Inverse Problems and Imaging
Volume 6, No. 4, 2012, 623–644
doi:10.3934/ipi.2012.6.623

GLOBAL MINIMIZATION OF MARKOV RANDOM FIELDS WITH APPLICATIONS TO OPTICAL FLOW

Tom Goldstein

Rice University, Department of Electrical and Computer Engineering
Houston, 77251, USA

Xavier Bresson

City University of Hong Kong, Department of Computer Science
Hong Kong, China

Stan Osher

UCLA, Department of Mathematics
Los Angeles, 90095, USA

(Communicated by Antonin Chambolle)

Abstract. Many problems in image processing can be posed as non-convex minimization problems. For certain classes of non-convex problems involving scalar-valued functions, it is possible to recast the problem in a convex form using a "functional lifting" technique. In this paper, we present a variational functional lifting technique that can be viewed as a generalization of previous works by Pock et. al and Ishikawa. We then generalize this technique to the case of minimization over vector-valued problems, and discuss a condition which allows us to determine when the solution to the convex problem corresponds to a global minimizer. This generalization allows functional lifting to be applied to a wider range of problems then previously considered. Finally, we present a numerical method for solving the convexified problems, and apply the technique to find global minimizers for optical flow image registration.

1. **Introduction.** Many problems in image processing are posed as regularized minimization problems. Classical examples include total-variation (TV) and non-local TV regularized models for denoising [53, 46, 31], deconvolution [45, 57], and segmentation [11, 43, 31]. When these models are convex, standard minimization methods provide reliable results. Simple and reliable variational techniques include gradient descent [53], dual formulations [18, 15, 1], split Bregman schemes [32, 55, 26]. In the discrete setting, state-of-the-art methods can be derived by embedding the problem into a graph or Markov random field (MRF), and solving the resulting problem using graph cuts [42, 20, 8].

When problems are non-convex, these techniques fail because they get "stuck" at local minima. A classical example of one such non-convex problem is image registration. For this problem, gradient descent techniques can be very ineffective, especially if the images involved suffer from strong noise contamination or repetitive details such as textures.

Recent advances in optimization theory have allowed certain classes of non-convex problems to be solved in polynomial time. Most of these techniques rely

2000 *Mathematics Subject Classification.* 46N10, 68U10, 90C26.
Key words and phrases. Optical flow, functional lifting, nonconvex optimization.

on the concept of "functional lifting," which converts a non-convex problem into a convex problem which can be easily minimized using standard techniques. The functional lifting concept has been used to solve segmentation and stereo problems within both the continuous and MRF framework [17, 11, 37, 50].

Unfortunately, the work that has been done on functional lifting methods pertains only to minimizing scalar-valued functions. Binary segmentation problems involve only one variable which determines which region a pixel is in, and stereo matching problems only involve image displacements in one direction. More sophisticated problems require the optimization of multi-variate functions. One of the most ubiquitous examples of this is image registration, in which the pixels of one image are mapped onto another. Each pixel in a registration problem is assigned both a horizontal and vertical displacement component. Because displacement functions are vector-valued, existing methods for non-convex optimization cannot tackle this problem.

In this work, we generalize the functional lifting framework to minimize multivariate functions over Markov random fields - allowing us to minimize a more general class of non-convex problems then has been previously considered. In doing so, we present a scheme for computing the global minimizer of image registration problems.

The organization of this paper is as follows: We first present a brief review of work that has been done on single variable functional lifting, focusing on the discrete theory of MRF's. We then discuss optical flow models for image registration, and explain how they can be formulated using first order MRF's. We then introduce a novel method for solving multi-variable non-convex problems in imaging via functional lifting, and explain how this method can be applied to image registration. Finally, we describe a numerical scheme to solve the resulting convexified problems, and present results of numerical experiments.

1.1. **Optimization on Markov Random Fields.** The Markov Random Field (MRF) has a long and important history within the physics and statistics communities, where it was first introduced as a formalization of the Ising-Potts model for particle interactions [3, 4, 51]. It was Geman and Geman [30] who first showed how MRF's could be used for image processing tasks.

We will now formalize the notion of Markov random field. An MRF is defined over a discrete lattice of points, Ω. In the context of image processing, Ω is a rectangular array of pixels. Each point, $x \in \Omega$, in the lattice takes on some value (e.g. the pixel intensity), denoted by $u(x)$. With each lattice point, we also identify its set of neighbors, $N_x \in \Omega$, with which the point x directly interacts.

Most MRF based image processing methods are concerned with minimizing energies of the form

$$E(u) = \sum_{x\in\Omega} \left\{ \underbrace{f(u(x),x)}_{fidelity} + \underbrace{\sum_{y\in N_x} r(u(x),u(y))}_{regularizer} \right\}. \tag{1}$$

Energy models on MRF's consist of two parts. The regularizer measures differences between neighboring nodes. In imaging applications, this term enforces some sort of "smoothness" of the recovered image. The fidelity term measures the relationship

between each individual node and some pre-defined quantity, such as the difference between the clean image, u, and a noisy image.

Energies of the form (1) can be minimized by standard techniques, such as gradient descent, simulated annealing, or other heuristic methods [5, 40, 6, 9, 8]. However, these techniques can be slow, and yield sub-optimal results.

For certain classes of problems with pairwise label interactions, exact minimizers can be computed very efficiently using a "graph cuts" approach. These techniques embed the energy function into a graph consisting of a large number of nodes connected with edges. Each edge is assigned a certain weight. The minimal graph cut problem is to partition the set of nodes into two disjoint sets in such a way that the sum of the weights of the "cut" edges is minimal [9, 41, 29, 4].

The first general class of problems solvable by graph cuts was identified by Picard and Ratliff [47]. This work was followed by Kolmogorov and Zabih [41], who present a regularity condition which is both necessary and sufficient for a problem to be solvable by graph cuts. This basic graph cuts framework was only applicable to binary problems (problems for which the function, u, is restricted to have 2 discrete values). Despite this restriction, Grieg et al. [33] realized that this framework could be used to solve TV regularized problems in imaging. This framework was powerful enough to solve problems on binary images, as well as 2-phase segmentation problems [7, 10, 60].

Later work on MRF optimization focused on generalizing these results to solve problems with more than two labels. For problems with TV regularizers and convex fidelity terms, Darbon and Sigelle show that an n-label problem can be exactly minimized by solving a sequence of $\log(n)$ binary minimization problems [23, 24, 22], allowing for fast minimization of the ROF denoising model [53].

For problems with non-convex regularizers, the problem becomes more difficult. In the next section, we present a modern approach to this problem.

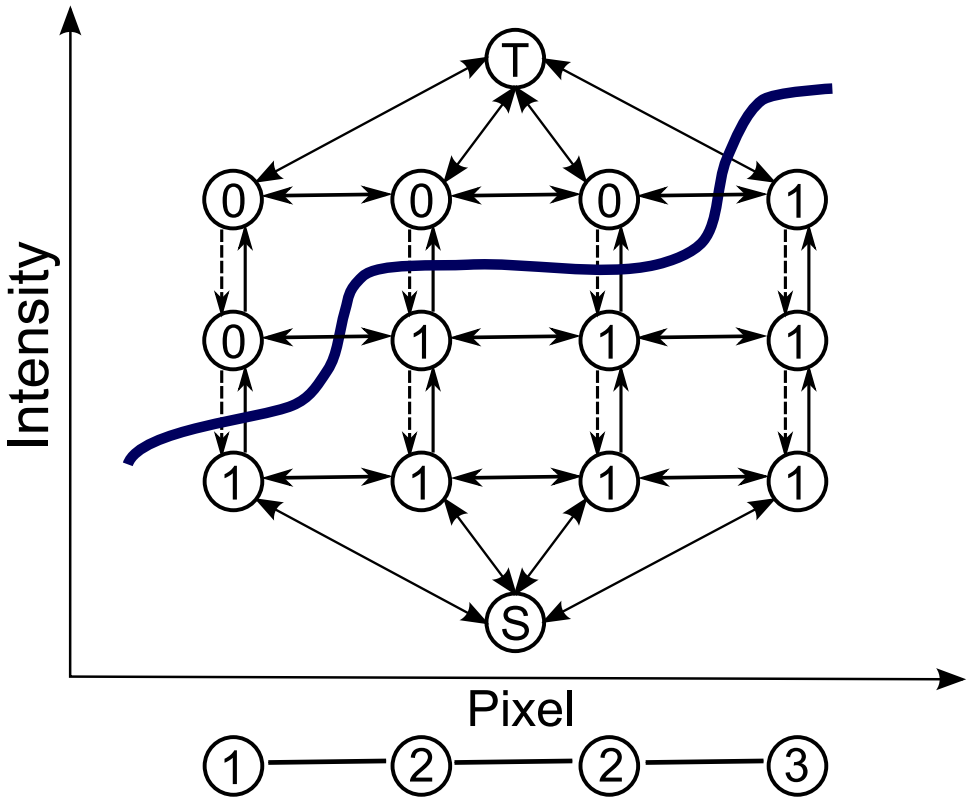
1.2. **Functional lifting for non-convex optimization.** One of the most influential frameworks for global minimization of non-convex functions was proposed by Ishikawa [37]. The simplest version of the Ishikawa method occurs when the regularizer is a linear combination of the pairwise differences between labels. In particular, it is assumed that the Markovian energy has the form

$$E(u) = \sum_{x\in\Omega} \left\{ f(u(x), x) + \sum_{y\in N_x} w^{x,y} |u(x) - u(y)| \right\} \tag{2}$$

for some set of positive weights $\{w^{x,y}\}$. Because the fidelity term $f(\cdot,\cdot)$ may be non-convex, standard minimization techniques fail to find a global minimum in general.

In [37], a technique is introduced to convert the non-convex, multi-label "primal" problem (2) into a regular, binary "lifted" problem that is solvable by graph cuts. This scheme is illustrated in figure 1. This is accomplished by replacing each node in the primal problem, which can take on Γ possible values, with a column of Γ binary valued nodes. By connecting adjacent nodes with properly weighted edges, we formulate the graph so that the value of a cut corresponds to a configuration of pixels in the primal problem. Suppose, for example, we choose each horizontal edge in figure 1 to have a weight of unity. If two adjacent pixels differ in intensity by d units, then the corresponding cut must sever d horizontal edges. This choice of horizontal edge weights gives us the familiar TV regularization for the simple

FIGURE 1. The Ishikawa method for embedding a non-convex Markovian energy. At the bottom of the figure, we show the primal representation which consists of a chain of 4 pixels. Above it is the lifted representation, in which each pixel corresponds to a column of 3 binary nodes. We obtain the intensity of each pixel by summing the binary values of the nodes in the corresponding column.



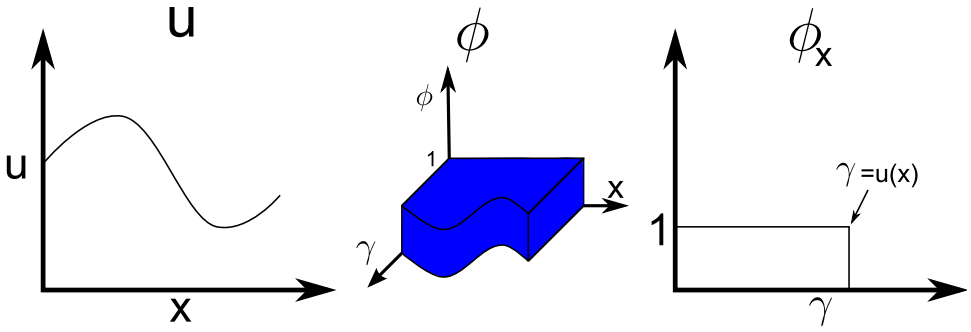
1-dimensional problem depicted here. Likewise, a cut in figure 1 must sever one solid vertical edges in each column. The energy associated with this severed edge corresponds to the fidelity energy contributed by that pixel.

In figure 1, there are dotted vertical edges between nodes. These edges have infinite weight, and only contribute to the value of a cut if the tail lies in group s. The purpose of these edges to to prevent a column of nodes from being cut more than once.

Although we focus on simple TV-like regularizers in this manuscript, the Ishikawa approach can be generalized to solve problems with more sophisticated priors [38, 54]. An approach applicable to arbitrary sub-modular priors is presented by Darbon [21].

The functional lifting approach to non-convex optimization has also been studied in the variation/PDE framework by Pock et al. [50], and further refined in

Figure 2. The continuous analog of the Ishikawa method. The one dimensional function u is lifted onto the two dimensional function ϕ by assigning an indicator function for each location x. A cross section of ϕ at location x is a step function with discontinuity at $u(x)$.



[48]. In this work, the authors re-interpret the Ishikawa method using a continuous formulation, and solve variational problems involving TV regularizers. Consider a continuous energy of the form

$$E_c(u) = \int_\Omega |\nabla u(x)| + f(u(x))\, dx. \tag{3}$$

We wish to minimize the above energy over all functions $u : \Omega \subset R^n \to [0, \Gamma] \subset R$. Rather than replacing each variable, $u(x)$, with a discrete column of nodes as was done above, Pock et. al replace each unknown with an indicator function defined on an interval. Consider the lifted function $\phi : \Omega \times [0, \Gamma] \to [0, 1]$ defined as follows

$$\phi(x, \gamma) = \begin{cases} 1 & \text{if } \gamma \leq u(x) \\ 0 & \text{otherwise} \end{cases}$$

Note that $\phi_x(\gamma) = \phi(x, \gamma)$ is an indicator function on an interval with a discontinuity at $\gamma = u(x)$. This relationship is depicted in figure 2.

From this definition of ϕ, we have that $|\partial_\gamma \phi_x(\gamma)| = |\partial_\gamma \phi(x, \gamma)| = \delta(\gamma - u(x))$. Using this result, we can write the fidelity portion of the energy (3) as

$$f(u(x)) = \int f(x, \gamma)\delta(\gamma - u(x))\, d\gamma = \int f(x, \gamma)|\partial_\gamma \phi(x, \gamma)|\, d\gamma. \tag{4}$$

The regularization portion of (3) can be written in terms of ϕ using the co-area formula as follows

$$\int_\Omega |\nabla u| = \int_{\gamma \in [0,\Gamma]} \int_\Omega |\nabla \phi(x, \gamma)|\, dx\, d\gamma \tag{5}$$

Using (4) and (5), we see that $E_c(u) = L_c(\phi)$ where

$$L_c(\phi) = \int_{\gamma \in [0,\Gamma]} \int_\Omega |\nabla \phi(x, \gamma)| + f(x, \gamma)|\partial_\gamma \phi(x, \gamma)|\, dx\, d\gamma \tag{6}$$

Note that, unlike $E_c(u)$, the energy $L_c(u)$ is convex. The idea presented in [50] is to compute a minimizer of the convex energy, (6), and then attempt to reconstruct a function u from the optimal function ϕ. Because we allow ϕ to take on any value

in the interval $[0,1]$, the minimizer of L_c may not be binary valued. In this case, we threshold ϕ at some value $\alpha \in (0,1)$, and use the resulting binary function to recover u. Interestingly, the authors propose to compute the convex minimization using a primal-dual scheme, rather than conventional graph cut minimization. The variational method has several advantages over graph cuts because it uses less memory, produces more accurate results, and can be easily parallelized.

The framework for functional lifting presented in [50] is rather informal. A rigorous definition of continuous functional lifting requires one to consider many technicalities. In particular, one must carefully define the appropriate function spaces for minimization. Also, in order to guarantee that L_c have a non-trivial minimizer, it is necessary to specify boundary conditions on ϕ involving limits at infinity. The interested reader should see [49] and [48], in which these issues are addressed formally. Furthermore, the authors of [48] address further generalizations of these method, such as applications to con-convex regularization and image segmentation. Rather than dwell on the complexities of the continuous problem, this paper shall focus on the case of functional lifting for discrete energies over MRF's.

The purpose of this work is to present a global optimization scheme for Markovian energies involving vector-valued functions. Our approach relies on a multi-dimensional analog on the functional lifting concept, and can thus be viewed as a generalization of the method of Ishikawa [37].

In particular, we allow the discrete Markovian function, $u : \Omega \to R^m$, to take on vector values at each node (i.e. the range of u has m components/channels), and we consider minimization problems of the following form:

$$E(u) = \sum_{x\in\Omega} \left\{ f(u(x),x) + \sum_{i=1}^{m} \sum_{y\in N_x} w^{x,y} |u_i(x) - u_i(y)| \right\} \tag{7}$$

Although we focus on discrete problems defined over MRF's, our approach to functional lifting is largely variational. Rather than forcing the lifted function ϕ to take on binary values, we parallel the continuous approach described above by allowing ϕ to range over the interval $[0,1]$. Furthermore, the minimization scheme we choose is a generalization of the continuous scheme presented in [50], and does not rely on graph-cuts. In this sense, the method presented here is a hybrid method – it relies on both discrete and variational techniques.

Later on, we will present several theoretical results pertaining to our functional lifting method. Our scheme converts vector-valued problems to a lifted problem with higher dimension. We show that, in order for the solution of the lifted problem to correspond to a solution of the original problem, the solution of the lifted problem must satisfy a certain "box function condition" after thresholding has been applied. We also present a proof that this condition is always satisfied when the method is applied to scalar-valued problems (i.e. $m = 1$). In the experimental results presented here, we found that the box function condition was also satisfied when the method was applied to vector-valued problems, however rigorous results will be a subject of future research.

1.3. **Optical flow methods for image registration.** Optical flow is an important concept in both computer vision and medical imaging fields. The optical flow problem can be phrased as follows: Given two images (a "source" and a "target") with corresponding features, compute the displacement field that maps the features of the source image onto the features of the target image. Optical flow models are

used in computer vision to track the movement of an object through consecutive images in a video segment. In this application, the displacement field represents the velocity field of the moving objects. In medical imaging, optical flow is used to remove motion artifacts from clinical images, as well as to map brain images onto reference images for quantitative comparison [39, 25, 34].

Variational models for optical flow were originally proposed by Horn and Schunck [36]. Given two images, I_0 and I_1, defined over some region Ω, these authors proposed to recover the displacement field, $u = (u_1, u_2)$, by minimizing a functional of the form

$$E_{HS}^{p,q} = \alpha \int |I_1(x + u(x)) - I_0(x)|^q \, d\Omega + \int |\nabla u_1|^p + |\nabla u_2|^p \, d\Omega \tag{8}$$

where α is a parameter that controls the strength of regularization term. In the original Horn-Schunck model, the authors choose $p = q = 2$, and assume the images to be smooth. Under these assumptions, the energy (8) can be simplified by replacing the images with a local linear approximation, and forming explicit optimality conditions.

The original Horn-Schunck model has several limitations. First, the quadratic regularizer does not allow for discontinuities. Second, the smoothness assumption does not hold for images with hard edges, and only allows for small displacements. For this reason, a myriad of authors have suggested non-quadratic regularizers [13, 35, 19, 52, 58, 44]. Of the non-quadratic regularizers that have been proposed for this problem, one of the most popular and successful is total-variation (TV) regularized optical flow [13, 35, 19, 52], which corresponds to the above model (8) with $p = 1$. The TV regularizer is known to enforce smoothness, while still allowing for jump discontinuities [53, 46, 15].

When problem (8) is discretized using an anisotropic TV regularizer, the energy can be written

$$\begin{aligned} E_{HS} = &\sum_{x \in \Omega} \alpha \left(I_1(x + u(x)) - I_0(x)\right)^2 \\ &+ \sum_{i=1}^{2} |u_i(x + e_1) - u_i(x)| + |u_i(x + e_2) - u_i(x)| \end{aligned} \tag{9}$$

where $e_1 = (1, 0)$ and $e_2 = (0, 1)$ denote elementary unit vectors. Note that the non-convex optical flow energy (9) is a Markovian energy of the form (7).

2. Theory.

2.1. Notation and preliminary results.

We now introduce some notation and conventions that are handy when working with discrete function spaces.

First, we use the conventional interval notations (a, b) and $[a, b]$ to denote continuous open and closed intervals, respectively. In the discrete setting, we use the notation $[a, b]_d = \{a, a + 1, \cdots, b - 1, b\}$ to denote a discrete interval. We also use the notation e_i to denote the elementary basis vector, which has its ith component equal to unity and all other components equal to zero.

We will frequently convert between multi- and binary-valued functions using a "threshold" operator. Given some scalar-valued function ϕ we define its thresholding at level α (also called the upper level set at level α) as

$$\phi^\alpha(\gamma) = \begin{cases} 1 & \text{if } \phi(\gamma) \geq \alpha \\ 0 & \text{otherwise} \end{cases}$$

Alternatively, we can represent the thresholding of ϕ as the indicator function $\phi^\alpha(\gamma) = \mathbb{1}_{\phi \geq \alpha}(\gamma)$ where $\mathbb{1}_A$ denotes the indicator function of the set A.

Many useful regularizers can be represented with level sets/thresholding using a coarea formula. Explanations of these formulas and their uses in imaging can be found in [28, 23, 24, 14]. In the continuous case, the coarea formula allows us to represent the total variation of a function $u : \Omega \subset R^n \to R$ as an integral of the total variations of its level sets as follows:

$$\int_\Omega |\nabla u| = \int_{\alpha \in R} \int_\Omega |\nabla u^\alpha| \, dx \, d\alpha.$$

The coarea formula also has a discrete analog. Given a discrete function, $u : \Omega \to [0, \Gamma - 1]_d$, and two points $x, y \in \Omega$, we have

$$|u(x) - u(y)| = \sum_{\alpha \in [0, \Gamma - 1]_d} |u(x)^\alpha - u(y)^\alpha| \, d\alpha. \tag{10}$$

To aid in the presentation of our theoretical results, we introduce the notion of a "box function." In the discrete setting, consider a rectangular domain $\Gamma = [0, \Gamma_1]_d \times [0, \Gamma_2]_d \times \cdots \times [0, \Gamma_m]_d$. We shall refer to such a domain as a "box" with "principle vertex" at $(\Gamma_1, \Gamma_2, \cdots, \Gamma_m)$. We say that a function $\phi : \Gamma \to R$ is a box function if it can be written in the form

$$\begin{aligned} \phi(\gamma) =& \mathbb{1}_{[1,\gamma_1]_d \times [1,\gamma_2]_d \times \cdots \times [1,\gamma_{N_\gamma}]_d}(\gamma) \\ =& \begin{cases} 1 \text{ if } \gamma \in [1,\gamma_1]_d \times [1,\gamma_2]_d \times \cdots \times [1,\gamma_{N_\gamma}]_d \\ 0 \text{ otherwise} \end{cases}. \end{aligned}$$

In plain English, the region Γ is shaped like a box with one vertex at the origin and one vertex (the principle vertex) in the first quadrant. The corresponding box function has value zero outside of this box, and unity inside the box.

Finally, we will make use of a discrete difference operator, which we shall denote D_γ^m. This operator is the discrete mixed-partial derivative operator $D_\gamma^m = (-1)^m D_1^+ D_2^+ \cdots D_m^+$ where D_i^+ represents the first order forward difference operator in the ith coordinate direction. In 2 dimensions, this operator has the stencil

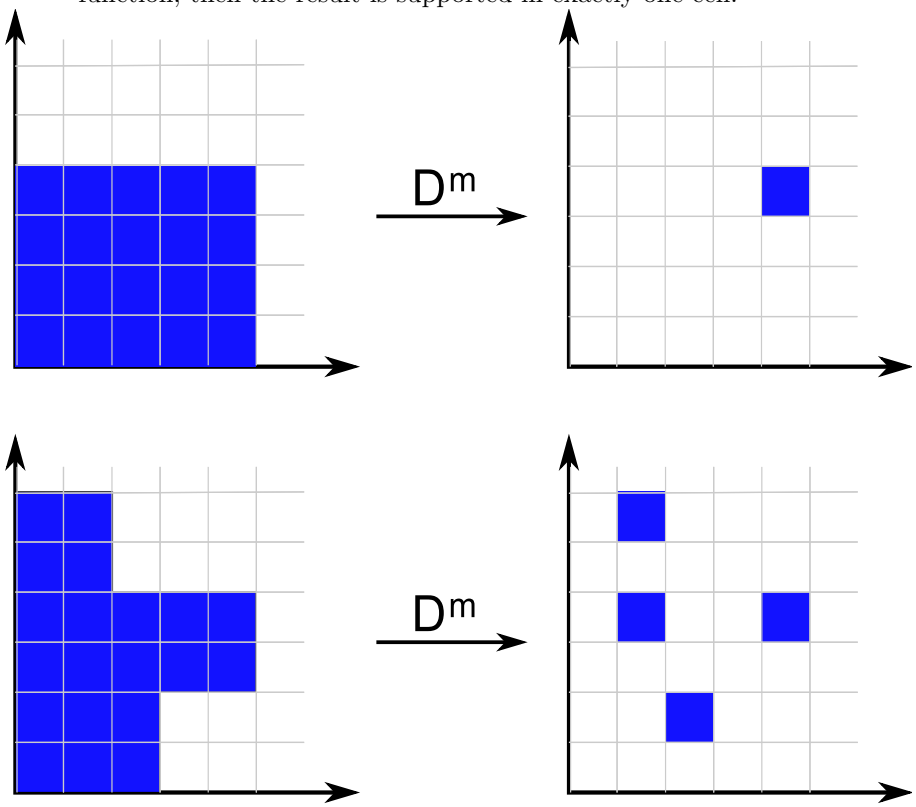
$$D_\gamma^2 = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

In the discussion below, we will frequently apply this difference operator to box functions. Note that, if ϕ represents a box function with principle vertex at $\hat{\gamma}$, then

$$D_\gamma^m \phi(\gamma) = \begin{cases} 1 \text{ if } \gamma = \hat{\gamma}, \\ 0 \text{ otherwise} \end{cases}$$

A more general understanding of this operator is gained by considering the following observation: When this operator is applied to a discrete binary/indicator function ϕ, the result has value 1 at every "corner" of the support of ϕ, and is zero elsewhere. This is illustrated in figure 3.

Figure 3. The effect of the operator D_γ^m on a discrete binary function. Binary functions are depicted in the $\gamma_1 - \gamma_2$ plane, with blue support. The image under D_γ^m has value 1 or -1 at every "corner" of the original binary function. If the input is a box function, then the result is supported in exactly one cell.



2.2. **Reformulation as a convex problem.** In this section, we describe a technique for reformulating the Markovian energy

$$E_f(u) = \sum_{x\in\Omega} \left\{ f(u(x), x) + \sum_{i=1}^{m} \sum_{y\in N_x} w^{x,y} |u_i(x) - u_i(y)| \right\} \tag{11}$$

as a convex problem. We also present theoretical results proving the equivalence between the non-convex minimization problem and its convex analog.

The energy (11) can also be written in the condensed form

$$E_f(u) = \sum_{x\in\Omega} \{f(u(x), x)\} + \sum_{i=1}^{m} r_i(u_i) \tag{12}$$

where

$$r_i(u_i) = \sum_{x\in\Omega} \sum_{y\in N_x} w_i^{x,y} |u_i(x) - u_i(y)| \tag{13}$$

is the regularizer of the *ith* component of u.

We begin by specifying the function space over which we intend to minimize (7). We let $u(x)$ range over the space of functions whose *ith* component lies in the set $[0, \Gamma_i - 1]_d = \{0, 1, 2, \cdots, \Gamma_i - 1\}$. For this purpose, we define the space of discrete functions $S_p = \{u : \Omega \to \Gamma\}$, where $\Gamma = [0, \Gamma_1 - 1]_d \times [0, \Gamma_2 - 1]_d \times \cdots \times [0, \Gamma_m - 1]_d$.

Note that (11) is in general non-convex, and so may have a large number of local minima. Rather than attack this non-convex problem directly, we will re-formulate this problem as a convex problem over the set of "lifted" functions $S_l = \{\phi : \Omega \times \Gamma \to [0, 1]\}$.

Consider the lifted function

$$\phi(x, \gamma) = \mathbb{1}_{[0, u(x)_1] \times [0, u(x)_2] \times \cdots \times [0, u(x)_m]}(\gamma) \tag{14}$$

Note that, if we fix x, then the function $\phi_x(\gamma) = \phi(x, \gamma)$ is a discrete box function. The support of $\phi(x, \cdot)$ is a "box" with a vertex at $u(x)$. We thus have that $D_\gamma \phi(x, \gamma) = 1$ if $\gamma = u(x)$, and $D_\gamma \phi(x, \gamma) = 0$ otherwise. This fact allows us to write the fidelity term in the energy (11) as follows:

$$f(x, u(x)) = \sum_{\gamma \in \Gamma} f(x, \gamma) D_\gamma \phi(x, \gamma).$$

We now consider the regularizing term in (12). Because this regularizer has the form (13), we can write it in terms of the level sets of u using the coarea formula as follows:

$$r_i(u_i) = \sum_{l \in [0, \Gamma_i - 1]} r_i(\mathbb{1}_{[0,\, l]} u_i).$$

If we simply note that $\mathbb{1}_{[0,\, l]}(u_i(x)) = \phi(x,\, l e_i)$, we then have

$$r_i(u_i) = \sum_{l \in [0, \Gamma_i - 1]} r_i(\phi(\cdot,\, l e_i)).$$

When we put these pieces together, we have that $E_f(u) = L_f(\phi)$, where

$$L_f(\phi) = \sum_{i=1}^{m} \sum_{0 < l < \Gamma_i - 1} r_i(\phi(\cdot,\, l e_i)) + \sum_{x \in \Omega,\, \gamma \in \Gamma} f(x, \gamma) D_\gamma^m \phi(x, \gamma). \tag{15}$$

We have just shown that, provided that the relationship (14) holds, the non-convex energy (11) is equivalent to the convex lifted energy (15). The set of lifted functions satisfying the relationship (14) is important enough that we introduce the following definition.

Definition 1. *We say that a binary function,*

$$\phi : \Omega \times \Gamma \to \{0, 1\}$$

satisfies the discrete box function condition *(BFC) if $\phi_x(\gamma) = \phi(x, \gamma)$ is a box function for every $x \in \Omega$.*

Note that a function ϕ satisfies a relationship of the form (14) if and only if it satisfies the box function condition. If ϕ does satisfy BFC, then we can recover the corresponding function u using the following multi-dimensional generalization of the the layer cake formula:

$$u_i(x) = \sum_{\gamma = 0}^{\Gamma_i - 1} \phi(x, \gamma e_i). \tag{16}$$

This observation suggests a method for finding the global minimum of (11): We could first find $\hat{\phi}$, the minimizer of the convex energy (15) over some appropriate function space. Provided that $\hat{\phi}$ satisfies the box function condition, this minimizer should correspond to some $u \in S_p$, which is the global minimizer of (11). But how can we ensure that $\hat{\phi}$ satisfies the box function condition? How can we even ensure that $\hat{\phi}$ is binary?

Before we answer these questions, we must clarify the function space we will use for minimization of the lifted energy (15). Ideally, we would like to minimize over the set of functions satisfying BFC, or even the set of binary functions. However, in order to guarantee that (15) have a unique minimizer, we must choose a convex function space. For this reason, we must relax the binary function condition, and allow our functions to take on values in the unit interval. We define the following space of functions:

Definition 1. *We say that $\phi \in S_l^+$ if the following two conditions are satisfied:*

1. $\phi : \Omega \times \Gamma \to [0,1]$
2. $D_\gamma^m \phi \geq 0$.

We now prove two key results that clarify when the minimizers of (11) and (15) are equivalent. The technique of proof has been adapted from [48, 50].

Theorem 1. *For any $f : \Omega \times \Gamma \to R$, let*

$$\hat{\phi} = arg \min_{\phi \in S_l^+} L_f(\phi).$$

Suppose further that for some $\alpha \in (0,1)$, the binary thresholding $\hat{\phi}^\alpha$ satisfies BFC. Then $\hat{\phi}^\alpha$ corresponds to a global minimizer of the original non-convex energy (11).

Proof. We begin by noting that if $\hat{\phi}^\alpha$ satisfies BFC for some $\alpha \in (0,1)$, then $\hat{\phi}^\alpha$ satisfies BFC for some set of α values of positive measure. This fact is easily seen by noting that $\hat{\phi}$ takes on only finitely many values over the discrete domain $\Omega \times \Gamma$. If we let α_1 be the smallest element in the set $\{\hat{\phi}(x,\gamma) : \hat{\phi}(x,\gamma) \geq \alpha\} \cup \{1\}$, and α_2 be the largest element in the set $\{\hat{\phi}(x,\gamma) : \hat{\phi}(x,\gamma) < \alpha\} \cup \{0\}$, then $\hat{\phi}^\beta = \hat{\phi}^\alpha$ for all $\beta \in (\alpha_2, \alpha_1)$.

We now wish to show that the function $\hat{\phi}^\alpha$ minimizes (15) over the set of all functions satisfying BFC. To this end, we decompose the energy (15) into an integral over the level sets of $\hat{\phi}$ using the discrete co-area formula (10). This formula tells us that

$$r_i(\phi(x, le_i)) = \int_0^1 r_i(\phi^\alpha(x, le_i))\, d\alpha. \tag{17}$$

We now consider the fidelity portion of the energy (15). Invoking the linearity of the operator D_γ^m, we get

$$D_\gamma^m \phi(x,\gamma) = D_\gamma^m \int_0^1 \phi^\alpha(x,\gamma)\, d\alpha = \int_0^1 D_\gamma^m \phi^\alpha(x,\gamma)\, d\alpha \tag{18}$$

Using (17) and (18), we get

$$L_f(\hat{\phi}) = \int_0^1 L_f(\hat{\phi}^\alpha)\, d\alpha.$$

Suppose for contradiction that there exists some BFC function, ϕ^*, with $L_f(\phi^*) < L_f(\hat{\phi}^\alpha)$. Then

$$L_f(\phi^*) = \int_0^1 L_f(\phi^*)\, d\beta < \int_0^1 L_f(\hat{\phi}^\beta)\, d\beta = L_f(\hat{\phi}),$$

which is impossible because of the optimality of $\hat{\phi}$. It follows that $\hat{\phi}^\alpha$ minimizes (15) over the set of BFC functions, and thus corresponds to a global minimizer of the original non-convex energy (11). □

The above theorem states that we can obtain a global minimizer of (11) from the minimizer of (15) provided that a BFC condition is satisfied. In the case $m = 1$ (e.g. the case that u is scalar–valued), the minimizer of (15) must always satisfy BFC. This trivial observation is formalized below.

Lemma 1. *Consider the case $m = 1$, and the minimizer*

$$\hat{\phi} = arg \min_{\phi \in S_l^+} L_f(\phi).$$

Then the thresholding $\hat{\phi}^\alpha$ satisfies BFC for every $\alpha \in [0, 1]$.

Proof. Since $m = 1$ the definition of S_l^+ guarantees that $D_\gamma^1 \hat{\phi} = -\partial_\gamma \hat{\phi} \geq 0$. This condition guarantees the $\hat{\phi}$ is monotonically decreasing in the γ direction. It follows that, for any $x \in \Omega$, $\hat{\phi}_x^\alpha$ is an indicator function of an interval, and thus satisfies the one dimensional box function condition. □

For $m > 1$, we do not at this time have a theoretical guarantee that the minimizer of (15) will satisfy BFC. This will be a subject of further research. However, in all of the numerical experiments presented below for $m = 2$, it was found that the minimizers did indeed satisfy the necessary box function condition. Furthermore, in case that the BFC is not satisfied at each node, the node can still be assigned a value using the generalized layer cake formula (16).

It is important to note, however, that it is certainly possible to recover non-BFC solutions that do not correspond to a global minimizer. One of the easiest ways to break this correspondence is simply to populate the function f with uniform random numbers, in which case it has been observed that global minimizers cannot be recovered. It seems that the good performance of this method (and the satisfaction of BFC) depends on the fact that image registration problems tend to have relatively smooth, well-defined solutions.

2.3. **Relationship to continuous models.** In this section, we explore the relationship between the convexification model presented above, and other schemes for continuous problems. t

We first address the scheme of Pock et. al [50] described in the introduction. The method presented in [50] was originally only applied to scalar valued minimizations, and so we restrict our discussion to the case $m = 1$. The pock model relies on a lifted minimization of the form

$$\phi^* = arg \min_{0 \leq \phi \leq 1} \int_{\gamma \in [0,\Gamma]} \int_\Omega |\nabla_x \phi(x,\gamma)| + f(x,\gamma) |\partial_\gamma \phi(x,\gamma)|\, dx\, d\gamma \tag{19}$$

The model proposed above involves a minimization of the form

$$\phi^* = arg \min_{\phi \in S_l^+} \int_{\gamma \in [0,\Gamma]} \int_\Omega |\nabla_x \phi(x,\gamma)| - f(x,\gamma) \partial_\gamma \phi(x,\gamma)\, dx\, d\gamma \tag{20}$$

The new model differs from the Pock model in 2 significant ways. First, the new model does not require the absolute value of the difference operator in the fidelity term. Second, the new model minimizes over the set S_l^+, which enforces that $D_\gamma^1\phi^* = -\partial_\gamma\phi^* \geq 0$. As discussed in lemma (1), this inequality constraint guarantees that the optimal function ϕ^* decays monotonically in the γ direction. This condition is necessary for us to be able to recover the primal function u using the layer cake formula. It should be noted that this positivity constraint is only necessary in the case that $d > 1$. It has been shown in [48], that this positivity constraint is automatically enforced in the case $d = 0$.

It is interesting to compare the method presented here to the method presented in [56]. The scheme presents a different representation of the data term f. The lifting proposed in [56] utilizes an ensemble of 1-dimensional functions, one for each component of Γ, rather than a $d-$dimensional relaxation. The result is an algorithm with much lower memory requirements, and higher computational efficiency than the algorithm presented here. However, this efficiency comes at a price. The method proposed in [56] is not a true relaxion of the original problem – i.e. even in cases where the minimizer happens to be binary, it may not correspond to a true solution of the original non-convex problem. The advantages of the present method over that of [56] are stronger theoretical guarantees of optimality – i.e. a global minimizer is recovered in cases where the solution is binary or satisfies BFC. The method [56], on the other hand, has no guarantee of global optimiality even in the case of a binary solution, however it is easier faster to minimize.

3. Numerical results.

We now present a numerical method for minimizing the convex energy (15), and results when this scheme is applied to the optical flow registration problem. The method we use is a variation-type method drawn from continuous optimization theory, and is a multi-dimensional generalization of the method presented in [50].

3.1. Why not graph cuts?

The method presented above allows us to indirectly solve a non-convex problem (2) by solving the corresponding convex problem (15). However, before the minimizer of (15) may be used to recover u, we must apply a thresholding operator to convert it to a binary function. Our reasons for choosing a variational approach to this problem (rather than constraining the solution to be binary valued) is that graph-cuts cannot be applied to the convex problem (15) for $m > 1$.

To show this, we appeal to the result of Boykov, Kolmogorov, and Zabih [41]. In this work, it is shown that sub-modularity is a necessary condition for an energy function to be graph representable. However, the convex energy (15) is not in general sub-modular. We will demonstrate this with a simple example for $m = 2$. In this case, the energy (15) contains terms of the form $D_\gamma^2 g(\gamma) = g(\gamma) + g(\gamma + (1,1)) - g(\gamma + (1,0)) - g(\gamma + (0,1))$. We now choose g_1, and g_2 such that

$$\begin{array}{cc} g_1(0,0) = 1 & g_1(0,1) = 0 \\ g_1(1,0) = 1 & g_1(1,1) = 0 \end{array} \quad \text{and} \quad \begin{array}{cc} g_2(0,0) = 1 & g_2(0,1) = 1 \\ g_2(1,0) = 0 & g_2(1,1) = 0 \end{array} \tag{21}$$

Letting $\vee$ and $\wedge$ denote component-wise maximum and minimum, we then have

$$D_\gamma^2 g_1 \vee g_2 + D_\gamma^2 g_1 \wedge g_2 = 2 > 0 = D_\gamma^2 g_1 + D_\gamma^2 g_2$$

showing that the energy (15) is not sub-modular.

3.2. **Minimization of the convex problem.** In this section, we present a primal-dual technique for minimizing the convex functional (15). Techniques of this type were first used to solve TV regularized problems in [59]. The technique presented here is a generalization of the technique presented in [50] for scalar-valued problems.

Although the primal-dual approach can minimize an arbitrary energy of the form (15), we shall focus on TV regularized problems for clarity of presentation. Such problems have the form

$$L_f(\phi) = \sum_{i=1}^{2} \sum_{0<l<\Gamma_i-1} \sum_{x\in\Omega} \Big\{ |\phi(x+e_1, le_i) - \phi(x, le_i)| \tag{22}$$

$$+|\phi(x+e_2, le_i) - \phi(x, le_i)| \Big\} + \sum_{x\in\Omega,\, \gamma\in\Gamma} f(x,\gamma)|D_\gamma^2 \phi(x,\gamma)| \tag{23}$$

which can be written more compactly as

$$L_f(\phi) = \sum_{i=1}^{2} \sum_{0<l<\Gamma_i-1} \sum_{x\in\Omega} |\nabla_x \phi(x, le_i)| + \sum_{x\in\Omega,\, \gamma\in\Gamma} f(x,\gamma)|D_\gamma^2 \phi(x,\gamma)|. \tag{24}$$

We derive the primal-dual minimization scheme by introducing the dual variable $p = \{p_x, p_\gamma\} = \{p_1, p_2, p_\gamma\}$. Using this variable, the energy (24) can be written

$$L_f(\phi) = \max_{p\in C} \sum_{i=1}^{2} \sum_{0<l<\Gamma_i-1} \sum_{x\in\Omega} p_x(x, l\, e_i) \cdot \nabla_x \phi(x, le_i) + \sum_{x\in\Omega,\, \gamma\in\Gamma} p_\gamma(x,\gamma) D_\gamma^2 \phi(x,\gamma) \tag{25}$$

where the dual variable may range over the set

$$C = \{p : |p_1(x,\gamma)| \le 1, |p_2(x,\gamma)| \le 1, p_\gamma(x,\gamma) \le f(x,\gamma)\}.$$

Minimizing the energy (24) then corresponds to solving the following min-max problem:

$$\min_{\phi\in S_l} \max_{p\in C} \sum_{i=1}^{2} \sum_{0<l<\Gamma_i-1} \sum_{x\in\Omega} p_x(x, l\, e_i) \cdot \nabla_x \phi(x, le_i) + \sum_{x\in\Omega,\, \gamma\in\Gamma} p_\gamma(x,\gamma) D_\gamma^2 \phi(x,\gamma). \tag{26}$$

Using summation by parts, this problem can be written in the equivalent form

$$\min_{\phi\in S_l} \max_{p\in C} \sum_{i=1}^{2} \sum_{0<l<\Gamma_i-1} \sum_{x\in\Omega} -\nabla_x \cdot p_x(x, l\, e_i) \phi(x, le_i) + \sum_{x\in\Omega,\, \gamma\in\Gamma} \bar{D}_\gamma^2 p_\gamma(x,\gamma) \phi(x,\gamma). \tag{27}$$

In the above formula, the discrete divergence operator is defined using backward differences as follows

$$\nabla \cdot p_x(x, l\, e_i) = p_x(x, l\, e_i) - p_x(x - e_1, l\, e_i) + p_x(x, l\, e_i) - p_x(x - e_2, l\, e_i)$$

and the adjoint of the mixed partial difference operator is

$$\bar{D}_\gamma^2 = D_1^- D_2^-$$

where D_i^- denotes the first order backward difference operator in the *ith* direction.

To compute the solution to the saddle-point problem (26), we use a 2 step gradient projection algorithm. On the first step, we minimize with respect to ϕ using a gradient descent step with timestep τ_ϕ. In the second step, we maximize with

respect to p using a gradient ascent step with timestep τ_p. After each step we reproject the variables back into their corresponding feasible sets. We formalize these two steps below.

1. **Primal Step**

   Gradient Descent:

$$\hat{\phi}^{k+1} = \phi^k + \tau_\phi\{\nabla_x \cdot p_x - \bar{D}^2_\gamma p_\gamma\}$$

   Reproject:

$$\phi^{k+1} = \min(1, \max(0, \hat{\phi}^{k+1}))$$

2. **Dual Step**

   Gradient Ascent:

$$\hat{p}^{k+1}_x = p^k_x + \tau_p \nabla_x \phi^{k+1}$$
$$\hat{p}^{k+1}_\gamma = p^k_\gamma + \tau_p D^2_\gamma \phi^{k+1}$$

   Reproject:

$$p^{k+1}_1 = \min(1, \max(-1, \hat{p}^{k+1}_1))$$
$$p^{k+1}_2 = \min(1, \max(-1, \hat{p}^{k+1}_2))$$
$$p^{k+1}_\gamma(x, \gamma) = \min(f(x, \gamma), \hat{p}^{k+1}_\gamma)$$

A derivation of this method using a proximal-point technique can be found in [50] and [48]. Note our algorithm differs from that presented in [50] in that the dual variable p^{k+1}_γ is reprojected into the interval $[-\infty, 1]$ rather than $[-1, 1]$. We have found the primal dual approach to this problem to very efficient and reliable solver for the convex problem (24). Primal dual algorithms of this type were proposed and analyzed by Esser [27], and simultaneously by Chambolle and Pock [16] . For $m = 2$, we have found the method to be convergent for $\tau_\phi \tau_p < \frac{1}{4}$. The numerical examples below were computed using $\tau_\phi = \tau_p = \frac{1}{2}$.

3.3. **Examples.** We now consider several examples which illustrate the effectiveness of our global optimization approach. We will consider synthetic examples, as well as stereo and object tracking examples from the Middlebury optical flow evaluation dataset [2].

The purpose of the synthetic examples is two-fold. First, these synthetic problems are pathologically difficult, and demonstrate the ability of our method to find the global minimum, even in the presence of many local minima. Second, these problems demonstrate the important of the positivity constraint, $D^2_\gamma \geq 0,$.

We first consider the synthetic stereo matching problem, depicted in figure (4). Note that this is a scalar-valued minimization problem because only horizontal shifts are considered. The top and bottom images in figure (4) depict the original and shifted image, respectively. For simplicity, periodic boundary conditions have been used for this problem. This is a pathologically difficult stereo matching problem due to the presence of many local minima, which occur whenever any of the rectangular outlines of the two images are aligned. Random initialization is used for the variables in the primal dual algorithm. The method presented in this paper perfectly maps the shifted image onto the original after 140 iterations of the primal-dual algorithm, and thus exactly recovers the global minimum. It has been shown in [48] that the positivity constraint is not strictly necessary in the scalar valued case. However, the unconstrained algorithm converged in 155 iterations - slightly slower than the constrained version that we propose.

FIGURE 4. A synthetic stereo matching problem. (top) The original/target image, I_0. (bottom) The shifted/source image, I_1. The convex model perfectly matches the source image onto the target after 140 iterations, even with random initialization.

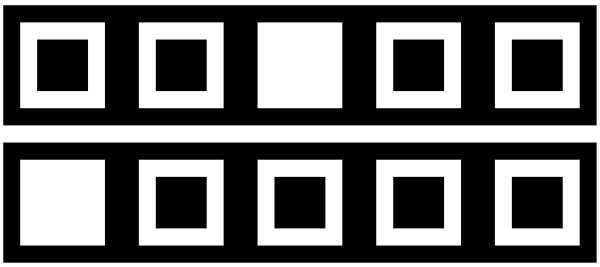
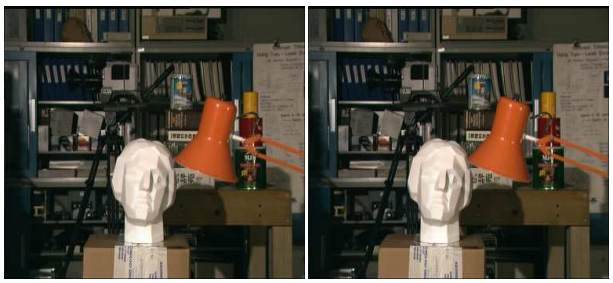


FIGURE 5. Two images taken from the "Tsukuba" stereo matching dataset. These images were taken from slightly different positions, and the deformation field relating them can be interpreted as a depth map.



We demonstrate the performance of this algorithm on stereo matching data using the "Tsukuba" images from the middle optical flow benchmark, depicted in figure 5. Figure 6 shows the resulting depth/displacement field map when the convex algorithm is used. This result was recovery in 454 iterations using our constrained formulation. The unconstrained formulation required 531 iterations to reach the same result.

We next consider the vector-valued analog of the above synthetic example. This is an optical flow registration problem involving both horizontal and vertical translations. Results are depicted in figure (7). The source image has been obtained from the target image by shifting the image in both the horizontal and vertical directions. For simplicity, periodic boundary conditions have been used for this problem. Like the stereo problem, this registration problem has many local minima, which occur whenever any of the rectangular outlines in the source and target images are aligned. Despite this fact, our algorithm obtains the global minimum after 800 iterations of the primal-dual algorithm. To demonstrate the necessity of the non-negativity constraint, $D_\gamma^2 \geq 0$, we compare our convex formulation to the 2-dimensional analog of the model presented in [50]. In this formulation, the mixed partial term is placed within an absolute value, as is done in (19), and minimization is carried out without an inequality constraint. Note that, without the inequality

Figure 6. Depth map for the "Tsukuba" dataset, recovered using the global minimization scheme.

constraint, the algorithm fails to find a global minimizer. Furthermore, results are dependent on initialization. By introducing the positivity constraint, our model finds the exact global minimizer, even when a random initialization is used.

We now apply the global minimization algorithm to two video tracking problems from the Middlebury benchmark dataset. Depicted in figure 8 are two images of a traffic intersection taken from the "dump truck" dataset. Figure 10 depicts the "basketball" dataset. The optical flow model (8) was applied to both datasets using $p = q = \alpha = 1$. The recovered displacement fields are shown in figures (9) and (11). Note that the vehicles correspond to regions of motion, while the background environment remains fixed. Note also that there are few speckle artifacts present in the displacement field because the global method is robust to noise.

4. **Conclusion.** We have presented a technique for minimizing non-convex, vector-valued energy functions defined on Markov random fields using a functional lifting approach. This technique can be viewed as an extension of previous work on functional lifting. However, while previous methods can only handle the case of scalar-valued functions, our technique can be used to minimize over vector-valued functions. This level of generality allows to solve a broad class of non-convex problems such as the image registration problem considered in this paper. There are many other areas where optimization of vector valued functions becomes important. Two such examples are multiphase segmentation problems [12, 48] and diffusion tensor

FIGURE 7. A synthetic registration problem. (left) The original/target image, I_0. (center) The shifted/source image, I_1. (right) Registration using an analog of the algorithm presented in [50] without positivity constraint, using random initialization, after 10000 iterations. If the proper global minimum were found, this image, which represents $I_1(x+u)$, should be identical to I_0. For this problem, registration using our model with positivity constraint perfectly registers the source image onto the target after 800 iterations, even with random initialization.

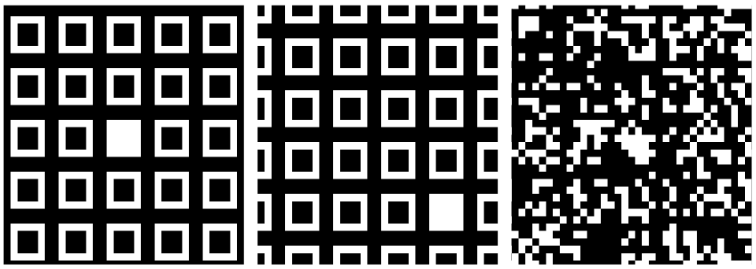


FIGURE 8. Two images of a traffic intersection taken at different times. The cars have moved, while the environment remains fixed. Using the global minimization algorithm, we can recover the displacement field describing the movement of the cars.

imaging. The application of our convexification technique to these problems will be a subject of future research.

REFERENCES

[1] Jean-François Aujol and Antonin Chambolle, *Dual norms and image decomposition models*, Int. J. Comput. Vision, **63** (2005), 85–104.

[2] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black and R. Szeliski, *A database and evaluation methodology for optical flow*, in "IEEE 11th International Conference on Computer Vision," Miami, Florida, 2007.

[3] Rodney J. Baxter, "Exactly Solved Models in Statistical Mechanics," Dover Publications, December, 2007.

[4] Julian Besag, *Spatial interaction and the statistical analysis of lattice systems*, Journal of the Royal Statistical Society Series B, **36** (1974), 192–236.

Figure 9. Displacement field for the "dump truck" images, recovered using the global minimization scheme.

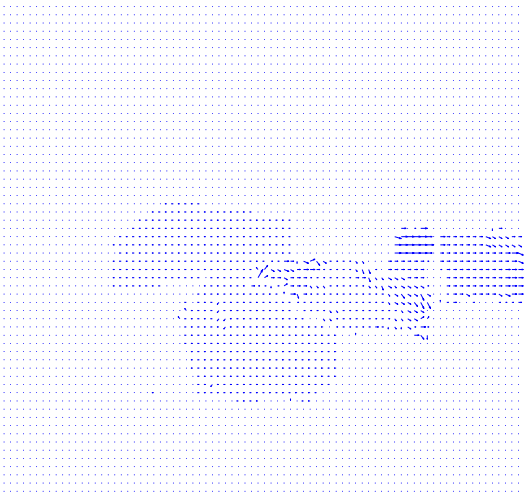
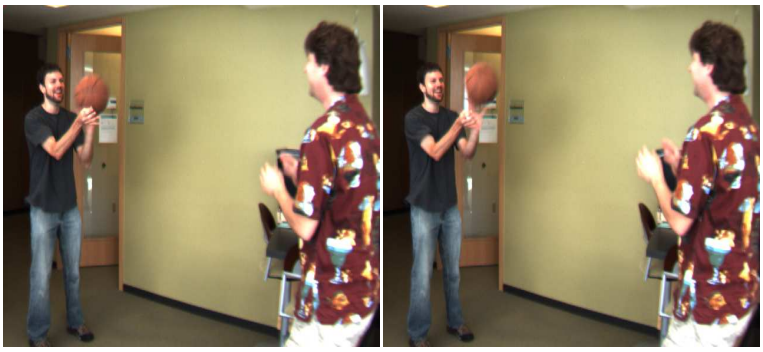


Figure 10. Two images from the "basketball" dataset. Using the global minimization algorithm, we can track the motion of the players and the ball.



[5] Julian Besag, *On the statistical analysis of dirty pictures*, Journal of the Royal Statistical Society Series B, **48** (1986), 259–302.
[6] Andrew Blake and Andrew Zisserman, "Visual Reconstruction," MIT Press Series in Artificial Intelligence, MIT Press, Cambridge, MA, 1987.
[7] Yuri Boykov and Gareth Funka-Lea, *Graph cuts and efficient N-D image segmentation*, Int. J. Comput. Vision, **70** (2006), 109–131.
[8] Yuri Boykov and Vladimir Kolmogorov, *An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision*, IEEE Trans. Pattern Anal. Mach. Intell., **26** (2004), 1124–1137.
[9] Yuri Boykov, Olga Veksler and Ramin Zabih, *Fast approximate energy minimization via graph cuts*, IEEE Trans. Pattern Anal. Mach. Intell., **23** (2001), 1222–1239.

FIGURE 11. Displacement field for the "basketball" dataset, recovered using the global minimization scheme.

[10] Y. Y. Boykov and M.-P. Jolly, *Interactive graph cuts for optimal boundary region segmentation of objects in N-D images*, in "Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference," Vol. 1, (2001), 105–112.

[11] Xavier Bresson, Selim Esedoḡlu, Pierre Vandergheynst, Jean-Philippe Thiran and Stanley Osher, *Fast global minimization of the active contour/snake model*, Journal of Mathematical Imaging and Vision, **28** (2007), 151–167.

[12] Ethan S. Brown, Tony F. Chan and Xavier Bresson, *A Convex Approach for Multi-phase Piecewise Constant Mumford-Shah Image Segmentation*, UCLA CAM Report 09-66, July, 2009.

[13] Andr Bruhn, Joachim Weickert, Timo Kohlberger and Christoph Schnr, *Discontinuity-preserving computation of variational optic flow in real-time*, in "Scale-Space and PDE Methods in Computer Vision," Lecture Notes in Computer Science, **3459**, Springer, (2005), 279–290.

[14] Antonin Chambolle, *An algorithm for mean curvature motion*, Interfaces Free Bound, **6** (2004), 195–218.

[15] Antonin Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vis., **20** (2004), 89–97.

[16] Antonin Chambolle and Thomas Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision, **40** (2011), 120–145.

[17] Tony Chan, Selim Esedoḡlu and Mila Nikolova, *Algorithms for finding global minimizers of image segmentation and denoising models*, SIAM Journal on Applied Mathematics, **66** (2006), 1632–1648.

[18] Tony F. Chan, Gene H. Golub and Pep Mulet, *A nonlinear primal-dual method for total variation-based image restoration*, SIAM J. Sci. Comput., **20** (1999), 1964–1977.

[19] Isaac Cohen, *Nonlinear variational method for optical flow computation*, in "Proc. Eighth Scandinavian Conference on Image Analysis," Vol. 1, Tromso, Norway, May, (1993), 523–530.

[20] J Darbon and M Sigelle, *A fast and exact algorithm for total variation minimization*, IbPRIA, **3522** (2005), 351–359.

[21] Jérôme Darbon, *Global optimization for first order Markov random fields with submodular priors*, Discrete Applied Mathematics, June, 2009.

[22] Jerome Darbon and Marc Sigelle, *A fast and exact algorithm for total variation minimization*, IbPRIA 2005, **3522** (2005), 351–359.
[23] Jérôme Darbon and Marc Sigelle, *Image restoration with discrete constrained total variation part I: Fast and exact optimization*, J. Math. Imaging Vis., **26** (2006), 261–276.
[24] Jérôme Darbon and Marc Sigelle, *Image restoration with discrete constrained total variation part II: Levelable functions, convex priors and non-convex cases,* J. Math. Imaging Vis., **26** (2006), 277–291.
[25] Q. Duan, E. Angelini, S. Homma and A. Laine, *Tracking endocardium using optical flow along iso-value curve*, in "Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE," (2006), 707–710.
[26] E. Esser, *Applications of lagrangian-based alternating direction methods and connections to split Bregman*, UCLA CAM technical report 09-31, 2009.
[27] Ernie Esser, Xiaoqun Zhang and Tony Chan, *A general framework for a class of first order primal-dual algorithms for TV minimization*, UCLA CAM Report 09-67, 2009.
[28] Herbert Federer, "Geometric Measure Theory," Die Grundlehren der mathematischen Wissenschaften, Band 153, Springer-Verlag New York Inc., New York, 1969.
[29] L. R. Ford, Jr. and D. R. Fulkerson, *Maximal flow through a network*, Canadian Journal of Mathematics, **8** (1956), 399–404.
[30] Stuart Geman, Donald Geman, K. Abend, T. J. Harley and L. N. Kanal, *Stochastic relaxation, gibbs distributions and the bayesian restoration of images*, Journal of Applied Statistics, **20** (1993), 25–62.
[31] Guy Gilboa and Stanley Osher, *Nonlocal linear image regularization and supervised segmentation*, Multiscale Model. Simul., **6** (2007), 595–630.
[32] Tom Goldstein and Stanley Osher, *The split Bregman method for l1 regularized problems*, UCLA CAM Report 08-29, 2008.
[33] D. M. Greig, B. T. Porteous and A. H. Seheult, *Exact maximum a posteriori estimation for binary images*, Journal of the Royal Statistics Society, **51** (1989), 271–279.
[34] N. Hata, A. Nabavi, W. Wells, S. Warfield, R. Kikinis, P. Black and F. Jolesz, *Three-dimensional optical flow method for measurement of volumetric brain deformation from intraoperative MR images,* J. Comput. Assist. Tomogr., 07 2000.
[35] F. Heitz and P. Bouthemy, *Multimodal estimation of discontinuous optical flow using Markov random fields*, IEEE Trans. Pattern Anal. Mach. Intell., **15** (1993), 1217–1232.
[36] Berthold K. P. Horn and Brian G. Schunck, *Determining optical flow*, Technical report, Massachusetts Institute of Technology, Cambridge, MA, 1980.
[37] Hiroshi Ishikawa, *Exact optimization for Markov random fields with convex priors,* IEEE Transactions on Pattern Analysis and Machine Intelligence, **25** (2003), 1333–1336.
[38] Satoru Iwata, Lisa Fleischer and Satoru Fujishige, *A combinatorial strongly polynomial algorithm for minimizing submodular functions*, Journal of the ACM, **48** (2000), 761–777.
[39] Stephen L. Keeling and Wolfgang Ring, *Medical image registration and interpolation by optical flow with maximal rigidity*, J. Math. Imaging Vis., **23** (2005), 47–65.
[40] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, *Optimization by simulated annealing*, Science, **220** (1983), 671–680.
[41] Vladimir Kolmogorov and Ramin Zabih, *What energy functions can be minimized via graph cuts?*, in "Computer Vision - ECCV 2002: 7th European Conference on Computer Vision," Copenhagen, Denmark, May 28-31, 2002, Proceedings, Part III, (2002), 185–208.
[42] Vladimir Kolmogorov and Ramin Zabih, *What energy functions can be minimized via graph cuts*, IEEE Trans. Pattern Anal. Mach. Intell., (2004), 147–159.
[43] R. Malladi, R. Kimmel, D. Adalsteinsson, G. Sapiro, V. Caselles and J. A. Sethian, *A geometric approach to segmentation and analysis of 3d medical images*, in "MMBIA '96: Proceedings of the 1996 Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA '96)," IEEE Computer Society, Washington, DC, USA, (1996), 244.
[44] H. Nagel and W. Enkelmann, *An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences*, IEEE Trans. Pattern Anal. Mach. Intell., **8** (1986), 565–593.
[45] J. P. Oliveira, J. M. Bioucas-Dias and M. Figueiredo, *Adaptive total variation image deblurring: A majorization-minimization approach*, Signal Process., **89** (2009), 1683–1693.
[46] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu and Wotao Yin, *An iterative regularization method for total variation-based image restoration*, MMS, **4** (2005), 460–489.

[47] J. C. Picard and H. D. Ratliff, *Minimum cuts and related problems*, Networks, **5** (1975), 357–370.
[48] Thomas Pock, Antonin Chambolle, H. Bischof and D. Cremers, *A convex relaxation approach for computing minimal partitions*, In "IEEE Conference on Computer Vision and Pattern Recognition (CVPR)," Miami, Florida, 2009.
[49] Thomas Pock, Daniel Cremers, Horst Bischof and Antonin Chambolle, *Global solutions of variational models with convex regularization*, SIAM J. Imaging Sci., **3** (2010), 1122–1145.
[50] Thomas Pock, Thomas Schoenemann, Gottfried Graber, Horst Bischof and Daniel Cremers, *A convex formulation of continuous multi-label problems*, in "ECCV '08: Proceedings of the 10th European Conference on Computer Vision," Springer-Verlag, Berlin, Heidelberg, 2008, 792–805.
[51] Renfrey B. Potts, *Some generalized order-disorder transformations,* Proceedings of the Cambridge Philosophical Society, **48** (1952), 106–109.
[52] Marc Proesmans, Luc J. Van Gool, Eric J. Pauwels and André Oosterlinck, *Determination of optical flow and its discontinuities using non-linear diffusion*, in "ECCV '94: Proceedings of the Third European Conference-Volume II on Computer Vision," Springer-Verlag, London, UK, (1994), 295–304.
[53] L. Rudin, S. Osher and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica. D., **60** (1992), 259–268.
[54] Alexander Schrijver, *A combinatorial algorithm minimizing submodular functions in strongly polynomial time*, J. Comb. Theory Ser. B, **80** (2000), 346–355.
[55] S. Setzer, *Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage*, in "Proceedings of the Second International Conference on Scale Space Methods and Variational Methods in Computer Vision," 2009.
[56] E. Strekalovskiy, B. Goldluecke and D. Cremers, *Tight convex relaxations for vector-valued labeling problems*, in "IEEE International Conference on Computer Vision (ICCV)," IEEE, Nov., (2011), 2328–2335.
[57] Y. Wang, W. Yin and Y. Zhang, *A fast algorithm for image deblurring with total variation regularization*, CAAM Technical Reports, 2007.
[58] Joachim Weickert and Christoph Schnörr, *A theoretical framework for convex regularizers in pde-based computation of image motion*, Int. J. Comput. Vision, **45** (2001), 245–264.
[59] Mingqiang Zhu and Tony Chan, *An efficient primal-dual hybrid gradient algorithm for total variation image restoration*, UCLA CAM technical report 08-34, 2008.
[60] Jie Zhu-Jacquot, *Graph cuts segmentation with geometric shape priors for medical images*, in "Image Analysis and Interpretation," SSIAI 2008, IEEE Southwest Symposium, March, (2008), 109–112.

Received September 2011; revised June 2012.

E-mail address: tag7@rice.edu
E-mail address: xbresson@cityu.edu.hk
E-mail address: sjo@math.ucla.edu