Linear stationary iterative methods

Nicholas Hu · Last updated on 2025-03-22

Let $A \in \mathbb{R}^{n \times n}$ be invertible and $b \in \mathbb{R}^n$. An **iterative method** for solving $Ax = b$ generates a sequence of iterates $(x^{(k)})_{k=1}^{\infty}$ approximating the exact solution x^*, given an initial guess $x^{(0)}$. We can express this as $x^{(k+1)} := \phi_{k+1}(x^{(0)}, \dots, x^{(k)}; A, b)$ for some functions ϕ_k; if these functions are eventually independent of k, the method is said to be **stationary**.

We will consider stationary iterative methods of the form $x^{(k+1)} := Gx^{(k)} + f$, called **(first-degree) linear stationary iterative methods**. Such methods are typically derived from a **splitting** $A = M - N$, where M is an invertible matrix that "approximates" A but is easier to solve linear systems with. Specifically, since $Mx^* = Nx^* + b$, we take $G := M^{-1}N = I - M^{-1}A$ and $f = M^{-1}b$ so that the exact solution is a fixed point of the iteration. Equivalently, we can view $x^{(k+1)} = x^{(k)} + M^{-1}(b - Ax^{(k)})$ as a correction of $x^{(k)}$ based on the **residual** $r^{(k)} := b - Ax^{(k)}$. We also note that the **error** $e^{(k)} := x^* - x^{(k)}$ satisfies $e^{(k+1)} = Ge^{(k)}$, so the convergence of a splitting method depends on the properties of its **iteration matrix** G.

Splitting methods

Let L, D, and U denote the *strictly* lower, diagonal, and *strictly* upper triangular parts of A. The splittings for some basic linear stationary iterative methods are as follows.

Method	M
Jacobi	D
ω-Jacobi ($\omega \neq 0$)	$\frac{1}{\omega}D$
Gauss–Seidel	$L + D$
ω-Gauss–Seidel/successive overrelaxation (SOR) ($\omega \neq 0$)	$L + \frac{1}{\omega}D$
Symmetric successive overrelaxation (SSOR) ($\omega \neq 0, 2$)	$\frac{\omega}{2-\omega}(L + \frac{1}{\omega}D)D^{-1}(U + \frac{1}{\omega}D)$
Richardson ($\alpha \neq 0$)	$\frac{1}{\alpha}I$

The parameter ω is known as the **relaxation**/**damping** parameter and arises from taking $x^{(k+1)} = (1 - \omega)x^{(k)} + \omega\hat{x}^{(k+1)}$, where $\hat{x}^{(k+1)}$ denotes the result of applying the corresponding nonparametrized ($\omega = 1$) method to $x^{(k)}$. If $\omega < 1$, the method is said to be **underrelaxed**/**underdamped**; if $\omega > 1$, it is said to be **overrelaxed**/**overdamped**.

The SSOR method arises from performing a "forward" ω-Gauss–Seidel step with $M = L + \frac{1}{\omega}D$ followed by a "backward" ω-Gauss–Seidel step with $M = U + \frac{1}{\omega}D$.

Convergence theorems

Clearly, since $e^{(k)} = G^k e^{(0)}$, if $\|G\| < 1$ for some operator norm, then the method **converges** in the sense that $x^{(k)} \to x^*$ for all $x^{(0)}$.[1] More generally, we see that the method is convergent if and only if $G^k \to 0$, which in turn depends on the **spectral radius** $\rho(G)$ of G, the maximum of the absolute values of its eigenvalues when regarded as a complex matrix.[2]

Namely, if (λ, v) is an eigenpair of G with $\|v\| = 1$, then $|\lambda|^k = |\lambda^k| = \|G^k v\| \leq \|G^k\|$ for the induced operator norm, so $\rho(G)^k = \rho(G^k) \leq \|G^k\|$. Thus, if $G^k \to 0$, then $\rho(G) < 1$. In fact, the converse is also true.

> Let $G \in \mathbb{C}^{n\times n}$. Then $G^k \to 0$ if and only if $\rho(G) < 1$ (such a matrix is called **convergent**).

Proof. It remains to show that $G^k \to 0$ if $\rho(G) < 1$. Let UTU^* be a Schur factorization of G and let D and N denote the diagonal and strictly upper triangular parts of T. Since the product of a diagonal matrix and a strictly upper triangular matrix is strictly upper triangular, and the product of n (or more) strictly upper triangular $n \times n$ matrices is zero, for all $k \geq n$, we have

$$\|G^k\|_2 = \|(D+N)^k\|_2 \leq \sum_{j=0}^{n-1} \binom{k}{j} \|D\|_2^{k-j} \|N\|_2^j = \sum_{j=0}^{n-1} \binom{k}{j} \rho(G)^{k-j} \|N\|_2^j \to 0. \quad \blacksquare$$

Using this fact, we can also prove a well-known formula for the spectral radius.[3]

> **Gelfand's formula**
>
> Let $G \in \mathbb{C}^{n\times n}$ and $\|\cdot\|$ be an operator norm. Then $\rho(G) = \lim_{k\to\infty} \|G^k\|^{1/k}$.

Proof. We previously saw that $\rho(G) \leq \|G^k\|^{1/k}$ for all k, so $\rho(G) \leq \liminf_{k\to\infty} \|G^k\|^{1/k}$. On the other hand, if $\varepsilon > 0$ is arbitrary and $\hat{G} := \frac{G}{\rho(G)+\varepsilon}$, then $\rho(\hat{G}) < 1$, so by the preceding result, $\|\hat{G}^k\| \leq 1$ for all sufficiently large k, which is to say that $\|G^k\| \leq (\rho(G)+\varepsilon)^k$. Hence we also have $\limsup_{k\to\infty} \|G^k\|^{1/k} \leq \rho(G) + \varepsilon$. ∎

As $\|e^{(k)}\| \leq \|G^k\| \|e^{(0)}\| \approx \rho(G)^k \|e^{(0)}\|$ for large k, the rate of convergence can often be estimated using the spectral radius of the iteration matrix. More precisely, a direct computation shows that if G is diagonalizable and has a unique dominant eigenvalue, then $\frac{\|e^{(k+1)}\|}{\|e^{(k)}\|} \sim \rho(G)$, provided that $e^{(0)}$ has a nonzero component in the corresponding eigenspace.

General matrices

> If the Jacobi method converges, then the ω-Jacobi method converges for $\omega \in (0, 1]$.

Proof. For the ω-Jacobi method, we have $G = (1-\omega)I + \omega G_{\mathrm{J}}$, where G_{J} denotes the iteration matrix for the Jacobi method. Hence $\rho(G) \leq (1-\omega) + \omega\rho(G_{\mathrm{J}}) < 1$. ∎

> If the ω-Gauss–Seidel method converges, then $\omega \in (0, 2)$.

Proof. For the ω-Gauss–Seidel method, we have $G = (L + \frac{1}{\omega}D)^{-1}((\frac{1}{\omega} - 1)D - U)$, so $\det(G) = \det(\frac{1}{\omega}D)^{-1} \det((\frac{1}{\omega} - 1)D) = \det((1-\omega)I) = (1-\omega)^n$. On the other hand, the determinant of G is the product of its eigenvalues, so $|1-\omega|^n \leq \rho(G)^n < 1$. ∎

> If the SSOR method converges, then $\omega \in (0, 2)$.

Proof. For the SSOR method, we have $G = G_b G_f$, where G_f and G_b denote the iteration matrices for the forward and backward ω-Gauss–Seidel methods. Arguing as in the preceding proof, we obtain $|1-\omega|^{2n} \le \rho(G)^n < 1$.

Symmetric positive definite matrices

If A is symmetric positive definite (SPD), then A is invertible and all the splitting methods above are applicable since its diagonal entries must be positive. Recall also that symmetric matrices are partially ordered by the Loewner order $\prec$ in which $A \prec B$ if and only if $B - A$ is SPD, and that an SPD matrix A defines an inner product $\langle x, y \rangle_A := \langle Ax, y \rangle_2$.

> If A is SPD and $A \prec M + M^\top$, then $\|G\|_A < 1$.

Proof. Let x be a vector with $\|x\|_A = 1$ such that $\|G\|_A = \|Gx\|_A$ (which exists by the extreme value theorem) and let $y := M^{-1}Ax$. Then

$$\begin{aligned} \|G\|_A^2 &= \|x - y\|_A^2 \\ &= 1 - \langle x, y \rangle_A - \langle y, x \rangle_A + \langle y, y \rangle_A \\ &= 1 - \langle (M + M^\top - A)y, y \rangle_2 < 1. \quad \blacksquare \end{aligned}$$

> **Convergence for SPD matrices**
>
> Let A be SPD.
>
> - If $A \prec \frac{2}{\omega} D$, then the ω-Jacobi method converges.
> - If $\omega \in (0, 2)$, then the ω-Gauss–Seidel method converges.
> - If $\omega \in (0, 2)$, then the SSOR method converges.
> - If $\alpha \in (0, \frac{2}{\rho(A)})$, then the Richardson method converges.
>   Moreover, if the eigenvalues of A are $\lambda_1 \ge \cdots \ge \lambda_n > 0$, then $\rho(G)$ is minimized when $\alpha = \alpha^* := \frac{2}{\lambda_1 + \lambda_n}$, in which case $\|e^{(k+1)}\|_2 \le (1 - \frac{2}{\kappa+1})\|e^{(k)}\|_2$, where κ is the 2-norm condition number of A.

Proof. The convergence statements follow immediately from the preceding result.
For the Richardson method, $\rho(G(\alpha)) = \max_i |1 - \alpha\lambda_i| = \max\{1 - \alpha\lambda_n, \alpha\lambda_1 - 1\}$, so $\rho(G(\alpha)) = 1 - \alpha\lambda_n \ge \rho(G(\alpha^*))$ when $\alpha \le \alpha^*$ and $\rho(G(\alpha)) = \alpha\lambda_1 - 1 \ge \rho(G(\alpha^*))$ when $\alpha \ge \alpha^*$.
Finally, since G and A are normal, we have $\|G(\alpha^*)\|_2 = \rho(G(\alpha^*)) = 1 - \frac{2}{\kappa+1}$. ∎

Diagonally dominant matrices

- A is **weakly diagonally dominant (WDD)** if every row i is WDD: $|a_{ii}| \ge \sum_{j \ne i} |a_{ij}|$.
- A is **strictly diagonally dominant (SDD)** if every row i is SDD: $|a_{ii}| > \sum_{j \ne i} |a_{ij}|$.
- The **directed graph** of A is $\mathcal{G}_A = (V, E)$ with $V = \{1, \ldots, n\}$ and $(i, j) \in E$ if and only if $a_{ij} \ne 0$.
  - A is **irreducible** if $\mathcal{G}_A$ is strongly connected.
- A is **irreducibly diagonally dominant (IDD)** if it is irreducible, WDD, *and some row is SDD*.
- A is **weakly chained diagonally dominant (WCDD)** if it is WDD and for every row i, there exists an SDD row j with a path from i to j in $\mathcal{G}_A$. (Thus, SDD and IDD matrices are both WCDD.)

> If A is WCDD, then A is invertible.

Proof. Suppose for the sake of contradiction that there exists an $x \in \ker(A)$ with $\|x\|_\infty = 1$, and let i_1 be such that $|x_{i_1}| = 1$. Let $(x_{i_1}, \dots, x_{i_k})$ be a path in $\mathcal{G}_A$ such that row i_k is SDD. Since $\sum_j a_{i_1 j} x_j = 0$, we have

$$|a_{i_1 i_1}| = |-a_{i_1 i_1} x_{i_1}| \le \sum_{j \neq i_1} |a_{i_1 j}||x_j| \le \sum_{j \neq i_1} |a_{i_1 j}|,$$

so row i_1 is not SDD. However, since it is WDD, equality must hold throughout, which implies that $|x_{i_2}| = 1$ because $a_{i_1 i_2} \neq 0$. Iterating this argument, we ultimately deduce that row i_k is not SDD, which is a contradiction. ∎

As a result, if A is WCDD, then all the splitting methods above are applicable since its diagonal entries must be nonzero (otherwise, it would have a zero row and fail to be invertible).

We also note that this immediately implies the **Levy–Desplanques theorem**: if A is SDD, then A is invertible. This, in turn, is equivalent to the **Gershgorin circle theorem**: if λ is an eigenvalue of A, then $|\lambda - a_{ii}| \le \sum_{j \neq i} |a_{ij}| =: r_i$ for *some* i (in other words, $\lambda \in B_{r_i}(a_{ii})$ for *some* i). Similarly, if A is IDD, then A is invertible; or equivalently, if A is irreducible and λ is an eigenvalue of A such that $|\lambda - a_{ii}| \ge r_i$ for *every* i, then $|\lambda - a_{ii}| = r_i$ for *every* i.

> **Convergence for WCDD matrices**
>
> If A is WCDD, then the ω-Jacobi and ω-Gauss–Seidel methods converge for $\omega \in (0, 1]$.

Proof. If $|\lambda| \ge 1$, then $|\frac{\lambda - 1}{\omega} + 1| \ge |\lambda|$, so $(\lambda - 1)M + A$ has the same WDD/SDD rows and directed graph as A. Hence $\lambda I - G = M^{-1}((\lambda - 1)M + A)$ is invertible for such λ, so $\rho(G) < 1$. ∎

---

1. In other words, if $\|G\| < 1$, then $G^k \to 0$ strongly (because $G^k \to 0$ in norm). ↩
2. In other words, $G^k \to 0$ strongly if and only if $G^k \to 0$ in norm (because G is an operator on a finite-dimensional space). ↩
3. This formula remains true if G is a continuous linear operator on a Banach space X. Consequently, in this setting, we still have that $G^k \to 0$ in norm if and only if $\rho(G) < 1$. These are in turn equivalent to the invertibility of $I - G$ and the convergence of the fixed-point iteration of $x \mapsto Gx + f$ for all $f \in X$ (to $(I - G)^{-1} f$). ↩