WILEY | Hindawi

Research Article

The Rate of Convergence of the SOR Method in the Positive Semidefinite Case

Achiya Dax (ID)

Hydrological Service, P.O.B. 36118, Jerusalem 91360, Israel

Correspondence should be addressed to Achiya Dax; dax20@water.gov.il

In this paper, we derive upper bounds that characterize the rate of convergence of the SOR method for solving a linear system of the form $Gx = b$, where G is a real symmetric positive semidefinite $n \times n$ matrix. The bounds are given in terms of the condition number of G, which is the ratio $\kappa = \alpha/\beta$, where α is the largest eigenvalue of G and β is the smallest nonzero eigenvalue of G. Let H denote the related iteration matrix. Then, since G has a zero eigenvalue, the spectral radius of H equals 1, and the rate of convergence is determined by the size of η, the largest eigenvalue of H whose modulus differs from 1. The bound has the form $|\eta|^2 \leq 1 - 1/(\kappa c)$, where $c = 2 + \log_2 n$. The main consequence from this bound is that small condition number forces fast convergence while large condition number allows slow convergence.

1. Introduction

The SOR method is one of the basic iterative algorithms for solving a large sparse linear system of the form

$$Gx = b, \tag{1}$$

where $G \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$, and $x \in \mathbb{R}^n$ denotes the vector of unknowns. That is, we need to solve a system of n linear equations with n unknowns. The abbreviation SOR stands for "successive overrelaxation." For a detailed description and discussion of this method, see [1–41]. In this paper, we investigate the SOR rate of convergence in the special case when G is a real symmetric positive semidefinite matrix. This means that G has at least one zero eigenvalue and that the system (1) can be inconsistent. It is also assumed that all the diagonal entries of G are positive. This has two justifications. First, it is not possible to apply the SOR iteration without this assumption. Second, since G is positive semidefinite, a zero diagonal entry implies that the corresponding row and column are null and can be deleted. The last assumption enables us to express G in the form

$$G = D - L - L^T, \tag{2}$$

where L is a strictly lower triangular matrix and D is a positive definite diagonal matrix. The kth SOR iteration, $k = 1, 2, \cdots$, starts $x_{k-1} \in \mathbb{R}^n$ and computes x_k by the rule

$$x_k = H_w x_{k-1} + d_w. \tag{3}$$

The matrix

$$H_w = I - w(D - wL)^{-1}G, \tag{4}$$

is called the SOR iteration matrix,

$$d_w = w(D - wL)^{-1}b, \tag{5}$$

and w is a relaxation parameter that satisfies

$$0 < w < 2. \tag{6}$$

Remark 1. In this paper, we study the asymptotic rate of convergence of the SOR iteration (3). Thus, when using simpler terms like "fast rate of convergence" or "fast convergence," we always refer to asymptotic behavior.

The role of H_w is seen from the following known observation. Assume for a moment that the linear system (1) has a solution $\widehat{x} \in \mathbb{R}^n$. Then, the iteration (3) satisfies

$$x_k - \widehat{x} = (H_w)^k (x_0 - \widehat{x}), \tag{7}$$

where x_0 denotes the starting point. The last equality reveals the importance of $\rho(H_w)$, the spectral radius of H_w. If $\rho(H_w) < 1$, then the sequence $\{x_k\}$ converges toward $\widehat{x}$, and the rate of convergence depends on the size of $\rho(H_w)$. The smaller $\rho(H_w)$ is, the faster is the convergence. Otherwise, when $\rho(H_w) > 1$, the SOR iteration diverges. Yet, when G has a zero eigenvalue, then $\rho(H_w) = 1$ and the situation depends on the consistency of (1). If the SOR iteration attempts to solve an inconsistent system, it diverges. Otherwise, when the system to solve is consistent, it converges. The question discussed in this paper is how the spectral properties of G affect the rate of convergence in this case.

The treatment of the positive semidefinite case becomes easier by noting the relation with Kaczmarz's method. For a detailed description and discussion of Kaczmarz's method, see [4–6, 8, 10, 11, 13, 15, 20, 21, 31, 33, 34, 36] and the references therein. Let r denote the rank of G. Then, $0 < r < n$ and G has $t = n - r$ zero eigenvalues. Moreover, using the spectral decomposition, it is possible to express G in the form

$$G = AA^T, \tag{8}$$

where $A \in \mathbb{R}^{n \times r}$ has r orthogonal columns. Let the sequence $y_k, \ k = 0, 1, 2, \cdots,$ be generated by Kaczmarz's method for solving the linear system

$$Ay = b, \tag{9}$$

where $y \in \mathbb{R}^r$ denotes the vector of unknowns. Then, the following observation is well known, e.g., [4, 8]. If the starting points satisfy

$$y_0 = A^T x_0, \tag{10}$$

then the equalities

$$y_k = A^T x_k, \tag{11}$$

hold for all k. This relation implies that Kaczmarz's method obeys the rule

$$y_k = F_w y_{k-1} + A^T d_w, \tag{12}$$

where

$$F_w = I - wA^T (D - wL)^{-1} A, \tag{13}$$

is the corresponding iteration matrix. The role of F_w will be clarified in the coming discussions.

The literature on the SOR method is huge and includes various types of convergence results. However, many of the early results are derived under certain assumptions on G, such as being an "M-matrix" or "consistently ordered," e.g., [1, 17–19, 35, 37, 41]. One of the first results without such assumptions is Ostrowski's theorem [29, 37], which ensures that $\rho(H_w) < 1$ whenever G is positive definite. Yet, it took about forty years until Oswald [30] lowered the bound on $\rho(H_w)$. See also Axelsson [1] (pp. 241-242) for a similar bound. The current paper extends Oswald's bound to the positive semidefinite case. The difficulty here is that $\rho(H_w) = 1$. So we need to see what dominates the rate of convergence in this case.

The interest in the semidefinite case was initiated in the work of Keller [22]. Then, the surprising behavior of the SOR method when solving inconsistent linear systems was studied in a small number of papers, e.g., [2, 7, 8, 28]. It is shown there that if the linear system (1) is inconsistent then although the SOR sequence $\{x_k\}$ diverges it obeys the rule

$$x_k = \widehat{x}_k + kv, \ \ k = 1, 2, \cdots, \tag{14}$$

where $\{\widehat{x}_k\}$ is a converging sequence and v is a fixed vector that belongs to $\mathrm{Null}(G) = \mathrm{Null}(A^T)$. Otherwise, when the linear system is consistent, $v = 0$ and $x_k = \widehat{x}_k$. This explains why the related Kaczmarz sequence (11) always converges.

The convergence properties of iterative methods for solving consistent positive semidefinite linear systems have attracted the attention of several authors. See, for example, [2, 3, 7, 8, 14, 22–26, 40], and the references therein. In particular, if G is positive semidefinite as above then the SOR iteration matrix, H_w, is known to be "semiconvergent". Let $\lambda_j, j = 1, \cdots, n$, denote the eigenvalues of H_w and assume that they are sorted to satisfy

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|, \tag{15}$$

which implies that $\rho(H_w) = |\lambda_1|$. Then "semiconvergent" means that $\lambda_1 = 1$, and that any eigenvalue of H_w that satisfies $|\lambda_j| = 1$ must equal 1 and has a 1×1 Jordan block. Furthermore, since G has t zero eigenvalues, the eigenvalues of H_w satisfy

$$\begin{aligned} \lambda_j &= 1 \text{ for } j = 1, \cdots, t, \\ 1 &> |\lambda_{t+1}| \geq \cdots \geq |\lambda_n|. \end{aligned} \tag{16}$$

Consequently, the Jordan canonical form of H_w shows that the rate of convergence is determined by the size of $|\lambda_{t+1}|$, which is sometimes called the "convergence factor," e.g., [26]. This situation means that we need an upper bound on $|\lambda_{t+1}|$.

The bound is gained in two stages. First, we show that

$$|\lambda_{t+1}| = \rho(F_w). \tag{17}$$

Then, we establish the inequality

$$(\rho(F_w))^2 \leq \rho(F_w^T F_w), \tag{18}$$

and derive a bound on $\rho(F_w^T F_w)$.

The plan of the paper is as follows. We start by exploring the relations between the eigenvalues of H_w and F_w, showing that the two matrices share all the eigenvalues that differ from 1, which proves (17). Then, we study the relations between the iteration matrices of the symmetric SOR (SSOR) method and the symmetric Kaczmarz method and use these relations to simplify the expression for $F_w^T F_w$. The bound on the spectral radius of this matrix is derived in the third section. The bound has the form

$$\rho\left(F_w^T F_w\right) \le 1 - \frac{1}{(\kappa c)}, \tag{19}$$

where κ denotes the condition number of G and

$$c = 2 + \log_2 n. \tag{20}$$

The condition number is defined as the ratio $\kappa = \alpha/\beta$, where α is the largest eigenvalue of G and β is the smallest nonzero eigenvalue of G. The bound shows that small condition number forces fast rate of convergence, while large condition number allows slow convergence. However, as this is an upper bound, a large condition number does not force slow convergence. Hence, it is worthwhile to have a close look at the reasons behind slow convergence. This issue is discussed in Section 4. It is shown there that small nonzero eigenvalues of G are likely to cause slow asymptotic rate of convergence. Finally, in the last section, we compare our approach with former attempts to derive such bounds.

2. Iteration Matrices and Their Eigenvalues

The assumption that G has positive diagonal entries allows us to make the following simplification. Consider the SOR iteration for solving the normalized system $\tilde{G}x = \tilde{b}$, where $\tilde{G} = D^{-1/2}GD^{-1/2}$ and $\tilde{b} = D^{-1/2}b$. Then, the related iteration matrix is similar to H_w. Thus, when studying the rate of convergence of the SOR method, it is possible to replace (1) with its normalized form. That is, there is no loss of generality in assuming that $D = I$. Hence, from now on, we assume that G has the form

$$G = I - L - L^T, \tag{21}$$

where I denotes the identity matrix and L is a strictly lower triangular matrix. As before, r denotes the rank of G and $r < n$. Consequently, G can be factorized in the form

$$G = AA^T, \tag{22}$$

where the matrix $A \in \mathbb{R}^{n \times r}$ has r orthogonal columns. Moreover, let a_i^T denote the ith row of A, $i = 1, \cdots, n$. Then, (21) implies

$$a_i^T a_i = 1 \text{ for } i = 1, \cdots, n. \tag{23}$$

That is, the rows of A have unit length. The SOR itera-tion splits G in the form

$$G = B_w - C_w, \tag{24}$$

where

$$B_w = \frac{(I - wL)}{w}, \tag{25}$$

and

$$C_w = \frac{\left[(1 - w)I + wL^T\right]}{w}. \tag{26}$$

Recall that w is a given relaxation parameter that satisfies $0 < w < 2$. The kth SOR iteration, $k = 1, 2, \cdots$, starts with x_{k-1} and ends with x_k, which is computed by solving the linear system

$$B_w x = C_w x_{k-1} + b. \tag{27}$$

In other words, x_k is obtained from x_{k-1} by the rule

$$x_k = H_w x_{k-1} + d_w, \tag{28}$$

where

$$H_w = B_w^{-1} C_w, \tag{29}$$

is the related iteration matrix, and

$$d_w = B_w^{-1} b. \tag{30}$$

Observe that (24) enables us to express H_w in the form

$$H_w = I - B_w^{-1} G = I - B_w^{-1} AA^T. \tag{31}$$

Multiplying (28) by A^T and using (11) gives

$$y_k = A^T H_w x_{k-1} + A^T d_w, \tag{32}$$

while substituting $I - B_w^{-1} AA^T$ instead of H_w shows that

$$y_k = \left(I - A^T B_w^{-1} A\right) y_{k-1} + A^T d_w. \tag{33}$$

This means that the iteration matrix of Kaczmarz's method has the form

$$F_w = I - A^T B_w^{-1} A. \tag{34}$$

Note that H_w is an $n \times n$ matrix while F_w is an $r \times r$ matrix. However, as the next theorem shows, these matrices share several eigenvalues.

Theorem 2. *Let λ be a nonzero eigenvalue of the matrix $B_w^{-1} AA^T$; then, λ is also an eigenvalue of the matrix $A^T B_w^{-1} A$. Conversely, let λ be a nonzero eigenvalue of $A^T B_w^{-1} A$; then, λ is also an eigenvalue of $B_w^{-1} AA^T$.*

Proof. Let $u \in \mathbb{R}^n$ be a unit eigenvector of $B_w^{-1}AA^T$ that corresponds to a nonzero eigenvalue λ. Then, the equality

$$B_w^{-1}AA^Tu = \lambda u, \tag{35}$$

implies $A^Tu \neq 0$, and multiplying this equality by A^T gives

$$A^TB_w^{-1}A(A^Tu) = \lambda(A^Tu), \tag{36}$$

which means that λ is an eigenvalue of $A^TB_w^{-1}A$.

The converse direction is proved in a similar way. Let $v \in \mathbb{R}^r$ be a unit eigenvector of $A^TB_w^{-1}A$ that corresponds to a nonzero eigenvalue λ. Then, the equality

$$A^TB_w^{-1}Av = \lambda v, \tag{37}$$

implies $Av \neq 0$,

$$AA^TB_w^{-1}(Av) = \lambda(Av), \tag{38}$$

and

$$B_w^{-1}AA^T(B_w^{-1}Av) = \lambda(B_w^{-1}Av), \tag{39}$$

which means that λ is an eigenvalue of $B_w^{-1}AA^T$. □

We have seen that the eigenvalues of H_w satisfy (16). The next theorem shows that $|\lambda_{t+1}|$ equals the spectral radius of F_w.

Theorem 3. *The eigenvalues of H_w satisfy (16) and (17) with $t = n - r$.*

Proof. Recall that $H_w = I - B_w^{-1}G$ and $F_w = I - A^TB_w^{-1}A$, where the matrix $A^TB_w^{-1}A$ is nonsingular. Now the last theorem implies that the largest eigenvalue of F_w equals the largest eigenvalue of H_w whose modulus differs from 1. □

The bounds which are derived in the next section are using the close relations between the Kaczmarz-SOR method and its symmetric version. The symmetric iteration is combined of two parts. The first one is the usual ("forward") iteration, while the second is a "backward" iteration in which the rows of the linear system are approached in the reverse order. See, for example, [1, 13, 14, 26, 33, 34, 37, 41]. The iteration matrix of the backward SOR method has the form

$$\tilde{H}_w = I - B_w^{-T}G, \tag{40}$$

where B_w^{-T} denotes the matrix $(B_w^{-1})^T = (B_w^T)^{-1}$. This implies that the iteration matrix of the backward Kaczmarz method is

$$\tilde{F}_w = I - A^TB_w^{-T}A = F_w^T. \tag{41}$$

Consequently, the iteration matrix of the symmetric

SOR method (SSOR in brief) has the form

$$\tilde{H}_wH_w = (I - B_w^{-T}G)(I - B_w^{-1}G), \tag{42}$$

while that of the symmetric Kaczmarz method is

$$F_w^TF_w = (I - A^TB_w^{-T}A)(I - A^TB_w^{-1}A). \tag{43}$$

The next assertion expresses these matrices in a useful form.

Theorem 4. *The iteration matrix of the SSOR method has the form*

$$\tilde{H}_wH_w = I - \left(\frac{2}{w} - 1\right)B_w^{-T}B_w^{-1}G, \tag{44}$$

while that of the symmetric Kaczmarz method is

$$F_w^TF_w = I - \left(\frac{2}{w} - 1\right)A^TB_w^{-T}B_w^{-1}A. \tag{45}$$

Proof. The second equality is a direct consequence of the first one, which is derived from the following identities.

$$\begin{aligned}
\tilde{H}_wH_w &= (I - B_w^{-T}G)(I - B_w^{-1}G) \\
&= I - B_w^{-1}G - B_w^{-T}G + B_w^{-T}GB_w^{-1}G \\
&= I - (B_w^{-1} + B_w^{-T} - B_w^{-T}GB_w^{-1})G \\
&= I - (B_w^{-T}B_w^TB_w^{-1} + B_w^{-T}B_wB_w^{-1} - B_w^{-T}GB_w^{-1})G \\
&= I - B_w^{-T}(B_w^T + B_w - G)B_w^{-1}G \\
&= I - B_w^{-T}\left(\left(\frac{2}{w} - 1\right)I\right)B_w^{-1}G \\
&= I - \left(\frac{2}{w} - 1\right)B_w^{-T}B_w^{-1}G.
\end{aligned} \tag{46}$$

□

The importance of the last theorem is that it gives a better insight into the eigenvalues of these matrices. In particular, by following the proof of Theorem 2, we obtain the following conclusions.

Theorem 5. *Let λ be a nonzero eigenvalue of the matrix $B_w^{-T}B_w^{-1}G$; then, λ is also an eigenvalue of the matrix $A^TB_w^{-T}B_w^{-1}A$ and vice versa. Moreover, since the last matrix is symmetric and positive definite, both matrices share the same r positive eigenvalues. The other $n - r$ eigenvalues of $B_w^{-T}B_w^{-1}G$ are zeros.*

Corollary 6. *Let θ denote the smallest eigenvalue of the matrix $A^TB_w^{-T}B_w^{-1}A$. Then*

$$\theta = \min\left\{x^TA^TB_w^{-T}B_w^{-1}Ax \mid x \in \mathbb{R}^r \text{ and}\|x\|_2 = 1\right\} > 0, \tag{47}$$

$$\rho\left(F_w^T F_w\right) = \rho\left(I - \left(\frac{2}{w} - 1\right) A^T B_w^{-T} B_w^{-1} A\right) = 1 - \left(\frac{2}{w} - 1\right)\theta < 1. \tag{48}$$

The next section uses these results to derive upper bounds on $\rho(F_w)$ and $\rho(F_w^T F_w)$.

3. Upper Bounds on the Spectral Radius

Let $T \in \mathbb{R}^{n \times n}$ be an arbitrary square matrix, and let $\|T\|_2$ denote the spectral norm of T. Then, it is well known that $\|T\|_2 = \sigma_1(T)$ where $\sigma_1(T)$ denotes the largest singular value of T. It is also well known that the spectral radius of T cannot exceed its spectral norm. That is, $\rho(T) \le \|T\|_2$ and

$$\rho(T)^2 \le \|T\|_2^2 = \rho\left(T^T T\right). \tag{49}$$

Combining these relations with Corollary 6 yields the following useful observation.

Theorem 7. *The Kaczmarz iteration matrix, F_w, and the symmetric Kaczmarz iteration matrix, $F_w^T F_w$, satisfy the relations*

$$\rho(F_w)^2 \le \rho\left(F_w^T F_w\right) = 1 - \left(\frac{2}{w} - 1\right)\theta. \tag{50}$$

The rest of this section is aimed at deriving an "effective" upper bound on the right hand side of (50). In particular, we are looking for a bound that shows how the condition number of A affects the rate of convergence. The first step is to establish a lower bound on the value of θ.

Let $\beta > 0$ denote the smallest nonzero eigenvalue of G. Then, β is also the smallest eigenvalue of the matrix $A^T A$. Note that the smallest eigenvalue of the matrix $B_w^{-T} B_w^{-1}$ is $1/\|B_w\|_2^2$. Now from (47), we see that

$$\theta \ge \beta/\|B_w\|_2^2 \ge \beta/((I/w) + \|L\|_2)^2, \tag{51}$$

where the last inequality follows from the triangle inequality for the matrix $B_w = I/w - L$. Observe that $\|L\|_2$ is not expected to be much larger than $\|G\|_2$. Indeed, using induction on n, one can verify that

$$\|L\|_2 \le \frac{1}{2}\log_2(2n)\|G\|_2. \tag{52}$$

For a detailed proof of this assertion, see Oswald [30]. Let $\alpha = \|G\|_2$ denote the largest eigenvalue of G. Then, (52) can be rewritten as

$$\|L\|_2 \le \gamma\alpha, \tag{53}$$

where

$$\gamma = \frac{1}{2}\log_2(2n). \tag{54}$$

Combining (51) with (53) gives

$$\theta \ge \frac{\beta}{((1/w) + \gamma\alpha)^2}, \tag{55}$$

and from (50), we obtain that

$$\rho(F_w)^2 \le \rho\left(F_w^T F_w\right) \le 1 - \frac{((2/w) - 1)\beta}{((1/w) + \gamma\alpha)^2}, \tag{56}$$

for any w from the interval $(0, 2)$. A further improvement is gained by noting that the bound function

$$\varphi(w) = 1 - \frac{((2/w) - 1)\beta}{((1/w) + \gamma\alpha)^2}, \tag{57}$$

has a unique minimizer in this interval. Computing the derivative of $\varphi(w)$ and eliminating w' from the equality $\varphi'(w') = 0$ shows that the minimizer lies at the point

$$w' = \frac{1}{1 + \gamma\alpha}. \tag{58}$$

It is also easy to verify that

$$\varphi(w') = 1 - 1/[(\alpha/\beta)(1/\alpha + 2\gamma)], \tag{59}$$

while the assumption that A has unit rows implies $\alpha \ge 1$ and

$$\varphi(w') \le 1 - 1/[(\alpha/\beta)(1 + 2\gamma)]. \tag{60}$$

In other words, the spectral radii of $F_{w'}$ and $F_{w'}^T F_{w'}$ satisfy the inequalities

$$\rho(F_{w'})^2 \le \rho\left(F_{w'}^T F_{w'}\right) \le 1 - 1/(\kappa c), \tag{61}$$

where

$$c = 1 + 2\gamma = 2 + \log_2 n, \tag{62}$$

and $\kappa = \alpha/\beta$ denotes the condition number of G. The bound on $\rho(F_{w'})$ can be simplified by using the inequality

$$1 - \frac{1}{(\kappa c)} \le [1 - 1/(2\kappa c)]^2, \tag{63}$$

which shows that

$$\rho(F_{w'}) \le 1 - 1/(2\kappa c). \tag{64}$$

Let $w'' \in (0, 2)$ be a value of w for which $\rho(F_w^T F_w)$ attains its smallest value. That is, w'' is an optimal relaxation parameter for the symmetric Kaczmarz method. Then, clearly,

$$\rho\left(F_{w''}^T F_{w''}\right) \le \rho\left(F_{w'}^T F_{w'}\right), \tag{65}$$

and the inequalities (61) and (64) remain valid when w'' replaces w'. Similarly, let w_{opt} denote the optimal relaxation parameter for Kaczmarz's method. Then, the inequality

$$\rho\left(F_{w_{\text{opt}}}\right) \leq \rho(F_{w'}), \tag{66}$$

implies

$$\rho\left(F_{w_{\text{opt}}}\right)^2 \leq 1 - \frac{1}{(kc)}, \tag{67}$$

and

$$\rho\left(F_{w_{\text{opt}}}\right) \leq 1 - \frac{1}{(2kc)}. \tag{68}$$

Since α is often considerably larger than 1, the point $w' = 1/(1 + \gamma\alpha)$ is often much smaller than 1. On the other hand, in many cases, w_{opt} is larger than 1, and the function $\psi(w) = \rho(F_w)$ is decreasing in the interval $[w', w_{\text{opt}}]$, which implies that the bound

$$\rho(F_w) \leq 1 - \frac{1}{(2kc)}, \tag{69}$$

is likely to hold for all $w \in [w', w_{\text{opt}}]$, including $w = 1$.

The main consequence from these bounds is that small condition number forces fast rate of convergence, while large condition number allows slow convergence. Yet the bounds are not tight in the sense that the actual rate of convergence is often considerably faster than the implied rate. This behavior is due to a number of reasons. First, in many symmetric positive semidefinite matrices, the ratio $\|L\|_2/\|G\|_2$ is considerably smaller than γ. Second, as noted above, w' is expected to be considerably smaller than 1, so the rate of convergence for w_{opt} (or $w = 1$) is expected to be much faster. Third, let $P \in \mathbb{R}^{n \times n}$ be an arbitrary permutation matrix and consider the SOR method for solving the linear system

$$\left(PGP^T\right)z = Pb. \tag{70}$$

Then, since the iteration matrix of PGP^T is not necessarily similar to that of G, we might get a different rate of convergence, e.g., [27, 32, 37]. On the other hand, since PGP^T has the same eigenvalues as G, both matrices share the same upper bound. This shows that the bound holds for the worst possible ordering.

Finally, we note that the above treatment of the positive semidefinite case is easily adapted to the positive definite case. In the latter case, A is an $n \times n$ invertible matrix, and F_w is similar to H_w, so the bounds on $\rho(F_w)$ apply to $\rho(H_w)$.

4. Slow Rate of Convergence

The bounds derived in the former section indicate that slow rate of convergence is possible only when G has a large condition number. On the other hand, the assumption that G has unit diagonal implies that the largest eigenvalue of G satisfies $1 \leq \alpha \leq n$. Consequently, a large condition number occurs whenever G has small positive eigenvalues. These observations raise the question of whether small positive eigenvalues are the reason behind slow convergence. Indeed, as explained below, a small positive eigenvalue may cause slow rate of convergence. The first two lemmas provide the tools for proving this claim.

Lemma 8. *Define $\tau = (2 - w)/w$. Then, G satisfies the equality*

$$G = H_w^T G H_w + \tau \left(B_w^{-1} G\right)^T \left(B_w^{-1} G\right). \tag{71}$$

Proof. The definition of B_w implies that $B_w = (1/w)I - L$ and

$$B_w^T + B_w - G = \tau I. \tag{72}$$

Now the equality $H_w = I - B_w^{-1}G$ shows that

$$\begin{aligned} G - H_w^T G H_w &= \left(B_w^{-1}G\right)^T G + G B_w^{-1} G - \left(B_w^{-1}G\right)^T G \left(B_w^{-1}G\right) \\ &= \left(B_w^{-1}G\right)^T B_w \left(B_w^{-1}G\right) + \left(B_w^{-1}G\right)^T B_w^T \left(B_w^{-1}G\right) \\ &\quad - \left(B_w^{-1}G\right)^T G \left(B_w^{-1}G\right) \\ &= \left(B_w^{-1}G\right)^T \left[B_w + B_w^T - G\right] \left(B_w^{-1}G\right) \\ &= \tau \left(B_w^{-1}G\right)^T \left(B_w^{-1}G\right). \end{aligned} \tag{73}$$

□

Lemma 9. *Let β denote a nonzero eigenvalue of G and let v denote the corresponding unit eigenvector. That is, $Gv = \beta v$, $v^T v = 1$, and $\beta > 0$. Then*

$$\left\|B_w^{-1}v\right\|_2 \leq 1/(\beta\tau)^{1/2}. \tag{74}$$

Proof. Using (71) and the fact that the matrix $H_w^T G H_w$ is positive semidefinite, we obtain the inequalities

$$\beta = v^T G v \geq \tau v^T G B_w^{-T} B_w^{-1} G v = \tau \beta^2 \|B_w^{-1}v\|_2^2, \tag{75}$$

and

$$\|B_w^{-1}v\|_2^2 \leq \frac{1}{(\tau\beta)}. \tag{76}$$

□

Theorem 10. *Let v be an eigenvector of G as above, and let u and η be defined by the equalities*

$$u = \frac{B_w^{-1} v}{\|B_w^{-1} v\|_2}, \tag{77}$$

$$\eta = \beta \|B_w^{-1} v\|_2. $$

Then

$$H_w v = v - \eta u, \tag{78}$$

where η satisfies

$$\eta \leq \left(\frac{\beta}{\tau}\right)^{1/2}, \tag{79}$$

and

$$\|H_w v\|_2 \geq 1 - \eta. \tag{80}$$

Proof. A further use of the equality $H_w = I - B_w^{-1} G$ gives

$$H_w v = v - B_w^{-1} G v = v - \beta B_w^{-1} v = v - \eta u, \tag{81}$$

while from (74) we see that

$$\eta = \beta \|B_w^{-1} v\| \leq \beta / (\beta \tau)^{1/2} = (\beta / \tau)^{1/2}. \tag{82}$$

Finally, the inequality

$$\|v - \eta u\|_2 \geq \|v\|_2 - \|\eta u\|_2 = 1 - \eta, \tag{83}$$

proves (80). □

One consequence of (79) is that small β implies small η, while (80) shows that for small η the error component in the direction of v decays slowly. That is, small β leads to slow error decay. Another consequence is that small τ (which means w close to 2) may compensate the slowing effect of small β.

5. Concluding Remarks

The SOR method and Kaczmarz's method have been intensively studied for many years. Thus, naturally, some of the mentioned results can be found elsewhere in different forms. In particular, the Kaczmarz iteration matrix (34) and the symmetric Kaczmarz iteration matrix (45) both easily come out as special cases of a more general iteration, see [13] (prop. 4 and 10). Also, the relation between the spectral radius of Kaczmarz and symmetric Kaczmarz (first part of Theorem 7) is already observed in [13] (§4).

Estimates of the rate of convergence in the semidefinite case are derived in a series of papers by Lee et al. [23, 24, 40]. However, these estimates have a different flavor, as they are not using the eigenvalues of G or its condition number.

More recently Oswald and Zhou [31] have used the concept of stable Hilbert splittings to develop a unified approach for studying the convergence of multiplicative Schwartz methods. This approach was used in [31] to derive upper bounds on the rate of convergence of Kaczmarz's method, and later in [32], it was modified to bound the SOR convergence in the semidefinite case.

The current treatment of the semidefinite case is quite different. It is based on direct arguments from linear algebra, such as the Jordan canonical form of H_w and the relations between the eigenvalues of H_w and F_w. This simplifies the proof and adds important insight into the semidefinite case.

The upper bounds on the convergence factor explain why small condition number ensures fast convergence. Another related question is whether and why large condition number leads to slow convergence. The analysis in Section 4 provides a convincing explanation. Yet, as this is the first attempt to resolve this enigma, there may be further ways to answer this question.

The relation between the condition number and the rate of convergence stands behind the "Kaczmarz anomaly" phenomenon [10, 11]. The Kaczmarz-SOR method is often considered as a prototype of more sophisticated methods from the families of Row-action methods [5, 6, 9], Projection methods [6, 33, 36], Column-action methods [8, 12, 38], and Coordinate-descent algorithms [39]. This suggests that other members of these families may share similar asymptotic behavior. Examples that illustrate this connection are described in [9].

Data Availability

No data were used to support this study.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, 1994.

[2] A. Berman and M. Neumann, "Consistency and splittings," *SIAM Journal on Numerical Analysis*, vol. 13, pp. 877–888, 1976.

[3] A. Berman and R. Plemmons, *Nonnegative Matrices in Mathematical Science, Academic Press, New York, 1979*, SIAM, Philadelphis, 1994.

[4] Å. Björck and T. Elfving, "Accelerated projection methods for computing pseudoinverse solutions of systems of linear equations," *BIT*, vol. 19, no. 2, pp. 145–163, 1979.

[5] Y. Censor, "Row-action methods for huge and sparse systems and their applications," *SIAM Review*, vol. 23, pp. 444–466, 1981.

[6] Y. Censor and S. Zenios, *Parallel Optimization: Theory, Algorithms and Applications*, Oxford Univ, Press, 1997.

[7] A. Dax, "A note on the convergence of linear stationary iterative processes," *Linear Algebra and its Applications*, vol. 129, pp. 131–142, 1990.

[8] A. Dax, "The convergence of linear stationary iterative processes for solving singular unstructured systems of linear equations," *SIAM Review*, vol. 32, pp. 611–635, 1990.

[9] A. Dax, "The adventures of a simple algorithm," *Linear algebra and its applications*, vol. 361, pp. 41–61, 2003.

[10] A. Dax, *Kaczmarz Anomaly: A Surprising Feature of Kaczmarz Method*, Hydrological Service of Israel, 2021.

[11] A. Dax, *Kaczmarz Anomaly in Tomography Problems*, Hydrological Service of Israel, 2021.

[12] T. Elfving, P. C. Hansen, and T. Nikazad, "Convergence analysis for column-action methods in image reconstruction," *Numerical Algorithms*, vol. 74, no. 3, pp. 905–924, 2017.

[13] T. Elfving and T. Nikazad, "Properties of a class of block-iterative methods," *Inverse Problems*, vol. 25, p. 115011, 2009, (13pp).

[14] A. Frommer, R. Nabben, and D. Szyld, "Convergence of stationary iterative methods for Hermitian semidefinite linear systems and applications to Schwarz methods," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 2, pp. 925–938, 2008.

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

[16] A. Greenbaum, *Iterative Methods for Solving Linear Systems*, SIAM, 1997.

[17] W. Hackbusch, *Iterative Solution of Large Sparse Systems of Equations*, Springer, 1994.

[18] A. Hadjidimos, "Successive overrelaxation (SOR) and related methods," *Journal of Computational and Applied Mathematics*, vol. 123, pp. 177–199, 2000.

[19] L. A. Hageman and D. M. Young, *Applied Iterative Methods*, Academic Press, New York, 1981.

[20] G. T. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*, Academic Press, New York, NY, 1980.

[21] S. Kaczmarz, "Angenäherte auflösung von systemen linearer gleichungen," *Bulletin de l'Académie Polonaise des Sciences et Lettres, A35*, pp. 355–357, 1937.

[22] H. B. Keller, "On the solution of singular and semidefinite linear systems by iteration," *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis*, vol. 2, no. 2, pp. 281–290, 1965.

[23] Y.-J. Lee, J. Wu, J. Xu, and L. Zikatanov, "On the convergence of iterative methods for semidefinite linear systems," *SIAM journal on matrix analysis and applications*, vol. 28, no. 3, pp. 634–641, 2006.

[24] Y.-J. Lee, J. Wu, J. Xu, and L. Zikatanov, "A sharp convergence estimate for the method of subspace corrections for singular systems of equations," *Mathematics of Computation*, vol. 77, no. 262, pp. 831–851, 2008.

[25] L. Lin, Y. Wei, C. Woo, and J. Zhou, "On the convergence of splittings for semidefinite linear systems," *Linear Algebra and its Applications*, vol. 429, no. 10, pp. 2555–2566, 2008.

[26] I. Marek and D. B. Szyld, "Comparison theorems for the convergence factor of iterative methods for singular matrices," *Linear Algebra and its Applications*, vol. 316, pp. 67–87, 2000.

[27] E. Natterer, *The Mathematics of Computerized Tomography*, Classics in Applied Mathematics, SIAM, Philadelphia, PA, 2001.

[28] M. J. O'Carroll, "Inconsistencies and S.O.R. convergence for the discrete Neumann problem," *Journal of the Institute of Mathematics and its Applications*, vol. 11, no. 3, pp. 343–350, 1973.

[29] A. M. Ostrowski, "On the linear iteration procedures for symmetric matrices," *Rendiconti Lincei - Matematica e Applicazioni*, vol. 14, pp. 140–163, 1954.

[30] P. Oswald, "On the convergence rate of SOR: a worst case estimate," *Computing*, vol. 52, pp. 245–255, 1994.

[31] P. Oswald and W. Zhou, "Convergence analysis for Kaczmarz-type methods in a Hilbert space framework," *Linear Algebra and its Applications*, vol. 478, pp. 131–161, 2015.

[32] P. Oswald and W. Zhou, "Random reordering in SOR-type methods," *Numerische Mathematik*, vol. 135, pp. 1207–1220, 2017.

[33] C. Popa, *Projection Algorithms - Classical Results and Developments*, Lambert Academic Publishing, Germany, Applications to image reconstruction, 2012.

[34] C. Popa, "Convergence rates for Kaczmarz-type algorithms," *Numerical Algorithms*, vol. 79, pp. 1–17, 2018.

[35] Y. Saad, *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, 2nd edition, 2003.

[36] K. Tanabe, "Projection method for solving a singular system of linear equations and its applications," *Numerische Mathematik*, vol. 17, pp. 203–214, 1971.

[37] R. S. Varga, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1962.

[38] D. W. Watt, "Column-relaxed algebraic reconstruction algorithm for tomography with noisy data," *Applied Optics*, vol. 33, pp. 4420–4427, 1994.

[39] S. J. Wright, "Coordinate descent algorithms," *Mathematical Programming*, vol. 151, pp. 3–34, 2015.

[40] J. Wu, Y.-J. Lee, J. Xu, and L. Zikatanov, "Convergence analysis on iterative methods for semidefinite systems," *Journal of Computational Mathematics*, vol. 26, pp. 797–815, 2008.

[41] D. M. Young, *Iterative Solution of Large Linear Systems*, Academic Press, New York, 1971.