NORTH-HOLLAND

Random-Walk Interpretations of Classical Iteration Methods

Jonathan Goodman

Courant Institute of Mathematical Sciences
New York University
251 Mercer Street
New York, New York, 10012

and

Neal Madras

Department of Mathematics and Statistics
York University
4700 Keele Street
North York, Ontario M3J 1P3, Canada

Submitted by Moshe Goldberg

ABSTRACT

We give a simple framework for computing relative convergence rates for relaxation methods with discrete Laplace operators (five point or nine point). This gives relations between the convergence rate for Jacobi, point Gauss Seidel, and various block relaxation strategies, essentially by inspection. The framework is a random walk interpretation of Jacobi relaxation that extends to these other relaxation methods.

1. INTRODUCTION

The purpose of this paper is to give simple probabilistic interpretations of classical Jacobi and Gauss-Seidel iteration methods for the discrete Laplacian. This leads to a quick and intuitive way to calculate the relative convergence rates of many point and block relaxation methods. All of these rates could also, in principle, be computed directly using Fourier analysis, but such computations can be very tedious and may not greatly aid one's intuition.

The simplest example is as follows. Corresponding to Jacobi iteration for the discrete Laplacian on a square domain there is a random walk process. The walker takes a random step for each Jacobi iteration. For Gauss-Seidel there

LINEAR ALGEBRA AND ITS APPLICATIONS 216:61–79 (1995)

0024-3795/95/$9.50
SSDI 0024-3795(93)00101-5

is another random walk. Each Gauss-Seidel iteration corresponds to a random number of random steps, the average number of which is 2. This will provide an intuitive explanation for the fact that Gauss-Seidel is twice as fact as Jacobi. The probabilistic reasoning will also explain why red-black Gauss-Seidel has the same convergence rate as ordinary Gauss-Seidel: the expected number of steps per sweep is two in both cases. (Adams and Jordan [1] have given another, very elegant explanation of this fact.) Several other examples are given in the text. We emphasize that many of the comparisons between convergence rates are not exactly true for a finite-size problem. Rather, they are asymptotically true as the size of the problem grows.

All of the convergence-rate calculations here fit into a general framework. A (block) relaxation method relaxes the grid points (or clusters of grid points) in some order. For a two-dimensional grid point (i,j), we define $S(i,j)$ to be the set of points which are relaxed simultaneously with or prior to (i,j). For example, for ordinary Gauss-Seidel, $S(i,j)$ contains all points that are lexicographically less than $S(i,j)$ [in particular, it includes the neighbors $(i-1,j)$ and $(i,j-1)$]. For block Jacobi, $S(i,j)$ contains all points in the same block as (i,j). And for red-black Gauss-Seidel, $S(i,j)$ is empty if $i+j$ is odd (a red point), and it contains only red points if $i+j$ is even (a black point).

Now consider a directed graph that has a directed edge from each point to each of its neighbors. The directed edge from (i,j) to a neighbor (k,ℓ) will be called an end if $(k,\ell) \notin S(i,j)$, that is, if (k,ℓ) is relaxed after (i,j). This definition is motivated by the probabilistic interpretation that is the subject of this paper. The general convergence-rate principle for the five-point Laplacian is essentially as follows. If the fraction of edges that are ends is σ, then the convergence rate of the iterative method is $1/\sigma$ times as fast as Jacobi. For example, in block line Jacobi, the horizontal edges are not ends, while the vertical edges are; so $\sigma = \frac{1}{2}$, and block line Jacobi is twice as fast as point Jacobi. For red-black Jacobi, any edge originating from a black point is an end, while those coming from red points are not, so again $\sigma = \frac{1}{2}$. Also, $\sigma = \frac{1}{2}$ for point Gauss-Seidel (with any ordering), since the edge from (i,j) to its neighbor (k,ℓ) is an end if and only if the reverse edge from (k,ℓ) to (i,j) is not. This confirms the fact that the convergence rate for Gauss-Seidel does not depend on the ordering [7]. (A simple variation is necessary in the case of the nine-point Laplacian, when diagonal edges have different weights; see Section 4.)

The main ideas of this paper require only some elementary probability theory to understand, although the technicalities of the proofs rely on some deeper results (see Appendices). To make the paper self-contained for linear algebraists who are not probabilists, we have included an outline of the probability theory that we require (more details and examples may be found in texts such as [6]). The probabilistic background and basic probabilistic interpretation of the Jacobi method are

given in Section 2. Section 3 discusses the main result for the five-point Laplacian, as outlined above, and has several examples. Some extensions—nine-point Laplacian, underrelaxation—appear in Section 4.

2. PROBABILITY BACKGROUND. BASIC EXAMPLES

We begin with a description of a random walk on the lattice $\mathbb{Z}^2$. For now we say that two lattice points $r_1 = (i_1, j_1)$ and $r_2 = (i_2, j_2)$ are neighbors if $|i_1 - i_2| + |j_1 - j_2| = 1$. Consider a ball that is sitting at some lattice point $r \in \mathbb{Z}^2$ at time $t = 0$. At time $t = 1$ the ball jumps to one of the neighbors of r, with each of the four neighbors equally likely to be chosen (see Figure 1). At each integer time this process is repeated. If the ball is at position s at time $t = n$, then it randomly selects one of the four neighbors of s and jumps there at time $t = n + 1$. The iterates $u_{ij}^{(n)}$ produced by iterative methods will later be expressed as expected values where the random variable is a function of a random walk.

We now review some facts about expected values. If Y is some random quantity whose possible values are $a_1, a_2 \ldots$, then the *expected value* of Y is the weighted average

$$E(Y) = \sum_i a_i \Pr(Y = a_i)$$

[writing $\Pr(B)$ to denote the probability of the even B]. The *law of total probability* often allows us to calculate $E(Y)$ without explicitly summing the above series. This is frequently true when $Y = f(X)$ and X is a random sequence or walk—e.g., $f(X)$

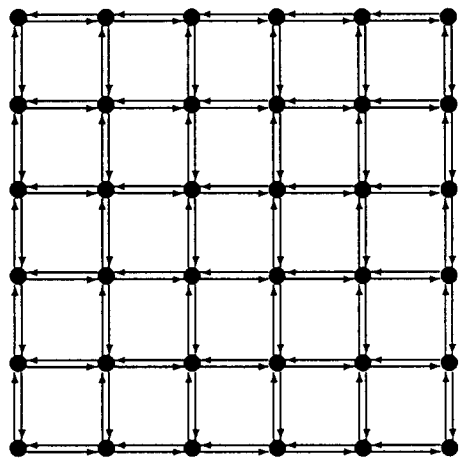

FIG. 1. A part of a grid with arrows indicating possible jumps between grid points.

is related to the first time something happens in X. This is how we will show that expected values related to random walks satisfy the recurrence relations of classical iterative methods. The simplest example of this method is:

EXAMPLE 1. Toss a fair coin until a head appears. Let Y be the number of tosses required. More formally, let $X = (X_1, X_2, \ldots)$, where $X_i = T$ or H, be the random sequence of tails and heads produced. Then $Y = f(X) = \min\{i : X_i = H\}$. We wish to calculate $E(Y)$. If the first toss is a head, then $Y = 1$; this occurs with probability $\frac{1}{2}$. If the first toss is a tail, then we start over again, with one toss already counted, and we expect that we will have to wait $E(Y)$ additional tosses. So

$$E(Y) = \tfrac{1}{2} \cdot 1 + \tfrac{1}{2} \cdot [1 + E(Y)],$$

that is, $E(Y) = 2$.

In this example we used the important principle (the *Markov property*) that the future is completely independent of the past (viz., a tail on the first toss doesn't affect the probabilities of future outcomes). We set up an equation by considering the first toss. In the next example we set up a system of equations by looking at the first step of a random walk.

EXAMPLE 2. Let D be a finite subset of $\mathbb{Z}^2$ (such as a rectangle). Consider a random walk that starts at some $r \in D$, and let R_n be the (random) position of the ball at time n. Let T be the *first exit time from* D:

$$T = \min\{n \geq 0 : R_n \notin D\}.$$

The expected value of T will depend on the starting position $R_0 = r$, so we write $E_r(T)$ to denote the expected exit time of a random walk starting at r. If $r = (i,j)$, we write $u_{ij} = E_r(T)$. Of course, $T = 0$ if $r \notin D$, so

$$u_{ij} = E_r(T) = 0 \qquad \text{if} \quad r \notin D,$$

and $T \geq 1$ if $r \in D$, so

$$u_{ij} = E_r(T) \geq 1 \qquad \text{if} \quad r \in D.$$

The first step of the random walk, R_1, could be any of the four neighbors of (i,j). If $R_1 \notin D$ then $T = 1$. If $R_1 \in D$, then the amount of time we still have to wait has expected value $E_{R_1}(T)$. Since each of the neighbors of r has probability

$\frac{1}{4}$ of being jumped to, we get

$$u_{ij} = E_r(T) = \tfrac{1}{4}\left[1 + E_{(i+1,j)}(T)\right] + \tfrac{1}{4}\left[1 + E_{(i-1,j)}(T)\right] + \tfrac{1}{4}\left[1 + E_{(i,j+1)}(T)\right] + \tfrac{1}{4}\left[1 + E_{(i,j-1)}(T)\right];$$

that is,

$$u_{ij} = 1 + \tfrac{1}{4}(u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1}),$$

with the convention that $u_{ij} = 0$ if $(i,j) \notin D$. If A is the matrix of the five-point Laplace operator on D with Dirichlet boundary conditions and $\vec{1}$ is the vector whose entries are all 1, then the above may be rewritten as

$$Au = \vec{1}.$$

EXAMPLE 3. Suppose we wish to solve a general discrete Poisson equation with Dirichlet boundary conditions:

$$Au = g, \tag{2.1}$$

where $g = \{g_{ij} : (i,j) \in D\}$ is given. We can interpret u using a random-walk construction that is a slight extension of Example 2. First observe that

$$T = \sum_{k=0}^{T-1} 1.$$

Writing analogously

$$U = \sum_{k=0}^{T-1} g(R_k),$$

we shall show that $u_{ij} = E_{(i,j)}(U)$ is the desired solution. We think of U as a cumulative reward. At each site (i,j) that it visits, the random walk collects a reward g_{ij}, until the walk first leaves D. A ball that starts at $r = (i,j)$ in D at time $t = 0$ first collects g_{ij}. If its first jump is to $(i+1,j)$, say, then it expects to collect $E_{(i+1,j)}(U)$ for the rest of the walk. Reasoning as in Example 2 shows that $u_{ij} = E_{(i,j)}(U)$ satisfies (2.1).

A closer look at the last example provides a ranodm-walk interpretation of the Jacobi method. Let $n \wedge T = \min\{n, T\}$. Then

$$u_{ij}^{(n)} = E_{(i,j)}\left[\sum_{k=0}^{(n \wedge T)-1} g(R_k)\right] \tag{2.2}$$

satisfies $u_{ij}^{(0)} = 0$, and we can show that

$$u_{ij}^{(n+1)} = g_{ij} + \tfrac{1}{4}\left(u_{i+1,j}^{(n)} + u_{i-1,j}^{(n)} + u_{i,j+1}^{(n)} + u_{i,j-1}^{(n)}\right) \tag{2.3}$$

if $(i,j) \in D$, with the usual convention that $u_{ij}^{(n)} = 0$ for $(i,j) \notin D$. Let $R_k = R_k(i,j)$ (for $k \geq 0$) be a random walk starting from (i,j), and let

$$U^{(n)}(i,j) = \sum_{k=0}^{(n\wedge T)-1} g(R_k(i,j)).$$

Then, for $r = (i,j) \in D$,

$$\begin{aligned} E(U^{(n+1)}(i,j)) &= g_{ij} + \tfrac{1}{4}E(U^{(n)}(i+1,j)) \\ &\quad + \tfrac{1}{4}E(U^{(n)}(i-1,j)) + \tfrac{1}{4}E(U^{(n)}(i,j+1)) \\ &\quad + \tfrac{1}{4}E(U^{(n)}(i,j-1)) \\ &= g_{ij} + \tfrac{1}{4}\left(u_{i+1,j}^{(n)} + u_{i-1,j}^{(n)} + u_{i,j+1}^{(n)} + u_{i,j-1}^{(n)}\right), \end{aligned}$$

as claimed.

We will now show how to discuss the convergence rate of the Jacobi method using the ranodm-walk interpretation. The key is that the difference between $U^{(n)}(i,j)$ and $U(i,j)$ is related to the number of additional steps beyond n needed to get to the boundary of D:

$$U(i,j) - U^{(n)}(i,j) = \sum_{k=T\wedge n}^{T-1} g(R_k).$$

Averaging over walk yields

$$\begin{aligned} |u(i,j) - u^{(n)}(i,j)| &= \left| E_{(i,j)}\left(\sum_{k=n\wedge T}^{T-1} g(R_k) \right) \right| \\ &\leq \|g\|_\infty E_{(i,j)}((T-n)^+), \end{aligned}$$

where $\|g\|_\infty = \max\{|g_{ij}| : (i,j) \in D\}$ and $a^+ = \max\{a, 0\}$. Now,

$$\begin{aligned} E_{(i,j)}((T-n)^+) &= \sum_{r\in D} \Pr_{(i,j)}\{T > n \text{ and } R_n = r\} E_r(T) \\ &\leq \left(\sum_{r\in D} \Pr_{(i,j)}\{T > n \text{ and } R_n = r\} \right) \cdot \|E_\cdot(T)\|_\infty \\ &= \Pr_{(i,j)}(T > n) \cdot \|E_\cdot(T)\|_\infty, \end{aligned}$$

where $\|E.(T)\|_\infty = \max\{E_{(i,j)}(T) : (i,j) \in D\}$. It can be shown that $\|E.(T)\|_\infty = O(m^2)$, where m is the diameter of D, and that $\Pr_{(i,j)}(T > n)$ decreases exponentially in n. In fact, the decay rate is precisely the first eigenvalue of D with respect to the Laplace operator A. Also, $E_{(i,j)}((T-n)^+) \geq \Pr_{(i,j)}(T > n)$, so the exponential-decay rates of $|u(i,j) - u^{(n)}(i,j)|$ and $\Pr_{(i,j)}(T > n)$ are the same. Thus the convergence rate of the iteration scheme depends on how quickly the random walk tends to escape from D.

3. RELATIVE CONVERGENCE RATES

In this section, we will show how random-walk interpretations allow us to compare convergence rates of different iteration methods for the five-point Laplacian. The nine-point Laplacian and underrelaxation are only slightly more complicated and will be treated in the next section.

We begin by formalizing the relation between the sets $S(r)$, defined in Section 1, and iterative methods. Recall that for each grid point $r = (i,j)$, the set $S(r)$ is the set of simultaneously or previously relaxed points. Thus $S(r) = \{r\}$ for point Jacobi, while for point Gauss-Seidel we have $S(r) = \{r' : r' \leq r\}$. Thus the iterative procedure can be written

$$u_{ij}^{(n+1)} = g_{ij} + \frac{1}{4} \sum_{(i',j') \in S(i,j)} u_{i'j'}^{(n+1)} + \frac{1}{4} \sum_{(i',j') \notin S(i,j)} u_{i'j'}^{(n)}, \tag{3.1}$$

where the sums include only those points (i',j') which are neighbors of (i,j). For Gauss-Seidel, this gives the familiar formulas

$$u_{ij}^{(n+1)} = g_{ij} + \tfrac{1}{4}\left(u_{i-1,j}^{(n+1)} + u_{i,j-1}^{(n+1)} + u_{i+1,j}^{(n)} + u_{i,j+1}^{(n)}\right). \tag{3.2}$$

These formulas can be understood as follows. Suppose that D is a square $N \times N$ grid. Start at the (1,1) grid point, and traverse the grid points in the "lexicographic" order $(1,1),(1,2),\ldots,(1,N),(2,1),\ldots,(N,N)$. At each grid point (i,j), update the variable according to its equation; i.e., average over the current values of the four neighboring variables and add g_{ij}.

Before we discuss the general random-walk interpretation, we will look at the Gauss-Seidel example in some detail. Suppose that a ball starts at some grid point (i,j), and we scan the grid points in reverse order $(N,N),(N,N-1),\ldots,(1,1)$ The ball moves according to the rule: whenever we scan a grid point occupied by the ball, the ball jumps to one of the four neighboring grid points at random. If the ball jumps to a grid point that has already been scanned [either $(i+1,j)$ or $(i,j+1)$], then the ball will not move again; but if the ball jumps to a grid point that has not

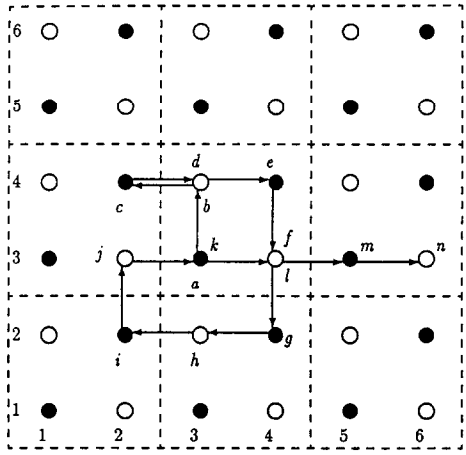

FIG. 2. An example of a random walk. The walk starts at a with coordinates (3, 3), goes up to b, and continues hopping until it finally exists at n.

yet been scanned [either $(i-1,j)$ or $(i,j-1)$], then the ball will make another jump when the scanner reaches that grid point. Of course, the ball stops jumping when it hits the boundary of D. If (i,j) is not close to the boundary, then the expected number of jumps of the ball in one complete scanning sweep is (approximately) two. This is an instance of Example 1: each time the ball jumps it is as likely to go ahead of the scanner as behind it.

Suppose we take repeated scanning sweeps, each sweep causing the ball to jump some random number of times. We will use the term "multijump" to refer to the sequence of jumps made during a single sweep. (See Figure 2 and the top of Table 1 for an example.) Let T_n be the number of jumps that the ball has taken after n sweeps (or equivalently, the number of jumps comprising the first n multijumps). If (i,j) is not near the boundary of D, then $E(T_1) \cong 2, E(T_2) \cong 4$, etc. Gauss-Seidel differs from Jacobi in that we consider the reward accumulated after n sweeps, not after n jumps. That is, Equation (2.2) is replaced by

$$u_{ij}^{(n)} = E_{(i,j)}\left[\sum_{k=0}^{(T_n \wedge T)-1} g(R_k)\right]. \tag{3.3}$$

To show that $u^{(n)}$ as given by (3.3) satisfies (3.2), consider the first jump away from the starting point (i,j) during the first of n sweeps. If the ball jumps ahead of the scanner [say, to $(i-1,j)$], then it will subsequently receive the same treatment as a ball starting at $(i-1,j)$. On the other hand, if the ball jumps behind the scanner [say, to $(i+1,j)$], then it will receive $n-1$ sweeps instead of n. This gives (3.2) (with n replaced by $n-1$).

TABLE 1.
Multijumps for the random walk in figure 2

Gauss-Seidel, Standard ordering	$a \to b$ $b \to c$ $c \to d \to e \to f \to g$ $g \to h$ $h \to i$ $i \to j$ $j \to k \to l \to m \to n$	7 multijumps
Gauss-Seidel, red-black ordering	$a \to b$ $b \to c \to d$ $d \to e \to f$ $f \to g \to h$ $h \to i \to j$ $j \to k \to l$ $l \to m \to n$	7 multijumps
2×2 Block Jacobi	$a \to b \to c$ $c \to d$ $d \to e \to f \to g$ $g \to h \to i$ $i \to j$ $j \to k$ $k \to l \to m \to n$ $l \to m \to n$	8 multijumps
2×2 block Gauss-Seidel, Standard ordering	$a \to b \to c$ $c \to d \to e \to f \to g \to h \to i$ $i \to j$ $j \to k \to l \to m \to n$ $k \to l \to m \to n$ $l \to m \to n$	6 multijumps

For any other iterative method, the sets $S(r)$ lead to a rule for multijumps: continue jumping as long as $R_{t+1} \in (R_t)$ but stop as soon as $R_{t+1} \notin S(R_t)$. Again, let T_n be the time of completion of the nth multijump, and let $u_{ij}^{(n)}$ be given by (3.3). Then reasoning as in the Gauss-Seidel example, we find that (3.1) holds.

Here is the intuition behind the convergence-rate comparison with Jacobi. The important fact is that the sequence of jumps in a random walk induced by another method is indistinguishable from that of the Jacobi method; the only difference appears in grouping some jumps together as multijumps. Let ρ be the expected number of jumps in a multijump (for starting points far from the boundary of D). For instance, $\rho = 1$ for point Jacobi and $\rho = 2$ for point Gauss-Seidel. So, if time is measured by the number of sweeps or multijumps, we expect the walker to reach the boundary "ρ times as fast" as in the single-step (Jacobi) case. We thus expect the convergence to be ρ times as fast as in the Jacobi case. In fact, as we saw in Section 2, the behavior of $|u^{(n)} - u|$ for Jacobi is governed by $\Pr(T > n)$ [which decays like e^{-Cn} for some $C = C(Jacobi) > 0$, depending on D], whereas for other methods

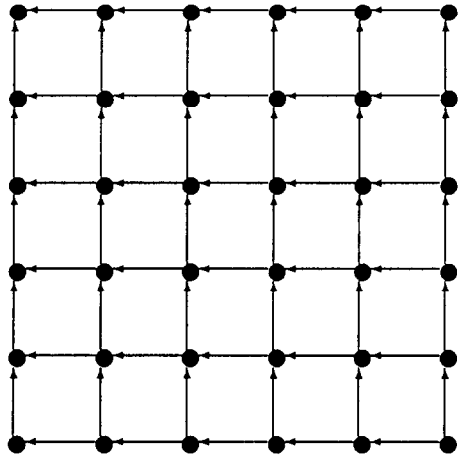

FIG. 3. The jumps that are ends for ordinary Gauss-Seidel. A jump is an end if it goes from a later to an earlier point in the sweep order.

it is governed by $\Pr(T > T_n)$, which approximately equals $\Pr(T > \rho n) = e^{-\rho C n}$. Thus for example we should have $C(\text{Gauss-Seidel}) = 2C(\text{Jacobi})$.

The following theorem makes the above precise and gives a rule for calculating ρ for more general iteration schemes. Given a region D, we consider a sequence of discretized domains $D(m)$. Here $D(m)$ is the set of points of $\mathbb{Z}^2$ inside the scaled region mD (i.e., D magnified by a factor of m). [Equivalently, we could consider a grid of mesh $1/m$, and think of $D(m)$ as the set of grid points inside D.] Let ρ_m be the convergence rate of the given scheme relative to point Jacobi on $D(m)$. Then this theorem says that as $m \to \infty$, ρ_m converges to a number ρ, which we call the (limiting) relative convergence rate of the given scheme to point Jacobi.

We define an edge to be a directed segment from a grid point r to a neighbor r'. The edge is said to be an end if a jump from r to r' would terminate a multijump [i.e., if $r' \notin S(r)$]. (See Figure 3.) Since we are considering a sequence of sets $D(m)$ which increases to $\mathbb{Z}^2$, this definition extends naturally to all edges of the infinite grid $\mathbb{Z}^2$ by saying that the edge from r to r' is an end if and only if the following holds: a jump from r to r' terminates a multijump whenever r and r' are both in $D(m)$. Observe that this definition of "end" depends only on the iteration scheme and not on the set D or $D(m)$.

THEOREM. *Consider an iteration scheme for the five-point Laplacian in which the ends form a periodic pattern on $\mathbb{Z}^2$. Let σ be a density of ends, i.e., the fraction of all edges that are ends. (Precise definitions of "periodic pattern" and "density' are given at the beginning of Appendix B.) Then for any plane region D, the limit $\rho = \lim_{m\to\infty} \rho(m)$ exists and equals σ^{-1}.*

We restrict ourselves here to explaining the basic idea of the proof; for more details, see Appendix B. As explained above, the relative convergence rate is given by ρ, the number of jumps in a "typical" multijump. Using standard methods of probability theory, it can be proven that ρ can be defined (on an infinite grid) by $\lim_{n\to\infty} E(T_n/n)$, and that $\rho = \lim_{n\to\infty} E(T_{n+1} - T_n) = \lim_{n\to\infty} T_n/n$. (The last limit exists with probability one.) After n multijumps (or T_n jumps), n/T_n is the fraction of jumps that have terminated some multijump. Since n/T_n converges to a limit as $n \to \infty$, the limit must be σ, the fraction of possible steps in the lattice that would terminate a multijump. Therefore $\rho = \sigma^{-1}$.

EXAMPLES.

(i) *Point Gauss-Seidel.* Here, half of all edges are ends, so $\sigma = \frac{1}{2}$ and $\rho = 2$. In fact, this should work for any ordering: suppose that the points of D are ordered $r_1, r_2, r_3, \ldots$ and $S(r_i) = \{r_1, r_2, \ldots, r_i\}$ for each i. Then for any pair of neighbors r and r' in D, either $r \in S(r')$ or $r' \in S(r)$, but not both: therefore $\sigma = \frac{1}{2}$, and $\rho = 2$. (Of course, the above theorem as stated only covers "periodic" orderings.)

(ii) *Block line Jacobi.* Each $S(r)$ is a horizontal line: $S(i,j) = \{(i',j) : i' = 1, \ldots, N\}$. Then

$$u_{ij}^{(n+1)} = g_{ij} + \tfrac{1}{4}\left(u_{i-1,j}^{(n+1)} + u_{i+1,j}^{(n+1)} + u_{i,j-1}^{(n)} + u_{i,j+1}^{(n)}\right).$$

Thus the ends are the vertical edges, so $\sigma = \frac{1}{2}$ and $\rho = 2$. (Alternatively, observe that a multijump ends when the vertical jump is made; the probability of a vertical jump at any given step is $\frac{1}{2}$, so the expected number of jumps in a multijump is 2.)

(iii) *Block line Gauss-Seidel.* $S(i,j) = \{(i',j') : j' \leq j\}$. An end is a bond from (i,j) to $(i,j+1)$, so $\sigma = \frac{1}{4}$ and $\rho = 4$.

(iv) $k \times k$ *block Jacobi* (Figure 4). Here, the grid is partitioned into an array of squares, each containing k^2 grid points, and $S(r)$ is the square that contains r. The reader can easily check that $\rho = 1/k$, and so $\rho = k$.

(v) *Red-black Gauss-Seidel* (Figure 5). Here

$$u_{ij}^{(n+1)} = \begin{cases} \frac{1}{4}\left[u_{i-1,j}^{(n)} + u_{i,j-1}^{(n)} + u_{i,j+1}^{(n)} + u_{i+1,j}^{(n)}\right] + g_{ij} & \text{for } i+j \text{ odd,} \\ \frac{1}{4}\left[u_{i-1,j}^{(n+1)} + u_{i,j-1}^{(n+1)} + u_{i,j+1}^{(n+1)} + u_{i+1,j}^{(n+1)}\right] + g_{ij} & \text{for } i+j \text{ even.} \end{cases}$$

If we refer to the points (i,j) for which $i+j$ is odd as "red," and the other points as "black," then one sweep updates all the red values, using the black values from the previous round, and then updates the black values, using the new red values. Using the observation that a random walk on a black point must jump to a red point (and vice versa), it is not hard to see that one sweep corresponds to two

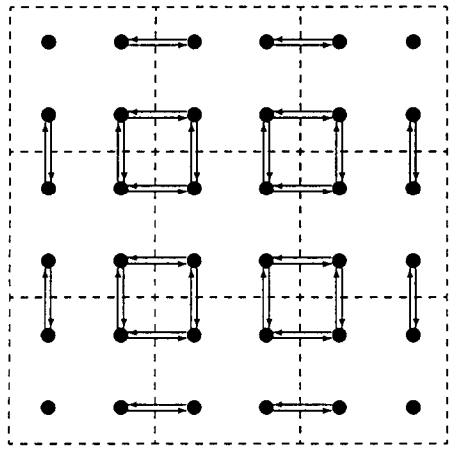

FIG. 4. In 2 × 2 block Jacobi, the jumps from one block to another are ends.

random-walk jumps: black to red and then red to black. In particular,

$$u_{ij}^{(n)} = \begin{cases} E_{(i,j)}\left(\sum_{k=0}^{(2n\wedge T)-1} g(R_k)\right) & \text{if } (i,j) \text{ is black,} \\ E_{(i,j)}\left(\sum_{k=0}^{[(2n-1)\wedge T]-1} g(R_k)\right) & \text{if } (i,j) \text{ is red.} \end{cases}$$

Thus *C(red-black Gauss-Seidel)* = 2*C(Jacobi)*. (This also follows from the theorem, since (i) above is applicable.)

4. VARIATIONS: NINE-POINT LAPLACIAN AND UNDERRELAXATION

Let $\overline{A}$ be the matrix of the nine-point Laplace operator on D:

$$\begin{aligned}(\overline{A}u)_{ij} &= u_{ij} - \tfrac{1}{5}(u_{i-1,j} + u_{i,j-1} + u_{i,j+1} + u_{i+1,j}) \\ &\quad -\tfrac{1}{20}(u_{i-1,j-1} + u_{i-1,j+1} + u_{i+1,j-1} + u_{i+1,j+1}).\end{aligned}$$

We can define a corresponding random walk $\overline{R}_0, \overline{R}_1, \ldots$ on the lattice $\mathbb{Z}^2$. We now say that two distinct lattice points (i_1,j_1) and (i_2,j_2) are neighbors of $|i_1 - i_2| \leq 1$ and $|j_1 - j_2| \leq 1$. If a ball is at point (i_1,j_1) at time $t = n$, then it jumps to one of its eight neighbors at time $t = n + 1$, but the eight neighbors are not all

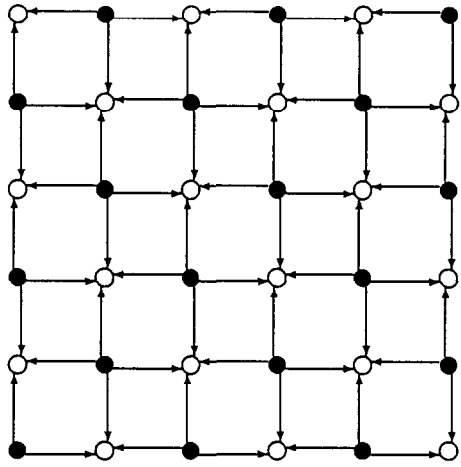

FIG. 5. For red-black Gauss-Seidel, the ends are the jumps that go from black sites to red ones (drawn here as hollow circles).

equally likely to be chosen. It chooses a neighbor (i_2, j_2) with probability $\frac{1}{5}$ if $|i_1 - i_2| + |j_1 - j_2| = 1$, and with probability $\frac{1}{20}$ if $|i_1 - i_2| = 1 = |j_1 - j_2|$.

Much of the preceding discussion of methods for the five-point Laplace operator is easy to extend to the nine-point case. As before, the nth iterate in the Jacobi scheme is the expected accumulated reward up the nth jump, $E_{(i,j)}(\sum_{k=0}^{(n\wedge T)-1} g(\overline{R}_k))$. Also unchanged are the definitions of multijumps and the interpretation of ρ as the relative convergence rate with respect to point Jacobi iteration. The only modification occurs in the definition of σ: when we compute the density of the ends, each edge must be weighted according to the Laplacian weights (i.e., with the corresponding jump probabilities in the random walk). Also, the theorem of Section 3 and its proof still hold.

EXAMPLES.

(i) *Gauss-Seidel.* As in Section 3, we have $\sigma = \frac{1}{2}$ and $\rho = 2$ for any ordering (since the weight of the edge from r to r' equals the weight of the edge from r' to r).

(ii) *Red-black Jacobi.* We have

$$
\begin{aligned}
u_{ij}^{(n+1)} &= u_{ij}^{(n)} - (\overline{A}u^{(n)})_{ij} + g_{ij} \qquad \text{if} \quad (i,j) \text{ is red,} \\
u_{ij}^{(n+1)} &= \tfrac{1}{5}\left(u_{i-1,j}^{(n+1)} + u_{i,j-1}^{(n+1)} + u_{i,j+1}^{(n+1)} + u_{i-1,j}^{(n+1)}\right) \\
&\quad + \tfrac{1}{20}\left(u_{i-1,j-1}^{(n)} + u_{i+1,j-1}^{(n)} + u_{i-1,j+1}^{(n)} + u_{i+1,j+1}^{(n)}\right) \\
&\quad + g_{ij} \qquad\qquad\qquad\qquad \text{if} \quad (i,j) \text{ is black.}
\end{aligned}
$$

Here, $S(r)$ is empty if r is red, and $S(r)$ is the set of all red points if r is black. The ends include all edges emanating from black points (density $\frac{1}{2}$) and all diagonal edges with two red endpoints (density $\frac{1}{2} \times 4 \times \frac{1}{20}$), so $\sigma = \frac{1}{2} + \frac{1}{10} = \frac{3}{5}$ and $\rho = \frac{5}{3}$.

(iii) Red-black Gauss-Seidel:[1]

$$\begin{aligned} u_{ij}^{(n+1)} &= \tfrac{1}{5}\left(u_{i-1,j}^{(n)} + u_{i,j-1}^{(n)} + u_{i,j+1}^{(n)} + u_{i+1,j}^{(n)}\right) \\ &\quad + \tfrac{1}{20}\left(u_{i-1,j-1}^{(n+1)} + u_{i-1,j+1}^{(n+1)} + u_{i+1,j-1}^{(n)} + u_{i+1,j+1}^{(n)}\right) \\ &\quad + g_{ij} \qquad \text{if} \quad (i,j) \text{ is red,} \\ u_{ij}^{(n+1)} &= \tfrac{1}{5}\left(u_{i-1,j}^{(n+1)} + u_{i,j-1}^{(n+1)} + u_{i,j+1}^{(n)} + u_{i+1,j}^{(n)}\right) \\ &\quad + \tfrac{1}{20}\left(u_{i-1,j-1}^{(n+1)} + u_{i-1,j+1}^{(n+1)} + u_{i+1,j-1}^{(n+1)} + u_{i+1,j+1}^{(n+1)}\right) \\ &\quad + g_{ij} \qquad \text{if} \quad (i,j) \text{ is black.} \end{aligned}$$

Here, $\rho = 2$, since this is an example of Gauss-Seidel in which all of the red points are listed first, in lexicographic order, and then all of the black points.

(iv) *Block line Jacobi.* Here, the end are all the diagonal and vertical edges, so $\rho = 2 \cdot \frac{1}{5} + 4 \cdot \frac{1}{20} = \frac{3}{5}$. Similarly, for block line Gauss-Seidel, $\sigma = \frac{3}{10}$. Note that block line Jacobi does less well for the nine-point than for the five-point operator. This is because the random walk is more likely to leave a line in a step of the nine-point walk than in a step of the five-point walk.

Now we will discuss the technique of underrelaxation. Fix a number $\omega, 0 < \omega < 1$:

(1) Jacobi:

$$u_{ij}^{(n+1)} = \omega\left(\tfrac{1}{4}\left[u_{i-1,j}^{(n)} + u_{i,j-1}^{(n)} + u_{i+1,j}^{(n)} + u_{i,j+1}^{(n)}\right] + g_{ij}\right) + (1-\omega)u_{ij}^{(n)}$$

(with an analogous expression for the nine-point operator). Here, our random walker either sits still and does nothing (with probability $1-\omega$), or else takes a single random-walk jump (with probability ω). Let Z_n = [time lapsed from $(n-1)$th to nth jump]-1 = amount of time spent sitting still. Let T_J = number of *jumps* required to exit D; and let T_ω = number of time units (iterations) required. So we expect $T_\omega / T_J \cong 1/\omega$. Observe that $T_\omega = T_J + \sum_{n=1}^{T_J} Z_n$. Now, as in Example 1, it is easy to see that $E(Z_1) = \omega \cdot 0 + (1-\omega)[1 + E(Z_1)]$, so $E(Z_1) = (1-\omega)/\omega$. By the strong law of large numbers, $\sum_{n=1}^{m} Z_n/m$ converges to $E(Z_1)$ as $m \to \infty$,

[1] For the five-point stencil, red-black Jacobi and red-black Gauss-Seidel are the same and are called red-black Gauss-Seidel.

so in particular $\sum_{n=1}^{T_J} Z_n/T_J$ is probably close to $(1-\omega)/\omega$. In fact, in the limit as the mesh size $\to 0$, it can be shown that $T\omega/T_J \to 1+(1-\omega)/\omega = 1/\omega$, and hence $\Pr(T_J > n) \cong \Pr(T_\omega > n/\omega)$, so $C(\omega\text{-}Jacobi) = \omega C(Jacobi)$.

(2) Gauss-Seidel.

$$u_{ij}^{(n+1)} = \omega\left(\tfrac{1}{4}\left[u_{i-1,j}^{(n+1)} + u_{i,j-1}^{(n+1)} + u_{i+1,j}^{(n)} + u_{i,j+1}^{(n)}\right] + g_{ij}\right) + (1-\omega)u_{i,j}^{(n)}.$$

The process here is very similar to the usual Gauss-Seidel setup, except that when the scanner moves to a point occupied by the walker, then the walker jumps with probability ω and sits still with probability $1-\omega$. Therefore if NJ = number of jumps in one sweep, we find that $E(NJ) = (1-\omega)0 + \omega[1 + \frac{1}{2}E(NJ)]$, so $E(NJ) = 2\omega/(2-\omega)$. Therefore

$$C(\omega\text{-Gauss-Seidel}) = \frac{2\omega}{2-\omega} C(\text{Jacobi}).$$

REMARK. The results for underrelaxation can also be obtained by the methods of the main theorem. The key is to add to the graph an edge from each point to itself, having weight $1-\omega$, which is an end for underrelaxation but not for the usual Jacobi.

Finally, we show how to compare the convergence rates of methods for the nine-point operator with methods for the five-point operator. It suffices to treat the respective point Jacobi methods.

Here we must compare the rates at which two different random walks escape from a region D. We consider a large problem in which the distance between nearest neighbors is λ^{-1}, for some large λ. As discussed in Appendix A, the convergence rate for point Jacobi iterations is $\lambda_1 a/\lambda^2$, where λ_1 depends only on D, and a is the expected squared length of a single random-walk step on the unscaled grid:

$$a = \sum_{(i,j)\in\mathbb{Z}^2} (i^2+j^2)\Pr(R_1 - R_0 = (i,j)).$$

Then $a = 1$ for the five-point scheme and $a = \frac{6}{5}$ for the nine-point scheme. We conclude that the convergence rate for nine-point Jacobi iteration in $\frac{6}{5}$ times the rate for five-point Jacobi.

APPENDIX A. THE CONTINUUM LIMIT

Consider a random walk $(R_0, R_1, R_2, \ldots)$ on a d-dimensional grid, where $E(R_{i+1} - R_i) = 0$ and $E(\|R_{i+1} - R_i\|^2) = a$. Since the steps $\{R_{i+1} - R_i : \geq 0\}$ are

all independent, it follows from basic probability that $E(\|R_n - R_0\|^2) = na$. Thus, for any $t \geq 0$, $E(\|R_{[t/a]} - R_0\|^2) = t$. If we reduce the mesh of the grid by a factor of λ, then the random walk $R_0^{(\lambda)}, R_1^{(\lambda)}, \ldots$ satisfies $E(\|R_{i+1}^{(\lambda)} - R_i^{(\lambda)}\|^2) = \lambda^{-2}a$, so $E(\|R_{[t\lambda^2/a]}^{(\lambda)} - R_0^{(\lambda)}\|^2) = t$. Thus if we rescale space so that the distance between neighboring grid points is λ times smaller than originally, and if we take λ^2 times more jumps per unit time, then we expect about the same overall displacement in the same total time by the rescaled random walk as by the original. In fact, in the limit $\lambda \to \infty$ (infinitely fine grid), we get an "infinitesimal random walk": a random process indexed by continuous time, $\{W_t : t \geq 0\}$, with very wiggly but *continuous* paths, and still satisfying $E(\|W_t - W_0\|^2) = t$. For $t > s \geq 0$, the coordinates of the displacement vector $W_t - W_s$ are independent random variables having normal (Gaussian) distributions with expected value 0 and expceted square equal to $(t - s)/d$. This limiting process is called a *Brownian motion* or a *Wiener process* (whence the notation W_t). It is a special case of a "diffusion process." A good introduction to the relations between Brownian motion and the equations of potential theory is [4].

For our purposes, consider a bounded open connected subset $D \subset \mathbb{R}^d$ and let g be a (nice) function on D. To obtain a discrete approximation to the solution u of $-\Delta u = g$, we look at a discrete Poisson equation $Au = g$ on a fine grid. In the body of the paper, we show that different iterative methods (and different A's) correspond to different random walks. Consider a scheme that corresponds to a random walk $R_0, R_1, \ldots$ with $E(\|R_1 - R_0\|^2) = a$. Let T (respectively T_W) be the first exit time from D for the random walk. (respectively for the Wiener process). The convergence rate of this scheme on a grid scaled down by λ is

$$\begin{aligned} -\lim_{n\to\infty} \frac{1}{n} \log \Pr(T > n) &\cong -\lim \frac{1}{n} \log \Pr\left(\frac{T_w \lambda^2}{a} > n\right) \\ &= \lambda_1 a/\lambda^2, \end{aligned}$$

where λ_1 is the first eigenvalue for D [5, p. 126]. (Note that in the context of the present paper, the above approximate equality has not been rigorously proven.) Thus, to compare convergence rates of different schemes, we need only compare the expected squared length of each jump (see Section 4).

APPENDIX B. PROOF OF THE THEOREM

In this appendix, we give the details of the proof of the main theorem, as stated in Section 3, for iteration schemes based on periodic patterns.

First we define *Periodic pattern* and *density*. We say that the set of ends forms a periodic pattern on $\mathbb{Z}^2$ if there is a finite set of edges of $\mathbb{Z}^2, B = \{b_1, \ldots, b_s\}$,

and a pair of vectors v_1 and v_2 such that (1) every edge e in $\mathbb{Z}^2$ is a translate of a unique edge $b(e)$ in B, in the form $e = k_1v_1 + k_2v_2 + b(e)$ for a unique pair of integers k_1 and k_2, and (2) if $b(e) = b(e')$ for two edges e and e', then e is an end if and only if e' is an end. Then the density σ is precisely the fraction of edges in B which are ends. Note that the ends are determined by the iteration scheme only, not by D or $D(m)$.

For example, let e_1, e_2, e_3, and e_4 be the edges from the origin to $(1,0), (-1,0)$, $(0,1)$, and $(0,-1)$ respectively. Then ordinary Gauss-Seidel corresponds to $B = \{e_1, e_2, e_3, e_4\}, v_1 = (1,0), v_2 = (0,1)$, with e_1 and $e_3$ends. Block line Jacobi corresponds to the same B, v_1, and v_2, but now e_3 and e_4 are ends. For $k \times k$ block Jacobi, we can choose $v_1 = (k,0), v_2 = (0,k)$, and B consisting of $4k^2$ edges (k^2 parallel to each e_i).

Let $R_0, R_1, \ldots$ be the associated random walk on all of $\mathbb{Z}^2$. Define $Y_n = b((R_n, R_{n+1}))$. Then $Y_0, Y_1, \ldots$ is a finite Markov chain with state space B. Let $I_n = 1$ if Y_n is an end, 0 otherwise. Observe that $T_{n+1} = \min\{k > T_n : I_k = 1\}$.

We shall first show that given $\varepsilon > 0$, there exists as $M > 0$ such that

$$\Pr\left\{\left|T_n - \frac{n}{\sigma}\right| > n\varepsilon\right\} \leq e^{-nM} \text{ for all sufficiently large } n. \tag{B.1}$$

Let $Q_j = \sum_{k=1}^{j} I_k$. Then the theory of large deviations [3,2] tells us that for every $\varepsilon > 0$ there exists an $L > 0$ such that

$$\lim_{n\to\infty}\left[\Pr\left\{\left|\frac{Q_n}{n} - E(I_1)\right| > \varepsilon\right\}\right]^{1/n} = e^{-L},$$

where $E(I_1)$ is the expectation with respect to the equilibrium measure of the Markov chain, which is uniform on B; and so $E(I_1) = \sigma$. [For the nine-point Laplacian of Section 4, the equilibrium distribution is no longer uniform on B, but it is not hard to check that $E(I_1) = \sigma$ still holds.]

Since $T_n \geq n$ and $Q_{T_n} = n$, we have

$$\begin{aligned}
\Pr\left\{\left|T_n - \frac{n}{\sigma}\right| > n\varepsilon\right\} &\leq \Pr\left\{\text{for some } j \geq n,\ \left|j - \frac{Q_j}{\sigma}\right| > Q_j\varepsilon\right\} \\
&\leq \sum_{j=n}^{\infty} \Pr\left\{\left|\frac{Q_j}{j} - \sigma\right| > \frac{Q_j}{j}\sigma\varepsilon\right\} \\
&\leq \sum_{j=n}^{\infty} \Pr\left\{\left|\frac{Q_j}{j} - \sigma\right| > \frac{\sigma^2\varepsilon}{1+\sigma\varepsilon}\right\}.
\end{aligned}$$

By the discussion of the preceding paragraph, there exists an $M' > 0$ such that

$$\Pr\left\{\left|\frac{Q_j}{j} - \sigma\right| > \frac{\sigma^2\varepsilon}{1+\sigma\varepsilon}\right\} \le e^{-jM'}$$

for all sufficiently large j. The result (B.1) follows.

We shall now prove that $\lim_{m\to\infty} \rho_m = \sigma^{-1}$. Given $\varepsilon > 0$, chosen m large enough so that $\lambda_m(\sigma^{-1} + \varepsilon) < M$, where M is given by (B.1), and λ_m is the first eigenvalue of $D(m)$ [recall $\lambda_m = O(m^{-2})$]. Then

$$\begin{aligned}\Pr\{T > T_n\} &= \Pr\left\{T > T_n, \left|T_n - \frac{n}{\sigma}\right| \le n\varepsilon\right\} \\ &\quad + \Pr\left\{T > T_n, \left|T_n - \frac{n}{\sigma}\right| > n\varepsilon\right\} \\ &\le \Pr\left\{T > \frac{n}{\sigma} - n\varepsilon\right\} + \Pr\left\{\left|T_n - \frac{n}{\sigma}\right| > n\varepsilon\right\}.\end{aligned}$$

Therefore

$$\begin{aligned}\limsup_{n\to\infty} \Pr\{T > T_n\}^{1/n} &\le \max\{\exp[-\lambda_m(\sigma^{-1} - \varepsilon)], \exp(-M)\} \\ &= \exp[-\lambda_m(\sigma^{-1} - \varepsilon)]. \qquad \text{(B.2)}\end{aligned}$$

Also

$$\Pr\{T > T_n\} \ge \Pr\left\{T > \frac{n}{\sigma} + n\varepsilon\right\} - \Pr\left\{\left|T_n - \frac{n}{\sigma}\right| > n\varepsilon\right\}$$

also so

$$\liminf_{n\to\infty} \Pr\{T > T_n\}^{1/n} \ge \exp[-\lambda_m(\sigma^{-1} + \varepsilon)]. \qquad \text{(B.3)}$$

Since λ_m is the convergence rate of point Jacobi on $D(m)$, we find from (B.2) and (B.3) that

$$\sigma^{-1} - \varepsilon \le \rho_m \le \sigma^{-1} + \varepsilon.$$

The desired result now follows.

The authors are grateful to S. R. S. Varadhan for helpful discussions and to a referee whose comments substantially improved the paper's correctness and readability. The first author was partially supported by NSF grant DMS-8501953, by an NSF Presidential Young Investigator award with matching funds from AT&T Bell Laboratories and Sun Microsystems, and by a Sloan Foundation Fellowship. The second author was partially supported by NSF grants DMS-8201599 and

DMS-8504033, and by the Natural Science and Engineering Research Council of Canada.

REFERENCES

1 L. Adams and H. Jordan, Is *SOR* color blind? *SIAM J. Sci. Statist. Comput.* 7:490–506 (1986).
2 L. Boza, Asymptotically optimal tests for finite Markov chains, *Ann. Math. Statist* 42:1992–2007 (1971).
3 R. S. Ellis, Large deviations for a general class of random vectors, *Ann. Probab.* 12:1–12 (1984).
4 R. Hersh and R. J. Griego, Brownian motion and potential theory, *Sci. Amer.* 220, Mar. pp. 66–74 (1969).
5 S. C. Port and C. J. Stone, *Brownian Motion and Classical Potential Theory*, Academic, New York, 1978.
6 S. M. Ross, *Introduction to Probability Models*, Academic, New York, 1980.
7 R. Varga, *Matrix Iterative Analysis*, Prentice Hall, Englewood Cliffs, N.J. 1962.

Received 28 December 1987; final manuscript accepted 22 April 1993