Dykstra's algorithm and robust stopping criteria

Ernesto G. Birgin * Marcos Raydan †

December 28, 2006

Keywords: Convex optimization, alternating projection methods, Dykstra's algorithm, stopping criteria.

MSC2000: 90C25, 65K05, 65G505.

1 Introduction

We consider the application of Dykstra's algorithm for solving the following optimization problem

$$\min_{x \in \Omega} \|x^0 - x\|, \tag{1}$$

where x^0 is a given point, Ω is a closed and convex set, and $\|z\|^2 = \langle z, z\rangle$ defines a real inner product in the space. The solution x^* is called the projection of x^0 onto Ω and is denoted by $P_\Omega(x^0)$. Dykstra's algorithm for solving (1) has been extensively studied since it fits in many different applications (see [5, 21, 22, 23, 27, 28, 29, 32, 34, 41, 42, 45]).

For simplicity, we consider the case

$$\Omega = \cap_{i=1}^{p} \Omega_i, \tag{2}$$

where Ω_i are closed and convex sets in $I\!R^n$, for $i = 1, 2, \ldots, p$, and $\Omega \neq \emptyset$. Moreover, we assume that for any $z \in I\!R^n$ the calculation of $P_\Omega(z)$ is not trivial; whereas, for each Ω_i, $P_{\Omega_i}(z)$ is easy to obtain as in the case of a box, an affine subspace, or

*Department of Computer Science IME-USP, University of São Paulo, Rua do Matão 1010, Cidade Universitária, 05508-090, São Paulo SP, Brazil. This author was supported by PRONEX-Optimization (PRONEX - CNPq / FAPERJ E-26 / 171.164/2003 - APQ1), FAPESP (Grant 06/53768-0) and CNPq (PROSUL 490333/2004-4). e-mail: egbirgin@ime.usp.br

†Departamento de Computación, Facultad de Ciencias, Universidad Central de Venezuela, Ap. 47002, Caracas 1041-A, Venezuela. This author was supported by the Center of Scientific Computing at UCV. e-mail: mraydan@kuaimare.ciens.ucv.ve

a sphere. For the not feasible case (i.e., when $\Omega = \emptyset$) the behavior of Dykstra's algorithm is treated in [2, 6, 37].

Dykstra's alternating projection algorithm is a cyclic scheme for finding asymptotically the projection of a given point onto the intersection of a finite number of closed convex sets. Roughly speaking, it iterates by projecting in a clever way onto each of the convex sets individually. The algorithm was originally proposed by Dykstra [20] for closed and convex cones in the Euclidean space $I\!R^n$, and later extended by Boyle and Dykstra [7] for closed and convex sets in a Hilbert space. It was rediscovered by Han [30] using duality theory, and the linear rate of convergence was established by Deutsch and Hundal [18] for the polyhedral case (see also [19, 43, 44]).

Dykstra's algorithm belongs to the general family of alternating projection methods, that dates back to von Neumann [46] who treated the problem of finding the projection of a given point in a Hilbert space onto the intersection of two closed subspaces. Later, Cheney and Goldstein [15] extended the analysis of von Neumann's alternating projection scheme to the case of two closed and convex sets. In particular, they established convergence under mild assumptions. However, the limit point need not be the closest in the intersection. Therefore, the alternating projection method, proposed by von Neumann, is not useful for problem (1). Fortunately, Dykstra [20] found the clever modification of von Neumann's scheme for which convergence to the solution point is guaranteed. For a complete discussion on alternating projection methods see Deutsch [17].

Dykstra's algorithm has been extended in several different ways. Gaffke and Mathar [24] proposed, via duality, a family of simultaneous Dykstra's algorithm in Hilbert space. Later Iusem and De Pierro [37] established the convergence of the simultaneous version considering also the inconsistent case in the Euclidean space $I\!R^n$. Bauschke and Borwein [2] further analyzed Dykstra's algorithm for two sets, that appears frequently in applications and in particular generalized the results in [37]. In [36] it was established that for linear inequality constraints the method of Dykstra reduces to the method proposed by Hildreth [33] in his pioneer work on dual alternating projections. See also [40] for further analysis and extensions.

Dykstra's algorithm has also been generalized by Deutsch and Hundal [35] to an infinite family of sets, and also to allow a random ordering, instead of cyclic, of the projections onto the closed convex sets. More recently, it has also been generalized by Bregman et al. [9] to avoid the projection onto each one of the convex sets in every cycle. Instead, projections onto either a suitable half space of the intersection of two half spaces are used. Further results concerning the connection between Bregman distances and Dykstra's algorithm can be found in [3, 4, 8, 14]. For the advantages of projecting cyclically onto suitable half spaces, see the previous work by Iusem and Svaiter [38, 39].

A computational experiment comparing Dykstra's algorithm and the Halpern-

Lions-Wittmann-Bauschke algorithm [1] on linear best approximation test problems can be found in [12].

2 Formulations

2.1 Dykstra's algorithm

Dykstra's algorithm solves (1)–(2) by generating two sequences: the iterates $\{x_i^k\}$ and the increments $\{y_i^k\}$. These sequences are defined by the following recursive formulae:

$$\begin{aligned} x_0^k &= x_p^{k-1}, \\ x_i^k &= P_{\Omega_i}(x_{i-1}^k - y_i^{k-1}), \quad i = 1, 2, \ldots, p, \\ y_i^k &= x_i^k - (x_{i-1}^k - y_i^{k-1}), \quad i = 1, 2, \ldots, p, \end{aligned} \tag{3}$$

for $k = 1, 2, \ldots$ with initial values $x_p^0 = x^0$ and $y_i^0 = 0$ for $i = 1, 2, \ldots, p$.

Remarks

1. For the sake of simplicity, the projecting control index $i(k)$ used in (3) is the most common one: $i(k) = k \bmod p + 1$, for all $k \geq 0$. However, more advanced control indices can also be used, as long as they satisfy some minimal theoretical requirements (see e.g., [35]).

2. The increment y_i^{k-1} associated with Ω_i in the previous cycle is always subtracted before projecting onto Ω_i. Only one increment (the last one) for each Ω_i needs to be stored.

3. If Ω_i is a closed affine subspace, then the operator P_{Ω_i} is linear and it is not required, in the k^{th} cycle, to subtract the increment y_i^{k-1} before projecting onto Ω_i. Thus, for affine subspaces, Dykstra's procedure reduces to the alternating projection method of von Neumann [46].

4. For $k = 1, 2, \ldots$ and $i = 1, 2, \ldots, p$, it is clear from (3) that the following relations hold

$$x_p^{k-1} - x_1^k = y_1^{k-1} - y_1^k, \tag{4}$$

$$x_{i-1}^k - x_i^k = y_i^{k-1} - y_i^k, \tag{5}$$

where $x_p^0 = x^0$ and $y_i^0 = 0$, for all $i = 1, 2, \ldots, p$.

For the sake of completeness we now present the key theorem associated with Dykstra's algorithm.

Theorem 2.1 (Boyle and Dykstra, 1986 [7]) *Let $\Omega_1, \ldots, \Omega_p$ be closed and convex sets of $I\!R^n$ such that $\Omega = \cap_{i=1}^{p} \Omega_i \neq \emptyset$. For any $i = 1, 2, \ldots, p$ and any $x^0 \in I\!R^n$, the sequence $\{x_i^k\}$ generated by (3) converges to $x^* = P_\Omega(x^0)$ (i.e., $\|x_i^k - x^*\| \to 0$ as $k \to \infty$).*

We now discuss the delicate issue of stopping Dykstra's algorithm within a certain previously established tolerance that indicates the distance of the current iterate to the unique solution.

2.2 Difficulties with some commonly used stopping criteria.

In some applications it is possible to obtain a *somehow* natural stopping rule, associated with the problem at hand. For example, when solving a linear system, $Ax = b$, by alternating projection methods [10, 25], the residual vector ($r(x) = b - Ax$) is usually available and yields some interesting and robust stopping rules. Another example appears in image reconstruction for which a *good and feasible* image tells the user that it is time to stop the process [13, 16]. Similar circumstances are present in some other specific applications (e.g. saddle point problems [31], and molecular biology [28, 29]).

However, in general, this is not the case, and we are left with the information produced only by the internal computations, i.e., the sequence of iterates and perhaps the sequence of increments, and some inner products. For this general case, a popular stopping rule is to monitor the subsequence of projections onto one particular convex set, Ω_i, and stop the process when the distance, in norm, of two consecutive projections is less than or equal to a previously established tolerance [26, 27, 32, 41].

Another commonly used criterion, that is claimed to improve the previous one (e.g. [7, 22, 28, 45]) is to somehow compute an average of all the projections at each cycle of projections, and then stop the process when the distance, in norm, of two consecutive of those average projections is less than or equal to a previously established tolerance.

Finally, we would like to mention that another criterion, that is also designed to improve any of the two criteria above, is to check any of the previously described rules during N consecutive cycles, where N is a fixed positive integer.

None of these stopping rules is a trustable choice. In [6], Birgin and Raydan presented the example below to establish that they can fail even for a two dimensional problem. (see Figures 1 and 2).

Consider the closed and convex set $\Omega = \Omega_1 \cap \Omega_2$, where $\Omega_1 = \{x \in I\!R^2 \mid x_1 + x_2 \geq 10\}$ is a half space and $\Omega_2 = \{x \in I\!R^2 \mid 3 \leq x_1 \leq 10, 0 \leq x_2 \leq 4\}$ is a box. This closed and convex set in $I\!R^2$ is shown in Figure 1.

Let $x^0 = (-49, 50)^T$ and let us use Dykstra's algorithm to find the closest point to x^0 in Ω. In Figure 2 we can see the first two cycles of this convergent process.

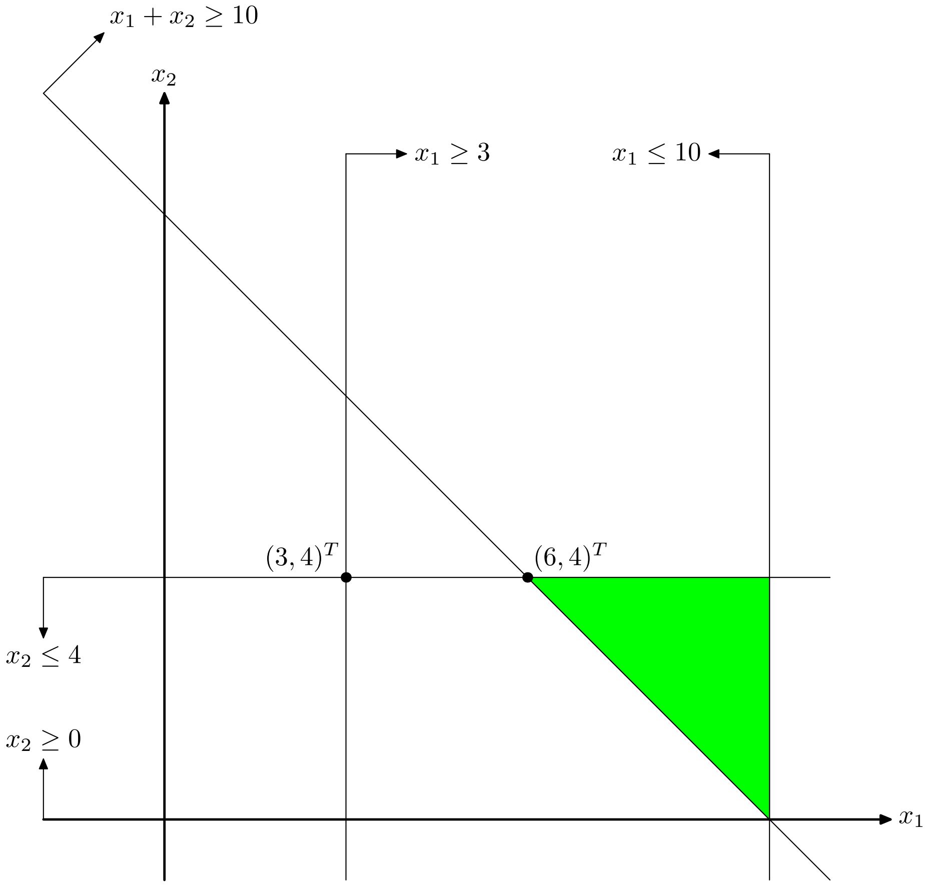

Figure 1: Feasible set $\Omega = \Omega_1 \cap \Omega_2$ in $I\!R^2$.

Since $y_1^0 = y_2^0 = 0$ (null initial increments) then for the first cycle we project x^0 onto Ω_1 to obtain $p_2 = x_1^1 = (-44.5, 54.5)^T$ and then we project p_2 onto Ω_2 to obtain $p_3 = x_2^1 = (3,4)^T$. For the second cycle, the increments are not null ($y_1^1 = (4.5, 4.5)^T$ and $y_2^1 = (47.5, -50.5)^T$), and we start from p_3. First we project $p_4 = p_3 - y_1^1$ onto Ω_1 to obtain $p_5 = x_1^2$. Then we project $p_6 = p_5 - y_2^1$ onto Ω_2 to obtain p_3 again. Hence $x_2^2 = x_2^1$. The increment associated with Ω_2 is large enough to take the iterate back to the quadrant where the projection onto the box is again p_3. As discussed in [6], this phenomenon will occur until cycle 32, i.e., $p_3 = x_2^1 = x_2^2 = \ldots = x_2^{32}$.

Moreover, by choosing x^0 far enough, this misleading event can be repeated for as many cycles as any previously established positive integer N. Eventually the size of the increments will be reduced and convergence to x^* will be observed.

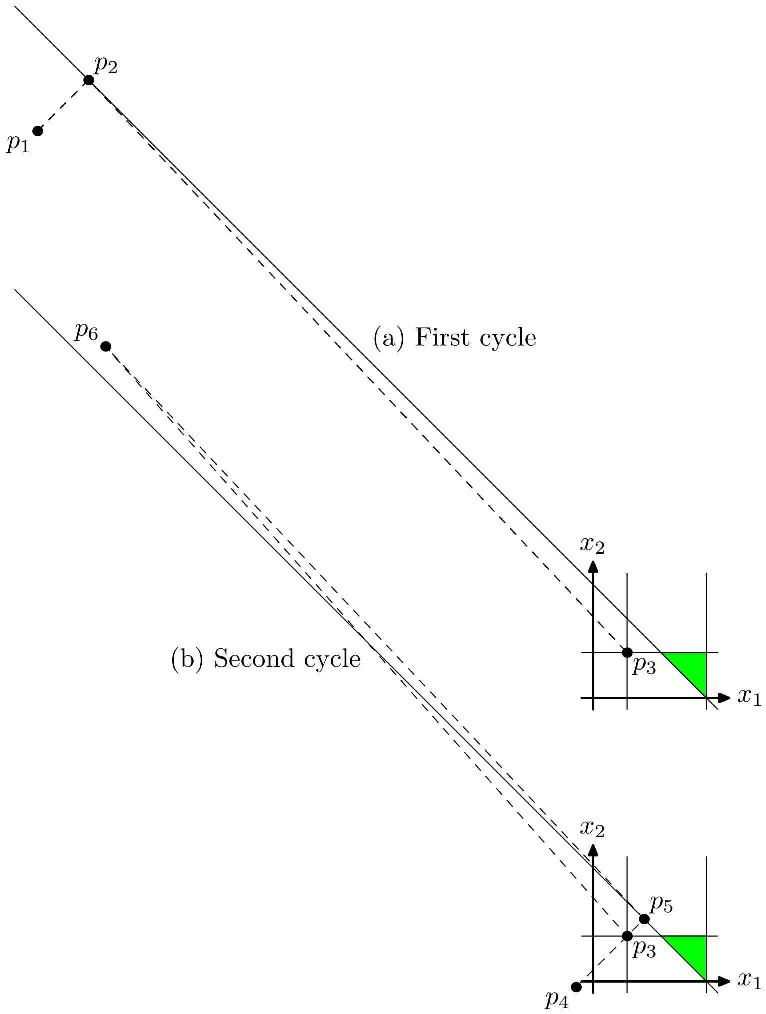

Figure 2: First two cycles of Dykstra's algorithm to find the projection of $x^0 = (-49, 50)^T$ onto $\Omega = \Omega_1 \cap \Omega_2$.

2.3 Robust stopping criteria

After a close inspection of the proof of the Boyle and Dykstra's theorem, Birgin and Raydan [6] proposed some robust stopping criteria for Dykstra's algorithm. For that they first established the following result.

Theorem 2.2 *Let x^0 be any element of $\mathbb{R}^n$. Consider the sequences $\{x_i^k\}$ and $\{y_i^k\}$ generated by (3) and define c^k as*

$$c^k = \sum_{m=1}^{k} \sum_{i=1}^{p} \|y_i^{m-1} - y_i^m\|^2 + 2 \sum_{m=1}^{k-1} \sum_{i=1}^{p} \langle y_i^m, x_i^{m+1} - x_i^m \rangle. \tag{6}$$

Then, in the k^{th} cycle of Dykstra's algorithm,

$$\|x^0 - x^*\|^2 \geq c^k. \tag{7}$$

Moreover, at the limit when k goes to infinity, equality is attained in (7).

Based on the previous theorem, let us now write c^k as follows:

$$c^k = c_L^k + c_S^k,$$

where

$$c_L^k = \sum_{m=1}^{k} c_I^m, \tag{8}$$

$$c_I^m = \sum_{i=1}^{p} \|y_i^{m-1} - y_i^m\|^2 \tag{9}$$

and

$$c_S^k = 2 \sum_{m=1}^{k-1} \sum_{i=1}^{p} \langle y_i^m, x_i^{m+1} - x_i^m \rangle.$$

Both c_L^k and c_S^k are monotonically nondecreasing by definition. Moreover in [6], the following theorem is also established.

Theorem 2.3 *Consider the sequences $\{x_i^k\}$ and $\{y_i^k\}$ generated by (3), and c^k, c_L^k and c_I^k as defined in (6), (8) and (9), respectively. For any $k \in I\!N$, if $x^k \neq x^*$ then $c_I^{k+1} > 0$ and, hence, $c_L^k < c_L^{k+1}$ and $c^k < c^{k+1}$.*

The results established in Theorems 2.2 and 2.3 are combined in [6] to propose robust stopping criteria. Notice that $\{c_L^k\}$ and $\{c^k\}$ are monotonically increasing and convergent, and also that $\{c_I^k\}$ converges to zero. Therefore we can stop the process when

$$c_I^k = \sum_{i=1}^{p} \|y_i^{k-1} - y_i^k\|^2 \leq \varepsilon$$

or, similarly, when

$$c^k - c^{k-1} = c_I^k + 2 \sum_{i=1}^{p} \langle y_i^{k-1}, x_i^k - x_i^{k-1} \rangle \leq \varepsilon, \tag{10}$$

where $\varepsilon > 0$ is a sufficiently small tolerance. As c^k may grow fast, computing $c^k - c^{k-1}$ may give inaccurate results due to loss of accuracy in floating point representation and, hence, cancellation. So, for the criterion in (10), it is recommendable to test convergence with the second expression.

The computation of c_I^k involves the squared-norm $\|y_i^{k-1} - y_i^k\|^2$, for $i = 1, 2, \ldots, p$. By (5), $y_i^k = y_i^{k-1} + v$, where $v = x_i^k - x_{i-1}^k$ is a temporary n-dimensional array needed in the computation of Dykstra's algorithm. So, the computational cost involved in the calculation of c_I^k is just the cost of the extra inner product $\langle v, v\rangle$ at each iteration.

The computation of c^k involves the calculation of c_I^k plus an extra term. The computational of this extra term is also small and involves an inner product and the difference of two vectors per iteration. But, in contrast with the computation of c_I^k which does not require additional savings, the computation of the extra term requires to save p extra n-dimensional arrays (the same amount of memory required in Dykstra's algorithm to save the increments). So, the computation of c^k requires some additional calculations and memory savings, and hence it is more expensive. However, it also has the advantage of revealing the optimal distance: $\|x^0 - x^*\|^2$, that could be of interest in some applications.

We close this section with some comments concerning the behavior of the stopping criteria when the problem is not feasible. In this case ($\Omega = \emptyset$), there is no solution and we know from Theorem 4.2 that the sequences $\{c_L^k\}$ and $\{c^k\}$ are monotonically increasing. Moreover, under some mild assumptions on the sets Ω_i, the sequences $\{x_i^k\}$ converge for $1 \leq i \leq p$, and there exists a real constant $\delta > 0$ such that $\sum_{i=1}^p \|x_{i-1}^k - x_i^k\|^2 \geq \delta$ for all k. A discussion on this topic is presented in [2, Section 6], including a notion of *distance* between all the sets Ω_i (see also [37]). Now using (5), we obtain

$$\sum_{i=1}^{p} \|x_{i-1}^k - x_i^k\|^2 = \sum_{i=1}^{p} \|y_i^{k-1} - y_i^k\|^2 = c_I^k.$$

Therefore, the sequence $\{c_I^k\}$ remains bounded away from zero, whereas $\{c_L^k\}$ and $\{c^k\}$ tend to infinity. Consequently, none of the proposed stopping criteria will be satisfied for any k, as expected.

References

[1] H. H. Bauschke [1996], The approximation of fixed points of compositions of nonexpansive mappings in Hilbert space, *Journal of Mathematical Analysis and Applications* 202, pp. 150–159.

[2] H. H. Bauschke and J.M. Borwein [1994], Dykstra's alternating projection algorithm for two sets, *J. Approx. Theory* 79, pp. 418–443.

[3] H. H. Bauschke and J.M. Borwein [1997], Legendre functions and the method of random Bregman projections, *Journal of Convex Analysis* 4, pp. 27–67.

[4] H. H. Bauschke and A. S. Lewis [2000], Dykstra's algorithm with Bregman projections: a convergence proof, *Optimization* 48, pp. 409–427.

[5] E. G. Birgin, J. M. Martínez and M. Raydan [2003], Inexact spectral gradient method for convex-constrained optimization, *IMA Journal on Numerical Analysis* 23, pp. 539-559.

[6] E. G. Birgin and M. Raydan [2005], Robust stopping criteria for Dykstra's algorithm, *SIAM Journal on Scientific Computing* 26, 1405–1414.

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

[8] L. M. Bregman and Y. Censor and S. Reich [1999], Dykstra's algorithm as the nonlinear extension of Bregman's optimization method, *Journal of Convex Analysis* 6, pp. 319–333.

[9] L. M. Bregman and Y. Censor and S. Reich and Y. Zepkowitz-Malachi [2003], Finding the projection of a point onto the intersection of convex sets via projections onto halfspaces, *Journal of Approximation Theory* 124, pp. 194–218.

[10] R. Bramley and A. Sameh [1992], Row projection methods for large non symmetric linear systems, *J. Sci. Statist. Comp.* 13, pp. 168–193.

[11] L.M. Bregman, Y. Censor and S. Reich [1999], Dykstra's algorithm as the nonlinear extension of Bregman's optimization method, *Journal of Convex Analysis* 6, pp. 319–333.

[12] Y. Censor [2006], Computational acceleration of projection algorithms for the linear best approximation problem, *Linear Algebra and its Applications* 416, pp. 111–123.

[13] Y. Censor and G.T. Herman [1987], On some optimization techniques in image reconstruction from projections, *Applied Numerical Mathematics* 3, pp. 365–391.

[14] Y. Censor and S. Reich [1998], The Dykstra's algorithm with Bregman projections, *Communications in Applied Analysis* 2, pp. 407–419.

[15] W. Cheney and A. Goldstein [1959], Proximity maps for convex sets, *Proc. Amer. Math. Soc.* 10, pp. 448–450.

[16] P.L. Combettes [1996], The convex feasibility problem in image recovery, *Advances in Imaging and Electron Physics* 95, pp. 155–270.

[17] F. Deutsch [2001], *Best Approximation in Inner Product Spaces*, Springer Verlag New York, Inc.

[18] F. Deutsch and H. Hundal [1994], The rate of convergence of Dykstra's cyclic projections algorithm: the polyhedral case, *Numer. Func. Anal. Optim.* 15, pp. 537–565.

[19] F. Deutsch [1995], Dykstra's cyclic projections algorithm: the rate of convergence, *Approximation Theory, Wavelets and Applications*, Kluwer Academic Publishers, Netherlands, Ed. S.P. Singh.

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

[21] M. G. Eberle and M. C. Maciel [2003], Finding the Closest Toeplitz Matrix, *Computational and Applied Mathematics* 22, pp. 1–18.

[22] R. Escalante and M. Raydan [1996], Dykstra's Algorithm for a Constrained Least-Squares Matrix Problem, *Numerical Linear Algebra and Applications* 3, pp. 459–471.

[23] R. Escalante and M. Raydan [1998], On Dykstra's algorithm for constrained least-squares rectangular matrix problems, *Computers and Mathematics with Applications* 35, pp. 73–79.

[24] N. Gaffke and R. Mathar [1986], A cyclic projection algorithm via duality, *Metrika* 36, pp. 29–54.

[25] U. M. Garcia-Palomares [1999], Preconditioning projection methods for solving algebraic linear systems, *Numerical Algorithms* 21, pp. 157–164.

[26] W. Glunt [1995], An alternating projection method for certain linear problems in a Hilbert space, *IMA J. of Numer. Anal.* 15, pp. 291–305.

[27] W. Glunt, T. L. Hayden, S. Hong and J. Wells [1990], An alternating projection algorithm for computing the nearest Euclidean distance matrix, *SIAM J. Matrix Anal. Appl.* 11, pp. 589–600.

[28] W. Glunt, T. L. Hayden and M. Raydan [1993], Molecular conformations from distance matrices, *Journal of Computational Chemistry* 14, pp. 114–120.

[29] W. Glunt, T. L. Hayden and W. M. Liu [1991], The embedding problem for predistance matrices, *Bulletin of Mathematical Biology* 53, pp. 769–796.

[30] S. P. Han [1988], A successive projection method, *Mathematical Programming* 40, pp. 1–14.

[31] L. M. Hernández-Ramos [2005], Alternating oblique projections for coupled linear systems, *Numerical Algorithms* 38, pp. 285–303.

[32] N. Higham [2002], Computing the nearest correlation matrix – a problem from finance, *IMA J. Numer. Anal.* 22, pp. 329–343.

[33] C. Hildreth [1957], A quadratic programming procedure, *Naval Res. Log. Quart.* 4, pp. 79–85.

[34] H. Hu and I. Olkin [1991], A numerical procedure for finding the positive definite matrix closest to a patterned matrix, *Statistics and Probability letters* 12, pp. 511–515.

[35] H. Hundal and F. Deutsch [1997], Two generalizations of Dykstra's cyclic projections algorithm, *Math. Program* 77, pp. 335–355.

[36] A. Iusem and A. De Pierro [1990], On the convergence properties of Hildreth's quadratic programming algorithm, *Math. Prog.* 47, pp. 37–51.

[37] A. Iusem and A. De Pierro [1991], On the convergence of Han's method for convex programming with quadratic objective, *Math. Prog.* 52, pp. 265–284.

[38] A. Iusem and B.F. Svaiter [1995], A row-action method for convex programming, *J. Math. Prog.* 64, pp. 149–171.

[39] A. Iusem and B.F. Svaiter [1995], Primal-dual row-action method for convex programming, *J. Optimization Theory Appl.* 86, pp. 73–112.

[40] A. Lent and Y. Censor [1980], Extensions of Hildreth's row-action method for quadratic programming, *SIAM Journal on Control and Optimization* 18, pp. 444–454.

[41] M. Mendoza, M. Raydan and P. Tarazaga [1998], Computing the nearest diagonally dominant matrix, *Numerical Linear Algebra with Applications* 5, pp. 461–474.

[42] M. Monsalve, J. Moreno, R. Escalante and M. Raydan [2003], Selective alternating projections to find the nearest SSD+ matrix, *Appl. Math. Comput.* 145, pp. 205–220.

[43] P. M. Morillas [2005], Dykstra's algorithm with strategies for projecting onto certain polyhedral cones, *Appl. math. comput.* 167, pp. 635–649

[44] C. Perkins [2002], A Convergence Analysis of Dykstra's Algorithm for Polyhedral Sets, *SIAM J. Numer. Anal.* 40, pp. 792–804.

[45] M. Raydan and P. Tarazaga [2002], Primal and polar approach for computing the symmetric diagonally dominant projection, *Numerical Linear Algebra with Applications* 9, pp. 333–345.

[46] J. von Neumann [1950], Functional operators vol. II. The geometry of orthogonal spaces, *Annals of Mathematical Studies* 22, Princeton University Press. This is a reprint of mimeographed lecture notes first distributed in 1933.