Introduction

Broadly understood, the term alternating direction method refers to a class of computational methods for solving stationary and dynamic operator equations which exploit a decomposition of the operator into simpler operators. It is frequently assumed that the operators are linear and almost always that the decomposition is as a sum. In this report we present an exposition of some of the basic concepts in the existant theory and also present a fully nonlinear alternating direction scheme, which dispenses with both of the above assumptions. Except briefly in §4 we discuss only stationary operator equations in $\mathbb{R}^m$.

In §1 we describe a basic alternating direction scheme, the Peaceman-Rachford method for solving a system of linear equations. This is the simplest and also one of the most effective alternating direction methods. There are a large number of variations of the scheme. We have chosen, however, not to discuss many variants, but rather to present the main ideas in light of the basic method. In §2 we give some fundamental convergence results, and in §3 discuss the rate of convergence and the selection of acceleration parameters. In §4 we present some alternate derivations of the basic method which are useful for the purpose of extending it. In §5 we discuss the extension to a sum of several linear operators and in §6 the extension to a sum of nonlinear operators. Then in §§7-8 we discuss the fully nonlinear case. The results of these sections are new and have not appeared elsewhere.

§1. The Peaceman-Rachford Alternating Direction Method

Alternating direction methods were developed in the mid 1950's by Douglas, Peaceman and Rachford to solve the large linear systems of equations arising from the discretization of elliptic partial differential equations [Douglas, 1955], [Peaceman and Rachford, 1955], [Douglas and Rachford, 1956]. The standard five point discretization of a PDE on a square gives rise to a linear system of N^2 equations where N is roughly the reciprocal of the mesh spacing. This system is sparse: a typical equation involves only five of the N^2 unknowns, namely the solution at the node where the operator is being discretized and at its immediate neighbors to the left and right and above and below. If the unknowns are ordered lexigraphically by rows then the matrix entrants corresponding to the first three of these unknowns occur on the diagonal and sub- and superdiagonals. The other two matrix entrants occur roughly N rows off the diagonal. Thus the matrix has bandwidth about $2N$ and direct solution by elimination requires $O(N^2 \times N) = O(N^3)$ operations.

The idea of Douglas, Peaceman, and Rachford was to exploit the fact that the differential operators with which they were concerned could be written as a sum of two operators, one involving partial derivatives in the horizontal direction, the other derivatives in the vertical direction. (For example, in the simplest case, the Laplace operator is a sum of $\frac{\partial^2}{\partial x^2}$ and $\frac{\partial^2}{\partial y^2}$.) Correspondingly the discretization matrix may be written as a sum $H + V$ where each row of H couples the corresponding unknown only with its neighbors to the left and right and each row of V couples the corresponding unknown only with its neighbors above and below. In a lexigraphic ordering of the nodes by rows H is a tridiagonal matrix, while in a lexigraphic ordering by columns V is tridiagonal.

The Peaceman-Rachford alternating direction algorithm is an iterative algorithm for the solution of the system

$$Hx + Vx = f \tag{1}$$

which alternately solves for the argument of H, while holding the argument of V fixed and vice versa. This suggests the algorithm

1. Choose initial approximation x_0.
2. For $n = 0,1,2,\ldots$ do:
   2.1. Solve for $x_{n+1/2}$

$$Hx_{n+1/2} + Vx_n = f. \tag{A1}$$

   2.2. Solve for x_{n+1}:

$$Hx_{n+1/2} + Vx_{n+1} = f.$$

For the discretized elliptic equations discussed above, both steps 2.1 and 2.2 involve solving a tridiagonal linear system of size N^2 which requires $O(N^2)$ operations. Hence the algorithm requires $O(N^2)$ operations per iteration. Thus if the algorithm converges rapidly, so that relatively few iterations are required to achieve the desired accuracy, it can be much more efficient than direct solution. However, it is immediately seen that algorithm (A1) is <u>not</u> convergent. If $x_{n+1/2}$ satisfies 2.1, then comparing steps 2.1 and 2.2 we see that the solution of 2.2 will be $x_{n+1} = x_n$, and then that $x_{n+3/2} = x_{n+1/2}$. Thus $x_0 = x_1 = x_2 = \ldots$, $x_{1/2} = x_{3/2} = \ldots$, and no convergence occurs. The way out of this difficulty is to introduce an iteration parameter $r_n \in \mathbb{R}$ to enforce in the limit that $x_n = x_{n+1/2}$. Consequently the Peaceman-Rachford method is

1. Choose initial approximation x_0.

2. For $n = 0,1,2,\ldots$ do:

   2.1. Choose r_n.

   2.2. Solve for $x_{n+1/2}$:

(A2)
$$Hx_{n+1/2} + Vx_n + r_n(x_{n+1/2}-x_n) = f.$$

   2.3. Solve for x_{n+1}:

$$Hx_{n+1/2} + Vx_{n+1} + r_n(x_{n+1}-x_{n+1/2}) = f.$$

Note that step 2.2 corresponds to the decomposition

$$(H+r_nI)x + (V-r_nI)x = f,$$

while 2.3 corresponds to

$$(H-r_nI)x + (V+r_nI)x = f,$$

(here I is the identity matrix).

The simplest possibility for the choice of the acceleration parameters is a nonzero constant, $r_0 = r_1 = \ldots = r$. Frequently, however, one will do much better by selecting a reasonably large finite set of parameters and then use these cyclically. It is possible to be more general than here by allowing different values of the acceleration parameters in equations 2.2 and 2.3 [Pearcy, 1962], but this rarely proves useful.

Finally, we remark that algorithm (A2) can be applied to any linear system for which the coefficient matrix decomposes as a sum. It is <u>not</u> necessary that the summands correspond in any sense to lower dimensional operators. (However the method is only justified if the two summand systems

are significantly less expensive to solve than the original system.) Thus the name "alternating direction method" is something of a misnomer and occasionally the term "splitting method" is used instead. We shall follow the more usual terminology.

§2. Convergence

In the applications to differential equations, the matrices H and V are usually positive definite (a matrix M is positive definite iff $Mx \cdot x > 0$ for all nonzero vectors x, or, equivalently, iff the symmetric part of M has positive spectrum). In this case, if the acceleration parameters are chosen to be constant and positive ($r_n = r > 0$), then the Peaceman-Rachford method converges (i.e., $x_n \to \bar{x}$ as $n \to \infty$, with $\bar{x}$ the unique solution of (1)). Since this result is basic we give a proof, following [Varga,1962] and [Kellogg,1963].

Our hypotheses imply that $rI + H$ and $rI + V$ are invertible, so we may solve for $x_{n+1/2}$ in (A2-2.2), substitute in (A2-2.3), and then solve for x_{n+1}. This gives

$$x_{n+1} = (rI+V)^{-1}(rI-H)(rI+H)^{-1}(rI-V)x_n + (rI+V)^{-1}(rI-H)(rI+H)^{-1}f + (rI+V)^{-1}f.$$

Let

$$G = (rI+V)^{-1}(rI-H)(rI+H)^{-1}(rI-V) \tag{2}$$

denote the transition function for this iterative process. Noting that

$$\begin{aligned}(rI+V)^{-1}(rI-H)(rI+H)^{-1}f + (rI+V)^{-1}f &= (rI+V)^{-1}[(rI-H)+(rI+H)](rI+H)^{-1}f \\ &= 2r(rI+V)^{-1}(rI+H)^{-1}f,\end{aligned}$$

we have

$$x_{n+1} = Gx_n + 2r(rI+V)^{-1}(rI+H)^{-1}f. \tag{3}$$

Setting $e_n = x_n - \bar{x}$, the error in the nth iterate, we have

$$e_{n+1} = Ge_n$$

so

$$e_n = G^n e_0, \qquad n = 0,1,2,\ldots .$$

Now

$$G = (rI+V)^{-1}\hat{G}(rI+V)$$

with

$$\hat{G} = (rI-H)(rI+H)^{-1}(rI-V)(rI+V)^{-1},$$

so

$$G^n = (rI+V)^{-1}\hat{G}^n(rI+V),$$

and

$$|e_n| \le \|(rI+V)^{-1}\|\,\|\hat{G}\|^n\,|(rI+V)e_0|.$$

Here $|\cdot|$ denotes the Euclidean vector norm and $\|\cdot\|$ the associated matrix norm:

$$\|M\| = \max_{x\neq 0} \frac{|Mx|}{|x|}.$$

Now

$$\|\hat{G}\| \le \|(rI-H)(rI+H)^{-1}\|\,\|(rI-V)(rI+V)^{-1}\|,$$

and so to complete the proof of convergence it suffices to show that for any positive definite matrix H

$$\|(rI-H)(rI+H)^{-1}\| < 1$$

(we apply the estimate to both H and V). To verify this, set $y = (rI+H)^{-1}x$ to get

$$\|(rI-H)(rI+H)^{-1}\|^2 = \max_{x\neq 0} \frac{|(rI-H)(rI+H)^{-1}x|^2}{|x|^2}$$

$$= \max_{y \neq 0} \frac{|(rI-H)y|^2}{|(rI+H)y|^2}$$

$$= \max_{y \neq 0} \frac{r^2|y|^2 - 2r(Hy \cdot y) + |Hy|^2}{r^2|y|^2 + 2r(Hy \cdot y) + |Hy|^2} < 1.$$

This completes the proof of convergence for fixed acceleration parameter.

For variable acceleration parameter we state without proof some convergence results.

1) If the operators H and V are symmetric, positive definite and commute with each other, then convergence holds for any sequence of acceleration parameters which is bounded from above and below by positive constants. (This follows from the analysis in the next section.)

2) This result remains valid without the assumption of commutativity if the lower bound on the parameters is sufficiently large. [Caspar, 1969, Theorem 2.2.2].

3) The condition on the lower bound of the parameter sequence may be relaxed for certain types of cyclically repeating sequences [Pearcy, 1962].