Notes on: Iterative Oblique Projection onto Convex Sets, the Split Feasibility Problem and the Variational Inequality Problem

Charles Byrne (Charles_Byrne@uml.edu)
Department of Mathematical Sciences
University of Massachusetts at Lowell
Lowell, MA 01854, USA

March 11, 2002

Abstract

Let C and Q be nonempty closed convex sets in R^N and R^M, respectively, and A an M by N real matrix. The *split feasibility problem* (SFP) is to find $x \in C$ with $Ax \in Q$, if such x exist. An iterative method for solving the SFP, called the CQ algorithm, has the following iterative step:

$$x^{k+1} = P_C(x^k + \gamma A^T(P_Q - I)Ax^k),$$

where $\gamma \in (0, 2/L)$ with L the largest eigenvalue of the matrix A^TA and P_C and P_Q denote the orthogonal projections onto C and Q, respectively; that is, P_Cx minimizes $||c - x||$, over all $c \in C$. The CQ algorithm converges to a solution of the SFP, or, more generally, to a minimizer of $||P_QAc - Ac||$ over c in C, whenever such exist.

The CQ algorithm involves only the orthogonal projections onto C and Q, which we shall assume are easily calcuated, and involves no matrix inverses. If A is normalized so that each row has length one, then L does not exceed the maximum number of nonzero entries in any column of A, which provides a helpful estimate of L for sparse matrices.

Particular cases of the CQ algorithm are the Landweber and projected Landweber methods for obtaining exact or approximate solutions of the linear equations $Ax = b$; the *algebraic reconstruction technique* (ART) of Gordon, Bender and Herman is a particular case of a block-iterative version of the CQ algorithm.

One application of the CQ algorithm that is the subject of ongoing work is dynamic emission tomographic image reconstruction, in which the vector x is the concatenation of several images corresponding to successive discrete times. The matrix A and the set Q can then be selected to impose constraints on the behavior over time of the intensities at fixed voxels, as well as to require consistency (or near consistency) with measured data.

A continuous operator $T : R^N \to R^N$ is called *monotone* if, for each x and y, we have

$$\langle Tx - Ty, x - y\rangle \geq 0.$$

The *variational inequality problem* (VIP) is to find $\hat{c} \in C$ such that, for all $c \in C$, we have

$$\langle T\hat{c}, c - \hat{c}\rangle \geq 0.$$

The VIP algorithm with the iterative step

$$x^{k+1} = P_C(x^k - \gamma T x^k)$$

is known to converge to a solution of the VIP, with certain fairly restrictive conditions on γ and T. This algorithm is similar to the CQ algorithm, with $T = A^T(I - P_Q)A$, but the conditions apparently needed for convergence do not hold. This suggests a strengthening of the VIP algorithm, which we show to converge whenever T is ν*-firmly monotone*, that is, there is $\nu > 0$ with

$$\langle Tx - Ty, x - y\rangle \geq \nu ||Tx - Ty||^2,$$

for all x and y, and $\gamma \in (0, 2\nu)$.

The operator $T = A^T(I - P_Q)A$ is ν-firmly monotone for $\nu = 1/L$ and the CQ algorithm is a special case of the strengthened VIP algorithm.

1

The Split Feasibility Problem

Let C and Q be nonempty closed convex sets in R^N and R^M, respectively, and A an M by N real matrix. Let $A^{-1}(Q) = \{x | Ax \in Q\}$ and $F = C \cap A^{-1}(Q)$. The problem, to find $x \in C$ with $Ax \in Q$, if such x exist, was called *the split feasibility problem* (SFP) by Censor and Elfving in [13]. Denote by P_C and P_Q the *orthogonal projection* (sometimes called the *proximity map*) onto C and Q, respectively; that is, $P_C x$ minimizes $||c - x||$, over all $c \in C$.

The CQ algorithm for the SFP

The CQ algorithm to solve the SFP is the following:

Algorithm 1.1 *Let x^0 be arbitrary. For $k = 0, 1, \ldots$ let*
The CQ algorithm:

$$x^{k+1} = P_C(x^k + \gamma A^T(P_Q - I)Ax^k), \tag{1.1}$$

where $\gamma \in (0, 2/L)$ and L denotes the largest eigenvalue of the matrix $A^T A$.

Note that the CQ algorithm involves only the orthogonal projections onto C and Q, which we shall assume are easily calculated, and involves no matrix inverses. Also, the relaxation parameter γ is independent of k and no line searchs are used at each step. Special cases of the CQ algorithm are the Landweber and projected Landweber methods for finding constrained solutions of systems of linear equations.

Projection onto convex sets - POCS

In [16] Cheney and Goldstein considered the case of two nonempty closed convex sets K_1 and K_2 in Hilbert space. With P_j denoting the orthogonal projection (proximity map) onto K_j, for $j = 1, 2$, respectively, and $T = P_1P_2$, they show that the sequence $x^k = T^k x^0$ obtained by alternating distance minimization converges to a fixed point of T if either a) one of the two sets is compact, or b) one set is finite dimensional and the distance between the two sets is attained. In particular, when the intersection of the two convex sets is nonempty, the sequence $\{x^k\}$ converges to a member of that intersection, in either of the two cases above.

Let K_j, $j = 1, ..., J$, be nonempty closed convex subsets of R^M, with nonempty intersection K. The *convex feasibility problem* (CFP) is to find an element of K. Solving the SFP is equivalent to finding a member of the intersection of the two sets Q and $A(C) = \{Ac|c \in C\}$, or of the intersection of the two sets $A^{-1}(Q)$ and C, and so the SFP can be viewed as a particular case of the CFP.

This result can be extended in several directions, to include finitely many convex sets with nonempty intersection and projections involving generalized distances; for details see the survey papers by Deutsch [17] and Bauschke and Borwein [2] and the recent books by Censor and Zenios [15] and Stark and Yang [26].

Oblique projection onto convex sets

If we try to use the Cheney-Goldstein approach to solve the SFP, we encounter the difficulty of calculating the orthogonal projection onto the set $A(C)$ or onto $A^{-1}(Q)$. In this paper we shall assume that the orthogonal projections P_C and P_Q are easily calculated; the main virtue of the CQ algorithm is that it involves only the maps P_C and P_Q at each step.

For the case in which $M = N$ and A is invertible, the map $P^{A^TA}_{A^{-1}(Q)} = A^{-1}P_QA$ is an oblique projection onto the set $A^{-1}(Q)$; that is, $A^{-1}P_Q(Ax)$ minimizes the function $f(z) = (z - x)^T A^T A(z - x)$ over all z in $A^{-1}(Q)$. This suggests the possibility of modifying the Cheney-Goldstein method to include one orthogonal and one oblique projection.

A counterexample

The following counterexample shows us that the iterative sequence

$$x^{k+1} = P_C P^{A^TA}_{A^{-1}(Q)}(x^k) \tag{1.2}$$

need not converge to a solution. Let $M = N = 2$, with C the horizontal axis, Q the vertical axis and A the matrix

$$A = \begin{bmatrix} 1 & -1 \\ -1 & 0 \end{bmatrix}.$$

Let $x^0 = (1,0)^T$. Then $Ax^0 = (1,-1)^T$, $P_Q(Ax^0) = (0,-1)^T$, $A^{-1}P_Q(Ax^0) = (1,1)^T$ and $P_C A^{-1}P_Q(Ax^0) = (1,0)^T = x^0$. The solution to the SFP in this case is $(0,0)$, to which the iterative sequence given by (1.2) fails to converge.

Ordinary and generalized relaxation

It is possible, however, to formulate convergent iterative algorithms using projections based on distinct generalized distances, provided a form of generalized relaxation is included at each step. Such an algorithm, called the *multidistance successive generalized projection* (MSGP) method, is the main topic of the papers [8] and [10].

Let $T : R^N \to R^N$ be an operator on R^N. An *ordinary relaxation* of T would then be an operator R of the form

$$Rx = (1-\gamma a^2)x + \gamma a^2 Tx,$$

for $\gamma > 0$ and $1 - \gamma a^2 > 0$. An operator of the following form could then be called a *generalized relaxation* of T:

$$Gx = (I - \gamma A^TA)x + \gamma A^TATx,$$

for $\gamma > 0$ and $I - \gamma A^TA$ positive-definite.

Generalized relaxation of the oblique projection

The CQ algorithm was originally obtained by incorporating in the Cheney-Goldstein method a generalized relaxation of the operator $P^{A^TA}_{A^{-1}(Q)}$. The CQ algorithm includes the mapping

$$Gx = x + \gamma A^T(P_Q - I)Ax,$$

which, for the case of invertible A, can be written as

$$Gx = (I - \gamma A^T A)x + \gamma A^T A P^{A^T A}_{A^{-1}(Q)} x.$$

The CQ algorithm allows for *overrelaxation* as well; that is, we require only that $2I - \gamma A^T A$ be positive-definite.

The CQ algorithm in the inconsistent case

Cheney and Goldstein consider the case in which the two convex sets have empty intersection. For the SFP, we are interested in what happens when the sets C and $A^{-1}(Q)$ have empty intersection. In this case, we show that the CQ algorithm converges to a minimizer of the function $||P_Q Ax - Ax||$, relative to the set C, whenever such relative minimizers exist.

The convergence theorem for the CQ algorithm

Let F be the (possibly empty) set of all $c \in C$ at which the function $||P_Q Ac - Ac||$ attains its minimum value over C. We have the following theorem concerning the CQ algorithm:

Theorem 1.1 *Let F be nonempty. Then the sequence $\{x^k\}$ defined by equation (1.1) converges to a member of F, for any starting vector x^0, for all $\gamma \in (0, 2/L)$.*

Corollary 1.1 *The sequence $\{x^k\}$ defined by equation (1.1) converges to a solution of the SFP, whenever such solutions exist, for any $\gamma \in (0, 2/L)$.*

An estimate of L useful for sparse matrices

We show that if A is normalized so that each row has length one, then L does not exceed the maximum number of nonzero entries in any column of A.

Convex functions and monotone operators

If $f : R^N \rightarrow R$ is a differentiable convex function [24],[25], then for all x and y we have

$$f(x) - f(y) \geq \langle \nabla f(y), x - y \rangle$$

and

$$f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle \, ;$$

adding the two inequalities we obtain

$$\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq 0$$

for all x and y. More generally, a continuous operator $T : R^N \to R^N$ is said to be *monotone* if, for all x and y,

$$\langle Tx - Ty, x - y\rangle \geq 0$$

and ν *-strongly monotone* if there is a positive constant ν such that

$$\langle Tx - Ty, x - y\rangle \geq \nu||x - y||^2.$$

The operator T is λ *-Lipschitz continuous* if there is a nonnegative constant λ such that

$$||Tx - Ty|| \leq \lambda||x - y||$$

for all x and y.

The Variational Inequality Problem

The vector $\hat{c} \in C$ will be a minimizer of f, relative to set C, if and only if

$$\langle \nabla f(\hat{c}), c - \hat{c}\rangle \geq 0,$$

for all $c \in C$. Here again f is assumed to be convex and differentiable. The analogous problem for monotone operator T is to find $\hat{c} \in C$ such that

$$\langle T\hat{c}, c - \hat{c}\rangle \geq 0,$$

for all $c \in C$. This is called the *variational inequality problem* (VIP).

An iterative algorithm for the VIP

The algorithm with the iterative step

$$x^{k+1} = P_C(x^k - \gamma Tx^k),$$

sometimes called the *projected gradient* method, converges to a unique solution of the VIP if the operator T is λ -Lipschitz continuous and ν -strongly monotone and $\gamma \in (0, 2\nu/\lambda^2)$ [4]. This is a consequence of the contraction mapping principle.

Although the CQ algorithm appears similar, with $T = A^T(I - P_Q)A$, we know it converges in situations in which there are non-unique solutions. This suggests a strengthening of the convergence result for the VIP algorithm.

Firmly monotone operators

The monotone operator T will be called ν *-firmly monotone* if there is a positive constant ν such that

$$\langle Tx - Ty, x - y\rangle \geq \nu ||Tx - Ty||^2.$$

From Cauchy's inequality we then have

$$||Tx - Ty|| \leq \frac{1}{\nu}||x - y||,$$

so T is Lipschitz continuous.

Strengthening the VIP convergence theorem

The so-called *projected gradient* algorithm with the iterative step

$$x^{k+1} = P_C(x^k - \gamma Tx^k)$$

converges to a solution of the VIP, if T is ν -firmly monotone and $\gamma \in (0, 2\nu)$, whenever solutions exist.

The CQ algorithm extended to the VIP

The operator $T = A^T(I - P_Q)A$ is ν -firmly monotone, with $\nu = 1/L$, so that the iterative CQ algorithm with the iterative step

$$x^{k+1} = P_C(x^k - \gamma A^T(I - P_Q)Ax^k)$$

converges to a solution of the SFP, or, more generally, to a minimizer, relative to the set C, of the function $||P_QAc - Ac||$, whenever such exist.

Acknowledgments

I wish to thank Professor James Graham-Eagle of the University of Massachusetts Lowell Department of Mathematical Sciences and Dr. Troy Farncombe of the University of Massachusetts Medical School for helpful discussions concerning these matters. This work was supported, in part, by the National Institutes of Health (NIH), under Grant CA23452. The contents of this paper are solely the responsibility of the author and do not necessarily represent the official views of NIH.

References

[1] Anderson, A. and Kak, A. (1984) Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm, *Ultrasonic Imaging*, **6**, pp. 81–94.

[2] Bauschke, H.H. and Borwein, J.M. (1996) On projection algorithms for solving convex feasibility problems, *SIAM Review*, **38**: 367–426.

[3] Bertero, M. and Boccacci, P. (1998) *Introduction to Inverse Problems in Imaging*, Institute of Physics Publishing, Bristol, UK.

[4] Bertsekas, D., and Tsitsiklis, J. (1989) *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, New Jersey.

[5] Bregman, L.M. (1967) The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, *USSR Computational Mathematics and Mathematical Physics*, **7**: 200–217.

[6] Byrne, C.L. (1993) Iterative image reconstruction algorithms based on cross-entropy minimization, *IEEE Transactions on Image Processing*, **2**, pp. 96–103.

[7] Byrne, C.L. (1996) Block-iterative methods for image reconstruction from projections, *IEEE Transactions on Image Processing*, **IP-5**, pp. 792–794.

[8] Byrne, C.L. (1999) Iterative projection onto convex sets using multiple Bregman distances, *Inverse Problems*, **15**, pp. 1295-1313.

[9] Byrne, C.L. (2000) Block-iterative interior point optimization methods for image reconstruction from limited data, *Inverse Problems*, **16**, pp. 1405–1419.

[10] Byrne, C.L. (2001) Bregman-Legendre multidistance projection algorithms for convex feasibility and optimization, in *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, Butnariu, D., Censor, Y. and Reich, S., editors, Elsevier Publ., pp. 87–100.

[11] Byrne, C. (2002) Iterative oblique projection onto convex sets and the split feasibility problem, *Inverse Problems*, to appear.

[12] Byrne, C., and Censor, Y. (2002) Proximity function minimization using multiple Bregman projections, with applications to split feasibility and Kullback-Leibler distance minimization, *Annals of Operations Research*, **105**, pp. 77–98.

[13] Censor, Y. and Elfving, T. (1994) A multiprojection algorithm using Bregman projections in a product space, *Numerical Algorithms*, **8**, pp. 221–239.

[14] Censor, Y., Iusem, A. and Zenios, S. (1998) An interior point method with Bregman functions for the variational inequality problem with paramonotone operators, *Mathematical Programming*, **81**, pp. 373-400 .

[15] Censor, Y. and Zenios, S.A. (1997) *Parallel Optimization: Theory, Algorithms and Applications*, Oxford University Press, New York.

[16] Cheney, W. and Goldstein, A. (1959) Proximity maps for convex sets, *Proceedings of the American Mathematical Society*, **10**, pp. 448–450.

[17] Deutsch, F. (1992) The method of alternating orthogonal projections, in *Approximation Theory, Spline Functions and Applications*, S.P. Singh, editor, Kluwer Academic Publishing, pp. 105–121.

[18] Gordon, R., Bender, R. and Herman, G. T. (1970) Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography, *Journal of Theoretical Biology*, **29**, pp. 471-481.

[19] Gubin, L.G., Polyak, B.T. and Raik, E.V. (1967) The method of projections for finding the common point of convex sets, *USSR Computational Mathematics and Mathematical Physics*, **7**: 1–24.

[20] Kaczmarz, S. (1937) Angenäherte Auflösung von Systemen linearer Gleichungen, *Bulletin de l'Academie Polonaise des Sciences et Lettres*, **A35**, 355-357.

[21] Landweber, L. (1951) An iterative formula for Fredholm integral equations of the first kind, *Amer. J. of Math.*, **73**, pp. 615-624.

[22] Lagendijk, R. and Biemond, J. (1991) *Iterative Identification and Restoration of Images*, Kluwer Publ.

[23] Luenberger, D. (1969) *Optimization by Vector Space Methods*, John Wiley and Sons, New York.

[24] Ortega, J., and Rheinboldt, W. (1970) *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, NY.

[25] Rockafellar, R. (1970) *Convex Analysis*, Princeton University Press.

[26] Stark, H. and Yang, Y. (1998) *Vector Space Projections: A Numerical Approach to Signal and Image Processing, Neural Nets and Optics*, John Wiley and Sons, New York.

[27] Tseng, P. (1992) On the convergence of the products of firmly nonexpansive mappings, *SIAM Journal of Optimization*, **2**, pp. 425–434.

[28] Yamada, I. (2001) The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings, *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications* (D. Butnariu, Y.Censor and S. Reich, editors), Elsevier, Amsterdam, pp. 473–504.

[29] Youla, D.C. (1987) Mathematical theory of image restoration by the method of convex projections, in: Stark, H. (Editor) (1987) *Image Recovery: Theory and Applications*, Academic Press, Orlando, FL, USA, pp. 29–78.

[30] Zeidler, E. (1990) *Nonlinear Functional Analysis and its Applications: II/B-Nonlinear Monotone Operators*, Springer.