Applications of Lagrangian-Based Alternating Direction Methods and Connections to Split Bregman

Ernie Esser

March 2009

Abstract

Analogous to the connection between Bregman iteration and the method of multipliers that was pointed out in [59], we show that a similar connection can be made between the split Bregman algorithm [32] and the alternating direction method of multipliers (ADMM) of ([29], [31]). Existing convergence theory for ADMM [23] can therefore be used to justify both the alternating step and inexact minimizations used in split Bregman for the cases in which the algorithms are equivalent. Application of these algorithms to different image processing problems is simplified by rewriting these problems in a general form that still includes constrained and unconstrained total variation, (TV), and l_1 minimization as was investigated in [32]. Numerical results for the application to TV-l_1 minimization [12] are presented. We also discuss applications of two related methods, the alternating minimization algorithm (AMA) of [56] and the Bregman operator splitting algorithm (BOS) of [61], which are sometimes better suited for problems where further decoupling of variables is useful.

1 Introduction

An important class of problems in image processing, and now also compressive sensing, is convex programs involving l_1 or TV regularization. Illustrative examples include ROF denoising [50] and basis pursuit [18]. Such problems have been notoriously slow to compute, but Bregman iteration techniques and variants such as linearized Bregman, split Bregman and Bregman operator splitting have been shown to yield simple, fast and effective algorithms for these types of problems. These recent algorithms also have many interesting connections to classical Lagrangian methods for the general problem of minimizing sums of convex functionals subject to linear equality constraints. There are close connections for example to the method of multipliers, the alternating direction method of multipliers, (ADMM) ([3], [23]), and the alternating minimization algorithm (AMA) [56]. These algorithms can be especially effective when the convex functionals are based on the l_1 norm and the l_2 norm squared.

Consider the problem

$$\min_{\substack{u \in \mathbb{R}^m \\ Ku = f}} J(u), \tag{1}$$

and assume $J(u)$ has separable structure in the sense that it can be written as

$$J(u) = H(u) + \sum_{i=1}^{N} G_i(A_i u + b_i),$$

1

where H and G_i are closed proper convex functions $G_i : \mathbb{R}^{n_i} \to (-\infty, \infty]$, $H : \mathbb{R}^m \to (-\infty, \infty]$, $f \in \mathbb{R}^s$, $b_i \in \mathbb{R}^{n_i}$, each A_i is a $n_i \times m$ matrix and K is a $s \times m$ matrix. An equivalent formulation that decouples the G_i is obtained by introducing new variables z_i and constraints $z_i = A_i u + b_i$. Now (1) can be rewritten as

$$\min_{\substack{z \in \mathbb{R}^n, u \in \mathbb{R}^m \\ Bz + Au = b}} F(z) + H(u), \tag{P0}$$

where $F(z) = \sum_{i=1}^N G_i(z_i)$, $n = \sum_{i=1}^N n_i$, $z = \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix}$, $B = \begin{bmatrix} -I \\ 0 \end{bmatrix}$, $A = \begin{bmatrix} A_1 \\ \vdots \\ A_N \\ K \end{bmatrix}$, and $b = \begin{bmatrix} -b_1 \\ \vdots \\ -b_N \\ f \end{bmatrix}$.

Letting $d = n + s$, note that A is a $d \times m$ matrix, B is a $d \times n$ matrix and $b \in \mathbb{R}^d$. Similar formulations are discussed for example in [3], [2], [48] and [56].

There is extensive literature in optimization and numerical analysis about splitting methods for solving convex programming problems that have separable structure as (P0) does. The goal is to produce algorithms that consist of simple, easy to compute steps that can deal with the terms of $J(u)$ one at a time. One approach based on duality leads to augmented Lagrangian type methods that can be interpreted as splitting methods applied to a dual formulation of the problem. A good summary of these methods can be found in chapter three of [30] and Eckstein's thesis [24]. Here we will focus mainly on ADMM because of its connection to the Split Bregman algorithm of Goldstein and Osher. They show in [32] how to simplify the minimization of convex functionals of u involving the l_1 norm of a convex function $\Phi(u)$. They replace $\Phi(u)$ with a new variable z, add a constraint $z = \Phi(u)$ and then use Bregman iteration [59] techniques to handle the resulting constrained optimization problem. A key application is functionals containing $\|u\|_{TV}$. A related splitting approach that uses continuation methods to handle the constraints has been studied by Wang, Yin and Zhang, [57] and applied to TV minimization problems including TV-l_1 ([34], [58]). The connection between Bregman iteration and the augmented Lagrangian for constrained optimization problems with linear equality constraints is discussed by Yin, Osher, Goldfarb and Darbon in [59]. They show Bregman iteration is equivalent to the method of multipliers of Hestenes [37] and Powell [46] when the constraints are linear. The augmented Lagrangian for problem (1) is

$$L_\alpha(u, \lambda) = J(u) + \langle \lambda, f - Ku \rangle + \frac{\alpha}{2} \|f - Ku\|^2,$$

where $\| \cdot \|$ and $\langle \cdot, \cdot \rangle$ denote the Euclidean norm and standard inner product. The method of multipliers is to iterate

$$u^{k+1} = \arg \min_{u \in \mathbb{R}^m} L_\alpha(u, \lambda^k) \tag{2}$$

$$\lambda^{k+1} = \lambda^k + \alpha(f - Ku^{k+1}),$$

whereas Bregman iteration yields

$$u^{k+1} = \arg \min_{u \in \mathbb{R}^m} J(u) - J(u^k) - \langle p^k, u - u^k \rangle + \frac{\alpha}{2} \|f - Ku\|^2 \tag{3}$$

$$p^{k+1} = p^k + \alpha K^T(f - Ku^{k+1}).$$

$J(u) - J(u^k) - \langle p^k, u - u^k \rangle$ is the Bregman distance between u and u^k, where p^k is a subgradient of J at u^k. Similarly, in the special case when Φ is linear, an interpretation of the split Bregman algorithm, explained in sections 3.1.1 and 3.2.1, is to alternately minimize with respect to u and z the augmented Lagrangian associated to the constrained problem and then to update a Lagrange multiplier. This procedure also describes ADMM, which goes back to Glowinski and Marocco [31], and Gabay and Mercier [29]. The augmented Lagrangian for problem (P0) is

$$L_\alpha(z, u, \lambda) = F(z) + H(u) + \langle \lambda, b - Au - Bz \rangle + \frac{\alpha}{2} \|b - Au - Bz\|^2,$$

and the ADMM iterations are given by

$$
\begin{aligned}
z^{k+1} &= \arg \min_{z \in \mathbb{R}^n} L_\alpha(z, u^k, \lambda^k) \\
u^{k+1} &= \arg \min_{u \in \mathbb{R}^m} L_\alpha(z^{k+1}, u, \lambda^k) \\
\lambda^{k+1} &= \lambda^k + \alpha(b - Au^{k+1} - Bz^{k+1}).
\end{aligned}
\tag{4}
$$

ADMM can also be interpreted as Douglas Rachford splitting [22] applied to the dual problem. Connections between these two interpretations were explored by Gabay [28] and Glowinski and Le Tallec [30]. The dual version of the algorithm was studied by Lions and Mercier [39]. The equivalence of ADMM to a proximal point method was studied by Eckstein and Bertsekas [23], who also generalized the convergence theory to allow for inexact minimizations. Techniques regarding applying ADMM to problems with separable structure are discussed in detail by Bertsekas and Tsitsiklis in ([3] section 3.4.4) and also by Glowinski and Fortin in [27]. The connection between split Bregman and Douglas Rachford splitting has also been made by Setzer [51].

Other splitting methods besides Douglas Rachford splitting can be applied to the dual problem, which is a special case of the well studied more general problem of finding a zero of the sum of two maximal monotone operators. See for example [25] and [39]. Some splitting methods applied to the dual problem can also be interpreted in terms of alternating minimization of the augmented Lagrangian. For example, Peaceman Rachford splitting [45] corresponds to an alternating minimization algorithm very similar to ADMM except that it updates the Lagrange multiplier twice, once after each minimization of the augmented Lagrangian [30].

Proximal forward backward splitting can also be effectively applied to the dual problem. This splitting procedure, which goes back to Lions and Mercier [39] and Passty [44], appears in many applications. Some examples include classical methods such as gradient projection and more recent ones such as the iterative thresholding algorithm FPC of Hale, Yin and Zhang [35] and the framelet inpainting algorithm of Cai, Chan and Shen [6].

The Lagrangian interpretation of the dual application of forward backward splitting was studied by Tseng in [56]. He shows that it corresponds to an algorithm with the same steps as ADMM except that one of the minimizations of the augmented Lagrangian, $L_\alpha(z, u, \lambda)$, is replaced by minimization of the Lagrangian, which for (P0) is

$$L(z, u, \lambda) = F(z) + H(u) + \langle \lambda, b - Au - Bz \rangle.$$

The resulting iterations are given by

$$u^{k+1} = \arg \min_{u \in \mathbb{R}^m} L(z^k, u, \lambda^k)$$
$$z^{k+1} = \arg \min_{z \in \mathbb{R}^n} L_\alpha(z, u^{k+1}, \lambda^k) \tag{5}$$
$$\lambda^{k+1} = \lambda^k + \alpha(b - Au^{k+1} - Bz^{k+1}).$$

Tseng called this the alternating minimization algorithm, referred to in shorthand as AMA. This method is useful for solving (P0) when H is strictly convex but including the augmented quadratic penalty leads to a minimization step that is more difficult to solve.

There are other methods for decoupling variables that don't require the functional to be strictly convex. An example is the predictor corrector proximal method (PCPM) by Chen and Teboulle [17], which alternates proximal steps for the primal and dual variables. The PCPM iterations are given by

$$u^{k+1} = \arg \min_{u \in \mathbb{R}^m} L(z^k, u, \lambda^k) + \frac{1}{2\alpha_k} \|u - u^k\|^2$$
$$z^{k+1} = \arg \min_{z \in \mathbb{R}^n} L(z, u^k, \lambda^k) + \frac{1}{2\alpha_k} \|z - z^k\|^2$$
$$\lambda^{k+1} = \lambda^k + (\alpha_{k+1} + \alpha_k)(b - Au^{k+1} - Bz^{k+1}) - \alpha_k(b - Au^k - Bz^k).$$

This method can require many iterations. Another technique to undo the coupling of variables that results from quadratic penalty terms of the form $\frac{\alpha_k}{2} \|Ku - f\|^2$ is to replace such a penalty with one of the form $\frac{1}{2\delta_k} \|u - u^k + \alpha_k \delta_k K^T(Ku^k - f)\|^2$, which instead penalizes the distance of u from a linearization of the original penalty. This was applied to the method of multipliers by Stephanopoulos and Westerberg in [52]. It was used in the derivation of the linearized Bregman algorithm in [59]. This technique is also used with Bregman iteration methods by Zhang, Burger, Bresson and Osher in [61], leading to the Bregman Operator Splitting (BOS) algorithm, which they apply for example to nonlocal TV minimization problems. They also show the connection to inexact Uzawa methods. Written as an inexact Uzawa method, the BOS algorithm applied to (1) yields the iterations

$$u^{k+1} = \arg \min_{u \in \mathbb{R}^m} J(u) + \langle \lambda^k, f - Ku \rangle + \frac{1}{2\delta_k} \|u - u^k + \alpha_k \delta_k K^T(Ku^k - f)\|^2 \tag{6}$$
$$\lambda^{k+1} = \lambda^k + \alpha_k(f - Ku^{k+1}).$$

Section 3.3.2 describes how this idea can be applied to (P0).

This paper consists of three parts. The first part discusses the Lagrangian formulation of the problem (P0) and the dual problem. The second part focuses on exploring the connection between split Bregman and ADMM, their application to (P0) and their dual interpretation. It also demonstrates how further decoupling of variables is possible using AMA and BOS. The third part shows how to apply these algorithms to some example image processing problems, focusing on applications that illustrate how to take advantage of problems' separable structure.

2 The Primal and Dual Problems

Lagrangian duality will play an important role in the analysis of (P0). In this section we define a Lagrangian formulation of (P0) and the dual problem. We also discuss conditions that guarantee

solutions to the primal and dual problems.

2.1 Lagrangian Formulation and Dual Problem

Associated to the primal problem (P0) is the Lagrangian

$$L(z, u, \lambda) = F(z) + H(u) + \langle \lambda, b - Au - Bz \rangle, \tag{7}$$

where the dual variable $\lambda \in \mathbb{R}^d$ can be thought of as a vector of Lagrange multipliers. The dual functional $q(\lambda)$ is a concave function $q : \mathbb{R}^d \to [-\infty, \infty)$ defined by

$$q(\lambda) = \inf_{u \in \mathbb{R}^m, z \in \mathbb{R}^n} L(z, u, \lambda). \tag{8}$$

The dual problem to (P0) is

$$\max_{\lambda \in \mathbb{R}^d} q(\lambda). \tag{Q0}$$

Since (P0) is a convex programming problem with linear constraints, if it has an optimal solution (z^*, u^*) then (Q0) also has an optimal solution λ^*, and

$$F(z^*) + H(u^*) = q(\lambda^*),$$

which is to say that the duality gap is zero, ([2] 5.2, [48] 28.2, 28.4). To guarantee existence of an optimal solution to (P0), assume that the set

$$\{(z, u) : F(z) + H(u) \le c , \ Au + Bz = b\}$$

is nonempty and bounded for some $c \in \mathbb{R}$. Alternatively, we could assume that $Ku = f$ has a solution, and if it's not unique, which it probably won't be, then assume $F(z) + H(u)$ is coercive on the affine subspace defined by $Au + Bz = b$. Either way, we can equivalently minimize over a compact subset. Since F and H are closed proper convex functions, which is to say lower semicontinuous convex functions not identically infinity, Weierstrass' theorem implies a minimum is attained [2].

2.2 Saddle Point Formulation and Optimality Conditions

Finding optimal solutions of (P0) and (Q0) is equivalent to finding a saddle point of L. More precisely ([48] 28.3), (z^*, u^*) is an optimal primal solution and λ^* is an optimal dual solution if and only if

$$L(z^*, u^*, \lambda) \le L(z^*, u^*, \lambda^*) \le L(z, u, \lambda^*) \qquad \forall \ z, u, \lambda. \tag{9}$$

From this it follows that

$$\max_{\lambda \in \mathbb{R}^d} F(z^*) + H(u^*) + \langle \lambda, b - Au^* - Bz^* \rangle = L(z^*, u^*, \lambda^*) = \min_{u \in \mathbb{R}^m, z \in \mathbb{R}^n} F(z) + H(u) + \langle \lambda^*, b - Au - Bz \rangle,$$

from which we can directly read off the Kuhn-Tucker optimality conditions.

$$Au^* + Bz^* = b \tag{10a}$$
$$B^T \lambda^* \in \partial F(z^*) \tag{10b}$$
$$A^T \lambda^* \in \partial H(u^*), \tag{10c}$$

where ∂ denotes the subgradient, defined by

$$\partial F(z^*) = \{p \in \mathbb{R}^n : F(v) \geq F(z^*) + \langle p, v - z^* \rangle \; \forall v \in \mathbb{R}^n\},$$

$$\partial H(u^*) = \{q \in \mathbb{R}^m : H(w) \geq H(u^*) + \langle q, w - u^* \rangle \; \forall w \in \mathbb{R}^m\}.$$

These optimality conditions (10) hold if and only if (z^*, u^*, λ^*) is a saddle point for L ([48] 28.3). Note also that $L(z^*, u^*, \lambda^*) = F(z^*) + H(u^*)$.

2.3   Dual Functional

The dual functional $q(\lambda)$ (8) can be written in terms of the Legendre-Fenchel transforms of F and H.

$$
\begin{aligned}
q(\lambda) &= \inf_{z \in \mathbb{R}^n, u \in \mathbb{R}^m} F(z) + \langle \lambda, b - Bz - Au \rangle + H(u) \\
&= \inf_{z \in \mathbb{R}^n} \left(F(z) - \langle \lambda, Bz \rangle \right) + \inf_{u \in \mathbb{R}^m} \left(H(u) - \langle \lambda, Au \rangle \right) + \langle \lambda, b \rangle \\
&= - \sup_{z \in \mathbb{R}^n} \left(\langle B^T \lambda, z \rangle - F(z) \right) - \sup_{u \in \mathbb{R}^m} \left(\langle A^T \lambda, u \rangle - H(u) \right) + \langle \lambda, b \rangle \\
&= -F^*(B^T \lambda) - H^*(A^T \lambda) + \langle \lambda, b \rangle,
\end{aligned}
$$

where F^* and H^* denote the Legendre-Fenchel transforms, or convex conjugates, of F and H defined by

$$F^*(B^T \lambda) = \sup_{z \in \mathbb{R}^n} \left(\langle B^T \lambda, z \rangle - F(z) \right),$$

$$H^*(A^T \lambda) = \sup_{u \in \mathbb{R}^m} \left(\langle A^T \lambda, u \rangle - H(u) \right).$$

2.4   Maximally Decoupled Case

An interesting special case of (P0), which will arise in many of the following examples, is when $H(u) = 0$. This corresponds to

$$\min_{\substack{u \in \mathbb{R}^m, z \in \mathbb{R}^n \\ Bz + Au = b}} F(z). \tag{P1}$$

As before, the dual functional is given by

$$q_1(\lambda) = -F^*(B^T \lambda) - H^*(A^T \lambda) + \langle \lambda, b \rangle,$$

except here $H^*(A^T \lambda)$ can be interpreted as an indicator function defined by

$$H^*(A^T \lambda) = \begin{cases} 0 & \text{if } A^T \lambda = 0, \\ \infty & \text{otherwise.} \end{cases}$$

This can be interpreted as the constraint $A^T \lambda = 0$, which is equivalent to $P\lambda = \lambda$, where P is the projection onto $\mathrm{Im}(A)^\perp$ defined by

$$P = I - A(A^T A)^{-1} A^T.$$

6

Therefore the dual problem for (P1) can be written as

$$\max_{\substack{\lambda \in \mathbb{R}^d \\ A^T \lambda = 0}} -F^*(B^T P \lambda) + \langle P \lambda, b \rangle. \tag{Q1}$$

The variable u can also be completely eliminated from the primal problem, which can be equivalently formulated as

$$\min_{\substack{z \in \mathbb{R}^n \\ P(b - Bz) = 0}} F(z). \tag{P2}$$

The associated dual functional is

$$q_2(\lambda) = -F^*(B^T P \lambda) + \langle P \lambda, b \rangle,$$

and the dual problem is therefore

$$\max_{\lambda \in \mathbb{R}^d} -F^*(B^T P \lambda) + \langle P \lambda, b \rangle, \tag{Q2}$$

which is identical to (Q1) without the constraint. However, since $q_2(\lambda) = q_2(P\lambda)$ the $A^T \lambda = 0$ constraint can be added to (Q2) without changing the maximum.

3   Algorithms

In this section we start by analyzing Bregman iteration (3) applied to (P0) because the first step in deriving the split Bregman algorithm in [32] was essentially to take advantage of the separable structure of (1) by rewriting it as (P0) and applying Bregman iteration. Then we show an equivalence between ADMM (4) and the split Bregman algorithm and present a convergence result by Eckstein and Bertsekas [23]. Next we interpret AMA (5) and BOS (6) as modifications of ADMM applied to (P0), and we discuss when they are applicable and why they are useful. Throughout, we also discuss the dual interpretations of Bregman iteration/method of multipliers as gradient ascent, split Bregman/ADMM as Douglas Rachford splitting and AMA as proximal forward backward splitting.

3.1   Bregman Iteration and Method of Multipliers

3.1.1   Application to Primal Problem

Bregman iteration applied to (P0) yields

$$
\begin{aligned}
(z^{k+1}, u^{k+1}) = \arg\min_{z \in \mathbb{R}^n, u \in \mathbb{R}^m} \; & F(z) - F(z^k) - \langle p_z^k, z - z^k \rangle + \\
& H(u) - H(u^k) - \langle p_u^k, u - u^k \rangle + \\
& \frac{\alpha}{2} \| b - Au - Bz \|^2 \\
p_z^{k+1} = p_z^k & + \alpha B^T (b - Au^{k+1} - Bz^{k+1}) \\
p_u^{k+1} = p_u^k & + \alpha A^T (b - Au^{k+1} - Bz^{k+1}).
\end{aligned}
\tag{11}
$$

For the initialization, p_z^0 and p_u^0 are set to zero while z^0 and u^0 are arbitrary. Note that for $k \geq 1$, $p_u^k \in \partial H(u^k)$ and $p_z^k \in \partial F(z^k)$. Now, following the argument in [59] that shows an equivalence between Bregman iteration and the method of multipliers (2) in the case of linear constraints, define λ^k for $k \geq 0$ by $\lambda^0 = 0$ and

$$\lambda^{k+1} = \lambda^k + \alpha(b - Au^{k+1} - Bz^{k+1}). \tag{12}$$

Notice that if $p_z^k = B^T\lambda^k$ and $p_u^k = A^T\lambda^k$ then $p_z^{k+1} = B^T\lambda^{k+1}$ and $p_u^{k+1} = A^T\lambda^{k+1}$. So by induction, it holds for all k. This implies that

$$-\langle p_z^k, z \rangle - \langle p_u^k, u \rangle = -\langle B^T\lambda^k, z \rangle - \langle A^T\lambda^k, u \rangle = \langle \lambda^k, -Au - Bz \rangle.$$

This means the objective function in (11) up to a constant is equivalent to the augmented Lagrangian at λ^k, defined by

$$L_\alpha(z, u, \lambda^k) = F(z) + H(u) + \langle \lambda^k, b - Au - Bz \rangle + \frac{\alpha}{2}\|b - Au - Bz\|^2. \tag{13}$$

Then (z^{k+1}, u^{k+1}) in (11) can be equivalently updated by

$$(z^{k+1}, u^{k+1}) = \arg \min_{z \in \mathbb{R}^n, u \in \mathbb{R}^m} L_\alpha(z, u, \lambda^k) \tag{14}$$

$$\lambda^{k+1} = \lambda^k + \alpha(b - Au^{k+1} - Bz^{k+1}), \tag{15}$$

which is the method of multipliers (2). This connection was also pointed out in [54].

Note that the same assumptions that guaranteed existence of a minimizer for (P0) also guarantee that (14) is well defined. Having assumed that there exists $c \in \mathbb{R}$ such that

$$Q = \{(z, u) : F(z) + H(u) \leq c , \ Au + Bz = b\}$$

is nonempty and bounded, it follows that

$$R = \left\{ (z, u) : F(z) + H(u) + \langle \lambda^k, b - Au - Bz \rangle + \frac{\alpha}{2}\|b - Au - Bz\|^2 \leq c \right\}$$

is nonempty and bounded. If not, then being an unbounded convex set, R must contain a half line. Because of the presence of the quadratic term, any such line must be parallel to the affine set defined by $Au + Bz = b$. But since R is also closed, by ([48] 8.3) a half line is also contained in that affine set, which contradicts the assumption that Q was bounded. Weierstrass' theorem can then be used to show that a minimum of (14) is attained.

3.1.2 Dual Interpretation

Since Bregman iteration with linear constraints is equivalent to the method of multipliers it also shares some of the interesting dual interpretations. In particular, it can be interpreted as a proximal point method for maximizing $q(\lambda)$ or as a gradient ascent method for maximizing $q_\alpha(\lambda)$, where $q_\alpha(\lambda)$ denotes the dual of the augmented Lagrangian L_α defined by

$$q_\alpha(\lambda) = \inf_{z \in \mathbb{R}^n, u \in \mathbb{R}^m} L_\alpha(z, u, \lambda). \tag{16}$$

8

Note that from previous assumptions guaranteeing existence of an optimal solution to (P0), and because the augmented term $\frac{\alpha}{2}\|b - Au - Bz\|^2$ is zero when the constraint is satisfied, the maximums of $q(\lambda)$ and $q_\alpha(\lambda)$ are attained and equal. Following an argument by Rockafellar in [47], note that

$$L_\alpha(z, u, \lambda^k) = \max_{y \in \mathbb{R}^d} L(z, u, y) - \frac{1}{2\alpha}\|y - \lambda^k\|^2$$

and that the maximum is attained at

$$y^* = \lambda^k + \alpha(b - Au - Bz).$$

Let (z^{k+1}, u^{k+1}) (possibly not unique) be where the minimum of $L_\alpha(z, u, \lambda^k)$ is attained. So

$$\inf_{z \in \mathbb{R}^n, u \in \mathbb{R}^m} \max_{y \in \mathbb{R}^d} L(z, u, y) - \frac{1}{2\alpha}\|y - \lambda^k\|^2$$

is attained at (z^{k+1}, u^{k+1}, y^*) where $y^* = \lambda^k + \alpha(b - Au^{k+1} - Bz^{k+1})$. Because we have convexity in (z, u) and strict concavity in y, the inf and max can be swapped ([48] 37.3). Thus we have

$$q_\alpha(\lambda^k) = \inf_{z,u} \max_y L(z, u, y) - \frac{1}{2\alpha}\|y - \lambda^k\|^2 \tag{17a}$$

$$= \max_y \inf_{z,u} L(z, u, y) - \frac{1}{2\alpha}\|y - \lambda^k\|^2 \tag{17b}$$

$$= \max_y q(y) - \frac{1}{2\alpha}\|y - \lambda^k\|^2 \tag{17c}$$

$$= q(y^*) - \frac{1}{2\alpha}\|y^* - \lambda^k\|^2 \tag{17d}$$

From the definition of the Lagrange multiplier update (15), we see that

$$\lambda^{k+1} = y^* = \arg\max_{y \in \mathbb{R}^d} q(y) - \frac{1}{2\alpha}\|y - \lambda^k\|^2, \tag{18}$$

which can be interpreted as a step in a proximal point method for maximizing $q(\lambda)$. The connection to the proximal point method is also derived for example in [3]. Since from (18), λ^{k+1} is uniquely determined given λ^k, that means that $Au^{k+1} + Bz^{k+1}$ is constant over all minimizers (z^{k+1}, u^{k+1}) of $L_\alpha(z, u, \lambda^k)$. Going back to the Bregman iteration (11), we also have that $p_z^{k+1} = B^T\lambda^{k+1}$ and $p_u^{k+1} = A^T\lambda^{k+1}$ were uniquely determined at each iteration.

One way to interpret (18) as a gradient ascent method applied to $q_\alpha(\lambda)$ is to note that from (17c), $q_\alpha(\lambda^k)$ is minus the Moreau envelope of index α of the closed proper convex function $-q$ at λ^k ([19] 2.3). The Moreau envelope can be shown to be differentiable, and there is a formula for its gradient [19], which when applied to (17c) yields

$$\nabla q_\alpha(\lambda^k) = -\left[\frac{\lambda^k - \arg\max_y \left(q(y) - \frac{1}{2\alpha}\|y - \lambda^k\|^2\right)}{\alpha}\right].$$

Substituting in λ^{k+1} we see that

$$\nabla q_\alpha(\lambda^k) = \frac{\lambda^{k+1} - \lambda^k}{\alpha},$$

which means we can interpret the Lagrange multiplier update as the gradient ascent step

$$\lambda^{k+1} = \lambda^k + \alpha \nabla q_\alpha(\lambda^k),$$

where $\nabla q_\alpha(\lambda^k) = (b - Au^{k+1} - Bz^{k+1})$.

3.2 Split Bregman and ADMM Equivalence

3.2.1 Alternating Minimization

The split Bregman algorithm uses an alternating minimization approach to minimize (14), namely iterating

$$z^{k+1} = \arg\min_{z\in\mathbb{R}^n} F(z) + \langle \lambda^k, -Bz \rangle + \frac{\alpha}{2}\|b - Au^k - Bz\|^2 \tag{19}$$

$$u^{k+1} = \arg\min_{u\in\mathbb{R}^m} H(u) + \langle \lambda^k, -Au \rangle + \frac{\alpha}{2}\|b - Au - Bz^{k+1}\|^2 \tag{20}$$

T times and then updating

$$\lambda^{k+1} = \lambda^k + \alpha(b - Au^{k+1} - Bz^{k+1}). \tag{21}$$

When $T = 1$, this becomes ADMM (4), which can be interpreted as alternately minimizing the augmented Lagrangian $L_\alpha(z, u, \lambda)$ with respect to z, then u and then updating the Lagrange multiplier λ. A similar derivation motivated by the augmented Lagrangian can be found in [3]. Note that this equivalence between split Bregman and ADMM is not in general true when the constraints are not linear.

Also note the asymmetry of the u and z updates. If we switch the order, first minimizing over u, then over z, we obtain a valid but different incarnation of ADMM, which we are not considering here.

3.2.2 Convergence Theory

In [23], Eckstein and Bertsekas demonstrate that ADMM can be interpreted as an application of the proximal point algorithm. They use this observation to prove a convergence result for ADMM that allows for approximate computation of z^{k+1} and u^{k+1}, as well some over or under relaxation. Their theorem as stated applies to (P0) in the case when $A = I$, $b = 0$ and B is an arbitrary full column rank matrix, but the same result also holds under slightly weaker assumptions. In particular, we can let b be nonzero and replace $A = I$ by the assumption that $H(u) + \|Au\|^2$ is strictly convex. Note the latter assumption holds in particular when A has full column rank. We restate their result as it applies to (P0) under the slightly weaker assumptions and in the case without over or under relaxation factors.

Theorem 3.1. *(Eckstein, Bertsekas [23]) Consider the problem (P0) where F and H are closed proper convex functions, B has full column rank and $H(u) + \|Au\|^2$ is strictly convex. Let $\lambda^0 \in \mathbb{R}^d$ and $u^0 \in \mathbb{R}^m$ be arbitrary and let $\alpha > 0$. Suppose we are also given sequences $\{\mu_k\}$ and $\{\nu_k\}$ such that $\mu_k \geq 0$, $\nu_k \geq 0$, $\sum_{k=0}^{\infty} \mu_k < \infty$ and $\sum_{k=0}^{\infty} \nu_k < \infty$. Suppose that*

$$\|z^{k+1} - \arg\min_{z\in\mathbb{R}^n} F(z) + \langle \lambda^k, -Bz \rangle + \frac{\alpha}{2}\|b - Au^k - Bz\|^2\| \leq \mu_k \tag{22}$$

$$\|u^{k+1} - \arg\min_{u\in\mathbb{R}^m} H(u) + \langle \lambda^k, -Au \rangle + \frac{\alpha}{2}\|b - Au - Bz^{k+1}\|^2\| \leq \nu_k \tag{23}$$

$$\lambda^{k+1} = \lambda^k + \alpha(b - Au^{k+1} - Bz^{k+1}). \tag{24}$$

If there exists a saddle point of $L(z, u, \lambda)$ (7), then $z^k \to z^*$, $u^k \to u^*$ and $\lambda^k \to \lambda^*$, where (z^*, u^*, λ^*) is such a saddle point. On the other hand, if no such saddle point exists, then at least one of the sequences $\{u^k\}$ or $\{\lambda^k\}$ must be unbounded.

The proof, which requires only very minor changes to the one presented in [23], is partially sketched in Appendix A.

Note that the convergence result carries over to the split Bregman algorithm in the case when the constraints are linear and when only one inner iteration is used.

3.2.3 Dual Interpretation

Some additional insight comes from the dual interpretation of ADMM as Douglas-Rachford [22] splitting applied to the dual problem (Q0), which we recall can be written as

$$
\max_{y \in \mathbb{R}^d} -F^*(B^T y) + \langle y, b \rangle - H^*(A^T y).
$$

Define operators Ψ and ϕ by

$$
\Psi(y) = B \partial F^*(B^T y) - b \tag{25}
$$
$$
\phi(y) = A \partial H^*(A^T y). \tag{26}
$$

Douglas Rachford splitting is a classical method for solving parabolic problems of the form

$$
\frac{d\lambda}{dt} + f(\lambda) + g(\lambda) = 0
$$

by iterating

$$
\frac{\hat{\lambda}^{k+1} - \lambda^k}{\Delta t} + f(\hat{\lambda}^{k+1}) + g(\lambda^k) = 0
$$
$$
\frac{\lambda^{k+1} - \lambda^k}{\Delta t} + f(\hat{\lambda}^{k+1}) + g(\lambda^{k+1}) = 0,
$$

where Δt is the time step. By iterating to steady state, this can also be used to solve

$$
f(\lambda) + g(\lambda) = 0.
$$

Solving the dual problem (Q0) is equivalent to finding λ such that zero is in the subdifferential of $-q$ at λ, which is equivalent to solving

$$
0 \in \Psi(\lambda) + \phi(\lambda). \tag{27}
$$

By formally applying Douglas Rachford splitting to (27) with α as the time step, we get

$$
0 \in \frac{\hat{\lambda}^{k+1} - \lambda^k}{\alpha} + \Psi(\hat{\lambda}^{k+1}) + \phi(\lambda^k), \tag{28a}
$$
$$
0 \in \frac{\lambda^{k+1} - \lambda^k}{\alpha} + \Psi(\hat{\lambda}^{k+1}) + \phi(\lambda^{k+1}). \tag{28b}
$$

Following the arguments by Glowinski and Le Tallec [30] and Eckstein and Bertsekas [23], we can show that ADMM satisfies (28). Define

$$
\hat{\lambda}^{k+1} = \lambda^k + \alpha(b - Bz^{k+1} - Au^k).
$$

Then from the optimality condition for (19),

$$B^T \hat{\lambda}^{k+1} \in \partial F(z^{k+1}).$$

Then from the definitions of subgradient and convex conjugate it follows that

$$z^{k+1} \in \partial F^*(B^T \hat{\lambda}^{k+1}).$$

Multiplying by B and subtracting b we have

$$Bz^{k+1} - b \in B\partial F^*(B^T \hat{\lambda}^{k+1}) - b = \Psi(\hat{\lambda}^{k+1}).$$

The analogous argument starting with the optimality condition for (20) yields

$$Au^{k+1} \in A\partial H^*(A^T \lambda^{k+1}) = \phi(\lambda^{k+1}).$$

With λ^{k+1} defined by (21) and noting that $Au^k \in \phi(\lambda^k)$, we see that the ADMM procedure satisfies (28).

It's important to note that Ψ and ϕ are not necessarily single valued, so there could possibly be multiple ways of formally satisfying the Douglas Rachford splitting as written in (28). For example, in the maximally decoupled case where $H(u) = 0$, ϕ can be defined by

$$\phi(y) = \begin{cases} \text{Im}(A) & \text{for } y \text{ such that } A^T y = 0 \\ \emptyset & \text{otherwise} \end{cases}.$$

The method of multipliers applied to either (P1) or (P2) with $P\lambda^0 = \lambda^0$ is equivalent to the proximal point method applied to the dual. This would yield

$$\lambda^{k+1} = \hat{\lambda}^{k+1} = \arg\max_{y \in \mathbb{R}^d} -F^*(B^T Py) + \langle Py, b \rangle - \frac{1}{2\alpha}\|y - \lambda^k\|^2$$

with $P\lambda^k = \lambda^k$. This also formally satisfies (28), but the λ^{k+1} updates are different from ADMM and ususally more difficult to compute. The particular way in which ADMM satisfies (28), rewritten in terms of the resolvents $(I + \alpha\Psi)^{-1}$ and $(I + \alpha\phi)^{-1}$ is

$$\hat{\lambda}^{k+1} = (I + \alpha\Psi)^{-1}(\lambda^k - \alpha Au^k) \tag{29}$$

$$\lambda^{k+1} = (I + \alpha\phi)^{-1}(\hat{\lambda}^{k+1} + \alpha Au^k) \tag{30}$$

Since u^k by assumption is uniquely determined, Au^k is well defined. One way to argue the resolvents are well defined is using monotone operator theory [25]. Briefly, a multivalued operator $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ is monotone if

$$\langle w - w', u - u' \rangle \geq 0 \text{ whenever } w \in \Phi(u) , \ w' \in \Phi(u') .$$

The operator Φ is maximal monotone if in addition to being monotone, its graph $\{(u, w) \in \mathbb{R}^d \times \mathbb{R}^d | w \in \Phi(u)\}$ is not strictly contained in the graph for any other monotone operator. From a result by Minty [40], if Φ is maximal monotone, then for any $\alpha > 0$, $(I + \alpha\Phi)^{-1}$ is single valued and defined on all of $\mathbb{R}^d$ ([23], [56]). Then from a result by Rockafellar ([48] 31.5.2), Φ is maximal monotone if it is the subdifferential of a closed proper convex function. Since $\Psi(y)$ and $\phi(y)$ were

defined to be subdifferentials of $F^*(B^T y) - \langle y, b \rangle$ and $H^*(A^T y)$ respectively, the resolvents in (29) are well defined.

It's possible to rewrite the updates in (29) completely in terms of the dual variable. Combining the two steps yields

$$\lambda^{k+1} = (I + \alpha\phi)^{-1} \left((I + \alpha\Psi)^{-1}(\lambda^k - \alpha A u^k) + \alpha A u^k \right). \tag{31}$$

Suppose

$$y^k = \lambda^k + \alpha A u^k.$$

Since $A u^k \in \phi(\lambda^k)$, $y^k \in (I + \alpha\phi)\lambda^k$. So $\lambda^k = (I + \alpha\phi)^{-1} y^k$. We can use this to rewrite (31) as

$$\lambda^{k+1} = (I + \alpha\phi)^{-1} \left[(I + \alpha\Psi)^{-1} \left(2(I + \alpha\phi)^{-1} - I \right) + \left(I - (I + \alpha\phi)^{-1} \right) \right] y^k.$$

Now let

$$y^{k+1} = \left[(I + \alpha\Psi)^{-1} \left(2(I + \alpha\phi)^{-1} - I \right) + \left(I - (I + \alpha\phi)^{-1} \right) \right] y^k. \tag{32}$$

Recalling the definition of $\hat{\lambda}^{k+1}$ and λ^{k+1}

$$\begin{aligned} y^{k+1} &= \left((I + \alpha\Psi)^{-1}(\lambda^k - \alpha A u^k) + \alpha A u^k \right) \\ &= \hat{\lambda}^{k+1} + \alpha A u^k \\ &= \lambda^k + \alpha(b - B z^{k+1}) \\ &= \lambda^{k+1} + \alpha A u^{k+1}. \end{aligned}$$

Thus assuming we initialize $y^0 = \lambda^0 + \alpha A u^0$ with $u^0 \in \partial H^*(A^T \lambda^0)$, $y^k = \lambda^k + \alpha A u^k$ and $\lambda^k = (I + \alpha\phi)^{-1} y^k$ hold for all $k \geq 0$. So ADMM is equivalent to iterating (32). This is the representation used by Eckstein and Bertsekas [23] and referred to as the Douglas-Rachford recursion. Note that in the maximally decoupled case, $(I + \alpha\phi)^{-1}$ reduces to the projection matrix P, which projects onto $\text{Im}(A)^\perp$.

3.3   Decoupling Variables Using AMA and BOS

The quadratic penalty terms of the form $\frac{\alpha}{2}\|Ku - f\|^2$ that appear in the ADMM iterations couple the variables in a way that can make the algorithm computationally expensive. If K has special structure, this may not be a problem. For example, K could be diagonal. Or it might be possible to diagonalize $K^T K$ using fast transforms like the FFT or the DCT. Alternatively, the ADMM iterations can be modified to avoid the difficulty caused by the $\|Ku\|^2$ term. In this section we show how AMA (5) and BOS (6) accomplish this by modifying the ADMM iterations in different ways. AMA essentially removes the offending quadratic penalty, while BOS adds an additional quadratic penalty chosen so that it cancels the $\|Ku\|^2$ term. A strict convexity assumption is required to apply AMA, but not for BOS.

3.3.1   AMA Applied to Primal Problem

In order to apply AMA to (P0), either F or H must be strictly convex. Assume for now that $H(u)$ is strictly convex with modulus $\sigma > 0$. The additional strict convexity assumption is needed so that the step of minimizing the non-augmented Lagrangian is well defined.

Recalling the definitions of Ψ and ϕ (25), proximal forward backward splitting applied to the dual problem (Q0) yields

$$\lambda^{k+1} = (I + \alpha\Psi)^{-1}(I - \alpha\phi)\lambda^k, \tag{33}$$

where λ^0 is arbitrary. Note that $\phi(\lambda^k)$ is single valued because of the strict convexity of $H(u)$. Also, $(I + \alpha\Psi)^{-1}$ is well defined because Ψ is maximal monotone. So (33) determines λ^{k+1} uniquely given λ^k.

As Tseng shows in [56], (33) is equivalent to

$$u^{k+1} = \arg\min_{u \in \mathbb{R}^m} H(u) - \langle A^T\lambda^k, u\rangle \tag{34a}$$

$$z^{k+1} = \arg\min_{z \in \mathbb{R}^n} F(z) - \langle B^T\lambda^k, z\rangle + \frac{\alpha}{2}\|b - Au^{k+1} - Bz\|^2 \tag{34b}$$

$$\lambda^{k+1} = \lambda^k + \alpha(b - Au^{k+1} - Bz^{k+1}). \tag{34c}$$

To see the equivalence, note that optimality of u^{k+1} implies $A^T\lambda^k \in \partial H(u^{k+1})$. It follows that

$$Au^{k+1} \in A\partial H^*(A^T\lambda^k) = \phi(\lambda^k).$$

Similarly, optimality of z^{k+1} implies

$$Bz^{k+1} - b \in \Psi(\lambda^{k+1}).$$

Since $\lambda^{k+1} = \lambda^k + \alpha(b - Au^{k+1} - Bz^{k+1})$,

$$0 \in \lambda^{k+1} + \alpha\Psi(\lambda^{k+1}) - \lambda^k + \alpha\phi(\lambda^k),$$

from which (33) follows.

Tseng shows that $\{u^k, z^k\}$ converges to a solution of (P0) and $\{\lambda^k\}$ converges to a solution of (Q0) if α, which he allows to depend on k, satisfies the time step restriction

$$\epsilon \le \alpha_k \le \frac{4\sigma}{\|A\|^2} - \epsilon \tag{35}$$

for some $\epsilon \in (0, \frac{2\sigma}{\|A\|^2})$.

It is tempting to try to extend AMA to the non strictly convex case by adding an extra variable. Consider applying AMA to (1) where J is closed proper convex but not strictly convex. A step in the method of multipliers applied this problem would require minimizing $J(u) + \langle\lambda^k, f - Ku\rangle + \frac{\alpha}{2}\|f - Ku\|^2$. To decouple the variables coupled by the matrix K, we can consider rewriting the problem as

$$\min_{\substack{z \in \mathbb{R}^m, u \in \mathbb{R}^m \\ Ku = f \\ z = u}} J(z) + \frac{c}{2}\|z - u\|^2,$$

Where $c > 0$. The Lagrangian for this problem is

$$L_c(z, u, \lambda, q) = J(z) + \frac{c}{2}\|z - u\|^2 + \langle\lambda, f - Ku\rangle + \langle q, z - u\rangle,$$

14

and the augmented Lagrangian is

$$L_{c,\alpha}(z,u,\lambda,q) = L_c(z,u,\lambda,q) + \frac{\alpha}{2}\|f - Ku\|^2 + \frac{\alpha}{2}\|z - u\|^2.$$

Since L_c is strictly convex in u, we can consider applying an AMA-like approach where we alternately minimize L_c with respect to u, then $L_{c,\alpha}$ with respect to z, and finally update the multipliers λ and q. Although empirically this works for α sufficiently small at least in the case where $J(z) = \|z\|_1$, it's important to note that this isn't actually an application of AMA. Because of the coupling of z and u, $J(z) + \frac{c}{2}\|z - u\|^2$ cannot be written as $F(z) + H(u)$ with $H(u)$ strictly convex. So the convergence theory for AMA doesn't immediately extend to this application.

3.3.2   BOS Applied to Primal Problem

The BOS algorithm applied to (1) was interpreted by Zhang, Burger, Bresson and Osher in [61] as an inexact Uzawa method. It modifies the augmented Lagrangian not by removing the quadratic penalty, but by adding an additional proximal-like penalty chosen so that the $\|Ku\|^2$ term cancels out. It simplifies the minimization step by decoupling the variables coupled by the constraint matrix K, and it doesn't require the functional J to be strictly convex. In a sense it combines the best advantages of Rockafellar's proximal method of multipliers [47] and Daubechies, Defrise and De Mol's surrogate functional technique [20]. Recall that the method of multipliers (2) applied to (1) requires solving

$$u^{k+1} = \arg\min_{u \in \mathbb{R}^m} J(u) + \langle \lambda^k, f - Ku \rangle + \frac{\alpha}{2}\|f - Ku\|^2.$$

The inexact Uzawa method in [61] modifies that objective functional by adding the term

$$\frac{1}{2}\langle u - u^k, (\frac{1}{\delta} - \alpha K^T K)(u - u^k)\rangle,$$

where δ is chosen such that $0 < \delta < \frac{1}{\alpha\|K^T K\|}$ in order that $(\frac{1}{\delta} - \alpha K^T K)$ is positive definite. Combining and rewriting terms yields

$$u^{k+1} = \arg\min_{u \in \mathbb{R}^m} J(u) + \frac{1}{2\delta}\|u - u^k + \alpha\delta K^T(Ku^k - f - \frac{\lambda^k}{\alpha})\|^2.$$

The new penalty keeps u^{k+1} close to a linear approximation of the old penalty evaluated at u^k, and the iteration is simplified because the variables u are no longer coupled together by K. An important example is the case when $J(u) = \|u\|_1$, in which case the decoupled functional can be explicitly minimized by a shrinkage formula discussed in section 4.2. In [61], the algorithm was combined with split Bregman and applied to more complicated problems such as one involving nonlocal total variation regularization. Applying the same decoupling trick to ADMM iterations means selectively replacing some quadratic penalties of the form $\frac{\alpha}{2}\|Ku - f\|^2$ with their linearized counterparts $\frac{1}{2\delta}\|u - u^k + \alpha\delta K^T(Ku^k - f)\|^2$. An example application to constrained TV minimization is given in section 4.7. The convergence theory from [61] has been extended by Zhang to this splitting application in [60].

4 Example Applications

Here we give a few examples of how to write several optimization problems from image processing in the form (P0) so that application of ADMM takes advantage of the separable structure of the problems and produces efficient, numerically stable methods. The example problems that follow involve minimizing combinations of the l_1 norm, the square of the l_2 norm, and a discretized version of the total variation seminorm. ADMM applied to these problems often requires solving a Poisson equation or l_1-l_2 minimization. So we first define the discretizations used, the discrete cosine transform, which can be used for solving the Poisson equations, and also the shrinkage formulas that solve the l_1-l_2 minimization problems.

4.1 Notation Regarding Discretizations Used

A straightforward way to define a discretized version of the total variation seminorm is by

$$\|u\|_{TV} = \sum_{p=1}^{M_r} \sum_{q=1}^{M_c} \sqrt{(D_1^+ u_{p,q})^2 + (D_2^+ u_{p,q})^2} \tag{36}$$

for $u \in \mathbb{R}^{M_r \times M_c}$. Here, D_k^+ represents a forward difference in the kth index and we assume Neumann boundary conditions. It will be useful to instead work with vectorized $u \in \mathbb{R}^{M_r M_c}$ and to rewrite $\|u\|_{TV}$. The convention for vectorizing an M_r by M_c matrix will be to associate the (p,q) element of the matrix with the $(q-1)M_r + p$ element of the vector. Consider a graph $G(\mathcal{E}, \mathcal{V})$ defined by an M_r by M_c grid with $\mathcal{V} = \{1, ..., M_r M_c\}$ the set of $m = M_r M_c$ nodes and $\mathcal{E}$ the set of $e = 2M_r M_c - M_r - M_c$ edges. Assume the nodes are indexed so that the node corresponding to element (p,q) is indexed by $(q-1)M_r + p$. The edges, which will correspond to forward differences, can be indexed arbitrarily. Define $D \in \mathbb{R}^{e \times m}$ to be the edge-node adjacency matrix for this graph. So for a particular edge $\eta \in \mathcal{E}$ with endpoint indices $i, j \in \mathcal{V}$ and $i < j$, we have

$$D_{\eta,k} = \begin{cases} -1 & \text{for } k = i, \\ 1 & \text{for } k = j, \\ 0 & \text{for } k \neq i, j. \end{cases} \tag{37}$$

Also define $E \in \mathbb{R}^{e \times m}$ such that

$$E_{\eta,k} = \begin{cases} 1 & \text{if } D_{\eta,k} = -1, \\ 0 & \text{otherwise.} \end{cases} \tag{38}$$

The matrix E will be used to identify the edges used in each forward difference. Now define a norm on $\mathbb{R}^e$ by

$$\|w\|_E = \sum_{k=1}^{m} \left(\sqrt{E^T(w^2)} \right)_k. \tag{39}$$

With this notation the discrete TV seminorm defined above (36) can be written as

$$\|u\|_{TV} = \|Du\|_E.$$

The matrix D is a discretization of the gradient and $-D^T$ is the corresponding discretization of the divergence. The product $-D^T D$ defines the discrete Laplacian $\triangle$ corresponding to Neumann

boundary conditions. It is diagonalized by the basis for the discrete cosine transform. Let $\tilde{g} \in \mathbb{R}^{M_r \times M_c}$ denote the discrete cosine transform of $g \in \mathbb{R}^{M_r \times M_c}$ defined by

$$\tilde{g}_{s,t} = \sum_{p=1}^{M_r} \sum_{q=1}^{M_c} g_{p,q} \cos\left(\frac{\pi}{M_r}(p - \frac{1}{2})s\right) \cos\left(\frac{\pi}{M_c}(q - \frac{1}{2})t\right)$$

Like the fast Fourier transform, this can be computed with $O(M_r M_c \log(M_r M_c))$ complexity. The discrete Laplacian can be computed by

$$\widetilde{(\triangle g)}_{s,t} = \left(2 \cos\left(\frac{\pi(s-1)}{M_r}\right) + 2 \cos\left(\frac{\pi(t-1)}{M_c}\right) - 4\right) \tilde{g}_{s,t}.$$

4.2 Shrinkage Formulas

When the original functional involves the l_1 norm or the TV seminorm, application of split Bregman or ADMM will result in l_1-l_2 minimization problems that can be solved by soft thresholding, or shrinkage formulas, which will be defined in this section. Consider

$$\min_w \sum_i \left(\mu\|w_i\| + \frac{1}{2}\|w_i - f_i\|^2\right),\tag{40}$$

where $w_i, f_i \in \mathbb{R}^{s_i}$. This decouples into separate problems of the form $\min_{w_i} \Theta_i(w_i)$ where

$$\Theta_i(w_i) = \mu\|w_i\| + \frac{1}{2}\|w_i - f_i\|^2.\tag{41}$$

Consider the case when $\|f_i\| \leq \mu$. Then

$$\begin{aligned}
\Theta_i(w_i) &= \mu\|w_i\| + \frac{1}{2}\|w_i\|^2 + \frac{1}{2}\|f_i\|^2 - \langle w_i, f_i \rangle \\
&\geq \mu\|w_i\| + \frac{1}{2}\|w_i\|^2 + \frac{1}{2}\|f_i\|^2 - \|w_i\|\|f_i\| \\
&= \frac{1}{2}\|w_i\|^2 + \frac{1}{2}\|f_i\|^2 + \|w_i\|(\mu - \|f_i\|) \\
&\geq \frac{1}{2}\|f_i\|^2 = \Theta_i(0),
\end{aligned}$$

which implies $w_i = 0$ is the minimizer when $\|f_i\| \leq \mu$. In the case where $\|f_i\| > \mu$, let

$$w_i = (\|f_i\| - \mu)\frac{f_i}{\|f_i\|},$$

which is nonzero by assumption. Then Θ is differentiable at w_i and

$$\nabla\Theta(w_i) = \mu\frac{w_i}{\|w_i\|} + w_i - f_i,$$

which equals zero because

$$\frac{w_i}{\|w_i\|} = \frac{f_i}{\|f_i\|}.$$

So altogether, the minimizer of (40) is given by

$$w_i = \begin{cases} w_i = (\|f_i\| - \mu)\frac{f_i}{\|f_i\|} & \text{if } \|f_i\| > \mu \\ 0 & \text{otherwise} \end{cases}. \tag{42}$$

When $f_i, w_i \in \mathbb{R}$ are the components of $f, w \in \mathbb{R}^s$, $\frac{f_i}{\|f_i\|}$ is just $\text{sign}(f_i)$. Define the scalar shrinkage operator S by

$$S_\mu(f)_\gamma = \begin{cases} f_\gamma - \mu\,\text{sign}(f_\gamma) & \text{if } |f_\gamma| > \mu \\ 0 & \text{otherwise} \end{cases}, \tag{43}$$

where $\gamma = 1, 2, ..., s$. This can be interpreted as solving the minimization problem,

$$S_\mu(f) = \arg\min_{w \in \mathbb{R}^s} \mu\|w\|_1 + \frac{1}{2}\|w - f\|^2.$$

The formula (42) can be interpreted as $w_i = S_\mu(\|f_i\|)\frac{f_i}{\|f_i\|}$, which is to say scalar shrinkage of $\|f_i\|$ in the direction of f_i. Note also that the problem of minimizing over $w \in \mathbb{R}^e$

$$\mu\|w\|_E + \frac{1}{2}\|w - z\|^2, \tag{44}$$

which arises in TV minimization problems, is of the form (40). In the notation of the previous section, it can be written as

$$\min_{w \in \mathbb{R}^e} \sum_{k=1}^m \left[\mu\left(\sqrt{E^T(w^2)}\right)_k + \frac{1}{2}\left(E^T(w - z)^2\right)_k \right].$$

Let

$$s = E\sqrt{E^T(z)^2}.$$

Similar to the scalar case, by applying (42) for $\gamma = 1, 2, ..., e$ we can define the operator $\tilde{S}_\mu(z)$ that solves (44) by

$$\tilde{S}_\mu(z)_\gamma = \begin{cases} z_\gamma - \mu\frac{z_\gamma}{s_\gamma} & \text{if } s_\gamma > \mu \\ 0 & \text{otherwise} \end{cases}. \tag{45}$$

4.3   ADMM Applied to Constrained TV Minimization

One of the example applications of split Bregman that was presented in [32] is constrained total variation minimization. Here we consider the same example but in the context of applying ADMM to (P0). Consider

$$\min_{\substack{u \in \mathbb{R}^m \\ Ku = f}} \|u\|_{TV},$$

which can be rewritten using the norm $\|\cdot\|_E$ defined in section 4.1 as

$$\min_{\substack{u \in \mathbb{R}^m \\ Ku = f}} \|Du\|_E. \tag{46}$$

Writing this in the form of (P0) while taking advantage of the separable structure, we let

$$z = Du \qquad B = \begin{bmatrix} -I \\ 0 \end{bmatrix} \qquad A = \begin{bmatrix} D \\ K \end{bmatrix} \qquad b = \begin{bmatrix} 0 \\ f \end{bmatrix}.$$

Now the problem can be written

$$\min_{\substack{z \in \mathbb{R}^n, u \in \mathbb{R}^m \\ Bz + Au = b}} \|z\|_E.$$

We assume that $\ker(D) \bigcap \ker(K) = \{0\}$, or equivalently that $\ker(K)$ does not contain the vector of all ones. This ensures that A has full column rank, so Theorem 3.1 can be used to guarantee convergence of ADMM applied to this problem. Introducing a dual variable λ, the augmented Lagrangian is

$$\|z\|_E + \langle \lambda, b - Bz - Au \rangle + \frac{\alpha}{2} \|b - Bz - Au\|^2.$$

Let $\lambda = \begin{bmatrix} p \\ q \end{bmatrix}$ and rewrite the augmented Lagrangian as

$$\|z\|_E + \langle p, z - Du \rangle + \langle q, f - Ku \rangle + \frac{\alpha}{2} \|z - Du\|^2 + \frac{\alpha}{2} \|f - Ku\|^2.$$

Moving linear terms into the quadratic terms, the ADMM iterations are given by

$$z^{k+1} = \arg\min_z \|z\|_E + \frac{\alpha}{2} \|z - Du^k + \frac{p^k}{\alpha}\|^2$$

$$u^{k+1} = \arg\min_u \frac{\alpha}{2} \|Du - z^{k+1} - \frac{p^k}{\alpha}\|^2 + \frac{\alpha}{2} \|Ku - f - \frac{q^k}{\alpha}\|^2$$

$$p^{k+1} = p^k + \alpha(z^{k+1} - Du^{k+1})$$

$$q^{k+1} = q^k + \alpha(f - Ku^{k+1}),$$

where $p^0 = q^0 = 0$, u^0 is arbitrary and $\alpha > 0$. Note that this example corresponds to the maximally decoupled case, in which the u update has the interesting interpretation of enforcing the constraint $A^T \lambda = 0$. Here, since $D^T p^0 + K^T q^0 = 0$ and by the optimality condition for u^{k+1}, it follows that $D^T p^k + K^T q^k = 0$ for all k. In particular, this makes the q^{k+1} update unnecessary. The explicit ADMM steps reduce to

$$z^{k+1} = \tilde{S}_{\frac{1}{\alpha}}(Du^k - \frac{p^k}{\alpha})$$

$$u^{k+1} = (-\triangle + K^T K)^{-1} \left(D^T z^{k+1} + \frac{D^T p^k}{\alpha} + K^T f + \frac{K^T q^k}{\alpha} \right)$$

$$= (-\triangle + K^T K)^{-1} \left(D^T z^{k+1} + K^T f \right)$$

$$p^{k+1} = p^k + \alpha(z^{k+1} - Du^{k+1}).$$

Since the discrete cosine basis diagonalizes the discrete Laplacian for Neumann boundary conditions, this can be efficiently solved whenever $K^T K$ can be simultaneously diagonalized.

4.4 ADMM Applied to TV-l_1

The same decomposition principle applied to constrained TV minimization also applies to the discrete TV-l_1 minimization problem ([11], [12]),

$$\min_{u \in \mathbb{R}^m} \|u\|_{TV} + \beta \|Ku - f\|_1,$$

which can be rewritten as

$$\min_{u \in \mathbb{R}^m} \|Du\|_E + \beta \|Ku - f\|_1. \tag{47}$$

Writing this in the form of (P0), we let

$$z = \begin{bmatrix} w \\ v \end{bmatrix} = \begin{bmatrix} Du \\ Ku - f \end{bmatrix} \qquad B = -I \qquad A = \begin{bmatrix} D \\ K \end{bmatrix} \qquad b = \begin{bmatrix} 0 \\ f \end{bmatrix}.$$

Again assume that $\ker(D) \bigcap \ker(K) = \{0\}$, or $\ker(K)$ does not contain the vector of all ones. With this assumption, Theorem 3.1 again applies. Introducing the dual variable λ, which we decompose into $\lambda = \begin{bmatrix} p \\ q \end{bmatrix}$, the augmented Lagrangian can be written

$$\|w\|_E + \beta\|v\|_1 + \langle p, w - Du \rangle + \langle q, v - Ku + f \rangle + \frac{\alpha}{2}\|w - Du\|^2 + \frac{\alpha}{2}\|v - Ku + f\|^2.$$

Minimizing over z would correspond to simultaneously minimizing over w and v. But no term in the augmented Lagrangian contains both w and v, so it is equivalent to separately minimizing over w and over v.

The ADMM iterations are given by

$$w^{k+1} = \arg\min_w \|w\|_E + \frac{\alpha}{2}\|w - Du^k + \frac{p^k}{\alpha}\|^2$$

$$v^{k+1} = \arg\min_v \beta\|v\|_1 + \frac{\alpha}{2}\|v - Ku^k + f + \frac{q^k}{\alpha}\|^2$$

$$u^{k+1} = \arg\min_u \frac{\alpha}{2}\|Du - w^{k+1} - \frac{p^k}{\alpha}\|^2 + \frac{\alpha}{2}\|Ku - v^{k+1} - f - \frac{q^k}{\alpha}\|^2$$

$$p^{k+1} = p^k + \alpha(w^{k+1} - Du^{k+1})$$

$$q^{k+1} = q^k + \alpha(v^{k+1} - Ku^{k+1} + f),$$

where $p^0 = q^0 = 0$, u^0 is arbitrary and $\alpha > 0$. Again, corresponding to the $A^T \lambda = 0$ constraint in the dual problem, since $D^T p^0 + K^T q^0 = 0$ and by the optimality condition for u^{k+1}, it follows that $D^T p^k + K^T q^k = 0$ for all k. The explicit formulas for w^{k+1}, v^{k+1} and u^{k+1} are given by

$$w^{k+1} = \tilde{S}_{\frac{1}{\alpha}}(Du^k - \frac{p^k}{\alpha})$$

$$v^{k+1} = S_{\frac{\beta}{\alpha}}(Ku^k - f - \frac{q^k}{\alpha})$$

$$u^{k+1} = (-\triangle + K^T K)^{-1}\left(D^T w^{k+1} + \frac{D^T p^k}{\alpha} + K^T(v^{k+1} + f) + \frac{K^T q^k}{\alpha}\right)$$

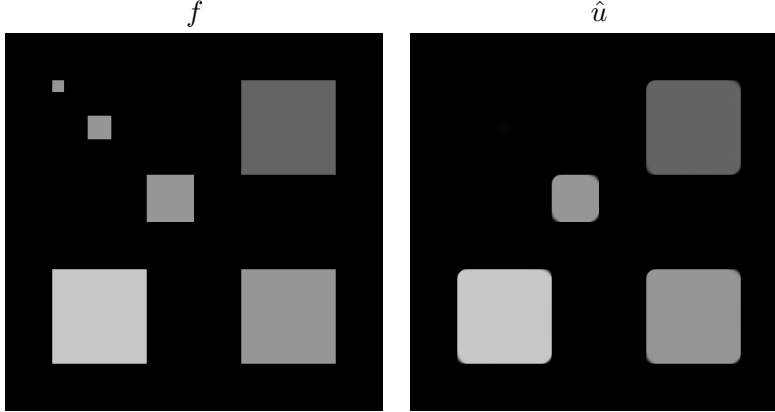$$= (-\triangle + K^T K)^{-1}\left(D^T w^{k+1} + K^T(v^{k+1} + f)\right).$$

Figure 1: TV-l_1 Minimization of 512×512 Synthetic Image

Image Size	Iterations	Time
64×64	40	1s
128×128	51	5s
256×256	136	78s
512×512	359	836s

Table 1: Iterations and Time Required for TV-l_1 Minimization

To get a sense of the speed of this algorithm, we let $K = I$ and test it numerically on a synthetic grayscale image similar to one from [11]. The intensities range from 0 to 255 and the image is scaled to sizes 64×64, 128×128, 256×256 and 512×512. Let $\beta = .6$, $.3$, $.15$ and $.075$ for the different sizes respectively. Similarly let $\alpha = .02$, $.01$, $.005$ and $.0025$. Let $\hat{u}$ denote u^k at the first iteration $k > 1$ such that $\|u^k - u^{k-1}\|_\infty \leq .5$, $\|Du^k - w^k\|_\infty \leq .5$ and $\|v^k - u^k + f\|_\infty \leq .5$. The original image f and the result $\hat{u}$ are shown in Figure 1. The number of iterations required and time to compute on an average PC running a MATLAB implementation are tabulated in Table 1.

4.5 ADMM Applied to TV-l_2

An example where there is more than one effective way to apply ADMM is the TV-l_2 minimization problem

$$\min_{u \in \mathbb{R}^m} \|u\|_{TV} + \frac{\lambda}{2}\|Ku - f\|^2,$$

which can be rewritten as

$$\min_{u \in \mathbb{R}^m} \|Du\|_E + \frac{\lambda}{2}\|Ku - f\|^2. \tag{48}$$

The splitting used by Goldstein and Osher for this problem in [32] can be written in the form of (P0) by letting

$$z = Du \qquad B = -I \qquad A = D \qquad b = 0.$$

Note that $F(z) = \|z\|_E$ and $H(u) = \frac{\lambda}{2}\|Ku - f\|^2$. Introducing the dual variable p, the augmented Lagrangian can be written

$$\|z\|_E + \frac{\lambda}{2}\|Ku - f\|^2 + \langle p, z - Du\rangle + \frac{\alpha}{2}\|z - Du\|^2.$$

Assume again that $\ker(D) \bigcap \ker(K) = \{0\}$, or $\ker(K)$ does not contain the vector of all ones. This ensures that $\frac{\lambda}{2}\|Ku - f\|^2 + \|Du\|^2$ is strictly convex, so Theorem 3.1 applies and guarantees the convergence of ADMM.

The ADMM iterations are given by

$$z^{k+1} = \arg\min_z \|z\|_E + \frac{\alpha}{2}\|z - Du^k + \frac{p^k}{\alpha}\|^2$$

$$u^{k+1} = \arg\min_u \frac{\lambda}{2}\|Ku - f\|^2 + \frac{\alpha}{2}\|Du - z^{k+1} - \frac{p^k}{\alpha}\|^2 \qquad (49)$$

$$p^{k+1} = p^k + \alpha(z^{k+1} - Du^{k+1}).$$

The explicit formulas for z^{k+1} and u^{k+1} are

$$z^{k+1} = \tilde{S}_{\frac{1}{\alpha}}(Du^k - \frac{p^k}{\alpha})$$

$$u^{k+1} = (-\alpha\triangle + \lambda K^T K)^{-1}\left(\lambda K^T f + \alpha D^T z^{k+1} + D^T p^k\right).$$

Another approach is to apply ADMM to TV-l_2 as it was applied to TV-l_1. This corresponds to the maximally decoupled case and involves adding new variables not just for the TV term but also for the l_2 term when rewriting (48) in the form of (P0). Let

$$z = \begin{bmatrix} w \\ v \end{bmatrix} = \begin{bmatrix} Du \\ Ku - f \end{bmatrix} \qquad B = -I \qquad A = \begin{bmatrix} D \\ K \end{bmatrix} \qquad b = \begin{bmatrix} 0 \\ f \end{bmatrix}.$$

Note that $F(z) = \|w\|_E + \frac{\lambda}{2}\|v\|^2$, $H(u) = 0$ and A has full column rank. The augmented Lagrangian can be written

$$\|w\|_E + \frac{\lambda}{2}\|v\|^2 + \langle p, w - Du\rangle + \langle q, v - Ku + f\rangle + \frac{\alpha}{2}\|w - Du\|^2 + \frac{\alpha}{2}\|v - Ku + f\|^2.$$

As with the TV-l_1 example, minimizing over z would correspond to simultaneously minimizing over w and v, which here is equivalent to separately minimizing over w and over v.

The ADMM iterations are then

$$w^{k+1} = \arg\min_w \|w\|_E + \frac{\lambda}{2}\|w - Du^k + \frac{p^k}{\alpha}\|^2$$

$$v^{k+1} = \arg\min_v \frac{\lambda}{2}\|v\|^2 + \frac{\alpha}{2}\|v - Ku^k + f + \frac{q^k}{\alpha}\|^2$$

$$u^{k+1} = \arg\min_u \frac{\alpha}{2}\|Du - w^{k+1} - \frac{p^k}{\alpha}\|^2 + \frac{\alpha}{2}\|Ku - v^{k+1} - f - \frac{q^k}{\alpha}\|^2$$

$$p^{k+1} = p^k + \alpha(w^{k+1} - Du^{k+1})$$

$$q^{k+1} = q^k + \alpha(v^{k+1} - Ku^{k+1} + f).$$

22

The formulas for w^{k+1}, v^{k+1} and u^{k+1} are

$$w^{k+1} = \tilde{S}_{\frac{1}{\alpha}}(Du^k - \frac{p^k}{\alpha})$$

$$v^{k+1} = \frac{1}{\lambda + \alpha}(\alpha Ku^k - \alpha f - q^k)$$

$$u^{k+1} = (-\triangle + K^T K)^{-1}\left(K^T f + D^T w^{k+1} + K^T v^{k+1}\right).$$

By substituting v^{k+1} into the update for u^{k+1} and using the fact that $D^T p^k + K^T q^k = 0$ for all k, the updates for q and v can be eliminated. The remaining iterations are

$$w^{k+1} = \tilde{S}_{\frac{1}{\alpha}}(Du^k - \frac{p^k}{\alpha})$$

$$u^{k+1} = (-\triangle + K^T K)^{-1}\left(\frac{\lambda K^T f}{\lambda + \alpha} + D^T w^{k+1} + \frac{D^T p^k}{\lambda + \alpha} + \frac{\alpha K^T Ku^k}{\lambda + \alpha}\right)$$

$$p^{k+1} = p^k + \alpha(w^{k+1} - Du^{k+1}).$$

This alternative application of ADMM to TVL2 is very similar to the first(49), differing only in the update for u^{k+1}. Empirically, at least in the denoising case for $K = I$, the two approaches perform similarly. But since the algorithm is neither simplified nor improved by the additional decoupling of the l_2 term, there is no compelling reason to do it.

An approach suggested in [32] for speeding up the iterations of (49) is to only approximately solve for u^{k+1} using several Gauss Seidel iterations instead of solving a Poisson equation. Convergence of the resulting approximate algorithm could be guaranteed by Theorem 3.1 if we knew that the sum of the norms of the errors was finite, but this is a difficult thing to know in advance. Since $H(u)$ was strictly convex in the first method for TV-l_2, an alternative approach to simplifying the iterations is to apply AMA.

4.6 AMA Applied to TV-l_2

Consider again the TV-l_2 problem (48) in the denoising case where $K = I$. Since $H(u)$ is strictly convex, we can apply AMA to obtain a similar algorithm that doesn't require solving the Poisson equation. Recall the Lagrangian for this problem is given by

$$\|z\|_E + \frac{\lambda}{2}\|u - f\|^2 + \langle p, z - Du\rangle.$$

The AMA iterations are

$$u^{k+1} = \arg\min_u \frac{\lambda}{2}\|u - f\|^2 - \langle D^T p^k, u\rangle$$

$$z^{k+1} = \arg\min_z \|z\|_E + \frac{\alpha}{2}\|z - Du^{k+1} + \frac{p^k}{\alpha}\|^2 \tag{50}$$

$$p^{k+1} = p^k + \alpha(z^{k+1} - Du^{k+1}).$$

$$\tag{51}$$

The explicit formulas for z^{k+1} and u^{k+1} are

$$u^{k+1} = f + \frac{D^T p^k}{\lambda}$$

$$z^{k+1} = \tilde{S}_{\frac{1}{\alpha}}(Du^{k+1} - \frac{p^k}{\alpha}).$$

Note that α must satisfy the time step restriction from (35). Since $H(u)$ is strictly convex with modulus $\frac{\lambda}{2}$, a safe choice for α is to let $\alpha \leq \frac{\lambda}{\|D\|^2}$. We can bound $\|D\|^2$ by the largest eigenvalue of $D^T D$, which is minus the discrete Laplacian corresponding to Neumann boundary conditions. The matrix $D^T D$ from its definition has only the numbers 2, 3 and 4 on its main diagonal. All the off diagonal entries are 0 or -1, and the rows sum to zero. Therefore, by the Gersgorin Circle Theorem, all eigenvalues of $D^T D$ are in the interval $[0, 8]$. Thus $\|D\|^2 \leq 8$, so we can take $\alpha = \frac{\lambda}{8}$.

For this example, since it is already efficient to solve the Poisson equation using the discrete cosine transform, the benefit of slightly faster iterations is slightly outweighed by the reduced stability and the additional iterations required.

4.7 BOS Applied to Constrained TV

Consider again the constrained TV minimization problem (46) but now with a more complicated matrix K that makes the update for u^{k+1}

$$u^{k+1} = \arg\min_u \frac{\alpha}{2}\|Du - z^{k+1} - \frac{p^k}{\alpha}\|^2 + \frac{\alpha}{2}\|Ku - f - \frac{q^k}{\alpha}\|^2$$

difficult to compute. Applying the main idea from the BOS algorithm, we can handle the $Ku = f$ constraint in a more explicit manner by adding $\frac{1}{2}\langle u - u^k, (\frac{1}{\delta} - \alpha K^T K)(u - u^k)\rangle$ to the objective functional for the u^{k+1} update, with $0 < \delta < \frac{1}{\alpha\|K^T K\|}$. This yields

$$u^{k+1} = \arg\min_u \frac{\alpha}{2}\|Du - w^{k+1} - \frac{p^k}{\alpha}\|^2 + \frac{1}{2\delta}\|u - u^k + \alpha\delta K^T(Ku^k - f - \frac{q^k}{\alpha})\|^2$$

$$= (\frac{1}{\delta} - \alpha\triangle)^{-1}\left(\alpha D^T w^{k+1} + D^T p^k + \frac{1}{\delta}u^k - \alpha K^T\left(Ku^k - f - \frac{q^k}{\alpha}\right)\right).$$

Altogether, the modified ADMM iterations are given by

$$z^{k+1} = \tilde{S}_{\frac{1}{\alpha}}(Du^k - \frac{p^k}{\alpha})$$

$$u^{k+1} = (\frac{1}{\delta} - \alpha\triangle)^{-1}\left(\alpha D^T w^{k+1} + D^T p^k + \frac{1}{\delta}u^k - \alpha K^T\left(Ku^k - f - \frac{q^k}{\alpha}\right)\right)$$

$$p^{k+1} = p^k + \alpha(z^{k+1} - Du^{k+1})$$

$$q^{k+1} = q^k + \alpha(f - Ku^{k+1}).$$

Although it no longer follows that $D^T p^k + K^T q^k = 0$ as it did for ADMM applied to constrained TV, all updates except for the u^{k+1} step remain the same.
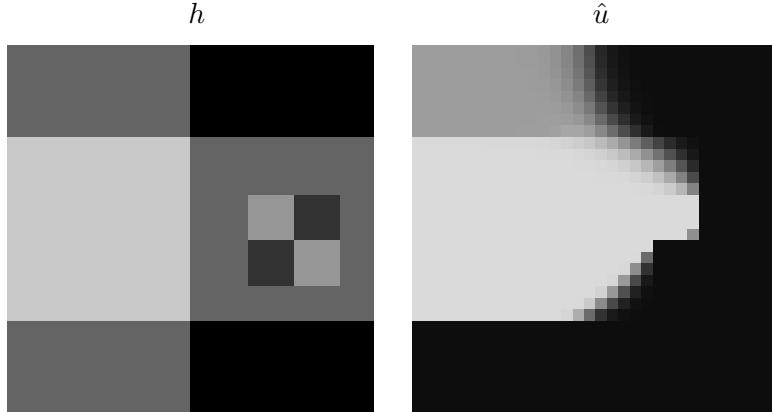
h $\hat{u}$

Figure 2: Constrained TV minimization of 32×32 image subject to constraints on 4 Haar wavelet coefficients

As a numerical test, we will apply this algorithm to a TV wavelet inpainting type problem [16]. Let $K = \mathcal{X}\psi$, where $\mathcal{X}$ is a row selector and ψ is the matrix corresponding to the translation invariant Haar wavelet transform. For a $2^r \times 2^r$ image, there are $(1 + 3r)2^{2r}$ Haar wavelets when all translations are included. The rows of the $(1 + 3r)2^{2r} \times 2^{2r}$ matrix ψ contain these wavelets weighted such that $\psi^T\psi = I$. $\mathcal{X}$ is a diagonal matrix with ones and zeros on the diagonal. For a simple example, let h be a 32×32 image that is a linear combination of four Haar wavelets. Let $\mathcal{X}$ select the corresponding wavelet coefficients and define $f = \mathcal{X}\psi h$. Also choose $\alpha = .01$ and $\delta = 50$. Let $\hat{u} = u^{10000}$, the result after 10000 iterations. Figure 2 shows h and $\hat{u}$. Although $\hat{u}$ may look unusual, it satisfies the four constraints and does indeed have smaller total variation. $\|h\|_{TV} = 1.25 \times 10^4$ whereas $\|\hat{u}\|_{TV} = 1.04 \times 10^4$.

A   ADMM Convergence Proof

This proof of theorem 3.1 is due to Eckstein and Bertsekas and is taken from their paper [23]. Only a few minor changes are needed to accomodate the slightly weaker assumptions made here. In other ways, however, this version is less general because it ignores the relaxation factors ρ_k in [23], which here we take to be one. The entire proof is not reproduced here. Just enough is sketched to make the changes clear.

Proof. Let $J_{\alpha\Psi}$ and $J_{\alpha\phi}$ be shorthand notation for the resolvents $(I + \alpha\Psi)^{-1}$ and $(I + \alpha\phi)^{-1}$ respectively. Also define

$$
\begin{aligned}
y^k &= \lambda^k + \alpha A u^k \qquad , k \geq 0 \\
\hat{\lambda}^k &= \lambda^k + \alpha(b - Bz^{z+1} - Au^k), \qquad k \geq 0 \\
a_k &= \alpha\|B\|\mu_k \qquad , k \geq 0 \\
\beta_0 &= \|\lambda^0 - J_{\alpha\phi}(\lambda^0 - \alpha Au^0)\| \\
\beta_k &= \alpha\|A\|\nu_k, \qquad k \geq 1
\end{aligned}
$$

The main outline of Eckstein and Bertsekas' proof is to show that

$$
\begin{aligned}
(Y1) \qquad & \|\lambda^k - J_{\alpha\phi}(y^k)\| \leq \beta_k \\
(Y2) \qquad & \|\hat{\lambda}^k - J_{\alpha\Psi}(2\lambda^k - y^k)\| \leq a_k \\
(Y3) \qquad & y^{k+1} = y^k + \hat{\lambda}^k - \lambda^k
\end{aligned}
$$

hold for all $k \geq 0$. Then assuming there exists a saddle point of $L(z, u, \lambda)$ (7), they apply an earlier theorem in their paper to say that $\{y^k\}$ converges. This theorem still applies here with the slightly different assumptions. Finally they argue that $z^k \to z^*$, $u^k \to u^*$ and $\lambda^k \to \lambda^*$, where (z^*, u^*, λ^*) is a saddle point of $L(z, u, \lambda)$. Some changes are made to this last part.

Noting that $(Y1)$ is true for $k = 0$, they suppose it is true at iteration k and show it follows that $(Y2)$ is true at k. Define

$$
\bar{z}^k = \arg\min_{z \in R^n} F(z) + \langle \lambda^k, -Bz \rangle + \frac{\alpha}{2}\|b - Bz - Au^k\|^2
$$

and

$$
\tilde{\lambda}^k = \lambda^k + \alpha(b - B\bar{z}^k - Au^k).
$$

Note that $\bar{z}^k$ is uniquely determined because B has full column rank. From the optimality conditions for the $\bar{z}^k$ update, it follows that

$$
\bar{z}^k \in \partial F^*(B^T\tilde{\lambda}^k),
$$

and therefore that

$$
B\bar{z}^k - b \in \Psi(\tilde{\lambda}^k).
$$

Since

$$
\tilde{\lambda}^k + \alpha(B\bar{z}^k - b) = \lambda^k - \alpha Au^k \in \tilde{\lambda}^k + \alpha\Psi(\tilde{\lambda}^k),
$$

it follows that

$$
\tilde{\lambda}^k = J_{\alpha\Psi}(\lambda^k - \alpha Au^k) = J_{\alpha\Psi}(2\lambda^k - y^k).
$$

Then

$$\|\hat{\lambda}^k - J_{\alpha\Psi}(2\lambda^k - y^k)\| = \|\hat{\lambda}^k - \tilde{\lambda}^k\| = \alpha\|B(z^{k+1} - \bar{z}^k)\| \le \alpha\|B\|\|z^{k+1} - \bar{z}^k\| \le \alpha\|B\|\mu_k = a_k.$$

Thus $(Y2)$ holds at iteration k. Next they assume $(Y1)$ and $(Y2)$ hold at k and define

$$s^k = y^k + \hat{\lambda}^k - \lambda^k$$
$$= \lambda^k + \alpha(b - Bz^{k+1})$$
$$\bar{u}^k = \arg\min_{u \in \mathbb{R}^m} H(u) + \langle \lambda^k, -Au \rangle + \frac{\alpha}{2}\|b - Bz^{k+1} - Au\|^2$$
$$\tilde{s}^k = \lambda^k + \alpha(b - Bz^{k+1} - A\bar{u}^k).$$

$(Y3)$ holds trivially since

$$y^{k+1} = \lambda^{k+1} + \alpha Au^{k+1} = \lambda^k + \alpha(b - Bz^{k+1}) = y^k + \hat{\lambda}^k - \lambda^k.$$

Next, from the assumption that $H(u) + \|Au\|^2$ is strictly convex, it follows that $\bar{u}^k$ is uniquely determined. The optimality condition for the $\bar{u}^k$ update yields

$$\bar{u}^k \in \partial H^*(A^T \tilde{s}^k)$$

from which it follows that

$$A\bar{u}^k \in \phi(\tilde{s}^k).$$

Since

$$s^k = \tilde{s}^k + \alpha A\bar{u}^k \in \tilde{s}^k + \alpha\phi(\tilde{s}^k),$$

we have that

$$\tilde{s}^k = J_{\alpha\phi}(s^k).$$

Noting that $y^{k+1} = s^k$,

$$\|\lambda^{k+1} - J_{\alpha\phi}(y^{k+1})\| = \|\lambda^{k+1} - J_{\alpha\phi}(s^k)\| = \|\lambda^{k+1} - \tilde{s}^k\| = \alpha\|A(u^{k+1} - \bar{u}^k)\| \le \alpha\|A\|\nu_k = \beta_k,$$

which means $(Y1)$ holds at $k+1$. By induction, $(Y1)$, $(Y2)$ and $(Y3)$ hold for all k. Moreover, the sequences $\{\beta_k\}$ and $\{a_k\}$ are summable by definition. Taken together this satisfies the requirements of a previous theorem in ([23] p. 307), Theorem 7. If there exists a saddle point $L(z, u, \lambda)$, then in particular there exists an optimal dual solution, in which case Theorem 7 implies that y^k converges to $y^* = \lambda^* + \alpha w^*$ such that $w^* \in \phi(\lambda^*)$ and $-w^* \in \Psi(\lambda^*)$. If there is no saddle point, Theorem 7 implies the sequence $\{y^k\}$ is unbounded, which means either $\{\lambda^k\}$ or $\{u^k\}$ is unbounded. In the case where y^k converges, note that

$$y^* \in \lambda^* + \alpha\phi(\lambda^*),$$

so

$$\lambda^* = J_{\alpha\phi}(y^*).$$

From $(Y1)$ and the continuity of $J_{\alpha\phi}$ it follows that $\lambda^k \to \lambda^*$. Let $w^k = Au^k$. Then $w^k = \frac{y^k - \lambda^k}{\alpha}$, which implies $w^k \to \frac{y^* - \lambda^*}{\alpha} = w^*$. If A had full column rank, we could immediately conclude the

convergence of $\{u^k\}$. Instead, define $S(u) = H(u) + \frac{\alpha}{2}\|Au\|^2$, which was assumed to be strictly convex. Rewrite the objective functional for the u minimization step

$$H(u) + \langle \lambda^k, -Au \rangle + \frac{\alpha}{2}\|b - Bz^{k+1} - Au\|^2 = S(u) + \langle \lambda^k, -Au \rangle + \frac{\alpha}{2}\|b - Bz^{k+1}\|^2 + \alpha\langle b - Bz^{k+1}, -Au \rangle.$$

The optimality condition for $\bar{u}^k$ then implies that

$$0 \in \partial S(\bar{u}^k) - A^T(\lambda^k + \alpha(b - Bz^{k+1}))$$
$$0 \in \partial S(\bar{u}^k) - A^T(\lambda^{k+1} + \alpha Au^{k+1})$$
$$A^T y^{k+1} \in \partial S(\bar{u}^k)$$
$$\bar{u}^k \in \partial S^*(A^T y^{k+1}).$$

Since S is strictly convex, S^* is continuously differentiable ([48] 26.3), so $\bar{u}^k = \nabla S^*(A^T y^{k+1})$. Since $\|u^{k+1} - \bar{u}^k\| \to 0$, this implies

$$u^k \to \nabla S^*(A^T y^*).$$

Let $u^* = \nabla S^*(A^T y^*)$. Since $Au^k \to w^*$, we have that $Au^* = w^*$. Now since $\lambda^{k+1} - \lambda^k = \alpha(b - Bz^{k+1} - Au^{k+1}) \to 0$, we have that

$$Bz^{k+1} \to b - Au^*.$$

Since B has full column rank, $z^k \to z^*$ where

$$Au^* + Bz^* = b.$$

Now note that we also have $\tilde{\lambda}^k \to \lambda^*$, $\tilde{s}^k \to \lambda^*$, $\bar{z}^k \to z^*$ and $\bar{u}^k \to u^*$. Recalling the optimality conditions for the u and z update steps,

$$\bar{z}^k \in \partial F^*(B^T \tilde{\lambda}^k) \qquad \text{and} \qquad \bar{u}^k \in \partial H^*(A^T \tilde{s}^k).$$

Citing a result by Brezis [5] regarding limits of maximal monotone operators, it then follows that

$$z^* \in \partial F^*(B^T \lambda^*) \qquad \text{and} \qquad u^* \in \partial H^*(A^T \lambda^*).$$

These together with $Au^* + Bz^* = b$ are exactly the optimality conditions (10) for (P0). Thus (z^*, u^*, λ^*) is a saddle point of $L(z, u, \lambda)$. $\qquad\qquad\square$

References

[1] Bertsekas, D., *Constrained Optimization and Lagrange Multiplier Methods*, Athena Scientific, 1996.

[2] Bertsekas, D., *Nonlinear Programming*, Second Edition, Athena Scientific, 1999.

[3] Bertsekas, D., and Tsitsiklis, J., *Parallel and Distributed Computation*, Prentice Hall, 1989.

[4] Boyd, S., and Vandenberghe, L., *Convex Analysis*, Cambridge University Press, 2006.

[5] Brézis, H., *Opérateurs Miximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert*, North-Holland, Amsterdam, 1973.

[6] Cai, J. F., Chan, R. H., Shen, Z., *A Framelet-based Image Inpainting Algorithm*, Applied and Computational Harmonic Analysis, Vol. 24, Issue 2, 2008, pp. 131-149.

[7] Cai, J., Candes, E., and Shen, Z., *A Singular Value Thresholding Algorithm for Matrix Completion*, UCLA CAM Report [08-77], 2008.

[8] Cai, J., Osher, S., and Shen, Z., *Linearized Bregman Iterations for Frame-Based Image Deblurring*, UCLA CAM Report [08-57], 2008.

[9] Candes, E., Romberg, J., *Practical Signal Recovery from Random Projections*, IEEE Trans. Signal Processing, 2005.

[10] Chambolle, A., *An Algorithm for Total Variation Minimization and Applications*, Journal of Mathematical Imaging and Vision, Vol. 20, pp. 89-97, 2004.

[11] Chan, T. F., and Esedoglu, S., *Aspects of Total Variation Regularized L^1 function approximation*, UCLA CAM Report [04-07], 2004.

[12] Chan, T. F., Esedoglu, S., Nikolova, M., *Algorithms for Finding Global Minimizers of Image Segmentation and Denoising Models*, UCLA CAM Report [04-54], 2004.

[13] Chan, T. F., and Glowinski, R., *Finite-Element Approximation and Iterative Solution of a Class of Mildly Nonlinear Elliptic Equations*, STAN-CS-78-674, Computer Science Department, Stanford, 1978.

[14] Chan, T. F., Golub. G. H., and Mulet, P., *A nonlinear primal dual method for total variation based image restoration*, SIAM J. Sci. Comput., 20, 1999.

[15] Chan, T. F., Shen, J., *Image Processing and Analysis: variational, PDE, wavelet and stochastic methods*, SIAM, Philadelphia, 2005.

[16] Chan, T. F., Shen, J., and Zhou, H., *Total Variation Wavelet Inpainting*, UCLA CAM Report [04-47], 2004.

[17] Chen, G., and Teboulle, M., *A Proximal-Based Decomposition Method for Convex Minimization Problems*, Mathematical Programming, Vol., 64, pp. 81-101, 1994.

[18] Chen, S., Donoho, D., and Saunders, M. A., *Atomic Decomposition by Basis Pursuit*, SIAM Journal on Scientific Computing, 20, 1998, pp. 33-61.

[19] Combettes, P., and Wajs, W., *Signal Recovery by Proximal Forward-Backward Splitting*, Multiscale Modelling and Simulation, 2006.

[20] Daubechies, I., Defrise, M., De Mol, C., *An Iterative Threshilding Algorithm for Linear Inverse Problems with a Sparsity Constraint*, Comm. Pure and Applied Math, Vol. 57, 2004.

[21] Daubechies, I., Teschke, G., *Wavelet Based Image Decomposition by Variational Functionals*, Wavelet Applications in Industrial Processing. Edited by Truchetet, Frdric. Proceedings of the SPIE, Volume 5266, pp. 94-105, 2004.

[22] Douglas, J., and Rachford, H. H., *On the Numerical Solution of Heat Conduction Problems in Two and Three Space Variables*, Transactions of the American mathematical Society 82, 1956, pp. 421-439.

[23] Eckstein, J., and Bertsekas, D., *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Mathematical Programming 55, North-Holland, 1992.

[24] Eckstein, J., *Nonlinear Proximal Point Algorithms Using Bregman Functions, with Applications to Convex Programming*, Mathematics of Operations Research, Vol. 18, No. 1, Feb. 1993.

[25] Eckstein, J., *Splitting Methods for Monotone Operators with Applications to Parallel Optimization*, Ph. D. Thesis, Massachusetts Institute of Technology, Dept. of Civil Engineering, http://hdl.handle.net/1721.1/14356, 1989.

[26] Elmoataz, A., Lezoray, O., and Bougleux, S., *Nonlocal Discrete Regularization on Weighted Graphs: A framework for Image and Manifold Processing*, IEEE, Vol. 17, No. 7, July 2008.

[27] Fortin, M., and Glowinski, R., *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, Trans-Inter-Scientia, 1983.

[28] Gabay, D., *Methodes numeriques pour l'optimisation non-lineaire*, These de Doctorat d'Etat et Sciences Mathematiques, Universite Pierre et Marie Curie, 1979.

[29] Gabay, D., and Mercier, B., *A dual algorithm for the solution of nonlinear variational problems via finite-element approximations*, Comp. Math. Appl., 2 (1976), pp. 17-40.

[30] Glowinski, R., and Le Tallec, P., *Augmented Lagrangian and Operator-splitting Methods in Nonlinear Mechanics*, SIAM 1989.

[31] Glowinski, R., and Marrocco, A., *Sur lapproximation par elements finis dordre un, et la resolution par penalisation-dualite dune classe de problemes de Dirichlet nonlineaires, Rev. Francaise dAut.* Inf. Rech. Oper., R-2 (1975), pp. 41-76.

[32] Goldstein, T., and Osher, S., *The Split Bregman Algorithm for L1 Regularized Problems*, UCLA CAM Report [08-29], April 2008.

[33] Goldstein, T., Bresson, X., Osher, S., *Geometric Applications of the Split Bregman Method: Segmentation and Surface Reconstruction*, UCLA CAM Report [09-06] 2009.

[34] Guo, X., Li, F., and Ng, M. K., *A Fast l_1-TV Algorithm for Image Restoration* http://www.math.hkbu.edu.hk/ICM/pdf/08-13.pdf 2008.

[35] Hale, E., Yin, W., and Zhang, Y., *A Fixed-Point Continuation Method for l_1-Regularized Minimization with Applications to Compressed Sensing*, CAAM Technical Report TR07-07, 2007.

[36] Haubruge, S., Nguyen, H., and Strodiot, J., *Convergence Analysis and Applications of the Glowinski-Le Tallec Splitting Method for Finding a Zero of the Sum of Two Maximal Monotone Operators*, Journal of Optimization Theory and Applications, Vol. 97, No. 3, pp. 645-673, June 1998.

[37] Hestenes, M. R., *Multiplier and Gradient Methods*, Journal of Optimization Theory and Applications, Vol. 4, 1969, pp. 303-320.

[38] Lieu, L. *Contribution to Problems in Image Restoration, Decomposition, and Segmentation by Variational Methods and Partial Differential Equations*, Ph.D Thesis. June 2006, [UCLA CAM Report 06-46]

[39] Lions, P., L., and Mercier, B., *Algorithms for the Sum of Two Nonlinear Operators*, SIAM Journal on Numerical Analysis, Vol. 16, No. 6, Dec., 1979, pp. 964-979.

[40] Minty, G.J., *Monotone (Nonlinear) Operators in Hilbert Space*, Duke Mathematics Journal 29, 1962, pp. 341-346.

[41] Moreau, J. J., *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93, 1965, pp. 273-299.

[42] Nocedal, J., and Wright, S., *Numerical Optimization*, Springer, 1999.

[43] Osher, S., Class Notes from UCLA Seminar, Math 285J, Winter 2009.

[44] Passty, G. B., *Ergodic Convergence to a Zero of the Sum of Monotone Operators in Hilbert Space*, Journal of Mathematical Analysis and Applications, 72, 1979, pp. 383-390.

[45] Peaceman, D. H., and Rachford, H. H., *The Numerical Solution of Parabolic Elliptic Differential Equations*, SIAM Journal on Applied Mathematics, Vol, 3., 1955, pp. 28-41.

[46] Powell, M. J. D., *A Method for nonlinear Constraints in Minimization Problems*, in Optimization, Fletcher, R., (Ed.), Academic Press, New York, 1969, pp. 283-298.

[47] Rockafellar, R. T, *Augmented Lagrangians and Applications of the Proximal Point Algorithm in Convex Programming*, Mathematics of Operations Research, Vol. 1, No. 2, 1976, pp. 97-116.

[48] Rockafellar, R., T., *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[49] Rockafellar, R. T., *Monotone Operators and the Proximal Point Algorithm*, SIAM J. Control and Optimization, Vol. 14, No. 5, 1976.

[50] Rudin, L., Osher, S., and Fatemi, E., *Nonlinear Total Variation Based Noise Removal Algorithms*, Physica D, 60, 1992, pp. 259-268.

[51] Setzer, S., *Split Bregman Algorithm, Douglas-Rachford Splitting and Frame Shrinkage*, http://kiwi.math.uni-mannheim.de/~ssetzer/pub/setzer_fba_fbs_frames08.pdf, 2008.

[52] Stephanopoulos, G., and Westerberg, A. W., *The Use of Hestenes' Method of Multipliers to Resolve Dual Gaps in Engineering System Optimization*, Journal of Optimization Theory and Applications, Vol. 15, No. 3, 1975.

[53] Strang, G., *The Discrete Cosine Transform*, SIAM Review, 1999, http://www-math.mit.edu/~gs/papers/dct.ps

[54] Tai, X,. and Wu, C., *Augmented Lagrangian Method, Dual Methods and Split Bregman Iteration for ROF Model*, UCLA CAM Report [09-05], 2009.

[55] Tseng, P., *Alternating Projection-Proximal Methods for Convex Programming and Variational Inequalities*, SIAM J. Optim., Vol. 7, No. 4, pp. 951-965, Nov. 1997.

[56] Tseng, P., *Applications of a Splitting Algorithm to Decomposition in Convex Programming and Variational Inequalities*, SIAM J. Control and Optimization, Vol. 29, No. 1, pp. 119-138, Jan. 1991.

[57] Wang, Y., Yin, W., Zhang, Y., *A Fast Algorithm for Image Deblurring with Total Variation Regularization*, July 2007, [UCLA CAM Report 07-22]

[58] Yang, J., Zhang, Y., Yin, W., *An Efficient TVL1 Algorithm for Deblurring Multichannel Images Corrupted by Impulsive Noise*, UCLA CAM Report [09-26], 2009.

[59] Yin, W., Osher, S., Goldfarb, D., Darbon, J., *Bregman Iterative Algorithms for l_1-Minimization with Applications to Compressed Sensing*, UCLA CAM Report [07-37], 2007.

[60] Zhang, X., *Convergence of Splitting Application of Inexact Uzawa Method*, (To Appear), 2009.

[61] Zhang, X., Burger, M., Bresson, X., Osher, S., *Bregmanized Nonlocal Regularization for Deconvolution and Sparse Reconstruction*, UCLA CAM Report [09-03] 2009.

[62] Zhu, M., and Chan, T., *An Efficient Primal-Dual Hybrid Gradient Algorithm for Total Variation Image Restoration*, UCLA CAM Report [08-34], May 2008.