Chambolle-Pock and Douglas-Rachford algorithms for non convex problems

Clement NSHIMIYIMANA
African Masters in Machine Intelligence(AMMI)
AIMS-Senegal
cnshimiyimana@aimsammi.org

Prof Dirk Lorenz, Emanuele Naldi
Technische Universität Braunschweig
Germany
d.lorenz@tu-braunschweig.de,
e.naldi@tu-braunschweig.de

Abstract

We introduce two proximal operator-based algorithms, namely Chambolle-Pock and Primal-Dual Douglas-Rachfold for non convex optimization problems with the objective functions $f(x) + g(Dx)$, where f is convex and g is non convex function. We implement our methods to signal reconstruction problem with ℓ_0 norm regularization. Numerical results show that both algorithms similarly converge to critical points and provide good results.

1 Introduction

In this work, we propose two primal-dual algorithms applied to non convex problems. In fact, we explore the Chambolle-Pock(CP) algorithm and Douglas-Rachford splitting applied to primal dual optimality conditions.
Chambolle-Pock algorithm was firstly introduced in[1], it is defined as a first-order primal dual hybrid gradient(PDGH) algorithm for convex problems with saddle point structure(refer to equation(4)). Several researchers demonstrated that PDGH methods have a state of the art in image processing to solving problems like total variation image denoising and deblurring, image and signal recovery [2, 3, 4].
Douglas-Rachford Splitting(DRS) algorithm was first published in 1956 by Douglas and Rachford in [5], where they used it to solve the heat flow problems. Later on, in [6] DRS was also implemented by Lions and Mercier to find among other things the intersection between two convex sets. Not only for the case of convex sets but also has proven that DRS can be used to solve different optimization problems in signal processing and machine learning as a regression task. Recently, DRS has been used to solve complex problems with the objective function split into two convex functions(see equation 2) [7, 8, 9, 10]. However, there are also various problems with the form (2) that are in the non convex settings, where either one of them is non convex or both of them are non convex. Hence, there are many works done on applying DRS in solving those problems. See for example [11, 12, 13, 14, 15]. Chambolle and Pock, O'Connor and Vandenberghe in [16, 17, 18] derived the relationship between DRS and primal-dual hybrid gradient(PDHG) algorithm in convex setting. They have shown that by relaxing some of the assumptions in PDHG, one can end up with DRS. On the other hand they demonstrated that PDHG can be derived from DRS.
In this work, we compare Chambolle pock algorithm and primal dual DRS in non convex setting. In fact, we modify Chambolle-Pock and the primal dual Douglas-Rachfold algorithms in such a way that they can suit the non convex setting. We consider a problem with the objective function which is composed of two functions, one is convex while another one is a non convex composite function(see equation 2). Particularly, we apply the two algorithms to sparse signal reconstruction problem with

ℓ_0-norm regularization. Here is the general form of the problems that are solved by DRS.

$$\min_x f(x) + g(x) \tag{1}$$

Whereas the primal-dual algorithms solve problems with the form

$$\min_x f(x) + g(Dx) \tag{2}$$

where f, g are proper convex functions with the structure that their proximal operators can be computed easily and D denotes a bounded linear operator. This problem can be reformulated as a constrained optimization problem in the following way

$$\begin{cases} \min_{x,y} f(x) + g(y) \\ Dx - y = 0 \end{cases} \tag{3}$$

On other hands the form 2 can also be written as saddle problem as

$$\max_y \min_x L(x, y) : f(x) + \langle y, Dx \rangle - g^*(y) \tag{4}$$

where g^* stands for the conjugate of g. Hence the optimality conditions are

$$0 \in \begin{pmatrix} \partial f(x^*) \\ \partial g^*(y^*) \end{pmatrix} + \begin{pmatrix} 0 & D^* \\ -D & 0 \end{pmatrix} \begin{pmatrix} x^* \\ y^* \end{pmatrix} \tag{5}$$

where (x^*, y^*) is a saddle point. This can be written as the sum of non linear operators

$$0 \in A(x^*, y^*) + B(x^*, y^*). \tag{6}$$

In our case, we consider problem(2) where f is a proper convex function and g a semi-lower continuous, non convex function. We assume that the proximal operators of those functions can be computed easily. Hence we have

$$0 \in \partial f(x^*) + D^* \partial g(Dx^*). \tag{7}$$

where x^* is a critical point. We can see that if we consider D as identity matrix, then the problem(2) can be reduced to problem(1). So, one can conclude that primal-dual algorithms generalize Douglas Rachford Splitting(DRS) algorithm.

2 Preliminaries

Most of the mathematical definitions we used are from the book of optimization[19].

Definition 1 (Convex set)**.** *A set C is convex if $\forall x, y \in C$ and $\lambda \in [0, 1]$ we have $\lambda x + (1 - \lambda)y \in C$.*

Definition 2 (Convex function)**.** *A function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is convex if $\forall x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$, we have $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$.*

Theorem 1. Suppose $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is twice differentiable over an open domain. Then the following are equivalent.

1. f is convex.
2. $f(y) \geq f(x) + \nabla f(x)^T (y - x), \forall x, y \in \mathbb{R}^n$.
3. $\nabla^2 f(x) \geq 0,\ \forall x \in \mathbb{R}^n$.

Definition 3. *A function $f : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$, is proper if $\forall x \in dom f, f(x) > -\infty$, and $\exists\ x_0, f(x_0) < \infty$. While f is said to be lower semicontinuous(lsc) at x_0 if $\liminf_{x \to x_0} f(x) \geq f(x_0)$.*

Definition 4. *Given a proper function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, Fréchet subdifferential of f at x is defined by*

$$\hat{\partial} f(x) = \left\{ v \in \mathbb{R}^n | \liminf_{\substack{z \to x \\ z \neq x}} \frac{f(z) - f(x) - \langle v, z - x \rangle}{\|z - x\|} \geq 0 \right\}.$$

If $x \notin dom f$, then $\hat{\partial} f(x) = \emptyset$. The sub-differential of f at x is defined as

$$\partial f(x) = \left\{ t \in \mathbb{R}^n | \exists x_k \to x,\ f(x_k) \to f(x),\ v_k \in \hat{\partial} f(x_k) \to t \right\}.$$

Recall that a necessary condition for $x \in \mathbb{R}^n$ to be a minimizer of f is that x is a critical point of f, which means that $0 \in \partial f(x)$. Moreover, if f is convex, this condition is sufficient.

Remark 1. *Definition4 implies that ∂f is closed. More precisely , we have the following property: Let $(x_k, y_k)_{k \in \mathbb{N}}$ be a sequence of Graph $\partial f(x) = \left\{ (x,t) \in \mathbb{R}^n \times \mathbb{R}^n | t \in \partial f(x) \right\}$. If (x_k, t_k) converges to (x,t) and $(f(x_k))$ converges to $f(x)$, then $(x,t) \in Graph \partial f$.*

Definition 5. *Let Ψ denote a vector space with norm $\|.\|_\Psi$. The proximal mapping associated with a function $f : \Psi \to \bar{\mathbb{R}}$, for every $v \in \Psi, t > 0$ is an operator that is defined as*

$$prox_{tf}(v) = \arg\min_{x \in \Psi} \{ \frac{1}{2t} \|x - v\|^2 + f(x) \}.$$

Note that if f is convex, then the proximal operator is a single valued operator. When f is non convex its proximal operator is a set valued operator. Moreover, it is possible to show that when f is convex, the proximal operator coincide with the resolvent of ∂f indeed:

$$prox_{tf}(v) = (I + tA)^{-1}(v)$$

where $A = \partial f$ is a maximal monotone operator.

Definition 6 (Moreau Identity)**.** *Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ be a convex function, then its Moreau identity is given by $prox_{\lambda f}(x) + \lambda prox_{\lambda^{-1} f^*}(\frac{x}{\lambda}) = x$. Where $f^*$is the conjugate of f.*

Definition 7. *(ℓ_0-norm) Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(x) = \lambda \|x\|_0$, where $\lambda > 0$.*

$$\|x\|_0 = \#\{i : x_i \neq 0\}$$

which means that ℓ_0- norm is the number of non zero elements of the set. Note that f is a non convex function.
The proximal operator of the ℓ_0-norm is defined as follows

$$prox_f(x_1, x_2, ..., x_n) = Tx_1, Tx_2, ..., Tx_n$$

where

$$Tx = \begin{cases} 0, & |x| < \sqrt{2\lambda} \\ x, & otherwise \end{cases} \tag{8}$$

The proximal operator of ℓ_0-norm is well known as hard-threshold operator.

3 Chambolle Pock algorithm

In this section, we are dealing with the Chambolle-Pock algorithm which is a first order primal-dual algorithm, and it is designed to solve the classical saddle points problems. Normally, Chambolle-Pock is employed to find the zero of problems of the form(2) where f, g are both proper convex functions. In our study we modify it to be applied on non convex problems.
The algorithm1 and algorithm2 represent the iterations of Chambolle-Pock used to solve convex problems, where $\tau > 0$ and $\sigma > 0$ are respectively the primal and dual step sizes, and ρ stands for the relaxation parameter.

As algorithm1 uses the proximal operator of the dual function g^*, it cannot be appropriate to non convex problems. Hence, using the Moreau identity(see definition(6)) we reformulate algorithm1 into algorithm3 that it is suitable for non convex setting.
For the rest of our work, we use the algorithm3 to minimise the problem(2) but considering f convex and g non convex.

Proposition 1. Let assume that x^k, y^k are sequences that are given by the iterates of algorithm3 and they converge to some points (x, y), then x is critical point of problem2.

Algorithm 1 Chambolle-Pock(CP)

Initialize:

Input: Primal and dual points (x, y), stepsizes τ, σ such that $\tau\sigma\|D\|^2 \leq 1$.

for k in 1,2,... maxit **do**

$$\begin{cases} x^{k+1} = prox_{\tau f}(x^k - \tau D^* y^k) \\ y^{k+1} = prox_{\sigma g^*}(y^k + \sigma D(2x^{k+1} - x^k)) \end{cases}$$

end for

Algorithm 2 Chambolle-Pock(CP) with relaxation

Initialize:

Input: Primal and dual points (x, y), stepsizes τ, σ such that $\tau\sigma\|D\|^2 \leq 1$ and $\rho \in (0, 2)$.

for k in 1,2,... maxit **do**

$$\begin{cases} x^{k+1} = x^k + \rho(prox_{\tau f}(x^k - \tau D^* y^k) - x^k) \\ y^{k+1} = y^k + \rho(prox_{\sigma g^*}(y^k + \sigma D(2x^{k+1} - x^k)) - y^k) \end{cases}$$

end for

Proof. Let suppose that the sequence (x^k, y^k) converges to some points (x, y). From the update of x in algorithm3, we have

$$\begin{aligned} x^{k+1} &= prox_{\tau f}(x^k - \tau D^* y^k) \\ &= \arg\min\{f(x^{k+1}) + \frac{1}{2\tau}||x^{k+1} - x^k + \tau D^* y^k||^2\} \\ \Rightarrow 0 &\in \tau\partial f(x^{k+1}) + x^{k+1} - x^k + \tau D^* y^k \\ x^k - x^{k+1} &\in \tau\partial f(x^{k+1}) + \tau D^* y^k \\ \frac{x^k - x^{k+1}}{\tau} - D^* y^k &\in \partial f(x^{k+1}). \end{aligned}$$

As $k \to \infty$ we have $x^k \to x$ and $y^k \to y$, so that

$$\frac{x^k - x^{k+1}}{\tau} - D^* y^k \to -D^* y.$$

By definition of proximal operator we have

$$f(x^{k+1}) + \frac{||x^{k+1} - x^k + \tau D^* y^k||^2}{2\tau} \leq f(x) + \frac{||x - x^k + \tau D^* y^k||^2}{2\tau} \tag{9}$$

so that $\limsup_k f(x^{k+1}) \leq f(x)$ and by lower semi-continuity of f we have

$$\lim_{k\to\infty} f(x^{k+1}) = f(x)$$

Algorithm 3 Chambolle Pock(CP) for non convex Problems

Initialize:

Input:Primal and dual points (x, y), stepsizes τ, σ such that $\tau\sigma\|D\|^2 \leq 1$.

for k in 1,2,... maxit **do**

$$\begin{cases} x^{k+1} = prox_{\tau f}(x^k - \tau D^* y^k) \\ y^{k+1} \in y^k + \sigma D(2x^{k+1} - x^k) - \sigma prox_{\sigma^{-1} g}\left(\dfrac{y^k + D(2x^{k+1} - x^k)}{\sigma}\right) \end{cases}$$

end for

, then we obtain by Remark1 that

$$-D^* y \in \partial f(x) \tag{10}$$

Similarly, by considering the iterate y, let take for every k an element

$$z_k^* = prox_{\sigma^{-1}g}\left(\frac{y^k}{\sigma} + D(2x^{k+1} - x^k)\right) \tag{11}$$

such that

$$y^{k+1} - y^k = \sigma D(2x^{k+1} - x^k) - \sigma z \tag{12}$$

$$z_k^* = \frac{y^k - y^{k+1}}{\sigma} + D(2x^{k+1} - x^k) \tag{13}$$

By necessary condition at the minimizer we have

$$z_k^* = prox_{\sigma^{-1}g}(\frac{y^k}{\sigma} + D(2x^{k+1} - x^k)) \tag{14}$$

$$z_k^* = \arg\min_z\{g(z) + \frac{\sigma}{2}\|z - \frac{y^k}{\sigma} - D(2x^{k+1} - x^k)\|^2\} \tag{15}$$

$$0 \in \partial g(z_k^*) + \sigma(z_k^* - \frac{y^k}{\sigma} - D(2x^{k+1} - x^k)) \tag{16}$$

By substituting the equation (13) in (16), we obtain

$$0 \in \partial g\left(\frac{y^k - y^{k+1}}{\sigma} + D(2x^{k+1} - x^k)\right) + \sigma(\frac{y^k - y^{k+1}}{\sigma} + D(2x^{k+1} - x^k)) -$$
$$\sigma(\frac{y^k}{\sigma} - D(2x^{k+1} - x^k)))$$
$$\Rightarrow 0 \in \partial g\left(\frac{y^k - y^{k+1}}{\sigma} + D(2x^{k+1} - x^k)\right) - y^{k+1}$$
$$\Rightarrow y^{k+1} \in \partial g\left(\frac{y^k - y^{k+1}}{\sigma} + D(2x^{k+1} - x^k)\right).$$

For $k \to \infty$, we have $x^k \to x$ and $y^k \to y$, thus $\frac{y^k - y^{k+1}}{\sigma} + D(2x^{k+1} - x^k) \to Dx$.

By definition of proximal map, we observe that

$$g\left(\frac{y^k - y^{k+1}}{\sigma} + D(2x^{k+1} - x^k)\right) + \frac{\sigma}{2}||\frac{y^{k+1}}{\sigma}||^2 \leq g(Dx) + \frac{\sigma}{2}||Dx - \frac{y^k}{\sigma} - D(2x^{k+1} - x^k)||^2$$

So that $\limsup_{k\to\infty} g\left(\frac{y^k - y^{k+1}}{\sigma} + D(2x^{k+1} - x^k)\right) \leq g(Dx)$ and by lower semicontinuity of g we have $\lim_{k\to\infty} g\left(\frac{y^k - y^{k+1}}{\sigma} + D(2x^{k+1} - x^k)\right) = g(Dx)$.
This implies that $y \in \partial g(Dx)$. From equation(10) we get

$$0 \in \partial f(x) + D^* y \Rightarrow 0 \in \partial f(x) + D^* \partial g(Dx).$$

Therefore, x is a critical point of problem(2). □

4 Primal dual DRS algorithm

In this section, we introduce Douglas-Rashford algorithm with primal-dual formulation applied to non convex setting. We get inspired by O'Connor and Vandenberghe in their works [17, 18], where they derived the algorithm4 from the relaxed form Primal-Dual Hybrid Gradient(PDGH) algorithm. They showed that by relaxing both iterates x and y,then apply the change of variables, they ended up with an algorithm4. So we take the primal-dual DRS formulation produced by O'Connor and Vandenberghe and apply the Moreau identity to get rid of the proximal operator of the dual function g^* which is not applicable to non convex problems.
Algorithm4 represents a primal dual Douglas Rachfold algorithm which is used to solve saddle point problems with the form(2) that are mainly convex. To make a use of it in non convex setting, we removed the dual formulation by applying the Moreau Identity(see Algorithm5).

Algorithm 4 Primal-Dual DR

Initialize:

Input: Initial values of primal and dual points $z^0 = (p^0, q^0)$, $\tau, \sigma, \rho > 0$ such that $\rho \in (0, 2)$.

for k in 1,2,... maxit **do**

$$\begin{cases} x^{k+1} = prox_{\tau f}(p^k) \\ y^{k+1} = prox_{\sigma g^*}(q^k) \\ \begin{pmatrix} u^{k+1} \\ v^{k+1} \end{pmatrix} = \begin{pmatrix} I & \tau D^* \\ -\sigma D & I \end{pmatrix}^{-1} \begin{pmatrix} 2x^{k+1} - p^k \\ 2y^{k+1} - q^k \end{pmatrix} \\ p^{k+1} = p^k + \rho(u^{k+1} - x^{k+1}) \\ q^{k+1} = q^k + \rho(v^{k+1} - y^{k+1}) \end{cases}$$

end for

Algorithm 5 Primal-Dual DR for non convex problem

Initialize:

Input:Initial values of primal and dual points $z^0 = (p^0, q^0)$, $\tau, \sigma, \rho > 0$ such that $\rho \in (0, 2)$.

for k in 1,2,... maxit **do**

$$\begin{cases} x^{k+1} = prox_{\tau f}(p^k) \\ y^{k+1} \in q^k - \sigma prox_{\sigma^{-1} g}(\sigma^{-1} q^k) \\ \begin{pmatrix} u^{k+1} \\ v^{k+1} \end{pmatrix} = \begin{pmatrix} I & \tau D^* \\ -\sigma D & I \end{pmatrix}^{-1} \begin{pmatrix} 2x^{k+1} - p^k \\ 2y^{k+1} - q^k \end{pmatrix} \\ p^{k+1} = p^k + \rho(u^{k+1} - x^{k+1}) \\ q^{k+1} = q^k + \rho(v^{k+1} - y^{k+1}) \end{cases}$$

end for

Proposition 2. If we suppose that the sequence $(x^k, y^k, u^k, v^k, p^k, q^k)$ generated by the algorithm5 is convergent to a point (x, y, u, v, p, q), then x is a critical point for $f + g \circ D$.

Proof. Suppose $(x^k, y^k, u^k, v^k, p^k, q^k) \to (x, y, u, v, p, q)$, then we notice that $u = x$ and $v = y$. From the iterate of x in algorithm4, we have

$$\frac{p^k - x^{k+1}}{\tau} \in \partial f(x^{k+1})$$

and note that for $k \to \infty$, $\frac{p^k - x^{k+1}}{\tau} \to \frac{p-x}{\tau}$.
By definition of proximal operator we have

$$f(x^{k+1}) + \frac{1}{2\tau} ||x^{k+1} - p^k||^2 \leq f(x) + \|x - p^k\|$$

so that $\limsup_k f(x^{k+1}) \leq f(x)$ and by lower semicontinuity of f, $\lim_k f(x^{k+1}) = f(x)$.
Applying Remark1 we obtain $\frac{p-x}{\tau} \in \partial f(x)$.
Similarly, from the update of y we get

$$\frac{q^k - y^{k+1}}{\sigma} \in prox_{\sigma^{-1} g}(\sigma^{-1} q^k). \tag{17}$$

This implies that

$$0 \in \partial g\left(\frac{q^k - y^{k+1}}{\sigma}\right) + \sigma\left(\frac{q^k - y^{k+1} - q^k}{\sigma}\right) \tag{18}$$

$$0 \in \partial g\left(\frac{q^k - y^{k+1}}{\sigma}\right) - \sigma\left(\frac{y^{k+1}}{\sigma}\right) \tag{19}$$

$$y^{k+1} \in \partial g\left(\frac{q^k - y^{k+1}}{\sigma}\right). \tag{20}$$

While $k \to \infty$, $\frac{q^k - y^{k+1}}{\sigma} \to \frac{q-y}{\sigma}$, by definition of proximal map we also have

$$g\left(\frac{q^k - y^{k+1}}{\sigma}\right) + \frac{\sigma}{2}\|\frac{y^{k+1}}{\sigma}\|^2 \le g(\frac{q-y}{\sigma}) + \frac{\sigma}{2}\|\frac{q-y}{\sigma} - \frac{q^k}{\sigma}\|^2 \tag{21}$$

So that $\limsup_{k\to\infty} g\left(\frac{q^k - y^{k+1}}{\tau}\right) \le g\left(\frac{q-y}{\tau}\right)$, by lower semi-continuity of g we obtain $y \in \partial g\left(\frac{q-y}{\tau}\right)$.

Now, the update of u^{k+1} and v^{k+1} gives $\begin{pmatrix} u^{k+1} + \tau D^* v^{k+1} \\ -\sigma D u^{k+1} + v^{k+1} \end{pmatrix} = \begin{pmatrix} 2x^{k+1} - p^k \\ 2y^{k+1} - q^k \end{pmatrix}$ so that by continuity of D and the fact that $u = x$ and $v = y$, we have

$$-D^* y = \frac{p-x}{\tau}, Dx = \frac{q-y}{\tau}$$

So that

$$0 \in \partial f(x) + D^* \partial g(Dx).$$

$\square$

5 Numerical experiments

In this section, we describe the experiments we did on both Chambolle-Pock and Primal-Dual Douglas Rachfold algorithms to solve a sparse signal reconstruction problem.

Signal reconstruction

In signal processing, signal reconstruction is defined as a technique of recovering the original signal. A signal $x \in \mathbb{R}^n$ is easily reconstructed from its measurement $y = Ax$ if it is sparse. In fact, a signal is said to be k-sparse if there are its $k << n$ non zero components. So ℓ_0 norm is considered as k-sparsity, and hence the sparse signal reconstruction problem is taken as ℓ_0 norm minimization problem[20, 21].

Signal reconstruction is applied in many different areas such as magnetic resonance imagery, radio communication, dictionary learning, radar signal recovery, seismic and genetics(Gene expression analysis)[22]. In our case, we considered the sparse signal reconstruction with ℓ_0 pseudo norm regularization of the forward finite differences. We used ℓ_0 as it is a non convex function and also induces more sparsity.

Problem formulation:

$$\min_x \frac{1}{2}\|Ax - b\|^2 + \lambda\|Dx\|_0 \tag{22}$$

Where $x \in \mathbb{R}^n$ represents the sparse signal. $A_{m\times n}, m \le n$ stands for a recovery matrix(measurement matrix) and it is generated randomly, It can also be generated differently based on the tasks or the problem to be solved, for example in compressed sensing it is generated as the coefficient of Fourier transform, while for total variation denoising is just identity and for deblurring, it is considered as a blurring matrix. b is given by the product of matrix A and the original signal. Refereed to [23], D is an $(n \times n)$ sparse matrix which is defined as

$$D = \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}$$

$\lambda > 0$ stands for a regularization parameter.

All experiments were done in python on computer with CPU, we used 10000 iterations. We used two different stepsizes namely $\tau, \sigma > 0$ which satisfy $\tau\sigma\|D\|^2 \leq 1$ [24]. We also used the relaxation parameter $\rho \in (0, 2)$ that helps to speed up the convergence. We utilised the relative error: $\frac{\|x_{opt} - x\|}{\|x\|}$ and the correlation to compare the original and recovered signals.

Based on the problem formulation(2) we assumed that $f(x) = \frac{1}{2}\|Ax - b\|^2$ and $g(Dx) = \lambda\|Dx\|_0$. As the primal dual DRS algorithm is based on proximal operators of f and g, we find them as follows

$$prox_{\tau f}(p^k) = (1 + \tau A^T A)^{-1}(\tau A^T b + p^k)$$

$$prox_{\sigma^{-1} g}(\frac{q}{\sigma}) = (T(\frac{q_1}{\sigma}), ..., T(\frac{q_n}{\sigma}))$$

where

$$T(\frac{q_i^k}{\sigma}) = \begin{cases} 0, & \frac{q_i^k}{\sigma} < \sqrt{\frac{2\lambda}{\sigma}} \\ \frac{q_i^k}{\sigma}, & \frac{q_i^k}{\sigma} > \sqrt{\frac{2\lambda}{\sigma}} \end{cases}$$

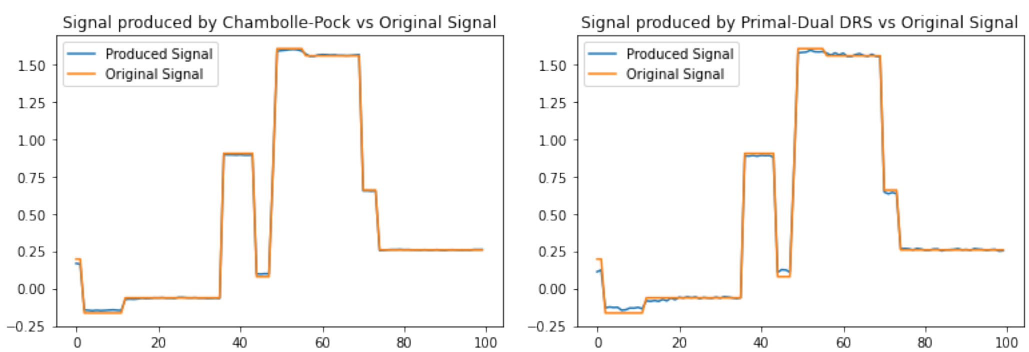

Figure 1: Original signal and recovered signals from PD-DRS, and CP algorithms by setting $\rho_k = \frac{1}{log(k+2)}, \tau = 0.1, \sigma = \frac{0.9}{\tau\|D\|^2}, \lambda = 1$.

6 Results and discussion

We realised that choosing $\rho \in (0, 2)$ and τ, σ which satisfy $\tau\sigma\|D\|^2 \leq 1$ leads to a better convergence. In addition, good results are obtained due to the choice of an appropriate(not very small) regularization parameter(λ) even if it causes the value of the objective function to be large. Particularly, we used the relaxation parameter(ρ_k) which is updated at each iteration, because we saw that using a constant value leads to high oscillation around the minimum. We have chosen moderate values of τ and λ because they provide good results and better convergence.

Figure1 indicates that approximately 99% of the original signal was recovered by using both Chambolle-Pock and Primal-Dual Douglas-Rachfold algorithms. We can also see that those algorithms provide quite similar results. While the figure2 shows how the relative error decreases over the iterations. We can say that the two algorithms behave a in similar way even if Chambolle-Pock has a small minimum error compared to that of Primal-Dual Douglas-Rachfold. On the right side of the figure, we see how the correlation between the original signal and a recovered signal gets

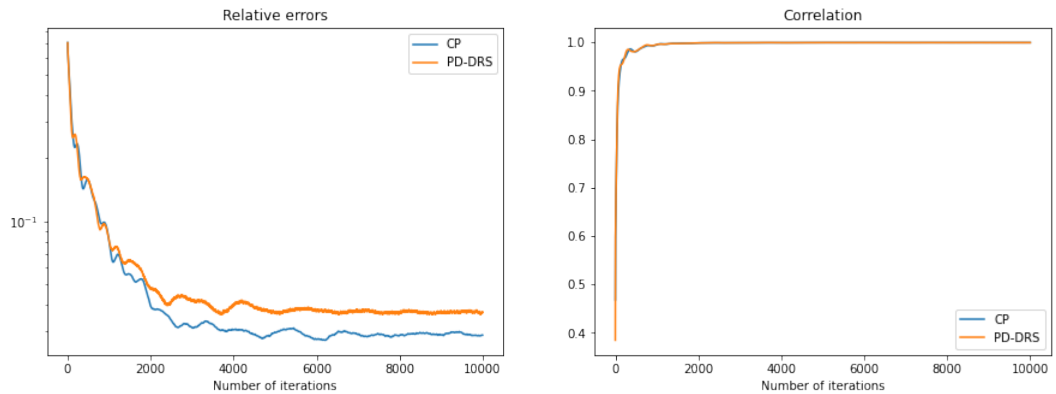

Figure 2: Relative errors and correlation between the original signal and the recovered signal by Chambolle-Pock and Primal-Dual Douglas Rachfold algorithms. When $\rho_k = \frac{1}{log(k+2)}, \tau = 0.1, \sigma = \frac{0.9}{\tau\|D\|^2}, \lambda = 1$.

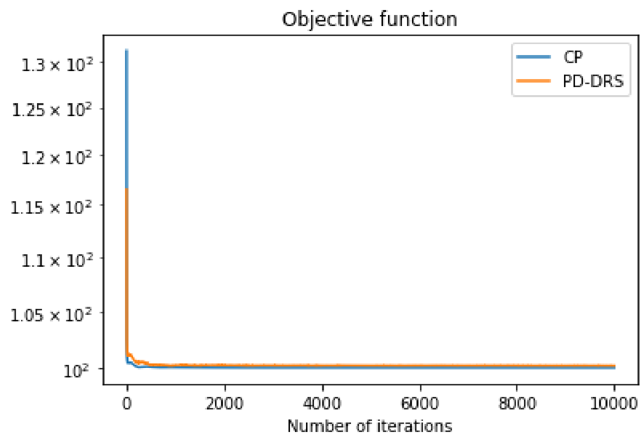

Figure 3: Objective function when $\rho_k = \frac{1}{log(k+2)}, \tau = 0.1, \sigma = \frac{0.9}{\tau\|D\|^2}, \lambda = 1$.

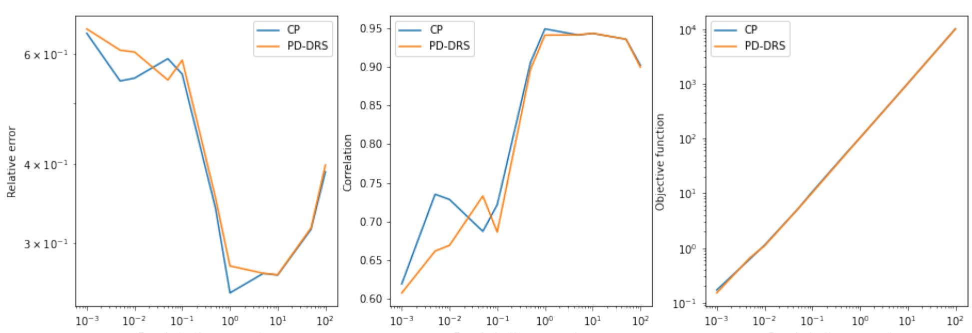

Figure 4: Effect of stepsizes(τ, σ) and regularization parameter(λ) on relative errors, correlation and objective function. We considered $\rho_k = \frac{1}{log(k+2)}, \tau = 0.001, \sigma = \frac{0.9}{\tau\|D\|^2}$.

stronger at each iteration. As it is shown in figure3, the objective function decreases in the same way of Chambolle-Pock and Primal-Douglas Rachfold algorithms.
Figures 4 - 7 represent the effect of different values of stepsize τ and regularization parameter λ. In

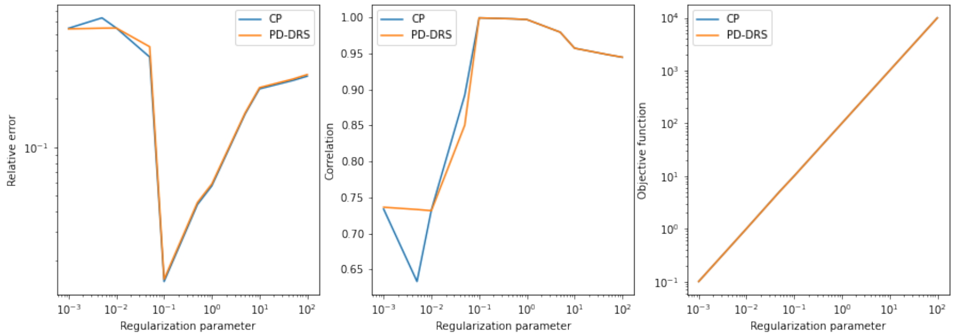

Figure 5: Effect of stepsizes(τ, σ) and regularization parameter(λ) on relative errors, correlation and objective function. We considered $\rho_k = \frac{1}{log(k+2)}, \tau = 0.01, \sigma = \frac{0.9}{\tau\|D\|^2}$.

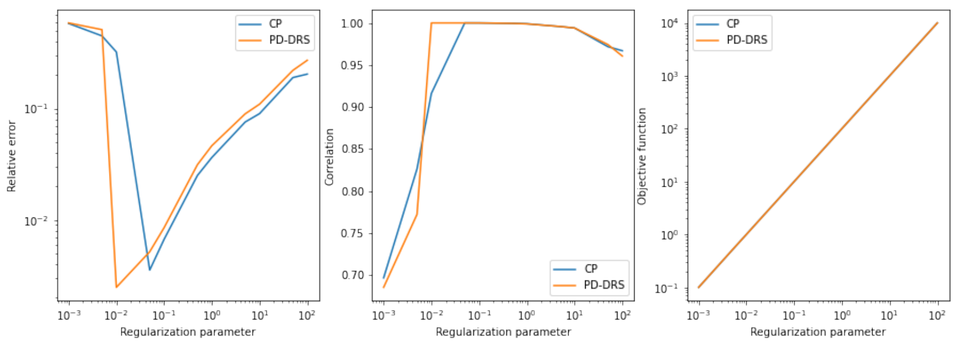

Figure 6: Effect of stepsizes(τ, σ) and regularization parameter(λ) on relative errors, correlation and objective function. We considered $\rho_k = \frac{1}{log(k+2)}, \tau = 0.1, \sigma = \frac{0.9}{\tau\|D\|^2}$.

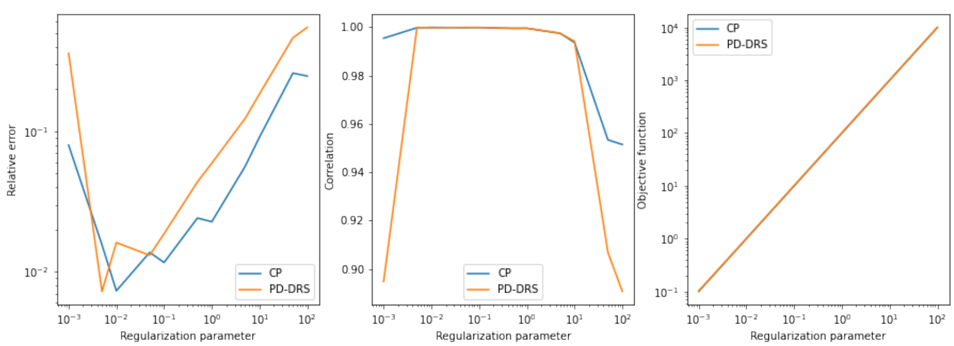

Figure 7: Effect of stepsizes(τ, σ) and regularization parameter(λ) on relative errors, correlation and objective function. We considered $\rho_k = \frac{1}{log(k+2)}, \tau = 1, \sigma = \frac{0.9}{\tau\|D\|^2}$.

each figure, there are three graphs(one left, one in the middle, and one right). One left represents the effects of τ, λ on relative error. The graph in middle is for correlation and the one on the right side shows the change of objective function according to different values of those parameters. In summary, those figures indicate that large values of stepsize and regularization parameter i.e $\tau = 0.1, \lambda \in [0.01, 1]$ imply a very small error and a very strong correlation. However, a very large regularization parameter leads to a very high value of the objective function.

7 Conclusion

We investigated the use of Chambolle-Pock and Primal-Dual-Douglas Rachfold algorithms to non convex problems of the form2 where f is convex function and g is a composite, lower semicontinous and non convex function. Particularly, we tested the performance of those algorithms on a sparse signal reconstruction, where we took a glance at ℓ_0 norm minimization. We found that by choosing the appropriate parameters such as step size, regularization, and relaxation Chambolle-Pock and Primal-Dual Douglas-Rachfold algorithms produce good results and behave in similarly.
In this work, we mainly focused on the implementation of Chambolle-Pock and Primal-Dual Douglas-Rachfold algorithms to solve a signal reconstruction problem with ℓ_0 regularization. So in the future, one can extend this work to solving other machine learning or signal processing problems and may be considering other regularizations such as $\ell_{2,p}, p \in \{0, 1\}$.

References

[1] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40(1):120–145, 2011.

[2] Ernie Esser, Xiaoqun Zhang, and Tony Chan. A general framework for a class of first order primal-dual algorithms for tv minimization. *Ucla Cam Report*, 9:67, 2009.

[3] Tuomo Valkonen. A primal–dual hybrid gradient method for nonlinear operators with applications to mri. *Inverse Problems*, 30(5):055012, 2014.

[4] Mingqiang Zhu and Tony Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. *UCLA Cam Report*, 34:8–34, 2008.

[5] Jim Douglas and Henry H Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American mathematical Society*, 82(2):421–439, 1956.

[6] Pierre-Louis Lions and Bertrand Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[7] Patrick L Combettes and Jean-Christophe Pesquet. A douglas–rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):564–574, 2007.

[8] Silvia Gandy, Benjamin Recht, and Isao Yamada. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse problems*, 27(2):025010, 2011.

[9] Bingsheng He and Xiaoming Yuan. On the o(1/n) convergence rate of the douglas–rachford alternating direction method. *SIAM Journal on Numerical Analysis*, 50(2):700–709, 2012.

[10] Panagiotis Patrinos, Lorenzo Stella, and Alberto Bemporad. Douglas-rachford splitting: Complexity estimates and accelerated variants. In *53rd IEEE Conference on Decision and Control*, pages 4234–4239. IEEE, 2014.

[11] Francisco J Aragón Artacho and Jonathan M Borwein. Global convergence of a non-convex douglas–rachford iteration. *Journal of Global Optimization*, 57(3):753–769, 2013.

[12] Hédy Attouch, Jérôme Bolte, Patrick Redont, and Antoine Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-łojasiewicz inequality. *Mathematics of operations research*, 35(2):438–457, 2010.

[13] Fengmiao Bian and Xiaoqun Zhang. A three-operator splitting algorithm for nonconvex sparsity regularization. *SIAM Journal on Scientific Computing*, 43(4):A2809–A2839, 2021.

[14] Guoyin Li and Ting Kei Pong. Douglas–rachford splitting for nonconvex optimization with application to nonconvex feasibility problems. *Mathematical programming*, 159(1):371–401, 2016.

[15] Andreas Themelis and Panagiotis Patrinos. Douglas–rachford splitting and admm for nonconvex optimization: Tight convergence results. *SIAM Journal on Optimization*, 30(1):149–181, 2020.

[16] Antonin Chambolle and Thomas Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319, 2016.

[17] Daniel O'Connor and Lieven Vandenberghe. Primal-dual decomposition by operator splitting and applications to image deblurring. *SIAM Journal on Imaging Sciences*, 7(3):1724–1754, 2014.

[18] Daniel O'Connor and Lieven Vandenberghe. On the equivalence of the primal-dual hybrid gradient method and douglas–rachford splitting. *Mathematical Programming*, 179(1):85–108, 2020.

[19] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[20] Milan Resetar, Gojko Ratkovic, and Svetlana Zecevic. Comparison of some commonly used algorithms for sparse signal reconstruction. *arXiv preprint arXiv:1902.07309*, 2019.

[21] Ljubiša Stanković, Ervin Sejdić, Srdjan Stanković, Miloš Daković, and Irena Orović. A tutorial on sparse signal reconstruction and its applications in signal processing. *Circuits, Systems, and Signal Processing*, 38(3):1206–1263, 2019.

[22] Prateek Jain and Purushottam Kar. Non-convex optimization for machine learning. *arXiv preprint arXiv:1712.07897*, 2017.

[23] Raja Giryes, Michael Elad, and Alfred M Bruckstein. Sparsity based methods for overparameterized variational problems. *SIAM journal on imaging sciences*, 8(3):2133–2159, 2015.

[24] Kristian Bredies, Enis Chenchene, Dirk A Lorenz, and Emanuele Naldi. Degenerate preconditioned proximal point algorithms. *arXiv preprint arXiv:2109.11481*, 2021.