note that in the first equation f is scalar, in the second it is vector, and, finally, in the last equation its values are $d \times d$ anti-symmetric matrices.

The actions of T and Σ on the three categories (scalar, vector, and bivector) remain the same in all cases ($f \mapsto f(\cdot - \tau)$ for the former and $f \mapsto f(\sigma^{-1}\cdot)$ for the latter, for all $\tau \in T$ and $\sigma \in \Sigma$).

As noted, we may, in more generality than previous sections, study the two families of *divergence-like* and *curl-like* operators and their adjoints. Operators in the former category go from d coordinates to 1, and back by their adjoint; whereas those in the latter go from d coordinates to $d(d-1)/2$ independent coordinates (forming a $d \times d$ anti-symmetric matrix) and back to d:

(V.1) *Divergences (R_{div}) and their adjoints (R^*_{div}):* These consist of maps from vector fields to scalar fields and vice versa. In the first case the invariance equation takes the form

$$R_{\mathrm{div}}\, [S]_v = [S]_s\, R_{\mathrm{div}};$$

and in the second case we require

$$R^*_{\mathrm{div}}\, [S]_s = [S]_v\, R^*_{\mathrm{div}}.$$

Given our focus on linear regularization operators and the shift-invariance assumption, we can restate the above properties as conditions on the Fourier multipliers of R_{div} and R^*_{div}. Scale-invariance in all cases leads to the same equation as (17). With regard to reflection-invariance, in place of (18) we have,

$$\hat{R}_{\mathrm{div}}(\omega\cdot) = \hat{R}_{\mathrm{div}}(\cdot)\omega^{\mathrm{T}}$$

for divergence-like operators and

$$\hat{R}^*_{\mathrm{div}}(\omega\cdot) = \omega^{\mathrm{T}}\hat{R}^*_{\mathrm{div}}(\cdot)$$

for their adjoints. These follow from (22) and (23).

(V.2) *Curls (R_{curl}) and their adjoints (R^*_{curl}):* Curl-like operators map vector fields to fields of bivectors. Accordingly, their adjoints map bivectors back to vectors. For the two we have, respectively,

$$R_{\mathrm{curl}}\,[S]_v = [S]_b\,R_{\mathrm{curl}};$$
$$R^*_{\mathrm{curl}}\,[S]_b = [S]_v\,R^*_{\mathrm{curl}}.$$

In the Fourier multipliers, scale-invariance is again reflected by (17). For orthogonal invariance the equivalents of the preceding pair of equations are, respectively,

$$[\hat{R}_{\mathrm{curl}}(\omega\cdot)]_{\alpha\beta\gamma} = \omega^{\mathrm{T}}_{\alpha i}\omega_{j\beta}\omega^{\mathrm{T}}_{k\gamma}[\hat{R}_{\mathrm{curl}}(\cdot)]_{ijk};$$
$$[\hat{R}^*_{\mathrm{curl}}(\omega\cdot)]_{\alpha\beta\gamma} = \omega_{i\alpha}\omega^{\mathrm{T}}_{\beta j}\omega_{\gamma k}[\hat{R}^*_{\mathrm{curl}}(\cdot)]_{ijk}.$$

These are consequences of (22) and (24) (we are using here a light form of Einstein's summation convention, whence repeated indices are summed upon; for instance, $c_{ij} = a_{ik}b_{kj}$ is the product of the matrices a_{ij} and b_{ij}). Notice that $\hat{R}_{\mathrm{curl}}$ and $\hat{R}^*_{\mathrm{curl}}$ are third-rank tensors (linear maps between vectors and matrices), acting, respectively, on vectors and matrices by

$$\big[R_{\mathrm{curl}}f\,\big]^\wedge_{ij} = [\hat{R}_{\mathrm{curl}}]_{ijk}\hat{f}_k \quad \text{and} \quad \big[R^*_{\mathrm{curl}}f\,\big]^\wedge_{k} = [\hat{R}_{\mathrm{curl}}]_{ijk}\hat{f}_{ij}.$$

Example 1: Fractional divergences and gradients: These are denoted, respectively, as div^γ and grad^γ, and are defined by their respective symbols

$$|\xi|^\gamma\frac{\mathrm{i}\xi^{\mathrm{T}}}{|\xi|} \quad \text{and} \quad |\xi|^\gamma\frac{\mathrm{i}\xi}{|\xi|}.$$

Fractional divergences act on vector fields, mapping them to scalars; gradients do the opposite, with div^γ and $\mathrm{div}^{\gamma*} = -\mathrm{grad}^\gamma$ forming and adjoint pair. The fractional gradient of order 0 (grad^0) is known as the Riesz transform [22]. □

Example 2: Fractional curls and adjoint curls: We shall denote the fractional curl and its adjoint by curl^γ and $\mathrm{curl}^{\gamma*}$ respectively. They are defined in the Fourier domain according to

$$(\mathrm{curl}^\gamma f)^\wedge = \tfrac{1}{\sqrt{2}}|\xi|^\gamma\Big(\frac{\mathrm{i}\xi}{|\xi|}\hat{f}^{\mathrm{T}} - \hat{f}\,\frac{\mathrm{i}\xi^{\mathrm{T}}}{\xi}\Big),$$
$$(\mathrm{curl}^{\gamma*} f)^\wedge = \tfrac{1}{\sqrt{2}}|\xi|^\gamma(\hat{f} - \hat{f}^{\mathrm{T}})\,\frac{\mathrm{i}\xi}{|\xi|}.$$

These definitions are valid in any number of dimensions ≥ 2 (they are trivial in one dimension). Fractional curls map d-dimensional vectors to $d\times d$ anti-symmetric bivectors; adjoint curls go in the opposite direction. □

One readily verifies that the above examples satisfy the invariances outlined in (V.1) and (V.2).

Our claim has been that the considerations of this section are more general than those of the previous two; and yet, until this point they seem to have been limited to operators mapping vectors to non-vector and vice-versa. Hence, at first sight, it might appear that for completeness we shall have to include in our discussion of regularization operators, additionally, the families considered in §II (scalar to scalar) and in §IV (vector to vector). However, we shall now show that the latter families can be decomposed in terms of fractional curls and divergences and their adjoints. Specifically, for the scalar fractional Laplacian we have

$$\begin{aligned}(-\Delta)^\gamma &= \mathrm{div}^\gamma(\mathrm{div}^\gamma)^* \\ &= \mathrm{div}^\gamma(-\mathrm{grad}^\gamma);\end{aligned}$$

and for the generalized fractional vector Laplacian of §IV,

$$\begin{aligned}(-\Delta)^\gamma_{(c_1,c_2)} &= c_1(\mathrm{div}^\gamma)^*\,\mathrm{div}^\gamma + c_2(\mathrm{curl}^\gamma)^*\,\mathrm{curl}^\gamma \\ &= c_1(-\mathrm{grad}^\gamma)\,\mathrm{div}^\gamma + c_2(\mathrm{curl}^\gamma)^*\,\mathrm{curl}^\gamma\,;\end{aligned}$$

or, what is the same,

$$\begin{aligned}(\mathrm{div}^\gamma)^*\,\mathrm{div}^\gamma &= (\mathrm{Id}-\mathbb{P})(-\Delta)^\gamma,\\ (\mathrm{curl}^\gamma)^*\,\mathrm{curl}^\gamma &= \mathbb{P}(-\Delta)^\gamma.\end{aligned}$$

In addition, we record the following factorization results that relate fractional curls and divergences to combinations of integer-order operators and the fractional vector Laplacian $(-\Delta)^\gamma$:

$$\begin{aligned}\mathrm{div}^\gamma &= \mathrm{div}\,(-\Delta)^{\gamma/2};\\ \mathrm{grad}^\gamma &= (-\Delta)^{\gamma/2}\,\mathrm{grad}\,;\\ \mathrm{curl}^\gamma &= \mathrm{curl}\,(-\Delta)^{\gamma/2};\\ \mathrm{curl}^{\gamma*} &= (-\Delta)^{\gamma/2}\,\mathrm{curl}^*\,.\end{aligned}$$

We shall not burden ourselves further by trying to find, in complete generality, the equivalents of Theorems 1 and 2 for curl-like and divergence-like families, as the cases covered by the above examples appear to us to be sufficiently versatile for applications.

Note, finally, that in order to form regularization functionals similar to (9) which involve curl-like operators, we shall need to define the equivalent of p-norms on $d\times d$ tensor fields. The matrix L_p norms defined in Appendix A work perfectly for this purpose. It is also easy to see that in the case of anti-symmetric matrices, the functional obtained in this ways equals the vector L_p norm of the upper diagonal elements of the matrix (in particular, in 3D this is effectively the same as the norm applied to vector fields). This means that we may alternatively define the same regularization functional in terms of the vector L_p norm of the $\star$-map of the curl (cf. (21)).

Given all this, a general vector cost functional with multiple regularizers can be written as

$$\mathfrak{J}_{(\lambda_1,\lambda_2)}(f;y) = \mathfrak{d}(f;y) + \sum_{1\le i\le N} \lambda_i \|(-\Delta)^{\gamma_i/2}_{(c_{1,i},c_{2,i})} f\|^{q_i}_{p_i} + \sum_{1\le i\le N'} \lambda'_i \|\operatorname{curl}^{\gamma'_i} f\|^{q'_i}_{p'_i} + \sum_{1\le i\le N''} \lambda''_i \|\operatorname{div}^{\gamma''_i} f\|^{q''_i}_{p''_i}.$$

(the three p-norms appearing in the above equation are those defined for vectors, bivectors, and scalars, in that order; cf. (7) and Appendix A).

An illustrative example is

$$\begin{aligned}\mathfrak{J}_{(\lambda_1,\lambda_2)}(f;y) = \mathfrak{d}(f;y) &+ \lambda_0 \|(-\Delta)^{\gamma/2} f\|^p_p \\ &+ \lambda_1 \|(\mathrm{Id}-\mathbb{P})(-\Delta)^{\gamma/2} f\|^p_p \\ &+ \lambda_2 \|\mathbb{P}(-\Delta)^{\gamma/2} f\|^p_p \\ &+ \lambda_c \|\operatorname{curl}^\gamma f\|^p_p \\ &+ \lambda_d \|\operatorname{div}^\gamma f\|^p_p,\end{aligned}$$

which incorporates independent regularization of curl-free and div-free subspaces (see the definition of $\mathbb{P}$ after (20)), as well as fractional curl and div terms. Note that some of the λ_is may be zero.

We conclude this section by the observation that, as we also saw in (14), in the quadratic case ($p=2$) the above functional reduces to the one given at the end of §IV, since in this particular case the norm is associated with an inner product, thus allowing us to equate $\|\operatorname{curl}^\gamma f\|_2$ with $\|\mathbb{P}(-\Delta)^{\gamma/2} f\|_2$ and $\|\operatorname{div}^\gamma f\|_2$ with $\|(\mathrm{Id}-\mathbb{P})(-\Delta)^{\gamma/2} f\|_2$, as can be readily verified using Parseval's identity. This is generally not true for other values of p (but it would have been, had we considered L_p norms in the Fourier domain in place of the usual spatial L_p norms).

VI. Illustration

For the purpose of illustration, we now consider the problem of reconstructing a vector field from noisy measurements—primarily in 3D ($d=3$) but also in 2D—using a quadratic fidelity criterion (consistent with a white Gaussian noise assumption). We shall focus on div-curl regularization with different (L_2 vs L_1) norms. We note in passing that in practical problems, higher-order regularization, such as the physically-motivated second-order div-curl regularization of Suter [1], can be of interest, especially in the context of motion estimation. In this section, our primary motivation is to demonstrate and compare the use of L_2 vs L_1 norms, in line with the similar comparison of quadratic vs total-variation type regularization of scalars that has frequently been made in image processing literature. For this reason we shall limit ourselves to first-order differential regularization operators. Specifically, we shall consider the cost functions

$$\mathfrak{J}^{(p)}(f;Y) = \sum_n |f(n) - Y[n]|^2 + \lambda_c \|\operatorname{curl} f\|^p_p + \lambda_d \|\operatorname{div} f\|^p_p \tag{25}$$

with $p=1,2$, where $Y[n]$s are the measurements (in this section, upper case letters will be used to denote discrete quantities such as $Y = Y[n] = (Y_1[n], Y_2[n], Y_3[n])$, for n in some subset of $\mathbb{Z}^d$). In interpreting the above formula when the number of samples and/or estimated values $f(x)$ goes to infinity, some form of normalization or limit argument may become necessary. But in practice the number of observations $Y[n]$ will be finite.

The norm applied to the curl in the former equation is a matrix L_p norm as defined in Appendix A; but in 3D, we may use the $\star$-map defined in (21) and rewrite it as a vector norm (cf. (7)):

$$\mathfrak{J}^{(p)}(x;Y) = \sum_i |f(i) - Y[i]|^2 + \lambda_c \int_{\mathbb{R}^d} \left(\sqrt{|\star \operatorname{curl} f|^2}\right)^p + \lambda_d \int_{\mathbb{R}^d} \left(\sqrt{|\operatorname{div} f|^2}\right)^p. \tag{26}$$

For $p=1$, the mixed L_2–L_1 functional proposed above is in the spirit of total variation (TV) regularization. It is of interest to compare it against its purely quadratic counterpart, if only to see whether the relative advantage of TV regularization to quadratic regularization in 2D image denoising carries over to the vector setting.

In three dimensions, the explicit definitions of curl and divergence are

$$\begin{aligned}\operatorname{div} f &= \sum_{1\le k\le n} \partial_k f_k = \partial_1 f_1 + \partial_2 f_2 + \partial_3 f_3; \\ \star \operatorname{curl} f &= \left[\partial_3 f_2 - \partial_2 f_3;\ \partial_1 f_3 - \partial_3 f_1;\ \partial_2 f_1 - \partial_1 f_2\right].\end{aligned}$$

While guided by the previous continuous formulation, our implementation on a digital computer is necessarily discrete. Although there is room for more sophistication, we shall discretize simply by taking finite differences in place of derivatives, while emphasizing that in practice, the discretization scheme used can play an important role in the numerical solution of inverse problems. It is therefore advisable, in real-world problems, to look at alternatives such as discrete orthogonal decompositions; see Yuan, Schnörr, and Mémin [23].

Let $F_j = F_j[n]$, $j = 1, 2, \ldots, d$, denote the reconstruction corresponding to samples of f_j over some discrete domain $\subset \mathbb{Z}^d$. Further, let us denote by $\delta_i F_j$ the finite difference associated with the partial derivative $\partial_i f_j$, that is,

$$\delta_i : F_j \mapsto F_j - F_j[\cdot - \hat{e}_i]$$

except at the boundaries where some preferred type of boundary conditions is applied ($\hat{e}_i$ is the i-th standard unit vector in $\mathbb{R}^d$). For future reference, also note the adjoint of δ_i:

$$\delta_i^* : F_j \mapsto F_j - F_j[\cdot + \hat{e}_i].$$

Discrete divergence and curl can be defined in 3D by the identities

$$\begin{aligned}\operatorname{div}_\delta F &= \delta_1 F_1 + \delta_2 F_2 + \delta_3 F_3; \\ \operatorname{curl}_\delta F &= \left[\delta_3 F_2 - \delta_2 F_3;\ \delta_1 F_3 - \delta_3 F_1;\ \delta_2 F_1 - \delta_1 F_2\right].\end{aligned}$$

The point-wise squared amplitudes of the curl and divergence that appear under the square root sign in (26) are then discretized as

$$|\operatorname{curl}_\delta F[m]|^2 = \sum_{1\le i<j\le 3} (\delta_i F_j[m] - \delta_j F_i[m])^2;$$
$$|\operatorname{div}_\delta F[m]|^2 = \sum_{1\le i,j\le 3} \delta_i F_i[m] \delta_j F_j[m].$$

Our discrete cost function can then be written as

$$\begin{aligned}\mathfrak{J}_\delta^{(p)}(F;Y) = &\sum_m |F[m]-Y[m]|^2 \\ &+ \lambda_c \sum_m \left(\sqrt{|\operatorname{curl}_\delta F[m]|^2}\right)^p \\ &+ \lambda_d \sum_m \left(\sqrt{|\operatorname{div}_\delta F[m]|^2}\right)^p\end{aligned} \tag{27}$$

(recall that $F[m], Y[m]$ are vectors, and $|\cdot|$ denotes the Euclidean length; the index m runs over the sampling/reconstruction grid in $\mathbb{Z}^d$).

For $p=2$, the problem is quadratic and can be efficiently solved using iterative linear methods. For the L_1 problem, following Figueiredo *et al.* [15], we shall now propose an iterative reweighted least squares (IRLS) approach belonging to the family of Majorize-Minimize (MM) algorithms.

Given some F' with $|RF'| > 0$, the L_1 terms of the functional can be upper-bounded as

$$\sum \sqrt{|RF|^2} \le \sum \sqrt{|RF'|^2} + \sum \left(|RF|^2 - |RF'|^2\right)/2\sqrt{|RF'|^2} \tag{28}$$

(this follows from the inequality $\sqrt{a} \le \sqrt{a'} + \frac{1}{2}(a-a')/\sqrt{a'}$). Let the sequence $(\tilde{F}_{(n)})$ be defined by

$$\tilde{F}_{(n)} = \arg\min_F \mathfrak{Q}_\delta(F, \tilde{F}_{(n-1)}; Y) \tag{29}$$

where

$$\begin{aligned}\mathfrak{Q}_\delta(F,F';Y) := &\sum_m \sum_{1\le i\le 3} F_i[m]^2 \\ &- \sum_m \sum_{1\le i\le 3} 2F_i[m]Y_i[m] \\ &+ \lambda_c \sum_m \frac{|\operatorname{curl}_\delta F[m]|^2}{\sqrt{|\operatorname{curl}_\delta F'[m]|^2}} \\ &+ \lambda_d \sum_m \frac{|\operatorname{div}_\delta F[m]|^2}{\sqrt{|\operatorname{div}_\delta F'[m]|^2}} \quad + \quad \mathfrak{K}_\delta(F';Y)\end{aligned}$$

is obtained by majorizing (27) using (28); we have collected all terms depending only on Y and F' in the scalar function $\mathfrak{K}_\delta$, which we may discard when solving (29).

Note that $\mathfrak{Q}(F,F;Y) = \mathfrak{J}_\delta^{(1)}(F;Y)$. Furthermore, we have

$$\begin{aligned}\mathfrak{J}_\delta^{(1)}(\tilde{F}_{(n)};Y) &\le \mathfrak{Q}_\delta(\tilde{F}_{(n)}, \tilde{F}_{(n-1)};Y) \\ &< \mathfrak{Q}_\delta(\tilde{F}_{(n-1)}, \tilde{F}_{(n-1)};Y) = \mathfrak{J}_\delta^{(1)}(\tilde{F}_{(n-1)};Y)\end{aligned}$$

which shows that, with increasing n, the $\mathfrak{J}_\delta^{(1)}(\tilde{F}_{(n)})$s form a decreasing sequence (in the second inequality we have used the strict convexity of $\mathfrak{Q}_\delta$, and assumed that $\tilde{F}_{(n)} \ne \tilde{F}_{(n-1)}$).

For fixed F', the minimizer of $\mathfrak{Q}_\delta(F,F';Y)$ over F is the solution of the linear system of equations obtained by setting all of the derivatives of $\mathfrak{Q}_\delta$ equal to zero. To see this, let us first define:

$$c'_m = \sqrt{|\operatorname{curl}_\delta F'[m]|^2}; \qquad d'_m = \sqrt{|\operatorname{div}_\delta F'[m]|^2}.$$

Further, let:

$$C'_k F[p] := \sum_{1\le i\le 3} \delta_i^* \frac{\delta_i F_k[p] - \delta_k F_i[p]}{c'_p};$$
$$D'_k F[p] := \delta_k^* \frac{\sum_{1\le i\le 3} \delta_i F_i[p]}{d'_p}.$$

After some algebraic simplification, one can write

$$\begin{aligned}\frac{\partial}{\partial F_k[p]} \mathfrak{Q}_\delta(F,F';Y) = 2(F_k[p] - Y_k[p]) &+ 2\lambda_C C'_k F[p] \\ &+ 2\lambda_D D'_k F[p].\end{aligned}$$

The system of equations

$$\frac{\partial}{\partial F_k[p]} \mathfrak{Q}_\delta(F,F';Y) = 0, \quad \text{for } k=1,2,3 \text{ and all } p, \tag{30}$$

thus corresponds to a linear system $AF = Y$ (shorthand for $(AF)_k[p] = Y_k[p]$, for all p). This system may then be solved using a variety of methods (conjugate gradient, multi-grid-preconditioned GMRES, *etc.*). In implementation, one may add a small ϵ to numerators and denominators to avoid division by zero.

To summarize, the complete algorithm for L_1 regularized denoising consists of a number of *outer* cycles in accordance with (29), which sequentially reduce the cost functional $\mathfrak{Q}_\delta$. The nth outer iteration takes the measurements (Y) and the output of the $(n-1)$th iteration ($\tilde{F}_{(n-1)}$) as inputs, and then moves in the direction of minimizing $\mathfrak{Q}_\delta(\cdot, \tilde{F}_{(n-1)}, Y)$. This local minimization corresponds to a linear system as specified in (30). Within each outer iteration, this system is then (approximately) solved using a number of *inner* iterations of some iterative linear solver.

Simulation and results

We implemented the scheme described above in MATLAB (The MathWorks, Inc., Natick, MA) in 2D and 3D. As experiments, we considered the denoising of phantoms corrupted by different levels of white Gaussian noise. λ_C and λ_D were optimized for best mean squared error (MSE) performance. In simulation, the true MSE for a given choice of λ_C and λ_D can be calculated using an oracle. In practice, even though the ground truth is not known and the true MSE is therefore not accessible, so long as the white Gaussian noise assumption remains valid, a highly accurate estimate of the MSE can be obtained using Monte Carlo techniques that approximate Stein's Unbiased Risk Estimate (SURE), as described in Ramani *et al.* [24] (see also Girard [25]). This estimate comes at the cost of solving an extra denoising problem for each choice of λ_C, λ_D, but in terms of effectiveness in predicting the best values of λ_C and λ_D we found it to be indistinguishable from the oracle in our experiments.

Results are reported in Table II and in Figures 1–5 (3D graphics were generated using ParaView 3.8.0 [26]). The phantoms, and high resolution images of their noisy and denoised versions, are available online, at the web address <http://bigwww.epfl.ch/tafti/gal/vreg/>.

The first 3D phantom, presented in Figure 1, consists of the gradient field of the potential

$$\phi_{3D}(x_1, x_2, x_3) = x_1 x_2 e^{-|x|^2}.$$

The second 3D phantom, depicted in Figure 2, models fully-developed laminar flow (with a parabolic profile) in a tube, encircled by constant flow inside a torus.

We solved the L_1 version of the denoising problem using the iterative reweighted least squares scheme described above, with 8 external cycles per (29) and 600 conjugate gradient (CG) inner iterations per cycle to solve the linearized problem in each step. The L_2 problem was solved to convergence using CG iterations.

In Figures 3 and 4 we show the amplitude profile of L_1 and L_2 reconstructions of the two 3D phantoms. These reconstructions are also compared in Table II in terms of SNR improvement after denoising (with λ_C, λ_D optimized for best SNR performance) and mean angular error. The latter performance measure is defined as the average pointwise angle between the ground truth and the reconstruction; see Barron, Fleet, and Beauchemin [27].

The point we wish to highlight here is that L_1 regularization performs remarkably well for the second phantom, which features discontinuities in the flow, while being almost comparable to L_2 regularization for the first (smooth) phantom. The former regularization also better preserves small details and discontinuities at flow boundaries, which are smoothed in L_2 denoising. On the other hand, not unexpectedly, L_2 denoising produces slightly higher SNRs in the case of the smooth 'gradient' phantom, although L_1 regularization is still quite comparable in terms of SNR and even yields smaller angular errors.

As hinted previously, we took advantage of the availability of the ground truth to optimize the parameters λ_C, λ_D for best SNR, for which purpose we used a bracketing search method (it also bears reminding that the parameters were therefore *not* optimized for our second quality criterion, the mean angular error). The parameter values obtained in our experiments are tabulated in Table III. We remark that the superior performance of the L_1 algorithm is in spite of the fact that, in contrast to the L_2 case, the experimentally-obtained parameters λ_C, λ_D for the L_1 problem may in fact be sub-optimal, primarily as a consequence of that the L_1 problem is typically solved only partially by fixing the number of iterations (computational budget) in advance, meaning that due to the variable state of convergence SNR performance fluctuates about its optimum, thus breaking the working assumptions of typical optimization algorithms used to optimize λ_C, λ_D. It is also worth noting that terminating the scheme before full convergence can itself be seen as an additional source of regularization; the optimal parameters λ_C, λ_D therefore depend also on the state of convergence of the problem.

As a further demonstration of potential, in Figure 5 we provide a sample output of 2D vector field denoising. The phantom used in this case was the gradient of the potential function

$$\phi_{2D}(x_1, x_2) = x_1 e^{-|x|^2}$$

(contour lines of ϕ_{2D} are superimposed in colour). For the example shown in Figure 5 we observed an SNR improvement of 12.74 dB with L_1 regularization, compared to an improvement of 12.58 dB when using quadratic (L_2) regularization. We note that reconstruction of 2D vector fields can have applications beyond denoising, for instance in image registration and motion estimation, although in the latter case temporal regularization also needs to be considered.

Table II: Comparison of denoising algorithms in 3D; algorithm parameters were optimized for best SNR for each regularizer and input SNR.

(a) gradient field

input SNR [dB] / angular error [deg.]	SNR improvement [dB] / angular error (mean ± stdev) [deg.] L_1	L_2
0	**11.70**	11.04
(59.12° ± 39.93°)	(**28.61°** ± 31.46°)	(31.84° ± 33.95°)
10	7.50	**7.78**
(37.81° ± 36.74°)	(**16.90°** ± 23.05°)	(20.87° ± 28.31°)
20	4.49	**4.89**
(20.22° ± 28.11°)	(**10.03°** ± 15.80°)	(12.40° ± 21.25°)

(b) tube and torus

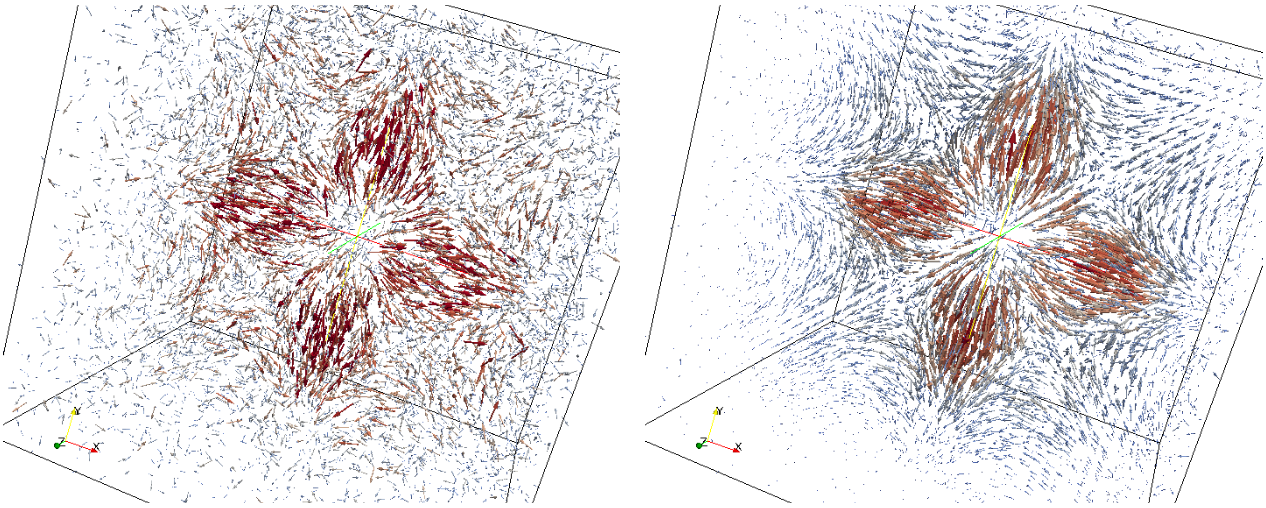
input SNR [dB] / angular error [deg.]	SNR improvement [dB] / angular error (mean ± stdev) [deg.] L_1	L_2
0	**8.03**	6.37
(12.11° ± 7.29°)	(5.97° ± 4.04°)	(**5.95°** ± 3.93°)
10	**7.96**	2.55
(3.82° ± 2.21°)	(**2.58°** ± 1.82°)	(3.16° ± 2.42°)
20	**6.67**	0.51
(1.21° ± 0.70°)	(**0.99°** ± 0.71°)	(1.25° ± 0.85°)

Table III: Optimal λ_C, λ_D pairs used to obtain the results of Table II.

input SNR [dB]	λ_C, λ_D gradient field L_1	gradient field L_2	tube and torus L_1	tube and torus L_2
0	0.5113, 0.7156	2.4355, 1.8028	0.3353, 1.4690	0.5364, 6.8902
10	0.2414, 0.4739	0.9549, 0.6777	0.0605, 0.0470	0.0958, 0.8582
20	0.0092, 0.0052	0.3848, 0.2434	0.0229, 0.0171	0.0122, 0.0587

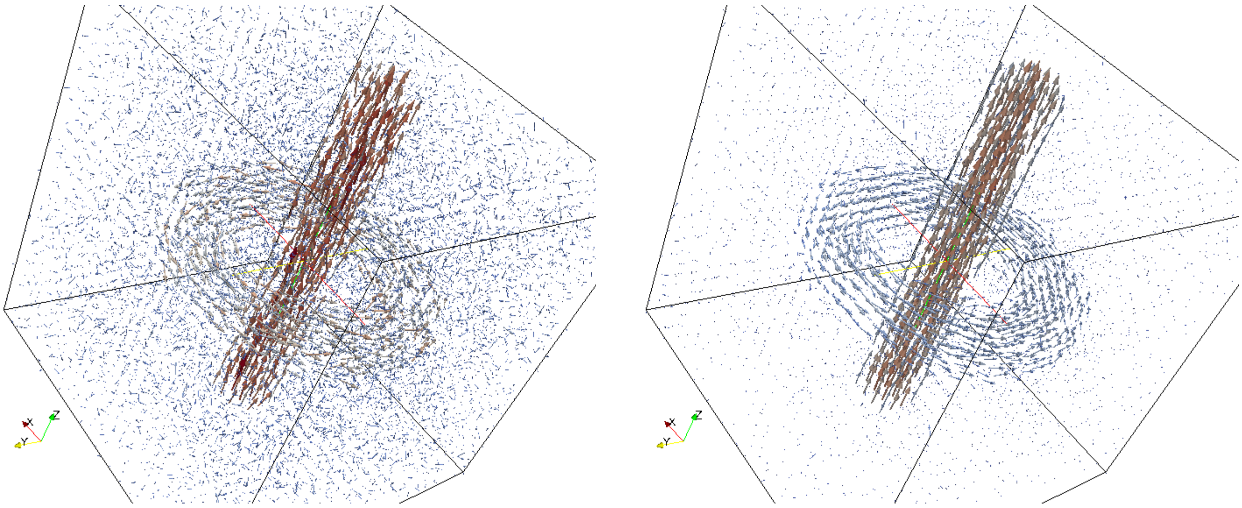
VII. Conclusion

In this paper we studied the question of designing regularization functionals for variational reconstruction of



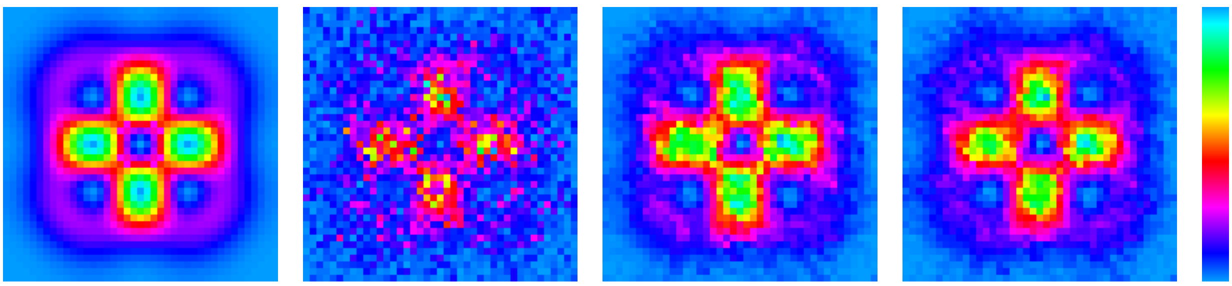
(a) Noisy field (0 dB SNR) (b) Denoised field, using L_1 regularization (11.70 dB SNR)

Figure 1: 'Gradient' phantom; see text for a description of the experiment.



(a) Noisy field (0 dB SNR) (b) Denoised field, using L_1 regularization (9.01 dB SNR)

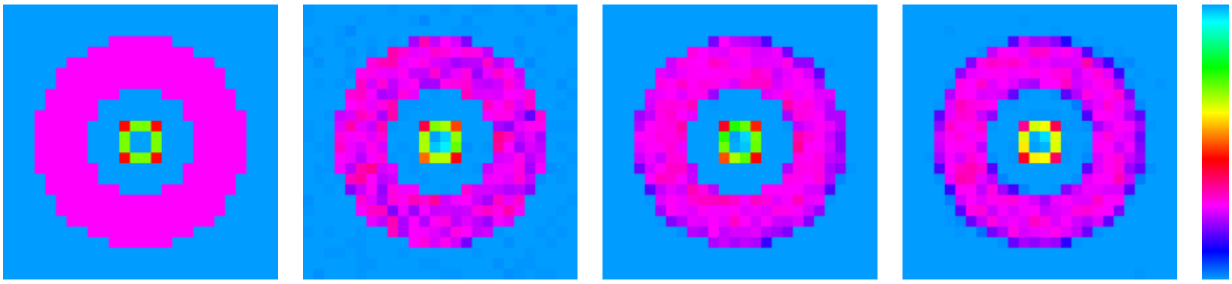
Figure 2: 'Tube and torus' phantom; see text for a description of the experiment.



(a) Original (b) Noisy (0 dB SNR) (c) L_1 denoised (11.70 dB improvement) (d) L_2 denoised (11.04 dB improvement)

Figure 3: Amplitude cross-sections, 'gradient' phantom, comparing L_1 and L_2 denoising.

vector fields. We approached this problem on the basis of requiring that the regularization functional satisfy certain geometric invariance properties, which we justified from different angles. To set the stage for our derivations, we first addressed some commonalities of invariant regularization in scalar and vector settings—followed by a derivation of the general form of invariant regularizers for scalar fields—before specializing to the problem of invariant vector regularization. The vector regularization functionals we derived consist of combinations of (possibly fractional) curl-



(a) Original (b) Noisy (10 dB SNR) (c) L_1 denoised (7.96 dB improvement) (d) L_2 denoised (2.55 dB improvement)

Figure 4: Amplitude cross-sections, 'tube and torus' phantom, comparing L_1 and L_2 denoising.

and divergence-like operators and their adjoints, wrapped in scalar, vector, and/or matrix L_p norms (also introduced in the paper). Generalized vector Laplacians of Tafti and Unser [20] also fall within this framework.

Finally, we presented an application of the proposed framework to the problem of vector field denoising in 2D and 3D, where we gave a natural generalization of L_2 (quadratic), as well as L_1 (TV-type) regularization for vector fields. While our preliminary results already show a systematic advantage of L_1- over L_2-regularization in the aforementioned problem, many interesting questions about the choice of higher-order regularization functionals for data with specific structure remain open. Also, in addition to vector denoising, the proposed construction can find applications in a variety of other problems which we did not study here. Examples include reconstruction of vector fields from non-uniform and incomplete (scalar) measurements, deconvolution, estimation of optical flow, and image registration. Another possible direction for future investigations is the incorporation of temporal regularization in the formulation.

Appendix

A. Matrix L_p norms and spaces

The vector L_p norms defined in (7) are special cases of the L_p norms for matrix-valued functions $\mathbb{R}^d \to \mathbb{R}^{n\times m}$ that we define below (ρ denotes the spectral radius).

$$\|f\|_p := \begin{cases} \left(\int_{\mathbb{R}^d} \mathrm{Tr}\left([f(x)^{\mathrm{H}} f(x)]^{p/2}\right) \mathrm{d}x\right)^{1/p}; & p < \infty, \\ \|\rho(f^{\mathrm{H}} f)^{1/2}\|_\infty. & p = \infty \end{cases}$$

This definition is motivated by a matrix Young inequality due to Ando [28]:

$$\mathrm{Tr}\left[(Y^{\mathrm{H}} X^{\mathrm{H}} X Y)^{1/2}\right] \le \tfrac{1}{p}\mathrm{Tr}\left[(X^{\mathrm{H}} X)^{p/2}\right] + \tfrac{1}{p'}\mathrm{Tr}\left[(Y^{\mathrm{H}} Y)^{p'/2}\right]$$

where $1/p + 1/p' = 1$. The preceding inequality can be used to prove a version of Hölder's inequality for L_p spaces of matrix-valued functions. Matrix L_p spaces are then defined in the standard manner. They have similar properties to scalar L_p spaces (completeness, inner product structure for $p = 2$, duality between L_p and $L_{p'}$ with $1/p + 1/p' = 1$ via the bilinear form $\langle f, g\rangle := \int_{\mathbb{R}^d} f^{\mathrm{T}} g$, *etc.*).

Definitions of matrix ℓ_q norms and spaces are obtained by replacing the integrals with sums.

B. Proof of Theorem 2

By (18) (this is the only place in this proof where we use invariance to improper rotations):

$$\hat{R}(-\xi) = (-I)\hat{R}(\xi)(-I) = \hat{R}(\xi) \quad . \tag{31}$$

Next, fix $\xi \neq 0$ and let $\xi_i^\perp$, $1 \le i \le d-1$, be $d-1$ pairwise orthogonal vectors in $\mathbb{R}^d$ all perpendicular to ξ and with $|\xi_i^\perp| = |\xi|$. We define the rotation matrices

$$\omega_i = I - 2\xi\xi^{\mathrm{T}}/|\xi|^2 - 2\xi_i^\perp(\xi_i^\perp)^{\mathrm{T}}/|\xi|^2, \qquad 1 \le i \le d-1.$$

Each ω_i is a simple rotation by 180° in the $\xi \wedge \xi_i^\perp$ plane. In particular, $\omega_i \xi = -\xi$. We also define, for each pair $i \neq j$, the 90° rotation matrix

$$\omega_{ij} = I - \frac{\xi_i^\perp(\xi_i^\perp)^{\mathrm{T}}}{|\xi|^2} - \frac{\xi_j^\perp(\xi_j^\perp)^{\mathrm{T}}}{|\xi|^2} + \frac{\xi_j^\perp(\xi_i^\perp)^{\mathrm{T}}}{|\xi|^2} - \frac{\xi_i^\perp(\xi_j^\perp)^{\mathrm{T}}}{|\xi|^2}.$$

ω_{ij} maps $\xi_i^\perp \mapsto \xi_j^\perp \mapsto -\xi_i^\perp$ and leaves ξ fixed (in this proof, ω_{ij} and $\xi_i^\perp$ denote, respectively, entire matrices and vectors and not the entries of some unspecified matrix ω or vector $\xi^\perp$).

Note that the matrices ω_i, $1 \le i \le d-1$, commute pairwise; also, by (31) and (18),

$$\hat{R}(\xi)\omega_i = \hat{R}(-\xi)\omega_i = \hat{R}(\omega_i\xi)\omega_i = \omega_i\hat{R}(\xi),$$

which shows that the ω_is commute with $\hat{R}(\xi)$ as well. Since, for $d > 2$, the vectors $\xi, \xi_1^\perp, \ldots, \xi_{d-1}^\perp$ are precisely the common eigenvectors of $\omega_1, \ldots, \omega_{d-1}$, they must also be eigenvectors of $\hat{R}(\xi)$, in particular ξ is an eigenvector of $\hat{R}(\xi)$. Denote its corresponding eigenvalue by $\mu_1 = \mu_1(\xi)$. By taking the transpose of (18) and applying the same argument, we can show that ξ is also an eigenvector of $\hat{R}(\xi)^{\mathrm{T}}$. Its corresponding eigenvalue, temporarily denoted as $\mu_1'(\xi)$, is equal to $\mu_1(\xi)$ since

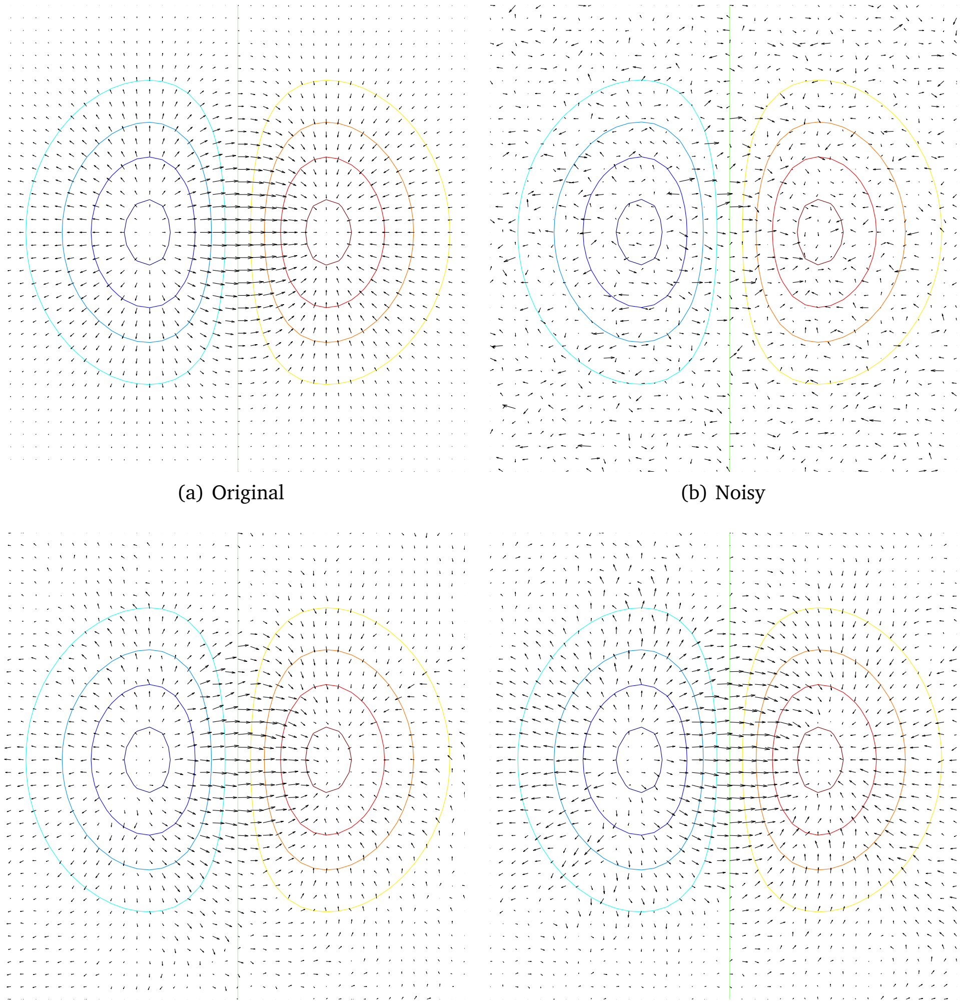
$$\mu_1'\xi^{\mathrm{T}}\xi = \xi^{\mathrm{T}}\hat{R}(\xi)\xi = \mu_1\xi^{\mathrm{T}}\xi.$$

We similarly denote the eigenvalue of $\xi_i^\perp$ by $\mu_{2,i}(\xi)$.

Alternatively, to find the eigenvectors of $\hat{R}(\xi)$ we might note that $\hat{R}(\xi)$ commutes with all ω_{ij}s:

$$\hat{R}(\xi)\omega_{ij} = \hat{R}(\omega_{ij}\xi)\omega_{ij} = \omega_{ij}\hat{R}(\xi),$$

and since ξ is an eigenvector of all ω_{ij}s with eigenvalue 1 (it is their only common eigenvector), $\hat{R}(\xi)\xi$ must be a common eigenvector of all ω_{ij}s, thus $\hat{R}(\xi)\xi = \mu_1(\xi)\xi$ for



(a) Original

(b) Noisy

(c) L_1 denoising output (12.74 dB SNR improvement)

(d) L_2 denoising output (12.58 dB SNR improvement)

Figure 5: Denoising in 2D with L_1 and L_2 regularization applied to the noisy gradient of ϕ_{2D} (see text) with contour lines of ϕ_{2D} overlaid in colour.

some scalar eigenvalue $\mu_1(\xi)$. Then, to show that the $\xi_i^\perp$s are also eigenvectors of $\hat{R}(\xi)$, we observe that

$$\hat{R}(\xi)\xi_i^\perp = \hat{R}(-\xi)\xi_i^\perp = \hat{R}(\omega_i\xi)\xi_i^\perp = \omega_i\hat{R}(\xi)\omega_i^{-1}\xi_i^\perp = \omega_i\hat{R}(\xi)(-\xi_i^\perp) = -\omega_i\hat{R}(\xi)\xi_i^\perp,$$

whereby, $(I+\omega_i)\hat{R}(\xi)\xi_i^\perp = 0$. This shows that $\hat{R}(\xi)\xi_i^\perp$ lies in the kernel of $I+\omega_i$. But the kernel of $I+\omega_i$ corresponds exactly to the span of $\{\xi,\xi_i^\perp\}$. We can therefore write $\hat{R}(\xi)\xi_i^\perp = \alpha_i\xi + \mu_{2,i}\xi_i^\perp$ for some $\alpha_i,\mu_{2,i}$. But then $\xi^{\mathrm{T}}\hat{R}(\xi)\xi_i^\perp = \alpha_i|\xi|^2$; we also have $\xi^{\mathrm{T}}\hat{R}(\xi)\xi_i^\perp = \mu_1\xi^{\mathrm{T}}\xi_i^\perp = 0$. The last two equations show that $\alpha_i = 0$, that is, we have $\hat{R}(\xi)\xi_i^\perp = \mu_{2,i}(\xi)\xi_i^\perp$. $\hat{R}(\xi)$ therefore has $\xi,\xi_1^\perp,\ldots,\xi_{d-1}^\perp$ as d eigenvectors with respective eigenvalues $\mu_1(\xi),\mu_{2,1}(\xi),\ldots,\mu_{2,d-1}(\xi)$.

Next, we show that all $\mu_{2,i}$s are equal to some $\mu_2 = \mu_2(\xi)$: by (18),

$$\mu_{2,j}\xi_j^\perp = \hat{R}(\xi)\xi_j^\perp = \hat{R}(\omega_{ij}\xi)\xi_j^\perp = \omega_{ij}\hat{R}(\xi)\omega_{ij}^{-1}\xi_j^\perp = \omega_{ij}\hat{R}(\xi)\xi_i^\perp = \mu_{2,i}\omega_{ij}\xi_i^\perp = \mu_{2,i}\xi_j^\perp,$$

proving that all $\mu_{2,i}(\xi)$s are equal as claimed. Putting everything together, we find that $\hat{R}(\xi)$ has the orthogonal eigenvectors $\xi,\xi_1^\perp,\ldots,\xi_{d-1}^\perp$ with eigenvalues $\mu_1(\xi)$ for ξ and $\mu_2(\xi)$ for the remaining vectors. Its eigen-decomposition is therefore of the form given in (19) (for $d=2$ we can make a similar demonstration of the above decomposition by working with the reflection matrix with axis ξ instead of the ω_{ij}s). Finally, for (18) and (17) to hold, μ_1,μ_2 must be rotation-invariant and homogeneous of degree γ. Thus, by Theorem 1 we have $\mu_i(\xi) = c_i|\xi|^\gamma$ for some c_i, $i=1,2$. ∎

References

[1] D. Suter, "Motion estimation and vector splines," in *Proc. 1994 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR94)*, June 1994, pp. 939–942.

[2] M. Arigovindan, M. Sühling, C. Jansen, P. Hunziker, and M. Unser, "Full motion and flow field recovery from echo Doppler data," *IEEE Trans. Med. Imaging*, vol. 26, no. 1, pp. 31–45, January 2007.

[3] A. F. Stalder, M. F. Russe, A. Frydrychowicz, J. Bock, J. Hennig, and M. Markl, "Quantitative 2D and 3D phase contrast MRI: Optimized analysis of blood flow and vessel wall parameters," *Magn. Reson. Med.*, vol. 60, no. 5, pp. 1218–1231, 2008.

[4] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, no. 6035, pp. 314–319, Sep. 1985.

[5] L. Alvarez, R. Deriche, T. Papadopoulo, and J. Sánchez, "Symmetrical dense optical flow estimation with occlusions detection," *Int. J. Comput. Vis.*, vol. 75, no. 3, pp. 371–385, dec 2007.

[6] G. Christensen, R. Rabbitt, and M. Miller, "Deformable templates using large deformation kinematics," *IEEE Trans. Image Process.*, vol. 5, no. 10, pp. 1435–1447, oct. 1996.

[7] C. Sánchez Sorzano, P. Thévenaz, and M. Unser, "Elastic registration of biological images using vector-spline regularization," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 4, pp. 652–663, April 2005.

[8] J. Duchon, "Splines minimizing rotation-invariant semi-norms in Sobolev spaces," in *Constructive Theory of Functions of Several Variables*, A. Dold and B. Eckmann, Eds. Springer-Verlag, 1977, pp. 85–100.

[9] D. Suter and F. Chen, "Left ventricular motion reconstruction based on elastic vector splines," *IEEE Trans. Med. Imaging*, vol. 19, no. 4, pp. 295–305, April 2000.

[10] M. Arigovindan, "Variational reconstruction of vector and scalar images from non-uniform samples," Ph.D. dissertation, EPFL, Biomedical Imaging Group, 2005. [Online]. Available: http://library.epfl.ch/theses/?nr=3329

[11] F. Dodu and C. Rabut, "Irrotational or divergence-free interpolation," *Numer. Math.*, vol. 98, no. 3, pp. 477–498, September 2004.

[12] L. Amodei and M. N. Benbourhim, "A vector spline approximation," *J. Approx. Theory*, vol. 67, no. 1, pp. 51–79, 1991.

[13] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenom.*, vol. 60, no. 1–4, pp. 259–268, 1992.

[14] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, mar 1992.

[15] M. A. T. Figueiredo, J. B. Dias, J. P. Oliveira, and R. D. Nowak, "On total variation denoising: A new majorization-minimization algorithm and an experimental comparison with wavalet denoising," in *Proc. 2006 IEEE Int. Conf. Image Process.*, Atlanta, GA, Oct. 2006, pp. 2633–2636.

[16] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak, "Majorization-minimization algorithms for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2980–2991, 2007.

[17] U. Grenander, *Abstract inference*. New York: John Wiley & Sons Inc., 1981.

[18] I. M. Gel'fand and G. E. Shilov, *Generalized Functions*. Academic Press, 1964, vol. I: Properties and Operations.

[19] J.-Y. Na and S.-Y. Chung, "Homogeneous generalized functions which are rotation invariant," *J. Math. Kyoto Univ.*, vol. 40, no. 1, pp. 155–163, 2000.

[20] P. D. Tafti and M. Unser, "Fractional Brownian vector fields," *SIAM Multiscale Model. Simul.*, vol. 8, no. 5, pp. 1645–1670, 2010.

[21] A. W. McDavid and C. D. McMullen, "Generalizing cross products and Maxwell's equations to universal extra dimensions," 2007. [Online]. Available: http://arxiv.org/abs/hep-ph/0609260

[22] M. Unser and D. Van De Ville, "Wavelet steerability and the higher-order Riesz transform," *IEEE Trans. Image Process.*, vol. 19, no. 3, pp. 636–652, March 2010.

[23] J. Yuan, C. Schnörr, and E. Mémin, "Discrete orthogonal decomposition and variational fluid flow estimation," *J. Math. Imaging Vision*, vol. 28, no. 1, pp. 67–80, 2007.

[24] S. Ramani, T. Blu, and M. Unser, "Monte-Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1540–1554, 2008.

[25] D. A. Girard, "The fast Monte-Carlo cross-validation and C_L procedures: comments, new results and application to image recovery problems," *Comput. Statist.*, vol. 10, no. 3, pp. 205–258, 1995.

[26] A. H. Squillacote, *The ParaView Guide: A Parallel Visualization Application*, 3rd ed. Kitware, Inc., 2008.

[27] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, pp. 43–77, 1994.

[28] T. Ando, "Matrix Young inequalities," in *Operator theory in function spaces and Banach lattices*. Basel: Birkhäuser, 1995, pp. 33–38.