HAL
open science

Dual constrained TV-based regularization

Camille Couprie, Hugues Talbot, Jean-Christophe Pesquet, Laurent Najman, Leo Grady

To cite this version:

Camille Couprie, Hugues Talbot, Jean-Christophe Pesquet, Laurent Najman, Leo Grady. Dual constrained TV-based regularization. Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, May 2011, Prague, Czech Republic. pp.945 - 948, 10.1109/ICASSP.2011.5946561 . hal-00744071

HAL Id: hal-00744071

https://hal.science/hal-00744071

Submitted on 22 Oct 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DUAL CONSTRAINED TV-BASED REGULARIZATION

Camille Couprie, Hugues Talbot
Jean-Christophe Pesquet, Laurent Najman

Université Paris-Est, Laboratoire
d'Informatique Gaspard-Monge,
Champs sur Marne, France

Leo Grady

Siemens Corporate Research
Dpt. Imaging Analytics and Informatics
Princeton, N.J. USA

ABSTRACT

Algorithms based on the minimization of the Total Variation are prevalent in computer vision. They are used in a variety of applications such as image denoising, compressive sensing and inverse problems in general. In this work, we extend the TV dual framework that includes Chambolle's and Gilboa-Osher's projection algorithms for TV minimization in a flexible graph data representation by generalizing the constraint on the projection variable. We show how this new formulation of the TV problem may be solved by means of a fast parallel proximal algorithm, which performs better than the classical TV approach for denoising, and is also applicable to inverse problems such as image deblurring.

***Index Terms*—** Proximal algorithm, inverse problems, image denoising, convex optimization, image restoration

1. INTRODUCTION

The Total Variation (TV) model was introduced as a regularizing criterion for image denoising by Rudin, Osher and Fatemi [9], and has been shown to be quite efficient for smoothing images while preserving contours. Moreover, a major advantage is that TV minimization is a convex problem, although coping with the non-differentiability of the involved objective function required the development of specific optimization techniques, particularly with respect to speed and efficiency. Much progress has been achieved by employing primal-dual approaches [4, 2], and more recently by the dual approach of Chambolle [3]. Qualitative improvements have also been obtained by introducing a weighted model [1, 7]. In this model, the discretization of the TV energy is performed by the use of an edge-weighted graph. Gilboa and Osher provided an efficient dual algorithm in [7], based on Chambolle's [3] method to address this problem. Currently, one of the fastest method for optimizing weighted TV is proposed by Zhang et. al [11].

The projection algorithms in [3, 7] are based on a relatively simple local constraint on the norm of the projection variable. In this work, we extend the constraint on this variable, as this allows us to better adapt the optimization procedure to the local information. We name this new approach "Dual-Constrained Total Variation" (DCTV) regularization.

The paper is organized as follows: in Section 2, we generalize TV models on a graph by extending the dual formulation under a constrained form. In Section 3, we demonstrate how our constrained TV-based optimization problem can be efficiently solved by using a parallel proximal algorithm. Finally, results obtained with the proposed approach are presented in three applications and compared to classical weighted TV.

2. GRAPH EXTENSION OF TV MODELS

Given an original corrupted image f, the purpose of variational methods for image restoration is to deduce a restored image u close to the observed image f under the assumption of smooth variations of intensity values inside objects. Let $\lambda \in]0, +\infty[$ be a real positive value representing a regularization parameter. In a continuous setting, given a plane domain Ω, and denoting by x and y two arbitrary points of Ω, the weighted anisotropic TV model [7] is given by

$$\min_u \int_\Omega \Big(\int_\Omega w_{x,y}(u_y - u_x)^2 dy \Big)^{1/2} dx + \frac{1}{2\lambda} \int_\Omega (u_x - f_x)^2 dx, \tag{1}$$

where w is a nonnegative valued function defined on Ω^2. As shown by Chan *et al* in [4] the TV minimization problem (1) is equivalent to the min-max problem

$$\min_u \max_{||p||_\infty \leq 1} \int_{\Omega^2} w_{x,y}^{1/2}(u_y - u_x) p_{x,y} dx dy + \frac{1}{2\lambda} \int_\Omega (u_x - f_x)^2 dx, \tag{2}$$

with p a projection vector field, $\|p\|_\infty = \sup_{x\in\Omega} \Big(\int_\Omega p_{x,y}^2 dy \Big)^{1/2}$.

Let us now establish the formulation of the DCTV energy within a discrete framework, by first specifying our notation. A graph consists of a pair $G = (V, E)$ with vertices $v \in V$ and edges $e \in E \subseteq V \times V$, with cardinalities $n = |V|$ and $m = |E|$. An edge, e, spanning two vertices, v_i and v_j, is denoted by $e_{i,j}$. In this paper we deal with weighted graphs

that include weights on both the edges and nodes. An edge weight is a value assigned to each edge $e_{i,j}$, and is denoted by $w_{i,j}$. We assume that $w_{i,j} \in]0, +\infty[$. The weight of a node v_i is denoted by g_i. We also assume that $g_i \in \mathbb{R}_+^*$. We note A the incidence matrix of the graph which is known to define the discrete calculus analogue of the gradient, while $A^\top$ similarly defines the discrete calculus analogue of the divergence (see [8] and the references therein). The incidence matrix $A \in \mathbb{R}^{m\times n}$ is defined, for every vertex v_k and edge e_{ij}, as $A_{e_{ij},v_k} = 1$ if $i = k$, $A_{e_{ij},v_k} = -1$ if $j = k$, and $A_{e_{ij},v_k} = 0$ otherwise. For any matrix M, we use $|M|$ for the matrix composed of the absolute value of each entry individually, and denote $\cdot$ the Hadamard product. As the step size of the graph tends toward zero, the discrete version of (1) and its dual (2) approximate the continuous versions. Now, f and u are vectors in $\mathbb{R}^n$. The discrete TV model as defined by [1, 7] is

$$\min_u \sum_{i=1}^n \Big(\sum_{j\in N_i} w_{i,j}(u_j - u_i)^2 \Big)^{1/2} + \frac{1}{2\lambda} \sum_{i=1}^n (u_i - f_i)^2 \tag{3}$$

where $N_i = \{j \in \{1, \dots, n\} \mid e_{i,j} \in E\}$ and the dual formulation optimized by a projection algorithm [7] is given by

$$\min_u \max_{\|p\|_\infty \le 1} p^\top((Au) \cdot \sqrt{w}) + \frac{1}{2\lambda}\|u - f\|^2 \tag{4}$$

where $\|\cdot\|$ denotes the Euclidean norm, $w \in \mathbb{R}^m$ is a vector with components $(w_{i,j})_{i,j}$ and denoting by $(p_{i,j})_{i,j}$ the components of $p \in \mathbb{R}^m$, $\|p\|_\infty = \max_{i\in\{1,\dots,n\}} \Big(\sum_{j\in N_i} p_{i,j}^2\Big)^{1/2}$. By introducing a vector $F = (F_{i,j})_{i,j} \in \mathbb{R}^m$ such that, for every $i \in \{1, \dots, n\}$ and $j \in N_i$, $F_{i,j} = p_{i,j}\sqrt{w_{i,j}}$, the problem can be reformulated as

$$\min_u \max_{F\in C} F^\top Au + \frac{1}{2\lambda}\|u - f\|^2, \tag{5}$$

$$C = \{(F_{i,j})_{i,j} \mid (\forall i \in \{1, \dots, n\}) \sum_{j\in N_i} \frac{F_{i,j}^2}{w_{i,j}} \le 1\}. \tag{6}$$

The objective of this work is to extend the discrete weighted TV variational formulation in (5) by investigating the following optimization problem

$$\min_u \max_{F\in C} F^\top Au + \frac{1}{2}(Hu - f)^\top \Lambda^{-1}(Hu - f) \tag{7}$$

where $f \in \mathbb{R}^q$ is an observed vector of data, $H \in \mathbb{R}^{q\times n}$ and Λ is weighting symmetric definite-positive matrix in $\mathbb{R}^{q\times q}$. H may simply be the identity matrix for image denoising; a convolution operator in restoration tasks; or a projection matrix in reconstruction problems. Λ may be a matrix proportional to the covariance matrix of the noise corrupting the data as commonly used in weighted least squares approaches. In addition, the main contribution of this work is to consider more general convex sets C than those given by (6). More precisely, the proposed optimization approach allows us to address nonempty convex sets C which can be decomposed as an intersection of closed convex subsets $(C_r)_{1\le r\le s}$ of $\mathbb{R}^m$, the projections onto which take closed forms. An example of a set C of interest is

$$C = \{F \mid (\forall i \in \{1, \dots, n\}) \ \|\theta_i \cdot F\|_\alpha \le g_i\} \tag{8}$$

where $\|\cdot\|_\alpha$ is the ℓ_α norm of $\mathbb{R}^m$ with $\alpha \in [1, +\infty]$ and, for every $i \in \{1, \dots, n\}$, $\theta_i \in]0, +\infty[^m$ is a vector of multiplicative constants. The form of C in (8) obviously includes (6) as a particular case while offering much more flexibility. In this paper, we are mostly interested in the case when θ_i is the i-th line vector of $|A^\top|$ and the i-th node weight g_i is a decreasing function of the image gradient. Given positive reals ϵ and β, we suggest using

$$g_i = \exp(-\beta\|\nabla\overline{u}_i\|_2) + \epsilon, \tag{9}$$

where $\overline{u}$ is some reference image (wich corresponds to some rough estimate of u) and $\|\nabla\overline{u}_i\|_2$ is the Euclidean norm of its discrete gradient $\nabla\overline{u}_i$ at node i. In addition to intensity information, g_i may be used to penalize changes in other relevant image quantities such as color or texture. In the absence of a contour, g_i takes large values, so are the components of F corresponding to nonzero values of θ_i, preventing large local variations of u in the minimization (7). Conversely, in the presence of a contour, $g_i \simeq \epsilon$, and the components of F corresponding to nonzero values of θ_i are small, thus allowing large local variations of u.

3. PROPOSED ALGORITHM

We show in this section that it is possible to efficiently solve Problem (7) by proximal methods [6]. To do so, define the support function σ_C of the closed convex constraint set C as

$$\sigma_C \colon \mathbb{R}^m \to]-\infty, +\infty] \colon a \mapsto \sup_{F\in C} F^\top a. \tag{10}$$

This is a proper lower-semicontinuous convex function, the conjuguate of which is the indicator function of C,

$$\imath_C \colon F \mapsto \begin{cases} 0 & \text{if } F \in C, \\ +\infty & \text{otherwise.} \end{cases} \tag{11}$$

This leads us to consider the following optimization problem:

$$\min_u \sigma_C(Au) + \frac{1}{2}(Hu - f)^\top \Lambda^{-1}(Hu - f) + \frac{\eta}{2}\|Ku\|^2 \tag{12}$$

where $\eta \in]0, +\infty[$ and $K \in \mathbb{R}^{n\times n}$ is the projection matrix onto the nullspace of H. When H is injective ($\operatorname{rank} H = n$), the last term vanishes and (12) is strictly equivalent to (7). The term $u \mapsto \eta\|Ku\|^2/2$ thus aims at introducing an additional regularization when H is not injective, so that the objective function remains strictly convex. The following holds:

Proposition 3.1 *Problem* (12) *admits a unique solution* $\widehat{u}$. *The dual Fenchel-Rockafellar form of the problem is*

$$\min_F \varphi(F) + \imath_C(F), \tag{13}$$

where $\varphi\colon F \mapsto \frac{1}{2}F^\top A\Gamma A^\top F - F^\top A\Gamma H^\top \Lambda^{-1} f$ *and* $\Gamma = (H^\top \Lambda^{-1} H + \eta K)^{-1}$. *The optimal solution to the primal problem* (12) *is deduced from any optimal solution* $\widehat{F}$ *of the dual problem by the relation*

$$\widehat{u} = \Gamma\left(H^\top \Lambda^{-1} f - A^\top \widehat{F}\right). \tag{14}$$

Note that the dual forward-backward algorithm proposed in [5] is not applicable here since the projection onto C is not explicit. In order to numerically solve (13), recall that $C = \cap_{r=1}^{s} C_r$, so that $\imath_C$ can be decomposed into the sum of the indicator functions of the convex subsets $(C_r)_{1\leq r\leq s}$. Hence, the problem is equivalent to solving

$$\min_F \sum_{r=1}^{s} \imath_{C_r}(F) + \varphi(F). \tag{15}$$

The above sum of $(s+1)$ convex functions can be efficiently optimized by resorting to a parallel proximal algorithm [6]. As shown in Algorithm 1, this requires to compute in parallel projections onto each set C_r with $r \in \{1,\ldots,s\}$, which are defined $\forall F \in \mathbb{R}^m$ as $\mathsf{P}_{C_r}(F) = \arg\min_{\Phi\in C_r} \|\Phi - F\|$. Note that the convergence of the sequence $(F_k)_k$ generated by this algorithm to a solution $\widehat{F}$ of (13) is guaranteed, which allows us to deduce a solution to (12) by using Relation (14).

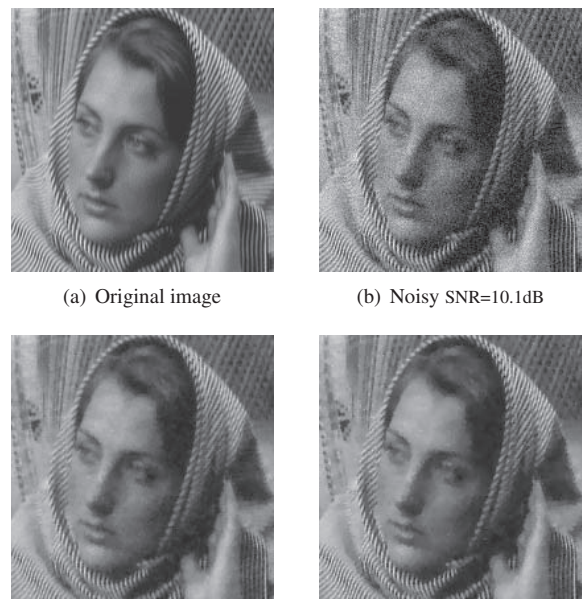
Algorithm 1: Parallel proximal algorithm solving (15)

Fix $\gamma > 0$ and $\nu \in]0,2[$. Set $k = 0$.
Choose $y_{1,0} = y_{2,0} = \ldots = y_{s+1,0} \in \mathbb{R}^m$ and F_0.
repeat
  for $r = 1,\ldots,s+1$ **do in parallel**
    $\pi_{r,k} = \begin{cases} \mathsf{P}_{C_r}(y_{r,k}) & \text{if } r \leq s \\ (\gamma A\Gamma A^\top + I)^{-1}(\gamma A\Gamma H^\top \Lambda^{-1} f + y_{s+1,k}) & \text{otherwise} \end{cases}$
  $z_k = \frac{2}{s+1}(\pi_{1,k} + \cdots + \pi_{s+1,k}) - F_k$
  for $r = 1,\ldots,s+1$ **do in parallel**
    $y_{r,k+1} = y_{r,k} + \nu(z_k - p_{r,k})$
  $F_{k+1} = F_k + \frac{\nu}{2}(z_k - F_k)$
until *convergence*

For a 4-connected lattice, when using (8) where, for every $i \in \{1,\ldots,n\}$, θ_i is the i-th line of matrix $|A^\top|$, the projection onto C is not explicit. However, a decomposition of C can be performed by setting $s = 2$ and, for every $r \in \{1,2\}$, $C_r = \{F \mid (\forall i \in S_r)\ \|\theta_i \cdot F\|_\alpha \leq g_i\}$. The partition $(S_r)_{1\leq r\leq 2}$ corresponds to two spatial disjoint sets each with a checkered pattern. The projection onto C_r for $\alpha = 2$ (resp. $\alpha = 1$ or $\alpha = \infty$) reduces to simple projections onto hyperspheres (resp. hypercubes [10]). Note that the computation of $\pi_{s+1,k}$ requires a matrix inversion which can be efficiently performed by Fast Fourier Transform when H and Λ are (or can be approximated by) circulant-block circulant matrices.

4. RESULTS

We now demonstrate the performance of DCTV with respect to weighted TV. In our experiments, we compare solutions to the TV problem (3) given by the state-of-the-art augmented Lagrangian algorithm (a.k.a. split Bregman) of [11], to solutions optimizing DCTV with our parallel proximal algorithm.



(a) Original image (b) Noisy SNR=10.1dB

(c) weighted TV SNR=13.4dB (d) DCTV SNR=13.8dB

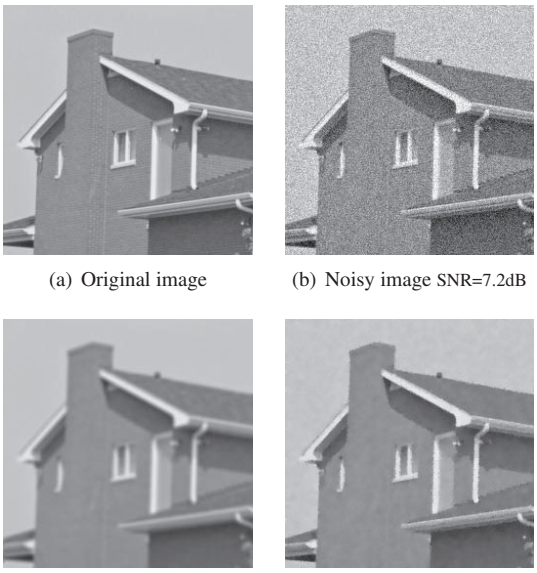
Fig. 1. Denoising 'Barbara' image corrupted with Gaussian noise of variance $\sigma^2 = 15$

We used four standard test images that we corrupted with synthetic Gaussian noise of variance σ^2. The weights have been set according to (9) with $\beta = 0.04$ for DCTV. The value of the image fidelity parameter λ was set according to an empirical rule depending of the variance of the noise. The value of ϵ was set to λ, and the stopping criteria for both algorithms is $\frac{||u^k - u^{k-1}||}{||u^k||} < 5.10^{-3}$. Signal to Noise Ratio (SNR) is used as performance measure in our quantitative evaluation. Table 1 reports SNR values for DCTV and weighted TV results obtained on each image corrupted with different values of noise variance. Examples of results are shown in Figure 1. Those experiments show that DCTV leads to improved results when the variance of the noise is lower than 50. Visually, DCTV are sharper and feature better contrast than the weighted TV results. This also explains the slight degradation of performance in presence of heavy noise (variances 50-100). Geometrically, the improvement of DCTV over the weighted TV can be interpreted in the following way: the

σ^2	5	10	15	20	25	50	100
SNR values obtained by optimizing weighted TV with Split Bregman							
house	22.7	18.9	17.1	15.8	15.1	11.8	7.8
man	23.3	19.6	17.8	16.5	15.5	12.3	8.3
lena	22.7	18.9	17.2	16.2	15.4	12.5	8.3
barbara	20.3	16.0	13.4	11.7	10.5	8.5	6.2
mean	**22.3**	**18.4**	**16.4**	**15.1**	**14.1**	**11.3**	**7.6**
SNR values obtained by optimizing DCTV							
house	22.7	19.3	17.4	15.8	15.2	11.8	7.4
man	23.7	20.3	18.2	16.7	15.6	12.1	8.1
lena	23.2	19.7	17.7	16.3	15.5	12.4	8.0
barbara	20.3	16.2	13.8	12.2	10.9	8.4	6.0
mean	**22.5**	**18.9**	**16.8**	**15.3**	**14.3**	**11.2**	**7.4**

Table 1. Quantitative denoising experiment on standard images corrupted with additive Gaussian noise of variance σ^2.

convex we use for projection adapts itself to the local neighborhood, and this is reducing blur as a result.



(a) Original image (b) Noisy image SNR=7.2dB

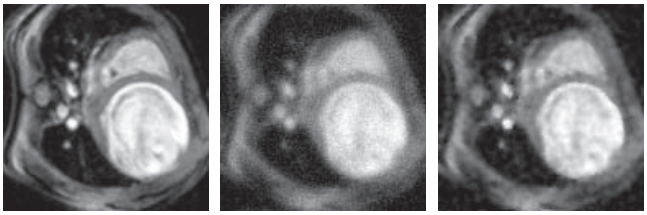
(c) Blurry image SNR=11.6dB (d) DCTV SNR=16.3dB

Fig. 2. Image fusion from a an image corrupted with Gaussian noise of variance $\sigma^2 = 20$ (b) and a blurry image (c).

In terms of computation time, DCTV is competitive with the most efficient weighted TV algorithm. Denoising the 512×512 Lena image corrupted with Gaussian noise ($\sigma^2 = 15$) requires 0.38 seconds for split Bregman, versus 0.7 seconds for PPXA to converge on an Intel Xeon 2.5GHz 8-core system. Figures 2 and 3 show the ability of DCTV to easily generalize to applications beyond denoising.

5. CONCLUSION

In this paper we have extended existing TV models by generalizing the constraint on the projection variable of the dual TV formulation. This new approach yields improved results compared with the weighted TV approach in image restoration



(a) Original image (b) Corrupt SNR=12.3dB (c) DCTV SNR=17.2dB

Fig. 3. Denoising and deblurring an MRI image corrupted with synthetic uniform 7×7 blur and Gaussian noise ($\sigma^2 = 10$)

applications. More generally, the proposed algorithm makes it possible to efficiently solve convex minimization problems involving the support function of an intersection of convex sets as a penalty term. It is also worth emphasizing that this approach can be applied on graph data structures such as those frequently employed in 3D modelling. Future work will further improve our results by using image patches in the weight computation before applying DCTV, and applications on embedded manifolds such as triangulated surfaces.

6. REFERENCES

[1] S. Bougleux, A. Elmoataz, and M. Melkemi. Discrete regularization on weighted graphs for image and mesh filtering. In *Proc. of SSVM 2007*, pages 128–139, Ischia, Italy, May 30 - June 2 2007.

[2] J. Carter. *Dual Methods for Total Variation-based Image Restoration*. PhD thesis, UCLA, Los Angeles, CA, USA, 2001.

[3] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.*, 20(1-2):89–97, 2004.

[4] T. F. Chan, G. H. Golub, and P. Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM J. Sci. Comput.*, 20(6):1964–1977, 1999.

[5] P. L. Combettes, D. Dũng, and B. C. Vũ. Dualization of signal recovery problems. *Set-Valued and Variational Analysis*, 18:373–404, 2010.

[6] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212, 2011.

[7] G. Giboa and S. Osher. Nonlocal operators with application to image processing. *SIAM Multiscale Model. Simul.*, 7(3):1005–1028, 2007.

[8] L. Grady and J. Polimeni. *Discrete Calculus: Applied Analysis on Graphs for Computational Science*. Springer, 2010.

[9] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4), 1992.

[10] E. van den Berg and M. P. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM J. on Scientific Computing*, 31(2):890–912, 2008.

[11] X. Zhang, M. Burger, X. Bresson, and S. Osher. Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. *SIAM Journal on Imaging Sciences*, 3(3):253–276, 2010.