p-NORM MINIMIZATION OVER INTERSECTIONS OF CONVEX SETS

İlker Bayram

Istanbul Technical University,
Department of Electronics and Telecommunications Engineering,
Istanbul, Turkey

ABSTRACT

We consider the minimization of the ℓ_p norm subject to convex constraints. The problem considered in this paper may be regarded as a relaxation of a similar problem that employs the ℓ_1 norm. We derive the dual problem, which is unconstrained and devise an algorithm for the dual problem by adapting the Douglas-Rachford algorithm. We demonstrate the utility of the algorithm on an experiment and discuss its differences with an existing algorithm.

***Index Terms*—** Basis pursuit, Douglas-Rachford algorithm, minimum norm solution, Dykstra's algorithm, bridge estimate.

1. INTRODUCTION

Let K_1, K_2, …, K_k be closed, convex sets (with nonempty intersection) in $\mathbb{R}^n$. In this paper, we consider the minimization problem,

$$\min_{\mathbf{x}\in K_1\cap K_2\ldots\cap K_k} \|\mathbf{x}\|_p \tag{1}$$

for $1 < p < \infty$. We derive a dual problem and making use of the dual problem, propose an algorithm to obtain the minimizer.

Although solving (1) with $p \neq 1$ could be of interest per se (see e.g. [8], Sec.6), we consider it as a variation of the basis pursuit (BP) problem [3],

$$\min \|\mathbf{x}\|_1 \quad \text{s.t. } A\,\mathbf{x} = \mathbf{d}\,. \tag{2}$$

BP is equivalent to a linear program and can be solved as such. However, if the observations '$\mathbf{d}$' are noisy, instead of BP, it is desirable to consider the problem

$$\min \|\mathbf{x}\|_1 \quad \text{s.t. } \|A\,\mathbf{x} - \mathbf{d}\|_2 \leq \epsilon, \tag{3}$$

where ϵ is related to the noise level. One approach to address this problem is to consider the unconstrained problem

$$\hat{x}_\lambda = \operatorname*{argmin}_x \|A\,\mathbf{x} - \mathbf{d}\|_2^2 + \lambda\,\|\mathbf{x}\|_1. \tag{4}$$

The author thanks TÜBİTAK for partially funding this research.

For each λ, if we set $\epsilon_\lambda = \|A\,\mathbf{x}_\lambda - \mathbf{d}\|_2$, then x_λ is the minimizer of (3) for $\epsilon = \epsilon_\lambda$. Moreover, for fixed ϵ, we can find λ_ϵ s.t. $\|A\,\mathbf{x}_{\lambda_\epsilon} - \mathbf{d}\| = \epsilon$ – that is, $\mathbf{x}_{\lambda_\epsilon}$ solves (3). There are methods (see, e.g. [7]) to trace the solutions x_λ as λ is varied. As a variant of this, [1] traces the solutions of the LASSO problem

$$\min_{\mathbf{x}} \|A\,\mathbf{x} - \mathbf{d}\|_2 \quad \text{s.t. } \|\mathbf{x}\|_1 \leq \tau \tag{5}$$

for different τ. We also refer to [8] for a relevant interesting approach, which obtains an algorithm for LASSO through a study of the so-called 'bridge estimators'.

In this paper, we replace

(i) the ℓ_1 norm with the ℓ_p norm,

(ii) consider a more general constraint of the form $\mathbf{x} \in K_1 \cap \ldots \cap K_k$ where K_i's are closed convex sets.

These changes lead to the following advantages.

(i) The solution of the dual-problem for the ℓ_1 norm does not directly lead to the solution of the primal problem. Relaxing the ℓ_1 norm allows us to link the primal and the dual solution directly.

(ii) Decomposing the constraint set K as $K = K_1 \cap \ldots \cap K_k$, allows us to utilize the projections onto K_i's which can be easier to realize than projections onto K.

As a final remark, we note that another approach to solving (1) might be to employ more general 'parallel proximal algorithms' as described in [4, 13]. These algorithms employ projections or proximal mappings that are as easy to realize as the ones in the proposed algorithm. Parallel proximal algorithms are also attractive as they can work with $p = 1$. However, they are different from the proposed algorithm in that they do not differentiate between the ℓ_p term and the constraint projections in (1). We provide a further discussion on the possible consequences in Section 4.

Outline

In Section 2 we derive the dual problem and Proposition 1. The Douglas-Rachford algorithm is briefly reviewed and adapted to the dual-problem in Section 3. In Section 4 we provide two different applications of the algorithm. Section 5 is the conclusion.

Notation

Throughout the paper, bold variables, like $\mathbf{z}$, denote vectors in $\mathbb{R}^n$. We denote the i^{th} component of a vector $\mathbf{z}$ as either z_i or $\mathbf{z}(i)$. We caution that, terms like $\mathbf{z}_i$ denote vectors in $\mathbb{R}^n$. For a set K, the 'support function' of K, denoted by $\sigma_K(x)$, is defined as $\sigma_K(x) = \sup_{z\in K}\langle x, z\rangle$. We refer to [12] for a detailed investigation.

2. THE DUAL PROBLEM

In this section, we derive the dual problem and discuss how the solution of the dual problem is related to the solution of the primal problem for $p \neq 1$. Specifically, we show that for $1/p + 1/q = 1$,

$$\min_{\mathbf{x}\in K_1\cap K_2\ldots\cap K_k} \frac{1}{p}\|\,\mathbf{x}\,\|_p^p \tag{6}$$

and

$$\min_{\mathbf{z}_1,\mathbf{z}_2,\ldots,\mathbf{z}_k} \frac{1}{q}\|\,\mathbf{z}_1+\mathbf{z}_2+\ldots+\mathbf{z}_k\,\|_q^q + \sup_{\mathbf{x}_1\in K_1}\langle \mathbf{x}_1,\mathbf{z}_1\rangle + \sup_{\mathbf{x}_2\in K_2}\langle \mathbf{x}_2,\mathbf{z}_2\rangle + \ldots + \sup_{\mathbf{x}_k\in K_k}\langle \mathbf{x}_k,\mathbf{z}_k\rangle. \tag{7}$$

are dual to each other. The minimizers of the two problems are related as,

Proposition 1. Suppose that $1 < p < \infty$. If $\{\mathbf{z}_1^*\ \mathbf{z}_2^*\ \ldots, \mathbf{z}_k^*\}$ minimize (7), then for $\mathbf{z}^* = \mathbf{z}_1^* + \mathbf{z}_2^* + \ldots + \mathbf{z}_k^*$,

$$\mathbf{x}^* = -\operatorname{sign}(\mathbf{z}^*)|\,\mathbf{z}^*\,|^{q-1} := \begin{pmatrix} -\operatorname{sign}(z_1^*)\,(z_1^*)^{q/p} \\ -\operatorname{sign}(z_2^*)\,(z_2^*)^{q/p} \\ \vdots \\ -\operatorname{sign}(z_n^*)\,(z_n^*)^{q/p} \end{pmatrix} \tag{8}$$

minimizes (6).

Remark 1. For $p = 2$ (and hence $q = 2$), the primal-dual pair is well known. Indeed, Dykstra's algorithm [5, 10] can be interpreted as a coordinate-descent type algorithm working on the dual problem [9]. □

Remark 2. For $p = 1$, the solution of the dual gives the signs of the solution for the primal problem.

The dual-problem consists of terms for which the 'proximal operators' are feasible to realize. This in turn makes the Douglas-Rachford algorithm feasible. Our plan is to adapt the Douglas-Rachford algorithm to the dual problem and obtain the minimizer of the primal problem by using Prop. 1.

Derivation of the Dual Problem

We start by noting that

$$P^* = \inf_{\mathbf{x}\in K}\frac{1}{p}\|\,\mathbf{x}\,\|_p^p = \inf_{\mathbf{x}}\sup_{\mathbf{z}} \underbrace{\frac{1}{p}\|\,\mathbf{x}\,\|_p^p + \langle \mathbf{x},\mathbf{z}\rangle - \sigma_K(\mathbf{z})}_{h(\mathbf{x},\mathbf{z})}$$

In words, the minimization problem (1) is equivalent to finding the saddle point of $h(\mathbf{x},\mathbf{z})$ (where we take $K = \cap_i K_i$). Now if we define the dual function $g(\mathbf{z})$ as,

$$g(\mathbf{z}) = \inf_{\mathbf{x}} \frac{1}{p}\|\,\mathbf{x}\,\|_p^p + \langle \mathbf{x},\mathbf{z}\rangle - \sigma_K(\mathbf{z}) \tag{9}$$

we also have $P^* = \sup_{\mathbf{z}} g(\mathbf{z})$ [12]. If the minimum of the lhs of (9) is achieved at $\mathbf{x}^*$, then, it can be shown that,

$$0 \in \operatorname{sign}(x_i^*)\,|x_i^*|^{p-1} + z_i \quad \text{for } i = 1,2,\ldots,n. \tag{10}$$

Here, $\operatorname{sign}(t)$ is a set valued mapping[1] defined as

$$\operatorname{sign}(t) = \begin{cases} -1 & \text{if } t < 0, \\ [-1,1] & \text{if } t = 0, \\ 1 & \text{if } t > 0. \end{cases} \tag{11}$$

From this, we obtain,

$$-z_i \in \operatorname{sign}(x_i^*)\,|x_i^*|^{p-1} \quad \text{for } i = 1,2,\ldots,n. \tag{12}$$

Noting that, $p - 1 = p/q$, this is true if

$$x_i^* = -\operatorname{sign}(z_i)\,|z_i|^{q/p} \quad \text{for } i = 1,2,\ldots,n. \tag{13}$$

Inserting this into the lhs of (9), and noting that $q/p = q - 1$, we get

$$\begin{aligned} g(\mathbf{z}) &= \frac{1}{p}\|\,\mathbf{z}\,\|_q^q + \langle -\operatorname{sign}(\mathbf{z})\,|\,\mathbf{z}\,|^{q/p}, \mathbf{z}\rangle - \sigma_K(\mathbf{z}) \\ &= \frac{1}{p}\|\,\mathbf{z}\,\|_q^q - \|\,\mathbf{z}\,\|_q^q - \sigma_K(\mathbf{z}) \\ &= -\left(\frac{1}{q}\|\,\mathbf{z}\,\|_q^q + \sigma_K(\mathbf{z})\right) \end{aligned} \tag{14}$$

The dual problem $\max_{\mathbf{z}} g(\mathbf{z})$, is therefore equivalent to,

$$\min_{\mathbf{z}} \frac{1}{q}\|\,\mathbf{z}\,\|_q^q + \sigma_K(\mathbf{z}). \tag{15}$$

Now if $K = K_1 \cap K_2 \ldots \cap K_k$ then we have [12],

$$\sigma_K(\mathbf{z}) = \min_{\mathbf{z}_1,\ldots,\mathbf{z}_k} \sigma_{K_1}(\mathbf{z}_1) + \ldots + \sigma_{K_k}(\mathbf{z}_k) \quad \text{subject to } \mathbf{z}_1 + \mathbf{z}_2 + \ldots + \mathbf{z}_k = \mathbf{z}\,. \tag{16}$$

Inserting (16) into (15), the dual problem becomes

$$\min_{\mathbf{z}} \frac{1}{q}\|\,\mathbf{z}\,\|_q^q + \sigma_{K_1}(\mathbf{z}_1) + \ldots + \sigma_{K_k}(\mathbf{z}_k) \quad \text{subject to } \mathbf{z}_1 + \mathbf{z}_2 + \ldots + \mathbf{z}_k = \mathbf{z}, \tag{17}$$

which is equivalent to (7). We see that if $\mathbf{z}^*$ is the minimizer of (17), the relation stated in Prop. 1 now follows from (12). Notice that, unlike the primal problem, the dual problem is unconstrained. This allows us to adapt known schemes to come up with an algorithm that converges to the minimizer. In the following, we adapt the Douglas-Rachford algorithm [11, 6] to this problem.

[1] Actually, '$\operatorname{sign}(x_i)\,|x_i|^{p-1}$' is the i^{th} entry of the subgradient [12] of $\|\,\mathbf{x}\,\|_p^p$.

3. ADAPTING THE DOUGLAS-RACHFORD ALGORITHM

Given a minimization problem of the form,

$$\min_{\mathbf{z}} f(\mathbf{z}) + g(\mathbf{z}) \tag{18}$$

the Douglas-Rachford algorithm finds the minimizer through successive application of some combination of proximal operators of f and g. For a function h, its proximal operator with parameter λ is a mapping that maps a point $\mathbf{z}$ to the minimizer of a functional defined in terms of h. More precisely, it is the operator $J_h^\lambda(\cdot)$ defined as,

$$J_h^\lambda(\mathbf{z}) = \operatorname*{argmin}_{\mathbf{u}} \frac{1}{2\lambda} \| \mathbf{z} - \mathbf{u} \|_2^2 + h(\mathbf{u}). \tag{19}$$

In terms of J_f^λ, J_g^λ, the proximal operators for f and g, the Douglas-Rachford algorithm for solving (18) is [11, 6],

Algorithm 1 The Douglas-Rachford Algorithm

repeat
  $\mathbf{z} \leftarrow J_f^\lambda\big(2\, J_g^\lambda(\mathbf{z}) - \mathbf{z}\big) + \big(\mathbf{z} - J_g^\lambda(\mathbf{z})\big)$
until convergence
$\mathbf{z} \leftarrow J_g^\lambda(\mathbf{z})$

Application of the Douglas-Rachford Algorithm to the Dual Problem

For our problem, we define f and g as,

$$f\big(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k\big) = \frac{1}{q} \left\|\mathbf{z}_1 + \mathbf{z}_2 + \ldots + \mathbf{z}_k\right\|_q^q, \tag{20a}$$

$$g\big(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k\big) = \sup_{\mathbf{x}_1 \in K_1} \langle \mathbf{x}_1, \mathbf{z}_1 \rangle + \sup_{\mathbf{x}_2 \in K_2} \langle \mathbf{x}_2, \mathbf{z}_2 \rangle + \ldots + \sup_{\mathbf{x}_k \in K_k} \langle \mathbf{x}_k, \mathbf{z}_k \rangle \tag{20b}$$

Algorithm 1 requires us to successively apply J_f^λ and J_g^λ. Let us now derive what these operators correspond to, in our case.

Computation of J_f^λ

To compute $J_f^\lambda(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k)$, we need to solve

$$\min_{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k} \frac{1}{2} \| \mathbf{z}_1 - \mathbf{u}_1 \|_2^2 + \ldots + \frac{1}{2} \| \mathbf{z}_k - \mathbf{u}_k \|_2^2 + \lambda/q \left\|\mathbf{u}_1 + \ldots + \mathbf{u}_k\right\|_q^q. \tag{21}$$

The point $J_f^\lambda(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k)$ is the unique point $\big(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\big)$ that satisfies,

$$0 \in \mathbf{u}_i^* - \mathbf{z}_i + \lambda \ \mathrm{sign}\big(\mathbf{u}_1^* + \ldots + \mathbf{u}_k^*\big) \left|\mathbf{u}_1 + \ldots + \mathbf{u}_k\right|^{q-1} \tag{22}$$

for $i = 1, \ldots, k$.[2] Summing these over i, we obtain,

$$0 \in \mathbf{u}^* - \mathbf{z} + \lambda\, k \ \mathrm{sign}(\mathbf{u}^*) | \mathbf{u}^* |^{q-1}, \tag{23}$$

where $\mathbf{u}^* = \mathbf{u}_1^* + \ldots + \mathbf{u}_k^*$ and $\mathbf{z} = \mathbf{z}_1 + \ldots + \mathbf{z}_k$. Notice that if we can solve for $\mathbf{u}^*$, then,

$$\mathbf{u}_i^* = \mathbf{z}_i - \lambda \ \mathrm{sign}(\mathbf{u}^*) | \mathbf{u}^* |^{q-1}. \tag{24}$$

Let us therefore look at (23). For the j^{th} entry of $\mathbf{u}^*$, namely u_j^*, we need to solve

$$u_j^* + (\lambda\, k) \, \mathrm{sign}(u_j^*) \, |u_j^*|^{q-1} = z_j. \tag{25}$$

To simplify notation, consider the problem of finding $u \in \mathbb{R}$ such that

$$u + d \ \mathrm{sign}(u) \, |u|^{q-1} = z. \tag{26}$$

for $d > 0$. We note that $\mathrm{sign}(u) = \mathrm{sign}(z)$. Now taking $z > 0$, again to further simplify the equality, we need to solve, for $u > 0$,

$$\underbrace{u + d\, u^{q-1}}_{s(u)} = z. \tag{27}$$

$s(u)$ is a strictly increasing function with $s(0) = 0 < z$, and $s(z) > z$. Therefore, there is a unique u that satisfies $s(u) = z$ with $u \in (0, z)$. This u can be found by an iterative algorithm.
An algorithm that solves (21) is thus,

Algorithm 2 Computation of J_f^λ

Input : $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k \in \mathbb{R}^n$
Output : $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k \in \mathbb{R}^n$ that minimizes (21)
$\mathbf{z} \leftarrow \sum_{i=1}^k \mathbf{z}_i$
$d \leftarrow \lambda\, k$
for $j = 1$ **to** n **do**
  $a \leftarrow 0$
  $b \leftarrow |\, \mathbf{z}(j)|$
  repeat
    $c \leftarrow (a + b)/2$
    if $c + d\, c^{q-1} > |\, \mathbf{z}(j)|$ **then**
      $b \leftarrow c$
    else
      $a \leftarrow c$
    end if
  until $a \approx b$
  $\mathbf{u}(j) \leftarrow \mathrm{sign}\big(\mathbf{z}(j)\big) \, (a + b)/2$
end for
for $i = 1$ **to** k **do**
  $\mathbf{u}_i \leftarrow \mathbf{z}_i - \lambda \ \mathrm{sign}(\mathbf{u}) \, | \, \mathbf{u} \, |^{q-1}$
end for

[2] Notice that, here the argument of the 'sign' function is a vector – the function is applied componentwise.

Computation of J_g^λ

To compute $J_g^\lambda(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k)$, we need to solve

$$\min_{\mathbf{u}_1,\ldots,\mathbf{u}_k} \frac{1}{2}\,\|\,\mathbf{z}_1 - \mathbf{u}_1\,\|_2^2 + \ldots + \frac{1}{2}\,\|\,\mathbf{z}_k - \mathbf{u}_k\,\|_2^2 + \lambda\,\sigma_{K_1}(\mathbf{u}_1) + \ldots + \lambda\,\sigma_{K_k}(\mathbf{u}_k). \quad (28)$$

This functional is separable with respect to $\mathbf{u}_i$'s. Therefore, we will only study the following minimization problem.

$$\mathbf{u}^* = \operatorname*{argmin}_{\mathbf{u}} \frac{1}{2}\,\|\,\mathbf{z} - \mathbf{u}\,\|_2^2 + \lambda\,\sigma_K(\mathbf{u}) \quad (29)$$

The following proposition (see [2] for a derivation of a similar result) will be used to obtain a description of the minimizer.

Proposition 2. Let K be a closed convex set. Then,

$$\operatorname*{argmin}_{\mathbf{u}} \frac{1}{2}\,\|\,\mathbf{z} - \mathbf{u}\,\|_2^2 + \sigma_K(\mathbf{u}) = \mathbf{z} - P_K(\mathbf{z}) \quad (30)$$

where $P_K(\mathbf{z})$ is the projection of $\mathbf{z}$ to K (i.e. the closest point in K to $\mathbf{z}$).

Using this proposition, and noting that $\lambda\,\sigma_K(\mathbf{u}) = \sigma_{\lambda\,K}(\mathbf{u})$, we can write

$$\mathbf{u}^* = \mathbf{z} - P_{\lambda\,K}(\mathbf{z}) \quad (31)$$

Here it is important that the projection operator be easily realizable.

Algorithm 3 Computation of J_g^λ

Input : $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k \in \mathbb{R}^n$
Output : $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k \in \mathbb{R}^n$ that minimizes (28)
for $i = 1$ **to** k **do**
  $\mathbf{u}_i \leftarrow \mathbf{z}_i - P_{\lambda\,K_i}(\mathbf{z}_i)$
end for

Remark 3. The projections in Algorithm 3 can be performed in parallel. Taking this into account leads to significant reduction in running times, especially when the number of constraint sets is high, as in the Experiment discussed below.

4. EXPERIMENTS

We consider a denoising application where side information is also available. We have at hand the noisy image, namely y shown in Fig. 1a. We also have 1-D (Radon) projections of the image along the angles $\theta = k\,\pi/18$ for $k = 0, \ldots, 17$. We denote the linear operator which computes the 1-D projection along the angle $k\,\pi/18$ as P_k and the given projection data at $k\,\pi/18$ as $\mathbf{d}_k$. Our primary variables are the wavelet coefficients of the image (we used an orthonormal wavelet transform with Daubechies filters that have 3 vanishing moments).

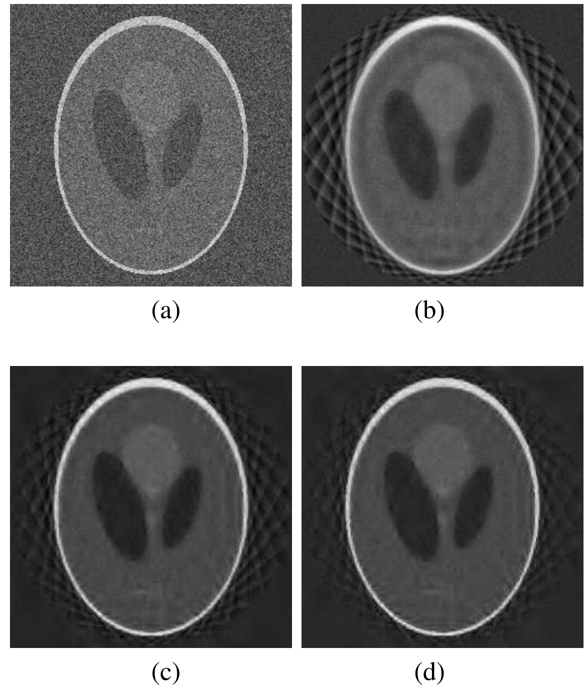

Fig. 1. Denoising with side information. The side information consists of Radon projections along the angles $k\,\pi/18$ for $k = 0, \ldots, 17$. (a) Noisy image, RMSE = 0.2. Solution for (b) $p = 2$, RMSE = 0.0967, (c) $p = 4/3$, RMSE = 0.0679, (d) $p = 8/7$, RMSE = 0.0615.

We denote the synthesis operator corresponding to the wavelet basis as W. We also define the constraint sets as,

$$K_i = \{x : P_i\,W\,x = d_i\} \quad \text{for } i = 0, \ldots, 17 \quad (32)$$

$$K_{18} = \{x : \|W\,x - y\| \leq \sigma\} \quad (33)$$

where σ is taken as the standard deviation of the noise. Based on these definitions, we consider the minimization problem, '$\min_{\mathbf{x}\in K}\,\|\,\mathbf{x}\,\|_p$', where $K = K_0 \cap \ldots \cap K_{18}$. The results are shown in Fig. 1(b,c) for $p = 2$, $p = 4/3$ and $p = 8/7$. We observe that $p \approx 1$ leads to an improved reconstruction. Note that the formulation allows us to easily take into account data

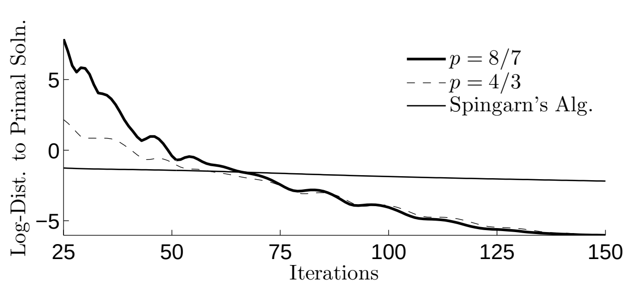

Fig. 2. Log-distance to the primal limit for the proposed algorithm and Spingarn's algorithm, derived from [13].

which cannot be expressed as an affine space.

Comparison with Parallel Proximal Algorithms

As we noted in the Introduction, there are other algorithms available for the primal problem (1). Specifically, let us consider Spingarn's method of partial inverses [13] applied to the problem at hand (see Algorithm 4 below). For another similar algorithm, see [4].

Algorithm 4 Spingarn's Algorithm for (1)

Initialize $x \leftarrow 0$, $y_i \leftarrow 0$ for $i = 0, 1, \ldots, m$.

repeat
$t_0 \leftarrow \operatorname{argmin}_z \frac{1}{2}\|x + y_0 - z\|_2^2 + \|z\|_p$
$u_0 \leftarrow x + y_0 - t_0$
for $i = 1$ to m **do**
$t_i \leftarrow P_{K_i}(x + y_i)$
$u_i \leftarrow x + y_i - t_i$
end for
$x \leftarrow \frac{1}{m+1}\sum_{i=0}^{m} t_i$
for $i = 1$ to m **do**
$y_i \leftarrow u_i - \frac{1}{m+1}\sum_{i=0}^{m} u_i$
end for
until convergence
$x^* \leftarrow x$

Algorithm 4 consists of simple substeps like the proposed algorithm. However, this algorithm treats the proximal mapping of the ℓ_p norm and the projections onto K_i's in the same manner. The updated information from the substeps are merged only at the end of each iteration. On the other hand, the proposed algorithm treats the two parts of the dual problem differently. Once the projections onto K_i's are performed, this information is fed into the proximal mapping of $\|\cdot\|_q^q$.

In order to compare the performances of the two algorithms we performed the following experiment on the phantom denoising problem described above. We ran both algorithms for 5000 iterations to obtain the 'limit' in each case. We note that there are two limit points of interest – one for the primal, one for the dual problem, related to each other through (12). Although the proposed algorithm works on the dual, we are interested in the ability of the algorithm to approach the minimizer of the primal problem in as few iterations as possible. In order to provide a fair comparison, we normalized distances by setting the norm of the corresponding limit to unity.

Fig.2 depicts the log-distance to the primal-limit for the proposed algorithm and Spingarn's algorithm. Since the proposed algorithm works on the dual-problem, we cannot claim that the distance to the primal-limit will be monotone decreasing. Indeed, the algorithm starts very far from the primal limit and the distance to the primal limit does increase in the first few iterations. However, interestingly, in 100 iterations, the distance to the primal-limit for the proposed algorithm is much lower, compared to Spingarn's method. Considering the parallel nature of Spingarn's method, the scheme presented in this paper allows, in a sense, to trade the parallel structure for faster convergence (given limited resources).

5. CONCLUSION

We considered a variation of the 'sparse reconstruction' problem. Specifically, we replaced the the ℓ_1 norm with the ℓ_p norm for $p > 1$, and studied a flexible formulation that allows multiple constraints to be placed on the reconstruction. Although minimum ℓ_p norm reconstructions, for $p > 1$, are no longer sparse (in the strict sense of the word), this formulation can also be feasible. For $p \approx 1$, the solution of the primal problem provides an approximation to the minimum ℓ_1 reconstruction. We also demonstrated that the proposed algorithm performs comparably with existing alternatives.

6. REFERENCES

[1] E. Van Den Berg and M. P. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comput.*, 31(2):890–912, 2008.

[2] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, January-March 2004.

[3] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. on Scientific Computing*, 20(1):33–61, 1998.

[4] P. L. Combettes and J.-C. Pesquet. A proximal decomposition method for solving convex variational inverse problems. *Inverse Problems*, 24(6), December 2008.

[5] R. L. Dykstra. An algorithm for restricted least squares regression. *Journal of the American Statistical Association*, 78:837–842, 1983.

[6] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(3):293–318, 1992.

[7] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Annals of Statistics*, 32(2):407–499, 2004.

[8] W. J. Fu. Penalized regressions : The bridge versus the lasso. *J. Comput. Graph. Statist.*, 7(3):397–416, 1998.

[9] N. Gaffke and R. Mathar. A cyclic projection algorithm via duality. *Metrika*, 36:29–54, 1989.

[10] S.-P. Han. A successive projection method. *Mathematical Programming*, 40:1–14, 1988.

[11] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. on Numer. Anal.*, 16(6):964–979, 1979.

[12] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1996.

[13] J. E. Spingarn. Partial inverse of a monotone operator. *Applied Math. and Optimization*, 10(1):247–265, 1983.