A new model for image regularization

Student: Cristóbal Guzmán
Supervisor: Arkadi Nemirovski
ARC Fellowship Application

Introduction. Image processing is an active area of reseach with connections to a number of applications, such as medical imaging, machine learning, security, among others. Since the seminal paper [ROF92], image reconstruction based on total variation (TV) has become an extensively used technique for inpainting, deblurring, and denoising images. The model proposes as true image candidate a signal x that minimizes a combination of an ℓ^2 fit of observations with a regularization term, given by the total variation of the image

$$\min_{x:\Omega\to\mathbb{R}} \left\{ \|\mathcal{A}x - b\|_2 + \lambda \mathrm{TV}[x] \right\}. \tag{1}$$

Here Ω is a discretized bounded domain in $\mathbb{R}^2$, which in most applications is simply an $m \times n$ grid (from now on we restrict to this case) that discretizes a rectangular domain; $\mathcal{A}$ is a blurring operator, i.e. a matrix representation of a discretized convolution with some kernel*, and physically it represents the inaccuracy of measurements by spatial perturbations on the observations; b are the actual measurements; and λ is a nonnegative penalization parameter. Finally, the TV can be measured by different metrics, being the most popular ones the anisotropic, which is the sum of the ℓ^1 norms of the discrete partial derivatives

$$\sum_{1\le i<m,\,1\le j\le n} |x_{i+1,j} - x_{i,j}| + \sum_{1\le i\le m,\,1\le j<n} |x_{i,j+1} - x_{i,j}|;$$

and the isotropic, which is the mixed $\ell^1 - \ell^2$ norm of the gradient, defined by

$$\sum_{\substack{1\le i<m \\ 1\le j<n}} \sqrt{(x_{i+1,j} - x_{i,j})^2 + (x_{i,j+1} - x_{i,j})^2} + \sum_{1\le i<m} |x_{i+1,n} - x_{i,n}| + \sum_{1\le j<n} |x_{m,j+1} - x_{m,j}|.$$

For these definitions we impose reflexive boundary conditions, although other choices can be made.

Intuitively, the TV tries to preserve sharp discontinuities, penalizing small levels of noise proportionally to their magnitude. This might look similar to the Compressed Sensing framework†; however, images are typically not sparse, but we assume near sparsity of their gradients instead.

State of the art. From a mathematical programming perspective, both the isotropic and anisotropic versions of TV regularization can be posed as second order cone programs (SOCP). For this reason, we are dealing with problems that are polynomial-time solvable, namely by interior point methods.

Nevertheless, for large-scale instances, interior point algorithms become impractical due to time consuming iterations. Even state of the art first order methods‡ show poor performance on this model. The reason why this regularization model is significantly harder than others (e.g. ℓ^1 regularization) is that the presence of the discrete gradient inside the norm gives a complex geometry for the nonsmooth regularizer, which makes projections and proximal mappings expensive to compute§.

*Some practical examples of kernels are Gaussian or uniform densities, supported on a small neighborhood of 0.

†Let me stress, however, that we are not aware of any results on the theoretical performance for TV regularization. The paper [CRT06] provides partial answers, in a somewhat simplified framework.

‡These are iterative methods that only require function values and subgradients to compute their next iterate.

§These operations are crucial for fast first order methods.

Some literature on recent approaches for this problem can be found in [Cha04], [GY04], [ZWC10] and [BT09a], and references therein. This body of work provides practical solvers for the pure denoising problem (i.e. where $\mathcal{A}$ is the identity), with provable complexity bounds. However, the combined deblurring/denoising problem is significantly less understood.

A major difference between these two problems, is that the pure denoising model can be easily dualized, leading to a smooth convex program, for which fast algorithms can be used[¶] (e.g. [BT09b], [Nes07]). In the deblurring/denoising case, just constructing the dual requires to invert the operator $\mathcal{A}^*\mathcal{A}$ which is for all practical purposes impossible, given the typical ill-conditioning of $\mathcal{A}$ (although heuristics exist [BT09a]).

Proposal. We propose to replace in (1) the TV norm[‖] in the role of regularizer by other norms, better suited for first order methods, and hopefully leading to image recovery routines of comparable quality. One option here is to parameterize a zero mean image x by its periodic Laplacian: $x = \Psi(\Delta x)$ and replace (1) by the problem

$$\min_{f\in E}\left\{\|\mathcal{A}\Psi(f)-b\|_2^2+\lambda\|f\|_1\right\}, \tag{2}$$

where $E := \{[g_{ij}]_{1\le i\le m, 1\le j\le n} : \sum_{i,j} g_{ij} = 0\}$. Another option is to extend the natural one-to-one parameterization $x = \Phi_*([\nabla_i x, \nabla_j x])$ of a zero mean image x by its gradient field (which is a *potential* vector field on the grid Ω) to the (redundant) linear parameterization $x = \Phi([g,h])$ of x by an *arbitrary* vector field $[g,h]$ on the grid, and instead of (1) consider the problem

$$\min_{[g,h]}\left\{\|\mathcal{A}\Phi([g,h])-b\|_2^2+\lambda\|[g,h]\|_1\right\}. \tag{3}$$

We have discovered a specific extension Φ of Φ_* such that

$$\forall x\in E: \frac{1}{2}\min_{[g,h]:\Phi([g,h])=x}\|[g,h]\|_1 \le \mathrm{TV}(x) \le O(1)\ln(m+n)\min_{[g,h]:\Phi([g,h])=x}\|[g,h]\|_1, \tag{4}$$

meaning that the norms of a zero mean image induced by the regularizers in (1) and (3) are within a quite moderate factor, which, we hope, will imply comparable performance of the associated image recovery techniques. Besides, it turns out that in (2) and (3) the proximal mappings are incomparably cheaper to compute than in (1), and Ψ, Φ happen to be simple (computable in nearly linear time); thus, (2) and (3) are much better suited for first order minimization.

To prove (4) we used basic techniques from discrete Fourier analysis, and a classical estimate of fundamental solutions of the Laplacian on a grid, compared to its continuous analogue (see [Man66]). We believe this new bound, and its proof, might be of independent interest, given the connection of these quantities to Spectral Graph theory, and the analysis of Markov Chain algorithms over graphs.

At this stage we are interested on conducting extensive computational experiments to compare these two models with TV. For this we plan to use standard datasets from the literature within the reach of interior point solvers, trying not only different regularizers, but also different formulations, such as ℓ^2 fit minimization with regularizer bounds, regularizer minimization with ℓ^2 bounds on the fit, or even data fitting with other norms than ℓ^2.

As a final stage of the project, we are interested on the complexity of first order algorithms for these models. Latest developments in the field give fast convergence results (see [BT09b] and [Nes07]). We hope that the simpler geometry of our problems could lead to significant improvements on the convergence rate, and to new image regularization models with better complexity bounds.

[¶] However, these smooth programs are defined in simple, but large domains, which affects the complexity bounds.

[‖] In this section we restrict to the anisotropic case, since the isotropic one is within a $\Theta(1)$ factor.

References

[BT09a] A. Beck and M. Teboulle, *Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems*, IEEE Transactions on Image Processing **18** (2009), no. 11, 2419–2434.

[BT09b] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Img. Sci. **2** (2009), no. 1, 183–202.

[Cha04] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vis. **20** (2004), no. 1-2, 89–97.

[CRT06] E. J. Candes, J. Romberg, and T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Transactions on Information Theory **52** (2006), no. 2, 489–509.

[GY04] D. Goldfarb and W. Yin, *Second-order cone programming methods for total variation-based image restoration*, SIAM J. Sci. Comput **27** (2004), 622–645.

[Man66] M. Mangad, *Bounds for the two-dimensional discrete harmonic Green's function*, Mathematics of Computation **20** (1966), no. 93, 60–67.

[Nes07] Y. Nesterov, *Gradient methods for minimizing composite objective function*, CORE Discussion Paper (2007), no. 76.

[ROF92] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D **60** (1992), no. 1-4, 259–268.

[ZWC10] M. Zhu, S. J. Wright, and F. Tony Chan, *Duality-based algorithms for total-variation-regularized image restoration*, Comput. Optim. Appl. **47** (2010), no. 3, 377–400.