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 $\Omega$, and denoting by $x$ and $y$ two arbitrary points of $\Omega$, the weighted anisotropic TV model [7] is given by

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

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

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

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

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 $e_{ij}$. 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 $e_{i,j}$, as $A_{e_{i,j},v_k} = 1$ if $i = k$, $A_{e_{i,j},v_k} = -1$ if $j = k$, and $A_{e_{i,j},v_k} = 0$ 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, if $f$ and $u$ are vectors in $\mathbb{R}^n$. The discrete TV model as defined by [1, 7] is
\[
\min_u \frac{1}{2} \sum_{i=1}^n \sum_{j \in N_i} (w_{i,j} (u_j - u_i)^2)^{1/2} + \frac{1}{2\lambda} \sum_{i=1}^n (u_i - f_i)^2
\]
where $N_i = \{ j \in \{1, \ldots, n\} \mid e_{i,j} \in E \}$ and the dual formulation optimized by a projection algorithm [7] is given by
\[
\min_u \max_{\|p\|_2 \leq 1} p^\top ((Au) \cdot \nabla u) + \frac{1}{2\lambda} \|u - f\|^2
\]
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 $p \in \mathbb{R}^m$, $\|p\|_\infty = \max_{i \in \{1, \ldots, n\}} \left( \sum_{j \in N_i} p_{i,j}^2 \right)^{1/2}$.

By introducing a vector $F = (F_{i,j})_{i,j} \in \mathbb{R}^m$ such that, for every $i \in \{1, \ldots, n\}$ and $j \in N_i$, $F_{i,j} = p_{i,j} \sqrt{w_{i,j}}$, the problem can be reformulated as
\[
\min_u \max_{F \in C} F^\top Au + \frac{1}{2\lambda} \|u - f\|^2,
\]
where $f \in \mathbb{R}^q$ is an observed vector of data, $H \in \mathbb{R}^{q \times n}$ and $\Lambda$ is weighting symmetric positive definite 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. $\Lambda$ 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 $(C_r)_{1 \leq r \leq s}$ of $\mathbb{R}^m$, the projections onto which take closed forms. An example of a set $C$ of interest is
\[
C = \{ F \mid \langle \theta_i \cdot F \rangle_\alpha \leq g_i \}
\]
where $\| \cdot \|_\alpha$ is the $\ell_\alpha$ norm of $\mathbb{R}^m$ with $\alpha \in [1, +\infty]$ and, for every $i \in \{1, \ldots, n\}$, $\theta_i \in [0, +\infty]^m$ 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 $\theta_i$ is the $i$-th node weight $g_i$ is a decreasing function of the image gradient. Given positive reals $\epsilon$ and $\beta$, we suggest using
\[
g_i = \exp(-\beta \|\nabla u_i\|_2) + \epsilon,
\]
where $\nabla$ is some reference image (wich corresponds to some rough estimate of $u$) and $\|\nabla u_i\|_2$ is the Euclidean norm of its discrete gradient $\nabla u_i$ at node $i$. In addition to intensity information, $g_i$ may be used to penalize image variations of 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 $\theta_i$, 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 $\theta_i$ 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 $\sigma_C$ of the closed convex constraint set $C$ as
\[
\sigma_C: \mathbb{R}^m \to [-\infty, +\infty]: a \mapsto \sup_{F \in C} F^\top a.
\]
This is a proper lower-semicontinuous convex function, the conjugate of which is the indicator function of $C$,
\[
\iota_C: F \mapsto \begin{cases} 0 & \text{if } F \in C, \\ +\infty & \text{otherwise}. \end{cases}
\]
This leads us to consider the following optimization problem:
\[
\min_u \sigma_C(Au) + \frac{1}{2} \|H u - f\|^2 + \eta \|K u\|^2
\]
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 $\text{rank} H = n$, the last term vanishes and (12) is strictly equivalent to (7). The term $u \mapsto \eta \|K u\|^2/2$ 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 \( \hat{u} \). The dual Fenchel-Rockafellar form of the problem is

\[
\min_F \varphi(F) + \tau_C(F),
\]

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

\[
\hat{u} = \Gamma (H^\top \Lambda^{-1} f - A^\top \hat{F}).
\]

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 \( \tau_C \) 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

\[
\min_F \sum_{r=1}^s \tau_{C_r}(F) + \varphi(F).
\]

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 \( P_{C_r}(F) = \arg \min_{F \in C_r} \|F - F\| \).

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

\[
\begin{align*}
\text{for } r = 1, \ldots, s + 1 \text{ do in parallel} & \quad \pi_r, k = \begin{cases} P_{C_r}(y_{r,k}) & \text{if } r \leq s \\ (\gamma A^\top \Lambda^r + I)^{-1} (\gamma A A^\top \Lambda^r + \eta K)^{-1} f + y_{s+1,k} & \text{otherwise} \end{cases} \\
& \quad z_k = \frac{2}{s+1} (\pi_1, k + \cdots + \pi_{s+1, k}) - F_k \\
& \quad \text{for } r = 1, \ldots, s + 1 \text{ do in parallel} & \quad y_{r,k+1} = y_{r,k} + \nu (z_k - p_{r,k}) \\
& \quad F_{k+1} = F_k + \frac{\gamma}{\nu} (z_k - F_k) \\
\end{align*}
\]

until convergence

For a 4-connected lattice, when using (8) where, for every \( i \in \{1, \ldots, n\} \), \( \theta_i \) 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 \( r \in \{1, 2\} \), \( C_r = \{ F \mid (\forall i \in S_r) \|\theta_i \cdot F\|_2 \leq g_i \} \). 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 \( A \) 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.

![Fig. 1. Denoising ‘Barbara’ image corrupted with Gaussian noise of variance \( \sigma^2 = 15 \)](image)

We used four standard test images that were corrupted with synthetic Gaussian noise of variance \( \sigma^2 \). The weights have been set according to (9) with \( \beta = 0.04 \) for DCTV. The value of the image fidelity parameter \( \lambda \) was set according to an empirical rule depending of the variance of the noise. The value of \( \epsilon \) was set to \( \lambda \) and the stopping criteria for both algorithms is \( \|u_k^\alpha - u_{k-1}^\alpha\|_{\|\cdot\|_1} < 5 \cdot 10^{-3} \). 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
Table 1. Quantitative denoising experiment on standard images corrupted with additive Gaussian noise of variance $\sigma^2$.  

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

In terms of computation time, DCTV is competitive with the most efficient weighted TV algorithm. Denoising the $512 \times 512$ 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 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