Fast and Exact Primal-Dual Iterations for Variational Problems in Computer Vision

Jan Lellmann, Dirk Breitenreicher, and Christoph Schnörr

Image and Pattern Analysis Group & HCI
Dept. of Mathematics and Computer Science, University of Heidelberg
{lellmann,breitenreicher,schnoerr}@math.uni-heidelberg.de

Abstract. The saddle point framework provides a convenient way to formulate many convex variational problems that occur in computer vision. The framework unifies a broad range of data and regularization terms, and is particularly suited for nonsmooth problems such as Total Variation-based approaches to image labeling. However, for many interesting problems the constraint sets involved are difficult to handle numerically. State-of-the-art methods rely on using nested iterative projections, which induces both theoretical and practical convergence issues. We present a dual multiple-constraint Douglas-Rachford splitting approach that is globally convergent, avoids inner iterative loops, enforces the constraints exactly, and requires only basic operations that can be easily parallelized. The method outperforms existing methods by a factor of $4-20$ while considerably increasing the numerical robustness.

1 Introduction

Overview and Motivation. In this work, we focus on algorithms for solving saddle point problems associated with variational formulations in image processing and analysis, which have recently become a very active research area. The output of a variational method is defined as the minimizer

$$u^* := \arg\min_{u \in \mathcal{C}} f(u) \,, \tag{1}$$

where $\mathcal{C}$ is some subset of a space of functions defined on some continuous domain, and f a functional depending on the input data. In contrast to "discretize first" approaches such as grid- or graph based methods, this "analyze first" approach allows to get a deeper insight into the underlying problem, and to abstract from inaccuracies caused by the discretization.

The interpretation of u is governed by the problem to be solved: for color denoising, $u : \Omega \to [0,1]^3$ could directly describe the colors of the output image on the image domain $\Omega \subseteq \mathbb{R}^d$; while for segmentation problems, $u : \Omega \to [0,1]$ could assign each point to the foreground $(u(x) = 1)$ or background $(u(x) = 0)$ class. Recently, interest has risen in a specific class of variational problems of the form

$$\inf_{u \in \mathcal{C}} \sup_{v \in \mathcal{D}} \left\{ \langle u, s \rangle + \langle Lu, v \rangle - \langle b, v \rangle \right\} \,, \tag{2}$$

where the *primal* and *dual constraint sets* $\mathcal{C} \subseteq X$ and $\mathcal{D} \subseteq Y$ are convex subsets

Fig. 1. Application of the proposed saddle point optimization method to multi-class color segmentation. **Left:** Input image. **Right:** Segmentation into 12 regions of constant color. The tight relaxation of the combinatorial labeling problem results in a saddle point problem with an intricate dual constraint set. In contrast to existing approaches, the method proposed in this work allows to compute global minimizers of such problems without requiring inaccurate and time-consuming iterative projections as subroutines.

of some function space X with dual space Y, $L : X \to Y$ is a linear operator, $s \in Y$ and $b \in X$. These *bilinear saddle point problems* are very useful in the context of labeling [4, 12, 14], and – using a "lifting" technique – can be used to minimize a large class of common variational problems [18].

As these problems are generally convex, they do not suffer from local minima, which allows to clearly separate modelling from optimization aspects. The inner problem turns out to be a convenient way of expressing objective functions f that contain *non-smooth* terms, such as Total Variation (TV) regularization, and allows to apply fast *primal-dual* optimization schemes that explicitly update the primal variables u as well as the dual variables v.

First-order methods of this kind have been shown to achieve a good performance for many problems while offering excellent parallelization characteristics [22, 17, 14]. These methods require to compute projections $\Pi_{\mathcal{C}}$ and $\Pi_{\mathcal{D}}$ on the sets $\mathcal{C}$ and $\mathcal{D}$. However, in many cases one faces discretized problems of the form

$$\min_{u \in \mathcal{C}} \max_{v \in \mathcal{D}_1 \cap \ldots \cap \mathcal{D}_r} \left\{ \langle u, s \rangle + \langle Lu, v \rangle - \langle b, v \rangle \right\} , \tag{3}$$

with $\mathcal{C} \subseteq \mathbb{R}^n$ and $\mathcal{D}_i \subseteq \mathbb{R}^m, i = 1, \ldots, r$. This occurs in particular in connection with relaxations of the combinatorial labeling problem (Fig. 1) and functional lifting [4, 18, 14]. Here the dual constraint set $\mathcal{D}$ is only given implicitly as an intersection, hence projections cannot be computed in closed form.

Current methods to solve such problems are based on *approximating* the projection on $\mathcal{D}$ by a series of projections on the simpler sets $\mathcal{D}_i$. However, this causes a number of issues. From a theoretical viewpoint, convergence of the outer algorithms usually requires the inner problem to be solved with an increasing accuracy at each step, which is impractical. Thus in practice convergence is no longer guaranteed. In addition, the projections become very slow, and raise many issues on how to choose suitable and matching stopping criteria.

Contribution. In this work, we propose a dual multiple-constraint Douglas-Rachford method for saddle point problems of the class (3), that *exactly* takes into account the dual constraint set $\mathcal{D}$ while still relying only on simple exact operations. The method is shown to converge to a global optimum and is suited for massive parallelization. While the method essentially solves the dual problem, we show that a primal solution can be recovered. As all steps in the proposed algorithm can be computed explicitly, the theoretical convergence results directly transfer to the actual implementation. The approach outperforms state-of-the-art methods on real-world problems with respect to computation time and numerical robustness by a factor of $4 - 20$.

Related Work. Continuous labeling approaches [21, 5] constitute a continuous equivalent to discrete graph cut methods [3]. These discrete methods are difficult to parallelize and suffer from anisotropy induced by the discretization. This *grid bias* can be reduced in some extent by using larger neighborhoods in the graph construction, but it cannot be completely eliminated and computational costs quickly increase in the process. In contrast, continuous methods can be used to construct discretizations that exactly represent the original metric in an infinitesimal sense [4]. The idea of functional lifting can be found in a discrete setting in [11] and in a continuous formulation in [4, 17], and has also proven to be useful in the context of optical flow [10].

Regarding optimization, our work extends the approach proposed in [9] for two-, and in [14] for multiclass labeling. The authors use a similar method, but require iterative projections at each step. The basic Douglas-Rachford iteration [6, 7] applied to the dual problem can be shown to be equivalent to the Alternating Direction Method of Multipliers [8] and the recently proposed Alternating Split Bregman method [9, 20], hence our results equally apply in these formulations.

2   Bilinear Saddle-Point Problems in Computer Vision

In the following, we will consider variational problems that can be stated in the specific saddle point form (3) when discretized. For $s \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $L \in \mathbb{R}^{m \times n}$, and some closed convex sets $\mathcal{C} \subseteq \mathbb{R}^n$ and $\mathcal{D}_i \subseteq \mathbb{R}^m, i = 1, \ldots, r$, define $\mathcal{D} := \mathcal{D}_1 \cap \ldots \cap \mathcal{D}_r$ and

$$g(u, v) := \langle u, s \rangle + \langle Lu, v \rangle - \langle b, v \rangle. \tag{4}$$

Then problem (3) consists in computing a minimizer of the *primal objective* $f(u) := \max_{v \in \mathcal{D}} g(u, v)$,

$$\min_{u \in \mathcal{C}} \max_{v \in \mathcal{D}} g(u, v) = \min_{u \in \mathcal{C}} f(u). \tag{5}$$

Under the assumption that at least one of the sets $\mathcal{C}$ and $\mathcal{D}$ is bounded, it can be shown that equivalently one may maximize the *dual objective* $f_d(v) := \min_{u \in \mathcal{C}} g(u, v)$ [19, Cor. 37.3.2],

$$\max_{v \in \mathcal{D}} \min_{u \in \mathcal{C}} g(u, v) = \max_{v \in \mathcal{D}} f_d(v). \tag{6}$$

In particular, pairs of primal resp. dual solutions (u^*, v^*) are *saddle points* of g,

$$\min_{u \in \mathcal{C}} f(u) = f(u^*) = g(u^*, v^*) = f_d(v^*) = \max_{v \in \mathcal{D}} f_d(v) \qquad (7)$$

We will now present two prototypical applications of the saddle point method: multiclass image labeling and generic scalar variational models with gradient-based regularizers.

Continuous Multiclass Labeling Approaches. Many problems in image analysis can be reduced to the basic problem of assigning to each point x in the image domain Ω one of l discrete labels $\{1, \ldots, l\}$, such as an object class in segmentation problems, a depth label in stereo reconstruction, or a displacement vector in image registration [16]. In order to reduce the influence of noise, some nonlocal spatial coherency constraints are required in addition to the local data fidelity measure based on the input image.

As for each point a discrete decision must be made, the problem is combinatorial and nonconvex, and in fact can be shown to be NP-hard even for relatively simple energies under a graph discretization [3]. However, by relaxing the original problem to a convex constraint set, good solutions for the original problem can be recovered using convex optimization [22, 4, 13, 14]. In the continuous setting, the labeling problem can be relaxed to the variational problem

$$\min_{u \in \mathcal{C}} \langle u, s \rangle + J(u), \quad \mathcal{C} := \{u \in \mathrm{BV}(\Omega, \mathbb{R}^l) | u(x) \geqslant 0, \sum_i u_i(x) = 1\}, \qquad (8)$$

where BV denotes the space of functions of bounded variation [2]. By embedding the original labels into a higher-dimensional space via the unit vectors $\{e^1, \ldots, e^l\}$, the local data fidelity can be completely encoded into the linear term, irrespective of the complexity of the original data term: assigning label i to the point x will locally be penalized by $s_i(x)$.

For the nonlocal regularizer J, we choose some metric $d : \{1, \ldots, l\}^2 \to \mathbb{R}$, denote by Du the (distributional) Jacobian of u, and set

$$J(u) := \sup_{v \in \mathcal{D}} \int_{\Omega} \langle Du, v \rangle, \quad \mathcal{D} := \{v \in (C_c^\infty)^{d \times l} | v(x) \in \mathcal{D}_{\mathrm{loc}} \forall x \in \Omega\}, \qquad (9)$$

$$\mathcal{D}_{\mathrm{loc}} := \{v = (v^1, \ldots, v^l) \in \mathbb{R}^{d \times l} | \|v^i - v^j\| \leqslant d(i, j), \sum_k v^k = 0\}. \qquad (10)$$

This is a tight relaxation of the requirement that switching from label i to label j along some curve should be penalized by the curve length, multiplied by a factor $d(i, j)$ *depending on the labels i and j*. In terms of graph-based approaches, this can be thought of as the potentials on the edges of the graph. The formulation (9) carries over this principle to the continuous domain Ω. By discretizing u, v and s on a rectangular grid and choosing a forward finite differences discretization L of the gradient operator D, the above variational formulation can be posed in the saddle point form (3) without introducing grid bias (cf. [4]).

The definition of $\mathcal{D}_{\mathrm{loc}}$ is derived by locally constructing the convex envelope of the desired regularizer restricted to the set of u that only assume the "hard"

labels $\{e^1, \ldots, e^l\}$. As a result, the minimizer of the *convex* problem (8) is often a unit vector in almost all points, and provides a very good approximation to the solution of the original *combinatorial* labeling problem.

The increased *approximation tightness* comes at the price of a more complicated *optimization* process. However, as $\mathcal{D}_{\mathrm{loc}}$ is the intersection

$$\mathcal{D}_{\mathrm{loc}} = \{v \in \mathbb{R}^{d \times l} | \sum_i v^i = 0\} \cap \bigcap_{i<j} \{v \in \mathbb{R}^{d \times l} | \|v^i - v^j\| \leqslant d(i,j)\}, \qquad (11)$$

the problem can be put into the form (3). Projections on the *individual* sets can be easily computed by subtracting the mean resp. by shrinkage-like operations.

Lifting Approach. For the case where the sought-after function takes scalar values, such as gray scale or depth, the saddle point formulation permits another interesting application. Assume we want to minimize over $\mathcal{C} \subseteq W^{1,1}(\Omega, \mathbb{R})$ some functional

$$\min_{u' \in \mathcal{C}} f'(u'), \quad f'(u') := \int_\Omega h(x, u'(x), \nabla u'(x)) dx \qquad (12)$$

with h convex in $\nabla u'(x)$, but not necessarily in $u'(x)$. Then, motivated by the "calibration" idea [1], it was shown in [18] that f' can be expressed in terms of the $\{0, 1\}$-indicator function $\chi_{H(u')}$ of the hypograph

$$H(u') := \left\{(x, t) \in \mathbb{R}^d \times \mathbb{R} | u'(x) \geqslant t\right\} \qquad (13)$$

of u', i.e. $\chi_{H(u')}(x, t) = 1$ iff $u'(x) \geqslant t$ and zero otherwise. Specifically,

$$f'(u') = \sup_{v \in \mathcal{D}} \int_{\Omega \times \mathbb{R}} \langle v, D\chi_{H(u')} \rangle, \quad \text{where} \qquad (14)$$

$$\mathcal{D} := \left\{(v^x, v^t) \in \mathcal{C}_c^\infty | \forall x \in \Omega, t \in \mathbb{R} : v^t(x, t) \geqslant h^*(x, t, v^x(x, t))\right\}. \qquad (15)$$

Here h^* denotes the convex conjugate of h with respect to the last argument. Intuitively, this *lifts* the problem to a higher-dimensional space and transforms it to the problem of finding the *set of points below the graph* of u'.

Again, the problem is transformed to a convex problem by replacing $\chi_{H(u')}$ with some function $u : \Omega \to [0, 1]$. This effectively linearizes the nonconvexity of h with respect to $u(x)$. The relaxed problem reads

$$\min_{u \in \mathcal{C}} \sup_{v \in \mathcal{D}} \int_{\Omega \times \mathbb{R}} \langle v, Du \rangle, \quad \mathcal{C} := \{u \in \mathrm{BV}(\Omega \times \mathbb{R}, [0, 1]) | u(x, t) \stackrel{t \to \pm \infty}{\Longrightarrow} 0/1\}, \quad (16)$$

which after discretization fits into the saddle point framework (3). Again, depending on the *integrand h*, the *dual constraint set* $\mathcal{D}$ may be very complicated. The approach can be extended to the full Mumford-Shah functional [15],

$$f'(u') = \lambda \int_\Omega (f - u)^2 dx + \int_{\Omega \setminus S_{u'}} \|\nabla u\|^2 dx + \nu \mathcal{H}^{d-1}(S_{u'}), \quad \lambda, \nu > 0, \quad (17)$$

where $\mathcal{H}^{d-1}$ is the $(d-1)$-dimensional Hausdorff measure (Fig. 2). The $W^{1,1}$ requirement above is relaxed to $u' \in \mathrm{SBV}(\Omega \times \mathbb{R})$, i.e. the set of special functions

Fig. 2. Application of the proposed optimization method to nonsmooth variational denoising. **Left:** Input image. **Right:** Result of variational denoising using the Mumford-Shah functional with 8 levels, $\lambda = 0.5$ and $\nu = 5$. Noise or fine details can be removed without blurring sharp edges. The lifting approach allows to minimize the full Mumford-Shah functional within the convex saddle point framework.

of bounded variation [2], such that u' may have a nonempty set of discontinuities $S_{u'}$. The dual constraint set then becomes [17]

$$\mathcal{D} = C_c^\infty(\Omega \times \mathbb{R}, \mathbb{R}^{d+1}) \cap \mathcal{R} \cap \bigcap_{p \leqslant q} \mathcal{S}_{p,q}\,, \tag{18}$$

$$\mathcal{R} := \left\{ (v^x, v^t) | v^t(x,t) + \lambda(t - f(x))^2 \geqslant \frac{v^x(x,t)^2}{4} \; \forall x, t \right\} \tag{19}$$

$$\mathcal{S}_{p,q} := \left\{ (v^x, v^t) | \left| \int_p^q v^x(x,t) dt \right| \leqslant \nu \; \forall x, t \right\}\,. \tag{20}$$

Again, projections on the discrete counterpart of $\mathcal{D}$ can only be *approximated*. On the other hand, projections on $\mathcal{S}_{p,q}$ and $\mathcal{R}$ can be computed explicitly by using a shrinkage-like method [4] resp. by solving a third-order polynomial using a solution formula. This motivates our optimization approach below that *exactly* takes $\mathcal{D}$ into account in terms of individual projections onto $\mathcal{S}_{p,q}$ and $\mathcal{R}$.

3   Dual Multiple-Constraint Douglas-Rachford Splitting

Based on the theory of set-valued operators applied to the subdifferential operators of convex functions, the Douglas-Rachford approach [6] provides a scheme to compute a minimizer of the problem

$$\min_{u \in \mathbb{R}^n} f(u), \quad f(u) := f_1(u) + f_2(u)\,, \tag{21}$$

by iterating a combination of backward (proximal) steps with step size $\tau > 0$,

$$u' \leftarrow \arg \min_{u' \in \mathbb{R}^n} \left\{ (2\tau)^{-1} \| u' - u \|_2^2 + f_i(u') \right\} \tag{22}$$

Algorithm 1 Dual Multiple-Constraint Douglas-Rachford Optimization for Saddle-Point Problems (DMDR)

1: Choose $\tau > 0, \bar{v}_i^0 \in \mathbb{R}^{n \times d \times l}, \bar{z}^0 \in \mathbb{R}^{n \times d}$. Set $k \leftarrow 0$.
2: **while** (not converged) **do**
3:     $v_i^k \leftarrow \Pi_{\mathcal{D}_i} \left(\bar{v}_i^k - \frac{\tau}{r} b \right)$.
4:     $z''^k \leftarrow \Pi_{\mathcal{C}} \left(\frac{1}{\tau} \left(\bar{z}^k - s \right) \right)$.
5:     $v'^k \leftarrow \left(rI + LL^\top \right)^{-1} \left(\sum_i \left(2v_i^k - \bar{v}_i^k \right) - L \left(\bar{z}^k - 2\tau z''^k \right) \right)$.
6:     $v_1'^k = \ldots = v_r'^k \leftarrow v'^k$.
7:     $z'^k \leftarrow \left(-L^\top \right) v'^k$.
8:     $\bar{v}_i^{k+1} \leftarrow \bar{v}_i^k + v_i'^k - v_i^k$.
9:     $\bar{z}^{k+1} \leftarrow z'^k + \tau z''^k$.
10:    $k \leftarrow k + 1$.
11: **end while**

---

on each of the f_i *individually*. More precisely, if both f_1 and f_2 are proper, convex, and lower semicontinuous functions, and the relative interiors of their domains have a nonempty intersection, the Douglas-Rachford iteration scheme converges to a minimizer of f [7, Thm. 3.15; Prop. 3.23, 3.20, 3.19]. A strong point of the method is that it does not require any part of the objective to be smooth or finite, which allows to introduce constraints into the f_i as required.

Algorithm and Convergence. We will now show how to add auxiliary variables before splitting the objective in order to avoid the iterative projections employed in [4, 18] and the associated accuracy and convergence issues. Instead of solving (5) directly, we solve the *dual* problem (6) and additionally introduce auxiliary variables z and $v_1, \ldots, v_r$, leading to the equivalent problem

$$\min_{v_i \in \mathbb{R}^m} \underbrace{\delta_{-L^\top \left(\frac{1}{r} \sum_i v_i \right) = z, v_1 = \ldots = v_r}}_{f_1} + \underbrace{\sum_i \delta_{v_i \in \mathcal{D}_i} + \left\langle \frac{1}{r} \sum_i v_i, b \right\rangle + \max_{u \in \mathcal{C}} \langle u, z - s \rangle}_{f_2} . \quad (23)$$

The extra constraints are represented as characteristic functions δ taking values $\{0, +\infty\}$. Applying the Douglas-Rachford method to the above splitting formulation leads to the complete algorithm as outlined in Alg. 1. Due to the auxiliary variables, the backward step for f_2 requires only separate projections on the $\mathcal{D}_i$ instead of the complete set $\mathcal{D}$. The backward step for f_1 amounts to solving a linear equation system. By the Woodbury identity, this can be transformed to

$$\left(rI + LL^\top \right)^{-1} x = r^{-1} x - r^{-1} L \left(rI + L^\top L \right)^{-1} L^\top x . \quad (24)$$

In all of the presented applications, L is a forward differences discretization of the gradient. Thus $LL^\top$ is the five-point Laplacian and diagonalizes with respect to the discrete cosine transform, allowing to solve (24) fast and exact using DCT and diagonal matrix multiplications. We now show convergence of Alg. 1 subject to a mild condition on the relative interiors ri of the domains.

Proposition 1. *Let $\mathcal{D}_1, \ldots, \mathcal{D}_r, \mathcal{C}$ be closed convex sets, $\mathcal{C}$ bounded such that $\mathrm{ri}(\mathcal{D}_1) \cap \ldots \cap \mathrm{ri}(\mathcal{D}_r) \neq \emptyset$ and $\mathrm{ri}(\mathcal{C}) \neq \emptyset$. Then Alg. 1 converges in $(v_1^k, \ldots, v_r^k, z''^k)$.*

Proof. As $\mathcal{C}$ is closed we have $\mathrm{ri}(\mathrm{dom}\, f_2) \cap \mathrm{ri}(\mathrm{dom}\, f_1) = \mathrm{ri}(\mathrm{dom}\, f_2) \cap \{v_1 = \ldots = v_r, -A^\top v_i = z\} = \{(v, \ldots, v, -A^\top v)^\top\, |v \in \mathrm{ri}(\mathcal{D}_1) \cap \ldots \cap \mathrm{ri}(\mathcal{D}_r)\}$. This set is nonempty by the assertion, which with the remarks at the beginning of the section implies convergence. $\square$

Duality Properties of the Proposed Method. In particular, the convergence property of the Douglas-Rachford approach guarantees that from some point on the *constraints hold exactly*. Then $v^k := v_1^k = \ldots = v_r^k$, and v^k converges to a solution v of the dual problem (6). Unfortunately, it is nontrivial to generate a primal solution u from a single dual solution, as both the dual and the primal problem are usually not strictly convex. However, it turns out that the above algorithm additionally returns a primal solution:

Proposition 2. *Let $(v := v_1 = \ldots = v_r, z'')$ be a fixed point of Alg. 1. Then z'' is a solution of the primal problem (5).*

Proof. We will only provide a sketch the proof as it is quite technical. The point is to show that the limit (z'', v) of Alg. 1 is a saddle point of $g(u, v)$ as defined in (4), i.e.

$$g(u, \tilde{v}) \leqslant g(z'', v) \leqslant g(\tilde{u}, v) \; \forall \tilde{u} \in \mathcal{C}, \tilde{v} \in \mathcal{D} \,. \tag{25}$$

Let $\bar{z}$ and $\bar{v}_i$ be the corresponding limits from Alg. 1, and substitute $z := \bar{z} - \tau z''$. Denoting by $\partial f(x)$ the subdifferential (i.e. the set of subgradients) of f in x, from the Douglas-Rachford convergence theorem [7, Prop. 3.19], it follows that

$$\tau^{-1} \left(\bar{v}_1 - v_1, \ldots, \bar{v}_r - v_r, \bar{z} - z\right)^\top \in \partial f_2(v_1, \ldots, v_r, z) \,. \tag{26}$$

Summing up and using the definition of the algorithm leads to

$$Lz'' = \tau^{-1} \sum_i (\bar{v}_i - v_i) \in \sum_i N_{\mathcal{D}_i}(v_i) + b = N_{\mathcal{D}}(v) + b, \tag{27}$$

where $N_{\mathcal{D}}$ denotes the normal cone of the set $\mathcal{D}$ from convex analysis. On the other hand, from (26) we get

$$\tau^{-1} \left(\bar{z} - z\right) \in \arg\max_{u \in \mathcal{C}} \langle u, z - s \rangle, \quad \text{i.e.} \quad z'' \in \arg\max_{u \in \mathcal{C}} \langle u, -L^\top v - s \rangle \,. \tag{28}$$

Together, (27) and (28) show the saddle point property of (z'', v). Thus z'' must be a primal solution. $\square$

By duality, the same scheme can be applied to solve problems where the *primal* constraint set is more complicated, i.e. $\mathcal{C} = \mathcal{C}_1 \cap \ldots \cap \mathcal{C}_r$. Also note that for $r = 1$, the algorithm reduces to the Douglas-Rachford method from [14]. In case both f and f_d can be numerically evaluated, the *gap* $f(z''^k) - f_d(v^k)$ provides a strong stopping criterion, as for any solution u^* and dual feasible point $v^k \in \mathcal{D}$,

$$f(z''^k) - f(u^*) \leqslant f(z''^k) - f_d(v^k) \,. \tag{29}$$

In practice, it is often better to stop depending on the *relative gap* $(f(u) - f_d(v))/f_d(v)$, which overestimates the actual gap and provides some scale invariance. However, in our case f usually cannot be evaluated due to the complexity of $\mathcal{D}$, and we must resort to a more elementary stopping criterion such as the difference between two consecutive iterates, $\|z''^k - z''^{k-1}\|$.
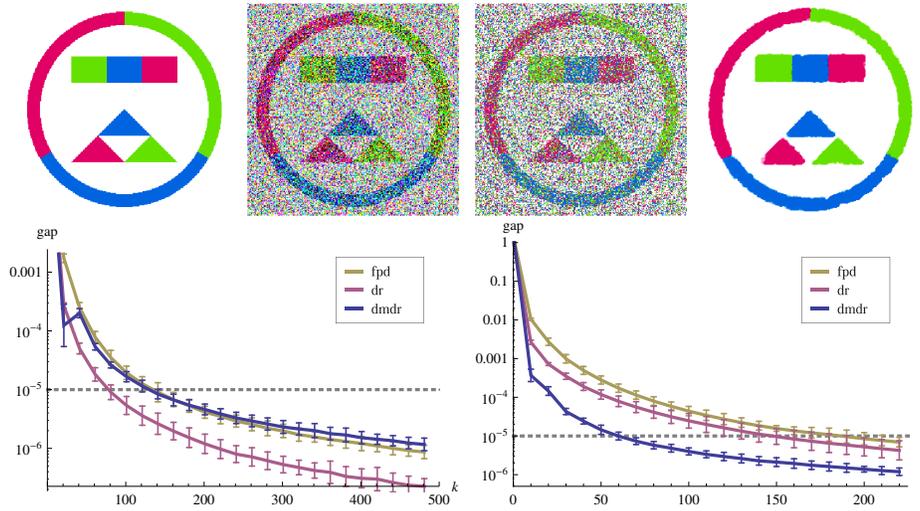
Fig. 3. Runtime comparison on a set of four-class labeling problems. **Top row, left to right:** Input image; input overlaid with heavy Gaussian noise; purely local labeling without regularizer; segmentation computed using the proposed method. The experiment was repeated 10 times with different noise. **Bottom row:** Gap vs. number of iterations (left) and time (right) with error indicators at 2σ. The proposed DMDR method performs comparable to FPD with respect to the number of iterations, but requires significantly less time per iteration, resulting in a total speedup of $2 - 3$.

4   Experimental Results

We implemented and evaluated the proposed DMDR method as well as the fast primal-dual (FPD) [17] and Douglas-Rachford (DR) [14] methods in Matlab on an Intel Core2 Duo 2.66 GHz with 4 GB of RAM and 64-bit Matlab 2009a. The full data set for the experiments is available at `ipa.iwr.uni-heidelberg.de`.

Runtime Comparison. We compared the performance of the above algorithms on a four-class color segmentation problem (Fig. 3). The input image was generated by overlaying the synthetical "four colors" image with Gaussian noise, $\sigma = 1$. The data term was set to the ℓ_1-RGB distance to the four prototypical color vectors. For the regularizer we chose the Potts distance, $d(i, j) = \lambda$ iff $i \neq j$ and $d(i, j) = 0$ otherwise, with $\lambda = \sqrt{2}$. A reference solution and optimal dual objective f_d were computed using 5000 iterations of the DR method. The experiment was repeated 10 times with varying noise.

In terms of the number of iterations, the proposed DMDR method converges as fast as FPD. However, as it requires significantly less effort per iteration, it outperforms FPD and DR by a factor of $2 - 3$ with respect to total runtime.

High Label Count and Improved Numerical Robustness. For a larger number of labels, the runtime advantage is expected to become more apparent
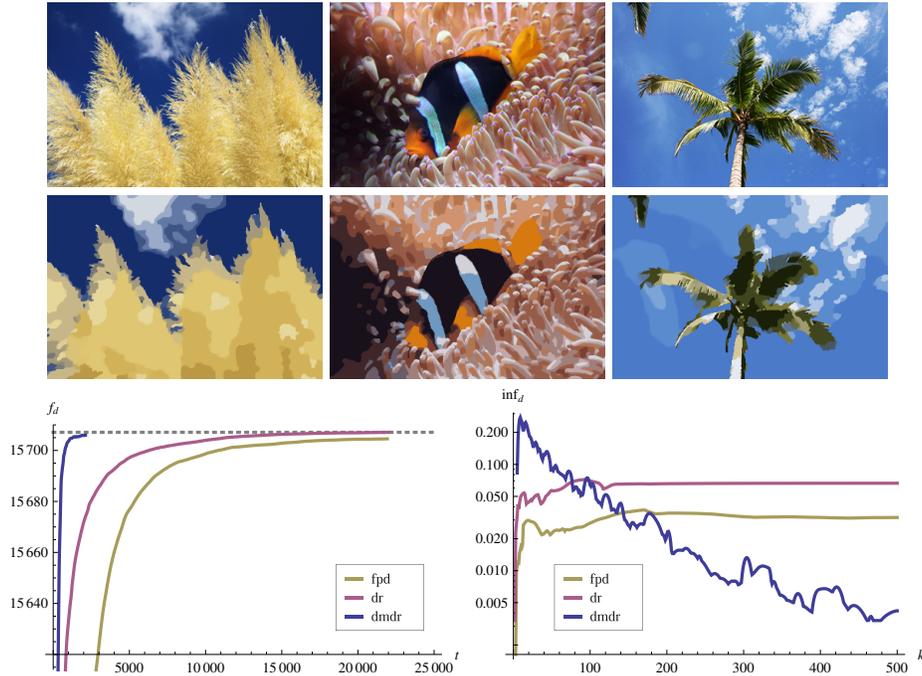
Fig. 4. Runtime performance on segmentation problems with a high label count. **Top row:** Input images (top) and segmentation into 12 classes (bottom) computed using the proposed DMDR method. **Bottom left:** Dual objective vs. time for 500 iterations on the "crop" image. The proposed method outperforms DR and FPD by a factor of 10 resp. 17. **Bottom right:** Infeasibility of the dual iterates vs. number of iterations. Due to the inexact projections, FPD and DR get stuck and converge to infeasible solutions. In contrast, DMDR gradually decreases the infeasibility to zero in theory and practice.

as the cost per iteration increases. We performed a 12-class segmentation of the real-world images in Fig. 1 and Fig. 4 with the same data term as above with $\lambda = 0.2$ for the lake and fish images, and $\lambda = 0.5$ for the palm and crop images.

For this moderate number of labels, the iterative projections for DR and FPD are already quite slow, so we fixed a maximum of 5 inner iterations per outer step in order to get a reasonable computation time. The proposed method is about $6-10$ times faster than DR, and $7-17$ times faster than FPD (Fig. 4).

Moreover, due to the inexact projections, DR and FPD converge to infeasible dual points, i.e. they generate dual solutions v that do not lie inside the dual constraint set $\mathcal{D}$. In contrast, using DMDR the infeasibility gradually decreases, and is guaranteed to eventually drop to zero given exact arithmetic (Sect. 3).

Histogram-Based Segmentation and Absolute Distance. Fig. 5 shows the application of our method to a histogram-based three-class segmentation where the data term is based on probabilities computed from histograms over regions preselected by the user. In order to preserve more details, we chose $\lambda = 0.025$.
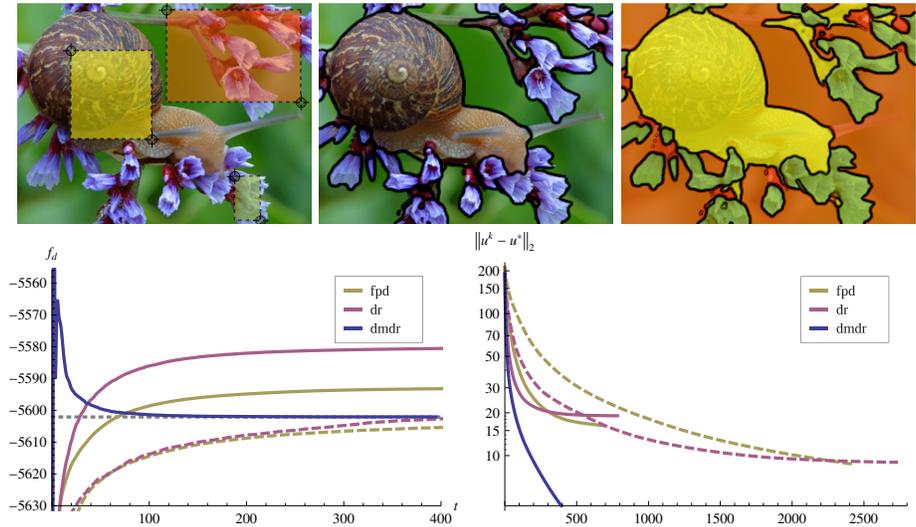
Fig. 5. Application to histogram-based segmentation. **Top row, left to right:** Input image with seed regions marked by the user; minimizer of the three-class variational segmentation using the proposed approach. **Bottom row:** Dual objectives (left) and ℓ_2 distance to the reference solution (right) vs. time. With low-accuracy approximate projections, FPD and DR get stuck in an infeasible solution (solid). Increasing the projection accuracy reduces the effect but slows down convergence (dashed). The proposed DMDR method avoids these problems and returns high-quality solutions after only a few iterations.

As above, it can be seen that FPD and DR get stuck at infeasible solutions, while DMDR converges smoothly. Increasing the accuracy of the approximate projections reduces the infeasibility, but leads to a much slower convergence.

It remains to ask how the dual gap relates to actual visual differences. Therefore at each step we evaluated the ℓ_2 distance of the current iterate to a reference solution computed using 5000 DMDR iterations (Fig. 5). Again it becomes clear that the inexact projections cause convergence issues for FPD and DR, while DMDR does not suffer from these problems. After 500 iterations, DMDR recovered a solution u^k with $\|u^k - u^*\|_2 \leqslant 10$, or $1.3 \cdot 10^{-4}$ per pixel, suggesting that only few iterations are required for visually high quality results.

Note that for all of the examples above, DMDR ran out of the box with $\tau = 1$, and did not require any parameter tuning.

Conclusion. We presented the DMDR method to efficiently solve saddle point problems with intricate dual constraints, as arise from tight relaxations of continuous multiclass labeling problems and general nonsmooth variational problems, using only simple operations that can easily be parallelized. Experiments indicate that it outperforms existing methods by a factor of $4 - 20$, and avoids the inaccuracies and convergence issues of the FPD and DR methods that rely on inexact projections.

References

1. Alberti, G., Bouchitté, G., Dal Maso, D.: The calibration method for the Mumford-Shah functional and free-discontinuity problems. Calculus of Variations and Partial Differential Equations 16(16), 299–333 (2003)
2. Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems. Clarendon Press (2000)
3. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. Patt. Anal. Mach. Intell. 23(11), 1222–1239 (2001)
4. Chambolle, A., Cremers, D., Pock, T.: A convex approach for computing minimal partitions. Tech. Rep. 649, Ecole Polytechnique CMAP (2008)
5. Chan, T.F., Esedoḡlu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. J. Appl. Math. 66(5), 1632–1648 (2006)
6. Douglas, J., Rachford, H.H.: On the numerical solution of heat conduction problems in two and three space variables. Trans. of the AMS 82(2), 421–439 (1956)
7. Eckstein, J.: Splitting Methods for Monotone Operators with Application to Parallel Optimization. Ph.D. thesis, MIT (1989)
8. Gabay, D.: Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, chap. IX. Applications of the Method of Multipliers to Variational Inequalities, p. 299. North-Holland (1983)
9. Goldstein, T., Bresson, X., Osher, S.: Geometric applications of the split Bregman method: Segmentation and surface reconstruction. CAM Report 09-06, UCLA (2009)
10. Goldstein, T., Bresson, X., Osher, S.: Global minimization of Markov random fields with applications to optical flow. CAM Report 09-77, UCLA (2009)
11. Ishikawa, H.: Exact optimization for Markov random fields with convex priors. Patt. Anal. Mach. Intell. 25(10), 1333–1336 (2003)
12. Lellmann, J., Becker, F., Schnörr, C.: Convex optimization for multi-class image labeling with a novel family of total variation based regularizers. In: Int. Conf. Comp. Vis. (2009)
13. Lellmann, J., Kappes, J., Yuan, J., Becker, F., Schnörr, C.: Convex multi-class image labeling by simplex-constrained Total Variation. In: Scale Space and Var. Meth. LNCS, vol. 5567, pp. 150–162 (2009)
14. Lellmann, J., Schnörr, C.: Continuous multiclass labeling approaches and algorithms. Tech. rep., Univ. of Heidelberg (Feb 2010), http://www.ub.uni-heidelberg.de/archiv/10460/
15. Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. Comm. Pure Appl. Math. 42, 577–685 (1989)
16. Paragios, N., Chen, Y., Faugeras, O. (eds.): The Handbook of Mathematical Models in Computer Vision. Springer (2006)
17. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: An algorithm for minimizing the Mumford-Shah functional. In: Int. Conf. Comp. Vis. (2009)
18. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: Global solutions of variational models with convex regularization. Tech. rep., Graz Univ. of Tech. (2009)
19. Rockafellar, R.: Convex Analysis. Princeton UP (1970)
20. Setzer, S.: Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage. In: Scale Space and Var. Meth. LCNS, vol. 5567, pp. 464–476 (2009)
21. Strang, G.: Maximal flow through a domain. Math. Prog. 26, 123–143 (1983)
22. Zach, C., Gallup, D., Frahm, J.M., Niethammer, M.: Fast global labeling for real-time stereo using multiple plane sweeps. In: Vis. Mod. Vis (2008)