ON THE STRUCTURE OF SOME PHASE RETRIEVAL ALGORITHMS

Heinz H. Bauschke,[1] *Patrick L. Combettes,*[2] *and D. Russell Luke*[3]

[1]Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario N1G 2W1, Canada
[2]Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie – Paris 6, 75005 Paris, France
[3]Institut für Numerische und Angewandte Mathematik, Universität Göttingen, 37083 Göttingen, Germany

ABSTRACT

The state of the art for solving the phase retrieval problem in two dimensions relies heavily on the algorithms proposed by Gerchberg, Saxton, and Fienup. Despite the widespread use of these algorithms, current mathematical theory cannot explain their remarkable success. It is already known that the Gerchberg-Saxton algorithm is a nonconvex version of method of alternating projections. In this paper, we show that two other prominent phase retrieval methods also have well known counterparts in the world of convex optimization algorithms: Fienup's basic input-output algorithm corresponds to Dykstra's algorithm, and Fienup's hybrid input-output algorithm can be viewed as an instance of the Douglas-Rachford algorithm. This work provides a theoretical framework to better understand and, potentially, improve existing phase recovery algorithms.

1. INTRODUCTION

The *phase retrieval problem* consists of estimating the phase of a complex-valued function from measurements of its modulus and additional *a priori* information. It is of fundamental importance in numerous areas of applied physics and engineering, and it has been studied for over forty years.

As in many inverse problems, a common formulation of the phase retrieval problem is to seek as a solution any function that is consistent with the measurements as well as with *a priori* constraints. The original Gerchberg-Saxton algorithm [8] and its descendent the error reduction algorithm [7] were the first widely used numerical scheme to solve this type of problem. While its intrinsic mechanism is clear physically — it consists of alternating back-substitutions of known information in the spatial and Fourier domains — it was not initially understood mathematically. In particular, failure of convergence and stagnation of the iterates away from solution points were observed from the outset but lacked a sound mathematical explanation.

In the early 1980s, with the work of Youla [14] and others, the application of Bregman's method of successive projections [3] to the recovery a signal described by convex constraints generated considerable interest in the signal recovery community [4, 6, 13]. Attempts were then made to extend this successful framework to a nonconvex setting in order to formalize a wider range of signal recovery schemes [4]. In [10], the error reduction algorithm was revealed to be a method of alternating projections between a vector subspace (support constraint) and a nonconvex set (Fourier modulus constraint). This study gave insightful geometrical interpretations of the stagnation problem as well as of other aspects of the error reduction iterative procedure.

In his seminal 1982 paper [7], Fienup introduced a broad framework for iterative algorithms. Three main classes of algorithms were presented: error reduction, Basic Input-Output (BIO), and Hybrid Input-Output (HIO). These algorithms still constitute the state of the art in phase retrieval.

While well-known in the field, the BIO and HIO algorithms lack a proper mathematical framework. The aim of this paper is to show that, just like the error reduction algorithm, the BIO and HIO algorithms also have powerful counterparts in the world of convex projection methods.

The paper is organized as follows. In Section 2, the phase retrieval problem is posed as a feasibility problem. Section 3 supplies the necessary review of nonlinear analysis. In Section 4, the classical phase retrieval methods are presented and new connections with convex optimization algorithms are established. Concluding remarks are formulated in Section 5.

2. PHASE RETRIEVAL AND FEASIBILITY

In its general form, the signal recovery problem is to estimate the original form of a signal x in a functional space $\mathcal{L}$ from the measurements of physically related signals and *a priori* information [6, 13]. In phase retrieval problems, the measurements consist of the modulus m of the Fourier transform $\widehat{x}$ of x. In other words, the imaging model is described by the relationship

$$|\widehat{x}| = m, \tag{1}$$

and x is commonly referred to as the *object* or *input* of the imaging model [12].

A general signal space that appropriately models the underlying physics is the complex Hilbert space

$$\mathcal{L} = L^2[\mathbb{R}^N, \mathbb{C}]. \tag{2}$$

Hence, a signal x in $\mathcal{L}$ is a square-integrable function mapping a continuous variable $t \in \mathbb{R}^N$ to a complex number $x(t) \in \mathbb{C}$. The set of signals that satisfy the *Fourier domain constraint* (1) is

$$M = \big\{ y \in \mathcal{L} \colon |\widehat{y}| = m \text{ a.e.} \big\}. \tag{3}$$

In addition to the imaging model, an important piece of information that is typically available in phase retrieval problems is that the support of x is contained in some set $D \subset \mathbb{R}^N$. If we let 1_E denote the characteristic function of a set $E \subset \mathbb{R}^N$ and $\complement E$ its complement, this *object domain constraint* confines x to the set

$$S = \big\{ y \in \mathcal{L} \colon y \cdot 1_{\complement D} = 0 \big\}. \tag{4}$$

The phase retrieval problem can be posed as that of finding a function $x \in \mathcal{L}$ that satisfies these two constraints, namely,

$$\text{find} \quad x \in S \cap M. \tag{5}$$

This formulation exhibits the phase retrieval problem as a problem of finding a point in the intersection of constraint sets, i.e., a *set theoretic estimation problem* in the sense of [4]. In mathematics (especially in optimization) problems of this kind are called *feasibility problems*. In this paper we shall restrict our attention to the case when (5) is consistent, i.e., $S \cap M \neq \varnothing$.

3. FUNDAMENTALS OF NUMERICAL THEORY

We review here the necessary mathematical background. $\mathcal{H}$ denotes a general Hilbert space with scalar product $\langle \cdot, \cdot \rangle$, norm $\| \cdot \|$, and distance d.

3.1. Projections

If Y is a nonempty set in $\mathcal{H}$, then the distance from x to Y is $d(x, Y) = \inf_{y \in Y} \|x - y\|$. The set of points in Y nearest to a point $x \in \mathcal{H}$, namely

$$\big\{ y \in Y \colon \|x - y\| = d(x, Y) \big\}, \tag{6}$$

is denoted $P_Y(x)$ and called the *projection* of x onto Y. The induced operator P_Y is called the *projection operator* or *projector* onto Y. If Y is closed and convex, then $P_Y(x)$ is always a singleton $\{y\}$ and it is common practice to simply write $P_Y(x) = y$, a slight abuse of notation.

3.2. Projections for the phase retrieval problem

In the setting of the phase retrieval problem, the abstract Hilbert space $\mathcal{H}$ is simply the function space $\mathcal{L}$ of (2). The most common approach for solving the phase retrieval problem is to enforce the known object domain and Fourier domain constraints in some alternating fashion. Thus, given a signal x, the support constraint is naturally enforced by setting x equal to zero outside the given domain D, via the transformation $x \mapsto x \cdot 1_{\complement D}$. This simple operation is actually a projection (see [1] for a formal proof):

Example 3.1 (support constraint) Suppose D is a measurable set in $\mathbb{R}^N$ and fix $x \in \mathcal{L}$. Then the projection of x onto the set S of (4) is $P_S(x) = x \cdot 1_D$.

The same observation is true for the Fourier modulus constraint. Approaches to enforce it are described below; again, these operations turn out to be projections (see [1] for details).

Example 3.2 (Fourier modulus constraint) Let m be a nonnegative function in $\mathcal{L}$ and fix $x \in \mathcal{L}$. Then the set M of (3) is closed and nonconvex (unless $m \equiv 0$). Moreover, $y \in \mathcal{L}$ belongs to the projection $P_M(x)$ of x onto M if and only if it satisfies a.e.

$$\widehat{y}(\omega) = \begin{cases} m(\omega) \frac{\widehat{x}(\omega)}{|\widehat{x}(\omega)|}, & \text{if } \widehat{x}(\omega) \neq 0; \\ m(\omega) \exp[i\varphi(\omega)], & \text{otherwise,} \end{cases} \tag{7}$$

for some measurable function $\varphi : \mathbb{R}^N \to \mathbb{R}$.

Example 3.2 shows that every function $y \in P_M(x)$ satisfies

$$d(\widehat{x}(\omega), m(\omega)\mathbb{S}) = d(\widehat{x}(\omega), \widehat{y}(\omega)) \quad \text{a.e. on } \mathbb{R}^N, \tag{8}$$

where $m(\omega)\mathbb{S} = \{u \in \mathbb{C} \colon |u| = m(\omega)\}$ denotes a circle in the complex plane, with radius $m(\omega)$ and centered at the origin. The *multi-valuedness* of the projection is now evident: whenever $\widehat{x}(\omega) = 0$, any phase φ will work.

In practice, one picks the *particular selection* $y_0 \in P_M(x)$ corresponding to zero phase $\varphi \equiv 0$:

$$\widehat{y_0}(\omega) = \begin{cases} m(\omega) \frac{\widehat{x}(\omega)}{|\widehat{x}(\omega)|}, & \text{if } \widehat{x}(\omega) \neq 0; \\ m(\omega), & \text{otherwise.} \end{cases} \tag{9}$$

Ultimately, the difficulty of the phase retrieval problem is caused by the lack of convexity of the Fourier domain constraint and the lack of good convex approximations to it.

4. CLASSICAL ALGORITHMS AND CONNECTIONS

We discuss three popular algorithms designed for solving the phase retrieval problem (5). We follow Fienup's framework [7]. To bring out the results as clearly as possible, we set Fienup's parameter β equal to 1 and assume that the object domain constraint is only a support constraint. Moreover, we identify the set-valued operator P_M with its selection defined in (9) and therefore regard it as a single-valued operator.

4.1. Error reduction algorithm

The error reduction algorithm, updates a current iterate x_n via

$$x_{n+1}(t) = \begin{cases} \big(P_M(x_n)\big)(t), & \text{if } t \in D; \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

Hence $x_{n+1} = 1_D \cdot P_M(x_n)$; equivalently, by Example 3.1,

$$x_{n+1} = (P_S P_M)(x_n). \tag{11}$$

4.2. Fienup's basic input-output (BIO) algorithm

The update x_{n+1} in the BIO algorithm is obtained from x_n by setting

$$x_{n+1}(t) = \begin{cases} x_n(t), & \text{if } t \in D; \\ x_n(t) - \big(P_M(x_n)\big)(t), & \text{otherwise.} \end{cases} \tag{12}$$

Note that $x_{n+1} = 1_D \cdot x_n + 1_{\complement D} \cdot \big(x_n - P_M(x_n)\big) = x_n - (1 - 1_D) \cdot P_M(x_n)$, which we rewrite as [1]

$$x_{n+1} = (P_S P_M + I - P_M)(x_n). \tag{13}$$

4.3. Fienup's hybrid input-output (HIO) algorithm

The HIO algorithm constructs the successor of x_n via

$$x_{n+1}(t) = \begin{cases} \big(P_M(x_n)\big)(t), & \text{if } t \in D; \\ x_n(t) - \big(P_M(x_n)\big)(t), & \text{otherwise.} \end{cases} \tag{14}$$

Thus $x_{n+1} = 1_D \cdot P_M(x_n) + 1_{\complement D} \cdot (x_n - P_M(x_n))$, which can be rewritten as [1]

$$x_{n+1} = \frac{1}{2}\big((2P_S - I)(2P_M - I) + I\big)(x_n). \tag{15}$$

4.4. Main results

We are now ready to establish the formal correspondence between classical algorithms for solving (5) and their counterparts for solving a two-set convex feasibility problem.

4.4.1. *The convex feasibility problem*

Assume A and B are two closed convex sets in a real Hilbert space $\mathcal{H}$. The associated *convex feasibility problem* is to

$$\text{find } x \in A \cap B. \tag{16}$$

We now revisit the three classical algorithms for solving the phase retrieval problem described above. It will turn out that each algorithm corresponds to a classical algorithm for solving (16).

4.4.2. *Error reduction algorithm and POCS*

The method of alternating *projections onto convex sets (POCS)* generates, for the present setting of two constraints, sequences (a_n) and (b_n) as follows: pick an arbitrary starting point $a_0 \in \mathcal{H}$, then update for $n \geq 0$ via

$$b_n = P_B(a_n) \quad \text{and} \quad a_{n+1} = P_A(b_n). \tag{17}$$

The following basic result shows that POCS does find a solution of (16):

Fact 4.1 [3] If $A \cap B \neq \varnothing$, then (a_n) and (b_n) converge weakly to a point in $A \cap B$.

Observation 4.2 Replace the set A with the (convex) object domain constraint set S and the set B with the (nonconvex) Fourier domain constraint set M. Then the sequence (a_n) generated by (17) corresponds to the sequence (x_n) generated by the error reduction algorithm (11). This connection was established by Levi and Stark [10] in 1984.

4.4.3. *Fienup's BIO algorithm and Dykstra's algorithm*

For two closed convex sets A and B, *Dykstra's algorithm* [2] produces four sequences (a_n), (b_n), (p_n), and (q_n) as follows. Fix a starting point a_0, set $q_{-1} = 0 = p_0$, and update for $n \geq 0$ via

$$\begin{cases} b_n = P_B(a_n + q_{n-1}); \\ q_n = (I - P_B)(a_n + q_{n-1}) = a_n + q_{n-1} - b_n; \\ a_{n+1} = P_A(b_n + p_n); \\ p_{n+1} = (I - P_A)(b_n + p_n) = b_n + p_n - a_{n+1}. \end{cases}$$

Clearly, Dykstra's algorithm is more involved than POCS and is more demanding in terms of storage; however, its convergence properties are superior.

Fact 4.3 [2] Suppose $A \cap B \neq \varnothing$. Then both sequences (a_n) and (b_n) converge in norm to $P_{A\cap B}(a_0)$, the point in $A \cap B$ closest to a_0.

Fact 4.3 is quite remarkable because the sequences converge in norm, and their limit is explicitly identified as the nearest feasible point to the starting point. For applications of Dykstra's algorithm to signal recovery, see [5].

For the rest of this subsection, we assume additionally that A is a *closed linear subspace*. Then (p_n) lies entirely in $A^\perp$, the orthogonal complement of A, and the computation of a_{n+1} becomes $a_{n+1} = P_A b_n + P_A p_n = P_A b_n$. Thus, the sequence (p_n) is not needed, and Dykstra's algorithm simplifies to:

$$\begin{cases} b_n = P_B(a_n + q_{n-1}); \\ q_n = (I - P_B)(a_n + q_{n-1}); \\ a_{n+1} = P_A(b_n). \end{cases} \tag{18}$$

The next observation identifies the BIO algorithm as a nonconvex Dykstra algorithm.

Observation 4.4 [1] Replace the set A with the (convex) object domain constraint set S and the set B with the (nonconvex) Fourier domain constraint set M. Then the sequence $(a_n + q_{n-1})$ generated by (18) corresponds to the sequence (x_n) generated by Fienup's BIO algorithm (13).

Remark 4.5 Even when $A \cap B \neq \emptyset$, it is possible that the sequences (p_n) and (q_n) generated by Dykstra's algorithm (in its general form) are both *unbounded*; see [9]. This suggests the pertinent sequence to monitor in Fienup's BIO algorithm is $\big(P_M(x_n)\big)$, rather than (x_n).

4.4.4. Fienup's HIO algorithm and the Douglas-Rachford algorithm

When specialized to the convex feasibility problem (16), the *Douglas-Rachford algorithm* generates a sequence (x_n), from an arbitrary starting point x_0, by

$$x_{n+1} = \frac{1}{2}\big(R_A R_B + I\big)(x_n), \tag{19}$$

where $R_A = 2P_A - I$ is the *reflector* with respect to A (and R_B is defined likewise).

Fact 4.6 [11] Suppose $A \cap B \neq \emptyset$. Then the sequence (x_n) converges weakly to some fixed point x of T and $P_B(x) \in A \cap B$. Moreover, the sequence $\big(P_B(x_n)\big)$ is bounded, and every weak cluster point of $\big(P_B(x_n)\big)$ lies in $A \cap B$. If $\mathcal{H}$ is finite-dimensional, then $x_n \to x$ and $P_B(x_n) \to P_B(x) \in A \cap B$.

The following connection identifies the HIO algorithm as a nonconvex Douglas-Rachford algorithm.

Observation 4.7 [1] Replace the set A with the (convex) object domain constraint set S and the set B with the (nonconvex) Fourier domain constraint set M. Then the sequence generated by the Douglas-Rachford algorithm (19) corresponds to the sequence generated by the HIO algorithm (15).

5. DISCUSSION

In this paper, new connections have been established between some classical phase retrieval methods and some standard convex optimization algorithms. While the mathematical theory remains unable to completely analyze the convergence behavior of these algorithms in nonconvex settings, the analogies drawn here open the door for experimentation with variations that are well understood in convex settings. We believe that the convex-analytical viewpoint adopted in this paper can be exploited further in order to develop alternative phase retrieval schemes.

6. REFERENCES

[1] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Phase retrieval, error reduction algorithm, and Fienup variants: A view from convex optimization," *J. Opt. Soc. Amer. A,* vol. 19, 2002.

[2] J. P. Boyle and R. L. Dykstra, "A method for finding projections onto the intersection of convex sets in Hilbert spaces," *Lect. Notes in Statistics*, vol. 37, pp. 28-47, 1986.

[3] L. M. Bregman, "The method of successive projection for finding a common point of convex sets," *Soviet Math. Doklady,* vol. 6, pp. 688-692, May 1965.

[4] P. L. Combettes, "The foundations of set theoretic estimation," *Proc. IEEE,* vol. 81, pp. 182-208, Feb. 1993.

[5] P. L. Combettes, "Signal recovery by best feasible approximation," *IEEE Trans. Image Processing,* vol. 2, pp. 269-271, Apr. 1993.

[6] P. L. Combettes, "The convex feasibility problem in image recovery," in *Advances in Imaging and Electron Physics,* vol. 95, pp. 155-270. New York: Academic, 1996.

[7] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Appl. Opt.,* vol. 21, pp. 2758-2769, Aug. 1982.

[8] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik,* vol. 35, pp. 237-246, 1972.

[9] S. P. Han, "A successive projection method," *Math. Programming (Ser. A)*, vol. 40, pp. 1-14, Jan. 1988.

[10] A. Levi and H. Stark, "Image restoration by the method of generalized projections with application to restoration from magnitude," *J. Opt. Soc. Amer. A,* vol. 1, pp. 932-943, Sept. 1984.

[11] P. L. Lions and B. Mercier, "Splitting algorithms for the sum of two nonlinear operators," *SIAM J. Numer. Anal.,* vol. 16, pp. 964-979, Dec. 1979.

[12] D. R. Luke, J. V. Burke, and R. Lyon. "Optical wavefront reconstruction: Theory and numerical methods," *SIAM Rev.,* to appear.

[13] H. Stark (Ed.), *Image Recovery: Theory and Application.* San Diego, CA: Academic, 1987.

[14] D. C. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1 – theory," *IEEE Trans. Medical Imaging,* vol. 1, pp. 81-94, Oct. 1982.