Total Variation Denoising of Probability Measures Using Iterated Function Systems With Probabilities

Davide La Torre[1,2,a)], Franklin Mendivil[3,b)] and Edward R. Vrscay[4,c)]

[1]*Department of Economics, Management and Quantitative Methods, University of Milan, Milan, Italy.*
[2]*Department of Applied Mathematics and Sciences, Khalifa University, Abu Dhabi, UAE.*
[3]*Department of Mathematics and Statistics, Acadua University, Wolfville, Canada.*
[4]*Department of Applied Mathematics, University of Waterloo, Waterloo, Canada.*

[a)]Corresponding author: davide.latorre@unimi.it
[b)]franklin.mendivil@acadiau.ca
[c)]ervrscay@uwaterloo.ca

Abstract. In this paper we present a total variation denoising problem for probability measures using the set of fixed point probability measures of iterated function systems with probabilities IFSP. By means of the Collage Theorem for contraction mappings, we provide an upper bound for this problem that can be solved by determining a set of probabilities.

Introduction

In image analysis, the notion of total variation (TV) or total variation regularization has applications in noise removal. The basic idea relies on the fact that that signals with spurious detail have high total variation or, more mathematically, the integral of the absolute gradient of the signal is high. It is well known that the process of reducing the total variation of the signal removes unwanted detail whilst preserving important details such as edges (see [11]). The total variation (TV) of a differentiable greyscale image $f : X \subset \mathbb{R}^n \to \mathbb{R}$ is defined as follows,

$$\|f\|_{TV} = \int_X \|\nabla f(x)\|_2 \, dx, \tag{1}$$

that is, the integral of the $\|\cdot\|_2$ norm of the gradient. Other definitions of total variation are available in the literature – the reader is referred to [4] for an overview of many of the most recently used ones.

A typical TV-based denoising problem will have the following form: Given a noisy image (function) f^*, solve the following optimization problem,

$$\min_{f \in \mathcal{F}} \left[\|f - f^*\| + \lambda \, \|f\|_{TV} \right],$$

where $\mathcal{F}$ denotes an approprate space of functions representing the images. The first term in the objective function is the the so-called *data fitting* term, which imposes the condition that the denoised image f should be close to the noisy data f^*. (Usually, the L^2 norm is employed.) The second term is the TV regularization term – higher values of the regularization parameter $\lambda > 0$ will, in general, yield solutions $f(\lambda)$ with lower TV.

In this paper, we examine the idea of TV-based denoising applied to *probability measures* instead of functions. The motivation comes from ongoing work which suggests that measure-valued approaches may be quite appropriate in the study of *diffusion spectral imaging* (DSI), a particular variation of diffusion magnetic resonance imaging (dMRI). In [7], we examined the following formulation of total variation denoising for measure-valued images: Given a noisy image measure μ^* (the "observed data"), find a solution to the following optimization problem,

$$\min_{\mu \in \mathcal{Y}} d_Y(\mu^*, \mu) + \lambda \|\mu\|_{TV} \, . \tag{2}$$

Here, we examine a variation of this denoising problem for probability measures which employs iterated function systems with probabilities (IFSP). (The natural connection between IFSP and probability measures [5] will be briefly reviewed below.) In [8], we proposed a TV-based denoising method for functions using iterated function systems on functions (IFSM). As such, this paper may be viewed as a kind of measure-based analog of [8].

Metrics on probability measures

In what follows we let (X, d) be a compact subset of $\mathbb{R}^p$ (tipically $X = [0, 1]^p$) and $\mathbb{B}$ be the Borel σ-algebra defined on X. In addition, let $\mathcal{M}(X)$ denote the set of Borel probability measures on X. There are many different metrics that can be defined on $\mathcal{M}(X)$, often with the goal of metrizing the weak topology. Here we shall use two of the most commonly employed ones, namely, the *total variation norm* and the *Monge-Kantorovich metric*. We mention at the outset that these two metrics yield different topologies – the topology given by the total variation is stronger than the one given by the Monge-Kantorovich metric.

Total variation norm

Given a finite signed measure μ, as usual we define the *total variation* of μ by

$$\|\mu\|_{TV} = \sup_{A \in \mathcal{B}} |\mu(A)|.$$

It is not difficult to see that this gives a norm on the (Banach) space, $ca(X)$, of all finite signed Borel measures on X (see page 160 of [3]). Moreover, $(ca(X), \|\cdot\|_{TV})$ is the Banach dual space to $(C(X), \|\cdot\|_\infty)$: This goes a long way towards ensuring the importance of $\|\|_{TV}$ as a norm on measures.

The induced total variation distance is given as $d_{TV}(\mu, \nu) = \|\mu - \nu\|_{TV}$. Since $\mathcal{M}(X) \subset ca(X)$ is d_{TV}-closed, it is also complete under d_{TV}.

If a measure μ is absolutely continuous with density f, then it is not hard to see that

$$\|\mu\|_{TV} = \int_X |f(x)|\, dx. \tag{3}$$

Monge-Kantorovich metric

The Monge-Kantorovich metric came out of considerations in the area of mass transportation problems [12, 5]. In our setting, it metrizes the weak* topology on the space of probability measures (weak* when $ca(X)$ is viewed as the dual space to $C(X)$).

Definition 1. *The* Monge-Kantorovich distance *on $\mathcal{M}(X)$ is defined as follows:*

$$d_{MK}(\mu, \nu) = \sup_{\|f\|_{Lip} \le 1} \left[\int_X f d\mu - \int_X f d\nu \right]$$

where $\|f\|_{Lip}$ is the Lipschitz constant of a function $f : X \to \mathbb{R}$.

It is simple to show that this definition gives a pseudo-metric on $\mathcal{M}(X)$ and a bit harder to show that it gives a metric. Since d_{MK} gives the weak* topology on $\mathcal{M}(X)$, this automatically implies that the topology it defines is weaker than that defined by d_{TV}. Since $\mathcal{M}(X)$ is weak* compact in $ca(X)$, it is compact (and thus complete) under the d_{MK} metric.

In the special case where $X \subset \mathbb{R}$, then it is known that (see [1, 2])

$$d_{MK}(\mu, \nu) = \int_X |F_\mu(x) - F_\nu(x)|\, dx, \tag{4}$$

where F_μ (respectively F_ν) is the CDF of μ (respectively ν).

IFS Markov Operator on $\mathcal{M}(X)$

Given a set $\mathbf{W}$ of N IFS contraction maps $w_i : X \to X$, an $(N+1)$-vector of probabilities $\mathbf{p} = (p_0, ..., p_N)$, $\sum_{i=0}^{N} p_i = 1$, and a probability measure $\mathbf{s} \in \mathcal{M}$, we construct the following IFSP Markov operator with condensation (see [5]),

$$M_{\mathbf{W},\mathbf{p},\mathbf{s}}\, \mu = \sum_{i=1}^{N} p_i \mu \cdot w_i^{-1} + p_0 \mathbf{s}.$$

If $p_0 = 0$, then the above definition collapses to the usual definition of an IFSP Markov operator. The action of $M_{\mathbf{W},\mathbf{p},\mathbf{s}}\, \mu$ on a set $A \in \mathcal{F}$ is then defined as

$$M_{\mathbf{W},\mathbf{p},\mathbf{s}}\, \mu(A) = \sum_{i=1}^{N} p_i \mu(w_i^{-1}(A)) + p_0 \mathbf{s}(A).$$

Proposition 1. *We have* $\|M_{\mathbf{W},\mathbf{p},\mathbf{s}}(\mu) - M_{\mathbf{W},\mathbf{p},\mathbf{s}}(\nu)\|_{TV} \leq \left(\sum_{i=1}^{N} p_i\right) \|\mu - \nu\|_{TV} = (1 - p_0)\|\mu - \nu\|_{TV}$.

Proof. Since $\mu(w_i^{-1}(w_i(A))) = \mu(A)$, it is easy to see that $\|\mu\|_{TV} = \|\mu \circ w_i^{-1}\|_{TV}$ for all i. Thus we have

$$\begin{aligned} \|M_{\mathbf{W},\mathbf{p},\mathbf{s}}(\mu) - M_{\mathbf{W},\mathbf{p},\mathbf{s}}(\nu)\|_{TV} &\leq \left\| \sum_{i=1}^{n} p_i (\mu \circ w_i^{-1} - \nu \circ w_i^{-1}) \right\| \\ &\leq \sum_{i=1}^{n} p_i \|\mu \circ w_i^{-1} - \nu \circ w_i^{-1}\|_{TV} \\ &= \left(\sum_{i=1}^{n} p_i \right) \|\mu - \nu\|_{TV}, \end{aligned}$$

as claimed. □ □

Corollary 2. *If $p_0 > 0$ then the Markov operator with condensation $M_{\mathbf{W},\mathbf{p},\mathbf{s}}$ has a unique fixed point $\bar{\mu} \in \mathcal{M}$.*

The operator $M_{\mathbf{W},\mathbf{p},\mathbf{s}}$ is also contractive with respect to the Monge-Kantorovich distance This is easy to show using the same arguments as the standard Markov operator without condensation (see [5, Theorem 2.60]). We let $0 \leq c_{MK} < 1$ denote the contraction factor of $M_{\mathbf{W},\mathbf{p},\mathbf{s}}$ with respect to the d_{MK} metric.

Proposition 3. *The Markov operator with condensation $M_{\mathbf{W},\mathbf{p},\mathbf{s}}$ is contractive with respect to the Monge-Kantorovich distance with*

$$d_{MK}(M_{\mathbf{W},\mathbf{p},\mathbf{s}}\mu, M_{\mathbf{W},\mathbf{p},\mathbf{s}}\nu) \leq (\max_i c_i) d_{MK}(\mu, \nu),$$

where c_i is the contractivity of w_i.

We now investigate an inverse problem involving the fixed point, μ, of $M_{\mathbf{W},\mathbf{p},\mathbf{s}}$. In these situations it is difficult to obtain estimates like $\|\mu - \nu\|_{TV}$ or $d_{MK}(\mu, \nu)$, where ν is some fixed target measure. A standard technique to avoid this difficulty is to use the *Collage Theorem*. This theorem is a simple but very useful consequence of Banach's contraction principle (see [5, Theorem 2.6]).

Theorem 4 (Collage Theorem)**.** *Let $(\mathbb{X}, d)$ be a complete metric space and $f : \mathbb{X} \to \mathbb{X}$ be a contraction with contractivity $c < 1$ and $\bar{x}$ be its unique fixed point. Then, for any $y \in \mathbb{X}$, we have*

$$d(y, \bar{x}) \leq \frac{d(y, f(y))}{1 - c}.$$

The benefit is that we replace the difficult (or impossible) $d(y, \bar{x})$ distance with the simpler $d(y, f(y))$ distance.

The Total Variation Denoising Problem

Given a set $\mathbf{W}$ of N IFS contraction maps $w_i : X \to X$, an $(N+1)$-vector of probabilities $\mathbf{p} = (p_0, ..., p_N)$, $\sum_{i=0}^{N} p_i = 1$, and a probability measure $\mathbf{s} \in \mathcal{M}$, define the feasible set,

$$\mathcal{F} = \left\{ \mu \in \mathcal{M} : M_{\mathbf{W},\mathbf{p},\mathbf{s}}\, \mu = \mu \right\} ,$$

that is, the set of all possible fixed points of $M_{\mathbf{W},\mathbf{p},\mathbf{s}}$ for a given choice of the triple $(\mathbf{W}, \mathbf{p}, \mathbf{s})$.

Now given a noisy measure μ^*, and an ideal de-noised measure ν, the Total Variation (TV) denoising problem for probability measures consists of finding the solution to the constrained minimization problem

$$\min_{\mu \in \mathcal{F}} \left[d_{MK}(\mu^*, \mu) + \xi d_{TV}(\mu, \nu) \right] ,$$

subject to the constraint

$$M_{\mathbf{W},\mathbf{p},\mathbf{s}}\, \mu = \mu ,$$

where $\xi > 0$ is a trade-off parameter. From the Collage Theorem, for a fixed condensation measure $\mathbf{s}$, and a fixed system of maps $\mathbf{W}$, an upper bound for the above problem can be found by solving the problem,

$$\min_{\mathbf{p}} \frac{1}{1 - c_{MK}} \left[d_{MK}(\mu^*, M_{\mathbf{W},\mathbf{p},\mathbf{s}}\mu^*) \right] + \frac{\xi}{1 - c_{TV}} \left[d_{TV}(M_{\mathbf{W},\mathbf{p},\mathbf{s}}\nu, \nu) \right] , \quad (5)$$

where $c_{MK} = \max_{i=1...N} c_i$ is the contractivity factor of the operator $M_{\mathbf{W},\mathbf{p},\mathbf{s}}$ w.r.t. the Monge Kantorovich distance, and $c_{TV} = \sum_{i=1}^{N} p_i$ is the contractivity factor of the operator $M_{\mathbf{W},\mathbf{p},\mathbf{s}}$ w.r.t. the Total Variation distance.

As a simple example, consider $X = [0, 1]$, the IFS contraction maps $w_1(x) = x/2$, $w_2(x) = x/2 + 1/2$, the measure $\mathbf{s}$ with density $f_{\mathbf{s}}$, and the probability $p_0 > 0$ fixed. Then if $\theta \in \mathcal{M}$ has density f_θ, it is easy to see that $M_{\mathbf{W},\mathbf{p},\mathbf{s}}\theta$ has density

$$f_{M_{\mathbf{W},\mathbf{p},\mathbf{s}}\theta} = p_0 f_{\mathbf{s}} + 2p_1 f_\theta \circ w_1^{-1} + 2p_2 f_\theta \circ w_2^{-1}.$$

Furthermore, if θ has CDF F_θ, then $M_{\mathbf{W},\mathbf{p},\mathbf{s}}\theta$ has CDF given by

$$F_{M_{\mathbf{W},\mathbf{p},\mathbf{s}}\theta}(x) = \begin{cases} p_0 F_{\mathbf{s}}(x) + p_1 F_\theta(2x), & \text{if } x \le 1/2, \\ p_0 F_{\mathbf{s}}(x) + p_1 + p_2 F_\theta(2x - 1), & \text{if } 1/2 \le x \le 1. \end{cases}$$

With these observations and equations (3) and (4), the minimization problem (5) becomes the non-differentiable but convex problem

$$\min_{p_1 + p_2 = 1 - p_0} 2 \int_0^1 |F_{\mu*}(x) - F_{M_{\mathbf{W},\mathbf{p},\mathbf{s}}\mu*}(x)|\, dx + \frac{\xi}{p_0} \int_0^1 |f_\nu(x) - f_{M_{\mathbf{W},\mathbf{p},\mathbf{s}}\nu}(x)|\, dx.$$

(Notice that the contraction factor for $M_{\mathbf{W},\mathbf{p},\mathbf{s}}$ in the Monge-Kantorovich metric is $1/2$ and the contraction factor for $M_{\mathbf{W},\mathbf{p},\mathbf{s}}$ in the total variation norm is $1 - p_0$.)

Conclusions

In this paper we have presented an innovative approach to deal with the problem of total variational denoising of probability measures using iterated function systems with probabilities. It has been shown that this problem can be approximated by means of the Collage Theorem and then solved by determining the set of probabilities p_i. This is a first attempt at solving this problem: future works will include a computational analysis of the above optimization model with different ideal de-noised measures.

ACKNOWLEDGMENTS

This work has been supported in part by Discovery Grants 238549-2012 (FM) and 106270-2012 (ERV) from the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] Brandt, J., Cabrelli, C.A., Molter, U., (1991), An algorithm for the computation of the Hutchinson distance, *Inform. Process. Lett.*, **40**, 113-117.

[2] Cabrelli, C.A. and Molter, U.M., The Kantorovich metric for probability measures on the circle, *J. Comput. Appl. Math.*, **57**, pp. 345-361, 1995.

[3] Dunford, N. and Schwartz J.T., (1958) *Linear Operators, Part I*, Interscience Publishers, New York.

[4] Goldluecke, B., Strekalovskiy, E., Cremers, D. (2012), The Natural Vectorial Total Variation Which Arises from Geometric Measure Theory, *SIAM J. Imaging Sci.*, 5 (2), pp. 537-563.

[5] Kunze, H., La Torre, D., Mendivil, F., and Vrscay, E.R. (2012), *Fractal-based methods in analysis* Springer Verlag.

[6] La Torre, D., Vrscay, E.R., Ebrahimi, M., Barnsley, M. (2009), Measure-Valued Images, Associated Fractal Transforms, and the Affine Self-Similarity of Images, *SIAM J. Imaging Sci.*, 2 (2), 470-507.

[7] La Torre, D., Mendivil, F., Michailovich, O. and Vrscay, E.R. (2016), Total Variation Minimization for Measure-Valued Images with Diffusion Spectrum Imaging as Motivation, in *Image Analysis and Recognition*, Lecture Notes in Computer Science **9730**, 131-137 (2016).

[8] La Torre, D., Mendivil, F. and Vrscay, E.R. (2016), Iterated Function Systems on Functions of Bounded Variation, *Fractals*, 24 (2) DOI 10.1142/S0218348X16500195.

[9] Michailovich, O., La Torre, D. and Vrscay, E.R., Function-Valued Mappings, Total Variation and Compressed Sensing for Diffusion MRI, in *Image Analysis and Recognition*, Lecture Notes in Computer Science **7325**, 286-295 (2012).

[10] Mendivil, F. (2016), Computing the Monge-Kantorovich distance, to appear in *Comp. Appl. Math.*.

[11] Strong, D., Chan, T., (2003), Edge-preserving and scale-dependent properties of total variation regularization, *Inverse Problems*, 19, 165-187.

[12] Vershik, A.M., (2006), Kantorovich metric: Initial history and little-known applications, *J. Math. Sci.*, 133 (4), 1410-1417.