The CQ Algorithm: Extensions and Applications

Charles Byrne
(Charles_Byrne@uml.edu)
http://faculty.uml.edu/cbyrne/cbyrne.html
Department of Mathematical Sciences
University of Massachusetts at Lowell
Lowell, MA 01854, USA

August 16, 2007
ICCOPT 07– MOPTA 07
McMaster University, Hamilton, Ontario
August 13–17, 2007

University of Massachusetts
UMASS Lowell

The Split Feasibility Problem

Let A be a real M by N matrix, and C and Q non-empty closed, convex sets in R^N and R^M, respectively.

The *split feasibility problem* (SFP) is to find a vector x in C, such that Ax is in Q.

The Inconsistent Case

When the SFP has no solution, it is sensible to seek a minimizer of the function

$$f(x) = \frac{1}{2}||P_Q Ax - Ax||_2^2, \tag{1}$$

over x in C; P_Q denotes the orthogonal projection onto Q.

The CQ Algorithm

For arbitrary x^0 and $k = 0, 1, ...,$ and γ in the interval $(0, 2/\rho(A^T A))$, where $\rho(A^T A)$ denotes the largest eigenvalue of the matrix $A^T A$, let

$$x^{k+1} = P_C(x^k - \gamma A^T(I - P_Q)Ax^k). \tag{2}$$

This is the CQ algorithm [9, 10]. The CQ algorithm converges to a solution of the SFP, whenever solutions exist. When there are no solutions of the SFP, the CQ algorithm converges to a minimizer, over x in C, of the function

$$f(x) = \frac{1}{2}||P_Q Ax - Ax||_2^2, \tag{3}$$

whenever such minimizers exist.

Topics for Discussion:

- The proof of convergence of the CQ algorithm;
- The algorithm in context- the CQ algorithm has particular cases, and is, itself, a particular case;
- Approximation of the orthogonal projections;
- Relation to proximity operators and forward-backward splitting;
- The CQ algorithm and successive orthogonal projection;
- Extension to incorporate multiple sets;
- Use in radiation therapy;
- Use in dynamic emission tomography.

Proof of Convergence

According to the Krasnoselskii-Mann Theorem [18], if $T : R^N \rightarrow R^N$ is an *averaged operator* then the sequence $\{T^k x^0\}$ converges to a fixed point of T, whenever fixed points exist. The operator P_C is averaged, since it is firmly non-expansive, and, for γ in the interval $(0, 2/\rho(A^T A))$, the operator $I - \gamma A^T(I - P_Q)A$ is also averaged. The product of averaged operators is averaged, so the CQ algorithm converges to a fixed point, if they exist, and these fixed points also minimize $f(x)$ over x in C.

Averaged Operators

An operator $T : R^N \to R^N$ is *non-expansive* (in the Euclidean norm) if, for all x and y, we have

$$||Tx - Ty||_2 \leq ||x - y||_2. \tag{4}$$

An operator $S : R^N \to R^N$ is *averaged* [4] if there is a non-expansive operator T and α in the interval $(0, 1)$, with

$$S = (1 - \alpha)I + \alpha T. \tag{5}$$

Details

- The function $f(x)$ is convex and differentiable on R^N and its derivative is the operator
$$\nabla f(x) = A^T(I - P_Q)Ax; \tag{6}$$
see Aubin [2].
- Let $h(x)$ be convex and differentiable and its derivative, $\nabla h(x)$, be non-expansive. Then $\nabla h(x)$ is firmly non-expansive; see Golshtein and Tretyakov [15].
- The derivative operator ∇f is λ-Lipschitz continuous for $\lambda = \rho(A^TA)$, therefore the operator $I - \gamma\nabla f$ is averaged, for γ in $(0, 2/\rho(A^TA))$.

Estimating $\rho(A^TA)$

The CQ algorithm employs the relaxation parameter γ in the interval $(0, 2/\rho(A^TA))$, where $\rho(A^TA)$ is the largest eigenvalue of the matrix A^TA. Choosing the best relaxation parameter in any algorithm is a nontrivial procedure. Generally speaking, we want to select γ near to $1/\rho(A^TA)$. A simple estimate for $\rho(A^TA)$ that is particularly useful when A is sparse is the following: if A is normalized so that each row has length one, then the spectral radius of A^TA does not exceed the maximum number of nonzero elements in any column of A. A similar upper bound on $\rho(A^TA)$ was obtained for non-normalized, ϵ-sparse A [9].

Particular Cases of the SFP

It is easy to find important examples of the SFP: if $C = R^J$ and $Q = \{b\}$ then solving the SFP amounts to solving the linear system of equations $Ax = b$; if C is a proper subset of R^J, such as the nonnegative cone, then we seek solutions of $Ax = b$ that lie within C, if there are any. Generally, we cannot solve the SFP in closed form and iterative methods are needed.

Particular Cases of the CQ Algorithm

A number of well known iterative algorithms, such as the Landweber [17] and projected Landweber methods (see [8]), are particular cases of the CQ algorithm.

The Landweber Algorithm

With x^0 arbitrary and $k = 0, 1, ...,$ the Landweber algorithm for finding a (possibly least-squares) solution of $Ax = b$ has the iterative step

$$x^{k+1} = x^k + \gamma A^T(b - Ax^k). \tag{7}$$

The Projected Landweber Algorithm

For a general nonempty closed convex C, x^0 arbitrary, and $k = 0, 1, ...,$ the projected Landweber algorithm for finding a (possibly constrained least-squares) solution of $Ax = b$ in C has the iterative step

$$x^{k+1} = P_C(x^k + \gamma A^T(b - Ax^k)). \tag{8}$$

The Simultaneous ART (SART)

Another example of the CQ algorithm is the *simultaneous algebraic reconstruction technique* (SART) of Anderson and Kak [1] for solving $Ax = b$, for nonnegative matrix A . Let A be an M by N matrix with nonnegative entries. Let $A_{m+} > 0$ be the sum of the entries in the mth row of A and $A_{+n} > 0$ be the sum of the entries in the nth column of A. Consider the (possibly inconsistent) system $Ax = b$. For x^0 arbitrary and $k = 0, 1, \ldots,$ let

$$x_n^{k+1} = x_n^k + \frac{1}{A_{+n}} \sum_{m=1}^{M} A_{mn}(b_m - (Ax^k)_m)/A_{m+}. \tag{9}$$

This is the SART algorithm. With a change of variables, the SART becomes a particular case of the Landweber iteration.

Changing Variables

We make the following changes of variables:

$$B_{mn} = A_{mn}/(A_{m+})^{1/2}(A_{+n})^{1/2}, \tag{10}$$

$$z_n = x_n(A_{+n})^{1/2}, \tag{11}$$

and

$$c_m = b_m/(A_{m+})^{1/2}. \tag{12}$$

Then the SART iterative step can be written as

$$z^{k+1} = z^k + B^T(c - Bz^k). \tag{13}$$

This is a particular case of the Landweber algorithm, with $\gamma = 1$. The convergence of SART follows, once we know that the largest eigenvalue of $B^T B$ is less than two; in fact, it is one [9].

Using the CQ Algorithm

We illustrate the use of the CQ algorithm to prove a convergence result for the successive orthogonal projections method for the case of two non-intersecting convex sets.

The Convex Feasibility Problem

The *convex feasibility problem*(CFP) is to find a point in the non-empty intersection C of finitely many closed, convex sets C_m, $m = 1, ..., M$, in R^N.
The *successive orthogonal projections* (SOP) method (see Gubin, Polyak and Raik [16]) is the following. Begin with an arbitrary x^0. For $k = 0, 1, ...,$ and $m = k(\mathrm{mod}\, M) + 1$, let

$$x^{k+1} = P_m x^k, \tag{14}$$

where $P_m x$ denotes the orthogonal projection of x onto the set C_m.

The SOP when C is not empty

Since each of the operators P_m is firmly non-expansive, the product

$$T = P_M P_{M-1} \cdots P_2 P_1 \tag{15}$$

is averaged. Since C is not empty, T has fixed points and the sequence $\{x^k\}$ converges to a member of C. It is useful to note that the limit of this sequence will not generally be the point in C closest to x^0.

When C is empty

When the intersection C of the convex sets C_m is empty, the SOP cannot converge. Drawing on our experience with two special cases of the SOP, the ART and the Agmon-Motzkin-Schoenberg algorithms, we conjecture that there is a *limit cycle*, that is, for each $m = 1, ..., M$, the subsequences $\{x^{nM+m}\}$ converge to $c^{*,m}$ in C_m, with $P_m c^{*,m-1} = c^{*,m}$ for $m = 2, 3, ..., M$, and $P_1 c^{*,M} = c^{*,1}$; the set $\{c^{*,m}\}$ is then a limit cycle. For the special case of $M = 2$ we can prove this. The proof here uses the CQ algorithm.

The Theorem

Theorem

Let C_1 and C_2 be nonempty, closed convex sets in R^J, with $C_1 \cap C_2 = \emptyset$. Assume that there is a unique $\hat{c}_2$ in C_2 minimizing the function $f(x) = ||c_2 - P_1 c_2||_2$, over all c_2 in C_2. Let $\hat{c}_1 = P_1 \hat{c}_2$. Then $P_2 \hat{c}_1 = \hat{c}_2$. Let z^0 be arbitrary and, for $n = 0, 1, ...,$ let

$$z^{2n+1} = P_1 z^{2n}, \; z^{2n+2} = P_2 z^{2n+1}. \tag{16}$$

Then

$$\{z^{2n+1}\} \to \hat{c}_1, \; \{z^{2n}\} \to \hat{c}_2. \tag{17}$$

The Proof

We apply the CQ algorithm, with $C = C_2$, $Q = C_1$, and the matrix $A = I$, the identity matrix. The CQ iterative step is now

$$x^{k+1} = P_2(x^k + \gamma(P_1 - I)x^k). \tag{18}$$

Using the acceptable choice of $\gamma = 1$, we have

$$x^{k+1} = P_2 P_1 x^k. \tag{19}$$

This CQ iterative sequence then converges to $\hat{c}_2$, the minimizer of the function $f(x)$. Since $z^{2n} = x^n$, we have $\{z^{2n}\} \to \hat{c}_2$. Because

$$||P_2\hat{c}_1 - \hat{c}_1||_2 \leq ||\hat{c}_2 - \hat{c}_1||_2, \tag{20}$$

it follows from the uniqueness of $\hat{c}_2$ that $P_2\hat{c}_1 = \hat{c}_2$. This completes the proof.

Extending the CQ Algorithm

The CQ algorithm has recently been extended in at least two different directions: approximating the projection operators; and taking the C and Q to be intersections of other convex sets. It can also be viewed as a particular case of forward-backward splitting.

Approximating the Projections

The orthogonal projections P_C and P_Q needed in the iterative step of the CQ algorithm need not be easy to implement. Extensions of the CQ algorithm that incorporate approximations of these projections have been presented by Qu and Xiu [19], Yang [22], and Zhao and Yang [23].

Proximal Minimization

The CQ algorithm is a particular case of an iterative algorithm based on Moreau's notion of proximity operator.

Proximity Operators

The Moreau envelope of a convex function f is the function

$$m_f(z) = \inf_x \{ f(x) + \frac{1}{2}||x - z||_2^2 \}, \qquad (21)$$

which is also the infimal convolution of the functions $f(x)$ and $\frac{1}{2}||x||_2^2$. It can be shown that the infimum is uniquely attained at the point denoted $x = \mathrm{prox}_f z$ (see Rockafellar [20]). The function $m_f(z)$ is differentiable and $\nabla m_f(z) = z - \mathrm{prox}_f z$. The point $x = \mathrm{prox}_f z$ is characterized by the property $z - x \in \partial f(x)$. Consequently, x is a global minimizer of f if and only if $x = \mathrm{prox}_f x$.

The Conjugate Function

The conjugate function associated with f is the function $f^*(x^*) = \sup_x(\langle x^*, x\rangle - f(x))$. In similar fashion, we can define $m_{f^*}z$ and $\mathrm{prox}_{f^*}z$. Both m_f and m_{f^*} are convex and differentiable.

Moreau's Theorem

Theorem

Let f be a closed, proper, convex function with conjugate f^*. Then

$$m_f z + m_{f^*} z = \frac{1}{2}||z||^2;$$

$$\mathrm{prox}_f z + \mathrm{prox}_{f^*} z = z;$$

$$\mathrm{prox}_{f^*} z \in \partial f(\mathrm{prox}_f z);$$

$$\mathrm{prox}_{f^*} z = \nabla m_f(z), \text{ and}$$

$$\mathrm{prox}_f z = \nabla m_{f^*}(z). \tag{22}$$

An Example

For example, consider the indicator function of the convex set C, $f(x) = \iota_C(x)$ that is zero if x is in the closed convex set C and $+\infty$ otherwise. Then $m_f z$ is the minimum of $\frac{1}{2}||x - z||_2^2$ over all x in C, and $\mathrm{prox}_f z = P_C z$, the orthogonal projection of z onto the set C. The operators $\mathrm{prox}_f : z \to \mathrm{prox}_f z$ are proximity operators. These operators generalize the projections onto convex sets, and, like those operators, are firmly non-expansive (see Combettes and Wajs [13]).
The support function of the convex set C is $\sigma_C(x) = \sup_{u \in C}\langle x, u\rangle$. It is easy to see that $\sigma_C = \iota_C^*$. For $f^*(z) = \sigma_C(z)$, we can find $m_{f^*} z$ using Moreau's Theorem:

$$\mathrm{prox}_{\sigma_C} z = z - \mathrm{prox}_{\iota_C} z = z - P_C z. \tag{23}$$

Using Moreau's Theorem

The minimizers of m_f and the minimizers of f are the same. From Moreau's Theorem we know that

$$\nabla m_f(z) = \mathrm{prox}_{f*} z = z - \mathrm{prox}_f z, \tag{24}$$

so $\nabla m_f z = 0$ is equivalent to $z = \mathrm{prox}_f z$.

Proximal Minimization

Because the minimizers of m_f are also minimizers of f, we can find global minimizers of f using gradient descent iterative methods on m_f.

Let x^0 be arbitrary. Then let

$$x^{k+1} = x^k - \gamma_k \nabla m_f(x^k). \tag{25}$$

We know from Moreau's Theorem that

$$\nabla m_f z = \mathrm{prox}_{f*} z = z - \mathrm{prox}_f z, \tag{26}$$

so that Equation (25) can be written as

$$x^{k+1} = x^k - \gamma_k (x^k - \mathrm{prox}_f x^k)$$

$$= (1 - \gamma_k) x^k + \gamma_k \mathrm{prox}_f x^k. \tag{27}$$

It follows from the definition of $\partial f(x^{k+1})$ that $f(x^k) \geq f(x^{k+1})$ for the iteration in Equation (27).

Minimizing $F(x) = f_1(x) + f_2(x)$

In [13] Combettes and Wajs consider the problem of minimizing the function $F(x) = f_1(x) + f_2(x)$, where $f_2(x)$ is differentiable and its gradient is λ-Lipschitz continuous. The function F is minimized at the point x if and only if

$$0 \in \partial F(x) = \partial f_1(x) + \nabla f_2(x), \tag{28}$$

so we have

$$-\gamma \nabla f_2(x) \in \gamma \partial f_1(x), \tag{29}$$

for any $\gamma > 0$. Therefore

$$x - \gamma \nabla f_2(x) - x \in \gamma \partial f_1(x). \tag{30}$$

From Equation (30) we conclude that

$$x = \mathrm{prox}_{\gamma f_1}(x - \gamma \nabla f_2(x)). \tag{31}$$

This suggests an algorithm, called the *forward-backward splitting* for minimizing the function $F(x)$.

Forward-Backward Splitting

Beginning with an arbitrary x^0, and having calculated x^k, we let

$$x^{k+1} = \mathrm{prox}_{\gamma f_1}(x^k - \gamma \nabla f_2(x^k)), \tag{32}$$

with γ chosen to lie in the interval $(0, 2/\lambda)$. The operator $I - \gamma \nabla f_2$ is then averaged. Since the operator $\mathrm{prox}_{\gamma f_1}$ is firmly non-expansive, the sequence $\{x^k\}$ converges to a minimizer of the function $F(x)$, whenever minimizers exist. It is also possible to allow γ to vary with the k.

The CQ Algorithm as Forward-Backward Splitting

Recall that the split-feasibility problem (SFP) is to find x in C with Ax in Q. The CQ algorithm minimizes the function

$$f(x) = ||P_Q Ax - Ax||_2^2, \tag{33}$$

over $x \in C$, whenever such minimizers exist, and so solves the SFP whenever it has solutions. The CQ algorithm therefore minimizes the function

$$F(x) = \iota_C(x) + f(x), \tag{34}$$

where ι_C is the indicator function of the set C. With $f_1(x) = \iota_C(x)$ and $f_2(x) = f(x)$, the function $F(x)$ has the form considered by Combettes and Wajs, and the CQ algorithm becomes a special case of their forward-backward splitting method.

The Multi-set SFP

Recently, Censor, Elfving, Kopf and Bortfeld [11] have extended the CQ algorithm to the case in which the sets C and Q are the intersections of finitely many other convex sets. The new algorithm employs the orthogonal projections onto these other convex sets.

Intensity-modulated Radiation Therapy

In [12] Censor, Bortfeld, Martin, and Trofimov use this new algorithm to determine intensity-modulation protocols for radiation therapy. The issue here is to determine the intensities of the radiation sources external to the patient, subject to constraints on how spatially varying the machinery permits these intensities to be, on the maximum dosage directed to healthy areas, and on the minimum dosage directly to the targets.

The CQ Algorithm in Dynamic ET

The CQ algorithm can be used to reconstruct the time-varying radionuclide distribution within a patient, from emission tomographic scanning data.

Emission Tomography

The objective in ET is to reconstruct the internal spatial distribution of intensity of a radionuclide from counts of photons detected outside the patient. In static ET the intensity distribution is assumed constant over the scanning time. Our data are photon counts at the detectors, forming the positive vector b and we have a matrix A of detection probabilities; our model is $Ax = b$, for x a nonnegative vector representing the radionuclide intensities at each pixel or voxel.

Dynamic ET

In *dynamic* ET (see, for example, the thesis of Farncombe [14]) the intensity levels at each voxel may vary with time. The observation time is subdivided into, say, T intervals and one static image, call it x^t, is associated with the time interval denoted by t, for $t = 1, ..., T$. The vector x is the concatenation of these T image vectors x^t. The discrete time interval at which each data value is collected is also recorded and the problem is to reconstruct this succession of images. Because the data associated with a single time interval is insufficient, by itself, to generate a useful image, one often uses prior information concerning the time history at each fixed voxel to devise a model of the behavior of the intensity levels at each voxel, as functions of time.

Constraining Behavior in Time

One may, for example, assume that the radionuclide intensities at a fixed voxel are increasing with time, or are concave (or convex) with time. The problem then is to find $x \geq 0$ with $Ax = b$ and $Dx \geq 0$, where D is a matrix chosen to describe this additional prior information. For example, we may wish to require that, for each fixed voxel, the intensity is an increasing function of (discrete) time; then we want

$$x_j^{t+1} - x_j^t \geq 0, \tag{35}$$

for each t and each voxel index j.

A Second Example of Constraint

We may wish to require that the intensity at each voxel describes a concave function of time, in which case nonnegative second differences would be imposed:

$$(x_j^{t+1} - x_j^t) - (x_j^{t+2} - x_j^{t+1}) \geq 0. \tag{36}$$

In either case, the matrix D can be selected to include the left sides of these inequalities, while the set Q can include the nonnegative cone as one factor.

Sequential Unconstrained Minimization

The problem is to minimize a function $f(x)$ over the set $C = \overline{D}$ in R^J. We assume that there is $\hat{x}$ in C with $f(\hat{x}) \leq f(x)$, for all x in C. In sequential unconstrained minimization, for $k = 1, 2, ...$, we minimize

$$G_k(x) = f(x) + g_k(x), \tag{37}$$

to get x^k, which we assume lies within the set D. The issue is how to select the auxiliary functions $g_k(x)$.

SUMMA

In SUMMA we minimize

$$G_k(x) = f(x) + g_k(x), \tag{38}$$

to get x^k, which we assume lies within the set D, with the $g_k(x)$ chosen so that

$$0 \leq g_{k+1}(x) \leq G_k(x) - G_k(x^k), \tag{39}$$

for $k = 1, 2, \ldots$.

Theorem

The sequence $\{f(x^k)\}$ converges to $f(\hat{x})$.

Induced Proximal Distance Method

In [3] Auslander and Teboulle consider the sequential unconstrained minimization method whereby, for each $k = 1, 2, ...$, we minimize

$$F_k(x) = f(x) + d(x, x^{k-1}), \tag{40}$$

to get x^k in D. They assume that the distance $d(x, y) \geq 0$ has an associated induced proximal distance $H(x, y)$ satisfying the inequality

$$\langle \nabla_1 d(b, a), c - b \rangle \leq H(c, a) - H(c, b). \tag{41}$$

If $d(x, y)$ is a Bregman distance, then $H(x, y) = d(x, y)$.

Theorem

The sequence $\{f(x^k)\}$ *converges to* $f(\hat{x})$.

If $d(x, y)$ is a Bregman distance, then this method is a particular case of SUMMA, but, in general, the two methods appear to be unrelated.

Bregman-Dominated Distances

Assume that, for each fixed a, the function $g(x) = d(x,a)$ is such that the associated Bregman distance $D_a(c,b)$ can be defined. Then

$$D_a(c,b) = g(c) - g(b) - \langle \nabla g(b), c-b \rangle \geq 0, \tag{42}$$

for all suitable b and c. Therefore,

$$D_a(c,b) = d(c,a) - d(b,a) - \langle \nabla_1 d(b,a), c-b \rangle \geq 0, \tag{43}$$

for all suitable b and c. Say that the distance d is *Bregman-dominated* if

$$D_a(c,b) \geq d(c,b), \tag{44}$$

for all suitable a, b, and c.

If d is Bregman-dominated, then

$$d(c,a) - d(b,a) - \langle \nabla_1 d(b,a), c-b \rangle \geq d(c,b), \tag{45}$$

or

$$\langle \nabla_1 d(b,a), c-b \rangle \leq d(c,a) - d(c,b) - d(b,a) \leq d(c,a) - d(c,b). (4$$

Consequently, the choice of $H = d$ satisfies the inequality in (41), and such a d fits the framework of Auslander and Teboulle.

For each k, let $D_k(x,y) = D_a(x,y)$, for the choice $a = x^{k-1}$. Since x^{k-1} minimizes the function $d(x, x^{k-1})$, we have

$$\nabla_1 d(x^{k-1}, x^{k-1}) = 0,$$

and so

$$D_k(x, x^{k-1}) = d(x, x^{k-1}).$$

Therefore, x^k minimizes the function

$$G_k(x) = f(x) + D_k(x, x^{k-1}).$$

It can be shown that

$$G_k(x) - G_k(x^k) = D_f(x, x^k) + D_k(x, x^k) \geq D_k(x, x^k), \tag{47}$$

assuming, of course, that f is convex.

If the distance d is Bregman-dominated, then we have

$$G_k(x) - G_k(x^k) \geq D_k(x, x^k) \geq d(x, x^k), \tag{48}$$

so the iteration is a particular case of SUMMA.

- Anderson, A. and Kak, A. (1984) "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm" , *Ultrasonic Imaging*, **6** 81–94.
- Aubin, J.-P., (1993) *Optima and Equilibria: An Introduction to Nonlinear Analysis*, Springer-Verlag.
- Auslander, A., and Teboulle, M. (2006) "Interior gradient and proximal methods for convex and conic optimization" *SIAM Journal on Optimization*, **16(3)**, pp. 697–725.
- Baillon, J.-B., Bruck, R.E., and Reich, S. (1978) "On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces" , *Houston Journal of Mathematics*, **4**, pp. 1–9.
- Baillon, J.-B., and Haddad, G. (1977) "Quelques proprietes des operateurs angle-bornes et n-cycliquement monotones" , *Israel J. of Mathematics*, **26** 137-150.

- Bauschke, H. (2001) Projection algorithms: results and open problems, in *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, Butnariu, D., Censor, Y. and Reich, S., editors, Elsevier Publ., pp. 11–22.
- Bauschke, H., and Borwein, J. (1996) On projection algorithms for solving convex feasibility problems, *SIAM Review*, **38 (3)**, pp. 367–426.
- Bertero, M., and Boccacci, P. (1998) *Introduction to Inverse Problems in Imaging* Bristol, UK: Institute of Physics Publishing.
- Byrne, C. (2002) "Iterative oblique projection onto convex sets and the split feasibility problem." *Inverse Problems* **18**, pp. 441–453.

- Byrne, C. (2004) “A unified treatment of some iterative algorithms in signal processing and image reconstruction.” *Inverse Problems* **20**, pp. 103–120.
- Y. Censor, T. Elfving, N. Kopf, and T. Bortfeld. “The Multiple-sets Split Feasibility Problem and its Application for Inverse Problems.” *Inverse Problems* 21 (2005), 2071-2084.
- Y. Censor, T. Bortfeld, B. Martin, and A. Trofimov, A. “A Unified Approach for Inversion Problems in Intensity-modulated Radiation Therapy.” *Physics in Medicine and Biology* 51 (2006), 2353-2365.
- Combettes, P., and Wajs, V. (2005) Signal recovery by proximal forward-backward splitting, *Multiscale Modeling and Simulation*, **4(4)**, pp. 1168–1200.

- Farncombe, T. (2000) “Functional dynamic SPECT imaging using a single slow camera rotation” , *Ph.D. thesis, Dept. of Physics, University of British Columbia*.
- Golshtein, E., and Tretyakov, N. (1996) *Modified Lagrangians and Monotone Maps in Optimization*. New York: John Wiley and Sons, Inc.
- 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.
- Landweber, L. (1951) “An iterative formula for Fredholm integral equations of the first kind.”*Amer. J. of Math.* **73**, pp. 615–624.
- Mann, W. (1953) “Mean value methods in iteration.”*Proc. Amer. Math. Soc.* **4**, pp. 506–510.

- Qu, B., and Xiu, N. (2005) "A note on the CQ algorithm for the split feasibility problem." *Inverse Problems* **21**, pp. 1–11.
- Rockafellar, R. (1970) *Convex Analysis*. Princeton, NJ: Princeton University Press.
- Teboulle, M. (1992) "Entropic proximal mappings with applications to nonlinear programming" *Mathematics of Operations Research*, **17(3)**, pp. 670–690.
- Yang, Q. (2004) "The relaxed CQ algorithm solving the split feasibility problem." *Inverse Problems* **20**, pp. 1261–1266.
- Zhao, J., and Yang, Q. (2005) "Several solution methods for the split feasibility problem." *Inverse Problems* **21**, pp. 1791–1799.