**MATH 5330: Computational Methods of Linear Algebra**

# Lecture Note 3: Stationary Iterative Methods

Xianyi Zeng
Department of Mathematical Sciences, UTEP

## 1 Stationary Iterative Methods

The Gaussian elimination (or in general most direct methods) requires $O(n^3)$ computational cost, which is not acceptable when $n$ is large. For example, let us consider the direct numerical simulation of the Navier-Stokes equation on a unit cube in 3D; we discretize the domain with $N^3$ cubes and compute the solutions up to $t=T$. The Navier-Stokes equation has five variables at each node, hence the solution vector is $\boldsymbol{u}\in\mathbb{R}^{5N^3}$. When an implicit/explicit (IMEX) method is used for the time-integration, the time step size scales with $1/N$ and the total number of time steps is $O(N)$. For each time step, there is a nonlinear equation to solve:

$$\boldsymbol{f}(\boldsymbol{u}^{n+1})=\boldsymbol{g}(\boldsymbol{u}^n)\,,$$

to update the solution from one time step ($\boldsymbol{u}^n$) to the next ($\boldsymbol{u}^{n+1}$). Let $K$ be the average number of iterations in the Newton method to solve this nonlinear system, in total we need to solve a linear system $O(KN)$ times, and each such linear system is of the size $5N^3\times 5N^3$. If a direct method is used for the linear solves, the total computational cost is:

$$O((5N^3)^3)\times O(KN)=O(KN^{10})\,,$$

which is prohibitive even for moderate value of $N$.

The target of stationary iterative methods[1] is to reduce the computational cost with linear solves to magnitudes smaller than $O(n^3)$. Particularly in solving $A\boldsymbol{x}=\boldsymbol{b}$, let us write $A=M+(A-M)$ for some matrix $M$ and the linear system as:

$$M\boldsymbol{x}=-(A-M)\boldsymbol{x}+\boldsymbol{b}\,. \tag{1.1}$$

In an iterative method, we start with an initial guess $\boldsymbol{x}_0$ and try to improve the result solving for $\boldsymbol{x}_{k+1}$, $k=0,1,\cdots$:

$$M\boldsymbol{x}_{k+1}=-(A-M)\boldsymbol{x}_k+\boldsymbol{b}\,,\quad \text{or equivalently}\quad \boldsymbol{x}_{k+1}=-M^{-1}(A-M)\boldsymbol{x}_k+M^{-1}\boldsymbol{b}\,. \tag{1.2}$$

Let's look at the last equation, clearly a requirement for the iterative method to make sense is that the linear system associated with $M$ should be easy to solve in the sense that the cost is no more than $O(n^2)$. Two such choices are diagonal matrices ($\sim O(n)$) and triangular matrices ($\sim O(n^2)$). Next, we also want to make sure that if there exists a solution $\boldsymbol{x}$, then $\boldsymbol{x}_k\to\boldsymbol{x}$ as $k\to+\infty$. Finally, providing that $\boldsymbol{x}_k\to\boldsymbol{x}$, we hope $||\boldsymbol{x}_k-\boldsymbol{x}||$ to be reasonably small for only a few number of iterations. These are the questions we'd like to answer for any iterative method in this lecture.

[1] Or simply "iterative methods" in this section.

Let $A$ be non-singular and $\boldsymbol{x}$ solves $A\boldsymbol{x}=\boldsymbol{b}$, we first look at the convergence. Define $\boldsymbol{\varepsilon}_k=\boldsymbol{x}_k-\boldsymbol{x}$ as the error vector in the $k$-th iteration, then:

$$M\boldsymbol{\varepsilon}_{k+1}=M(\boldsymbol{x}_{k+1}-\boldsymbol{x})=[-(A-M)\boldsymbol{x}_k+\boldsymbol{b}]-[-(A-M)\boldsymbol{x}+\boldsymbol{b}]=-(A-M)\boldsymbol{\varepsilon}_k\,,$$

or equivalently:

$$\boldsymbol{\varepsilon}_{k+1}=G\boldsymbol{\varepsilon}_k\,,\quad G=-M^{-1}(A-M)\,. \tag{1.3}$$

The growth matrix $G$ remains the same for all iterations (hence the name "stationary iterative methods"), thus we obtain an estimate on the error $\boldsymbol{\varepsilon}_k$:

$$\boldsymbol{\varepsilon}_k=G^k\boldsymbol{\varepsilon}_0\quad\Longrightarrow\quad \big|\big|\boldsymbol{\varepsilon}_k\big|\big|\le\big|\big|G^k\big|\big|\,||\boldsymbol{\varepsilon}_0||\,. \tag{1.4}$$

Thus we have $\boldsymbol{\varepsilon}_k\to 0$ for any $\boldsymbol{\varepsilon}_0$ if $G^k\to 0$ when $k\to\infty$, for which a sufficient and necessary condition is given by Theorem 1.1.

**Remark 1.** *Strictly speaking, we do not need $G^k\to 0$ to deduce $\boldsymbol{\varepsilon}_k$ if we can choose $\boldsymbol{\varepsilon}_0$ carefully. In an extreme case if $\boldsymbol{\varepsilon}_0$ is in the null space of some $G^{k_0}$ (which happens with $k_0=1$ in the foolish case when $M=A$), the error becomes zero after $k_0$ iterations.*

**Theorem 1.1.** *$G^k\to 0$ as $k\to\infty$ if and only if $\rho(G)<1$, where the spectral radius $\rho(G)$ is the maximum absolute value of all the eigenvalues of $G$.*

*Proof.* We consider the Jordan canonical form $G=QJQ^{-1}$, where $Q$ is invertible and $J$ is given by:

$$J=\begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_m \end{bmatrix}_{n\times n},\quad J_l=\begin{bmatrix} \lambda_l & 1 & & \\ & \lambda_l & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_l \end{bmatrix}_{n_l\times n_l}. \; l=1,\cdots,m \tag{1.5}$$

Here $\lambda_1,\cdots,\lambda_m$ are the eigenvalues of $G$, which may not be different from each other; and $n_1+\cdots+n_m=n$.

To this end $G^k=QJ^kQ^{-1}$ and we just need to show $J^k\to 0$ if and only if $\rho(G)<1$ or equivalently $\big|\lambda_l\big|<1$ for all $l$. Because $J^k=\mathrm{diag}(J_1^k\,,\,J_2^k\,,\,\cdots\,,\,J_m^k)$, we only need to show $J_l^k\to 0$ if and only if $\big|\lambda_l\big|<1$. The last point is straightforward as:

$$J_l^k=\begin{bmatrix} \binom{k}{k}\lambda_l^k & \binom{k}{k-1}\lambda_l^{k-1} & \binom{k}{k-2}\lambda_l^{k-2} & \cdots & \binom{k}{k-n_l+1}\lambda_l^{k-n_l+1} \\ & \binom{k}{k}\lambda_l^k & \binom{k}{k-1}\lambda_l^{k-1} & \cdots & \binom{k}{k-n_l}\lambda_l^{k-n_l} \\ & & \binom{k}{k}\lambda_l^k & \cdots & \binom{k}{k-n_l-1}\lambda_l^{k-n_l-1} \\ & & & \ddots & \vdots \\ & & & & \binom{k}{k}\lambda_l^k \end{bmatrix}_{n_l\times n_l}, \tag{1.6}$$

and the fact that each entry converges to zero as $k\to\infty$. □

In fact, if $G=-M^{-1}(A-M)$ has spectral radius smaller than 1, we can drop the assumption that $A$ is non-singular (it remains true, though), as stated by the next theorem.

**Theorem 1.2.** *If $\rho(G)<1$, then $A$ is invertible.*

*Proof.* Because $A=M(I-G)$, we just need to show that $I-G$ is invertible for the first part. Indeed, if $(I-G)\boldsymbol{x}=0$ for some vector $\boldsymbol{x}\in\mathbb{R}^n$, then

$$\boldsymbol{x}=I\boldsymbol{x}=G\boldsymbol{x}=G^2\boldsymbol{x}=\cdots=G^k\boldsymbol{x}\to 0$$

by Theorem 1.1; hence the null space of $I-G$ contains only the zero vector and $I-G$ is invertible. □

The condition of the preceding theorem is in general very difficult to check; and a more convenient one is based on the inequality $\left|\left|G^k\right|\right|\le||G||^k$. Hence a necessary condition is given by $||G||<1$ for some induced matrix norm.

## 2 Jacobi Method

A simplest iterative method is given by choosing $M$ as the diagonal part of $A$, this is called the *Jacobi method*. Let $A=L+D+U$, where $L$, $D$, and $U$ are the lower-tridiagonal part, diagonal part, and upper-tridiagonal part, respectively (not to be confused with the LU or LDU decomposition!). Then in the Jacobi method, $M=D$ and (1.2) reduces to:

$$\boldsymbol{x}_{k+1}=-D^{-1}(L+U)\boldsymbol{x}_k+D^{-1}\boldsymbol{b}\,. \tag{2.1}$$

The growth matrix $G=-D^{-1}(L+U)$ is:

$$G=-\begin{bmatrix}\frac{1}{a_{11}}&&&\\&\frac{1}{a_{22}}&&\\&&\ddots&\\&&&\frac{1}{a_{nn}}\end{bmatrix}\begin{bmatrix}0&a_{12}&\cdots&a_{1n}\\a_{21}&0&\cdots&a_{2n}\\\vdots&\vdots&\ddots&\vdots\\a_{n1}&a_{n2}&\cdots&0\end{bmatrix}=-\begin{bmatrix}0&\frac{a_{12}}{a_{11}}&\cdots&\frac{a_{1n}}{a_{11}}\\\frac{a_{21}}{a_{22}}&0&\cdots&\frac{a_{2n}}{a_{22}}\\\vdots&\vdots&\ddots&\vdots\\\frac{a_{n1}}{a_{nn}}&\frac{a_{n2}}{a_{nn}}&\cdots&0\end{bmatrix}. \tag{2.2}$$

If the diagonal elements of $A$ are sufficiently large, say:

$$\left|a_{ii}\right|>\sum_{j\neq i}\left|a_{ij}\right|\,,\quad\forall i=1\,,\,2\,,\,\cdots\,,\,n\,, \tag{2.3}$$

then we have $||G||_\infty<1$. By the argument at the end of Section 1, we see that the Jacobi method converges if $A$ is *diagonally dominant*, i.e., if $A$ satisfies (2.3).

(2.3) actually provides a way to estimate the number of iterations needed in order to achieve certain accuracy. Let:

$$\rho_{\text{jac}}=\max_i\frac{\sum_{j\neq i}\left|a_{ij}\right|}{\left|a_{ii}\right|}<1\,, \tag{2.4}$$

then by (1.4) $\left|\left|\boldsymbol{\varepsilon}_k\right|\right|_\infty\le\rho_{\text{jac}}^k\left|\left|\boldsymbol{\varepsilon}_0\right|\right|_\infty$.

To see an example of diagonal-dominate linear systems, we consider the implicit method to solve the system of ordinary differential equations:

$$\frac{d\boldsymbol{x}}{dt}=\boldsymbol{f}(\boldsymbol{x})\,.$$

Let $\boldsymbol{x}_m$ and $\boldsymbol{x}_{m+1}$ be the solutions at $t_m$ and $t_{m+1} = t_m + \Delta t_m$, respectively; then we update the solution from $\boldsymbol{x}_m$ to $\boldsymbol{x}_{m+1}$ by:

$$\frac{\boldsymbol{x}_{m+1} - \boldsymbol{x}_m}{\Delta t_m} = \boldsymbol{f}(\boldsymbol{x}_m) + \frac{\partial \boldsymbol{f}(\boldsymbol{x}_m)}{\partial \boldsymbol{x}}(\boldsymbol{x}_{m+1} - \boldsymbol{x}_m)\,,$$

which involves the linear system with $A_m = I + \Delta t_m \frac{\partial \boldsymbol{f}(\boldsymbol{x}_m)}{\partial \boldsymbol{x}}$. Thus we can always find a diagonal-dominate matrix $A_m$ by choosing a small $\Delta t_m$.

# 3 Gauss-Seidel Method

The Jacobi method can be written component-by-component as:

$$\text{for } i = 1\,, 2\,, \cdots,\, n: \tag{3.1}$$
$$x_{k+1;i} = \frac{1}{a_{ii}} \left( -\sum_{j \neq i} a_{ij} x_{k;j} + b_j \right).$$

Here we denote $\boldsymbol{x}_k = [x_{k;i}]$. One argument about the Jacobi method is that we're not using the most updated information in the solution, i.e. it is always the components of $\boldsymbol{x}_k$ that appear on the right hand side of the updating formula.

A modification that always use the most recent data and saves some storage is the following:

$$\text{for } i = 1\,, 2\,, \cdots,\, n: \tag{3.2}$$
$$x_{k+1;i} = \frac{1}{a_{ii}} \left( -\sum_{j<i} a_{ij} x_{k+1;j} - \sum_{j>i} a_{ij} x_{k;j} + b_j \right).$$

This algorithm is known as the *Gauss-Seidel method*; and it is equivalent to the choice $M = D + L$:

$$\boldsymbol{x}_{k+1} = -(D+L)^{-1}(U\boldsymbol{x}_k - \boldsymbol{b})\,. \tag{3.3}$$

The Gauss-Seidel method also guarantees convergence for arbitrary initial data for diagonally dominant matrices. Particularly, similar to (2.4) we can derive a decay rate:

$$\rho_{\text{gs}} = \max_i \frac{\sum_{j>i} |a_{ij}|}{|a_{ii}| - \sum_{j<i} |a_{ij}|}\,. \tag{3.4}$$

We prove in the exercises $\left|\left|-(D+L)^{-1}U\right|\right|_\infty \leq \rho_{\text{gs}}$ and hence deduce that $\left|\left|\boldsymbol{\varepsilon}_k\right|\right|_\infty \leq \rho_{\text{gs}}^k \left|\left|\boldsymbol{\varepsilon}_0\right|\right|_\infty$. Comparing (2.4) and (3.4) we see for the same diagonally dominant matrix $A$, $\rho_{\text{gs}} \leq \rho_{\text{jac}}$; hence the Gauss-Seidel method in general converges faster than the Jacobi method, at the cost of solving a triangular system instead of a diagonal one at each iteration.

Because solving the linear system $L + D$ involves forward substitution, the method described before is also called the *forward Gauss-Seidel method*. Similarly, we can choose $M = D + U$ and establishing similar convergent result for diagonally dominant matrices; and this method is called the *backward Gauss-Seidel method*.

Finally, we show the convergence of the forward Gauss-Seidel method for another type of very important matrices, namely the symmetric positive-definite ones. This is a direct result of the following theorem and Theorem 1.1.

**Theorem 3.1.** *Let $A$ be symmetric positive-definite, then $G=-(L+D)^{-1}L^t$ satisfies $\rho(G)<1$.*

*Proof.* Clearly $D$ has all its diagonal entries positive and it is non-singular; thus we may write:

$$-G=[D^{\frac{1}{2}}(D^{-\frac{1}{2}}LD^{-\frac{1}{2}}+I)D^{\frac{1}{2}}]^{-1}L^t=D^{-\frac{1}{2}}(\tilde{L}+I)^{-1}\tilde{L}^tD^{\frac{1}{2}}\,,\quad \text{where}\quad \tilde{L}=D^{-\frac{1}{2}}LD^{-\frac{1}{2}}\,.$$

Let $\lambda\in\mathbb{C}$ be any eigenvalue of $-G$, since:

$$(\tilde{L}+I)^{-1}\tilde{L}^t=-D^{\frac{1}{2}}GD^{-\frac{1}{2}}\,,$$

$\lambda$ is also an eigenvalue of $-\tilde{G}=(\tilde{L}+I)^{-1}\tilde{L}^t$. Choose $\boldsymbol{z}\in\mathbb{C}^n$ as a unit eigenvector of $-\tilde{G}$ corresponding to $\lambda$, i.e.

$$\tilde{L}^t\boldsymbol{z}=\lambda(\tilde{L}+I)\boldsymbol{z}\,.$$

Define $\tilde{A}=\tilde{L}+I+\tilde{L}^t=D^{-\frac{1}{2}}(L+D+L^t)D^{-\frac{1}{2}}=D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$, then $\tilde{A}$ is also symmetric positive-definite. Let us define $\alpha=\boldsymbol{z}^*\tilde{L}\boldsymbol{z}$ and denote $\alpha=a+ib$, $a,b\in\mathbb{R}$; we also denote $\boldsymbol{z}=\boldsymbol{x}+i\boldsymbol{y}$ where $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$. Then we have:

$$\overline{\alpha}=\alpha^*=\boldsymbol{z}^*\tilde{L}^t\boldsymbol{z}=\lambda\boldsymbol{z}^*(\tilde{L}+I)\boldsymbol{z}=\lambda(1+\alpha)\,,$$

where we used the assumption $\boldsymbol{z}^*\boldsymbol{z}=1$.

Next, we compute $\boldsymbol{z}^*\tilde{A}\boldsymbol{z}$:

$$\boldsymbol{z}^*\tilde{A}\boldsymbol{z}=\boldsymbol{z}^*(\tilde{L}+I+\tilde{L}^t)\boldsymbol{z}=(1+\lambda)\boldsymbol{z}^*(\tilde{L}+I)\boldsymbol{z}=(1+\lambda)(1+\alpha)=1+\alpha+\overline{\alpha}=1+2a\,.$$

However, by the positive definiteness of $\tilde{A}$, we have:

$$\boldsymbol{z}^*\tilde{A}\boldsymbol{z}=(\boldsymbol{x}^t-i\boldsymbol{y}^t)\tilde{A}(\boldsymbol{x}+i\boldsymbol{y})=\boldsymbol{x}^t\tilde{A}\boldsymbol{x}+\boldsymbol{y}^t\tilde{A}\boldsymbol{y}>0\quad\Rightarrow\quad 1+2a>0\,.$$

Thus by $\lambda=\overline{\alpha}/(1+\alpha)$ we have:

$$|\lambda|^2=\left|\frac{a-ib}{1+a+ib}\right|^2=\frac{a^2+b^2}{1+2a+a^2+b^2}<1\,.$$

The proof is completed by noting that $\rho(G)=\rho(-G)$ and the choice of $\lambda$ is arbitrary. □

# 4 SOR Method

The successive over-relaxation method (SOR) takes a "linear combination" of the Jacobi method and the Gauss-Seidel method to provide more control over the convergence rate. Particularly, we choose $M=M_\omega=\frac{1}{\omega}D+L$ for some $\omega>0$ – Letting $\omega=\infty$ the method tens to the Jacobi method and setting $\omega=1$ the method corresponds to the forward Gauss-Seidel.

For the SOR with $\omega>0$, we have:

$$G=G_\omega=-\left(\frac{1}{\omega}D+L\right)^{-1}\left(\frac{\omega-1}{\omega}D+U\right).$$

It can be shown that if $A$ is symmetric positive-definite and $0<\omega<2$, then the SOR method converges. The proof is similar to that of Theorem 3.1, and the details are left as an exercise.

Note that due to a theorem of by Kahan [1], $\rho(G_\omega) \geq |\omega - 1|$; hence a necessary condition for the SOR method to converge is also $0 < \omega < 2$.

A major purpose of the SOR method is that it allows people to tune $\omega$ in order to minimize $\rho(G_\omega)$ for some special matrices. For example, if $A$ is symmetric positive-definite and also tridiagonal, then $\rho(G_{\rm gs}) = \rho(G_{\rm jac})^2 < 1$ and the optimal choice for SOR is:

$$\omega = \frac{2}{1 + \sqrt{1 - \rho(G_{\rm jac})^2}} \,.$$

In this case, $\rho(G_\omega) = \omega - 1$, which is optimal by the Kahan theorem.

# 5 Related Topics: Acceleration and Preconditioners

The acceleration technique tries to improve the convergence of an existing iterative method. Suppose we obtained $\boldsymbol{x}_1 \,,\, \boldsymbol{x}_2 \,,\, \cdots \,,\, \boldsymbol{x}_k$ from the standard iterative method, then the plan is to compute a linear combination:

$$\boldsymbol{y}_k = \sum_{i=0}^{k} \nu_i(k) \boldsymbol{x}_i \,, \tag{5.1}$$

so that $\boldsymbol{y}_k$ represents a better approximation to the exact solution. Note that a natural condition on the coefficients is $\sum_{i=0}^{k} \nu_i(k) = 1$, so that if all iterates are exact, so is $\boldsymbol{y}_k$. If we define a polynomial:

$$p_k(x) = \nu_0(k) + \nu_1(k) x + \cdots + \nu_k(k) x^k \tag{5.2}$$

and extend its definition to matrices naturally, we have:

$$\boldsymbol{y}_k - \boldsymbol{x} = p_k(G) \boldsymbol{\varepsilon}_0 \,,$$

where $\boldsymbol{x}$ is the exact solution. Hence the target is to minimize $||p_k(G)||$ in a certain norm.

The Chebyshev semi-iterative method for symmetric matrices makes use of the fact that the eigenvalues of $p_k(G)$ are $p_k(\lambda)$, where $\lambda$ is any eigenvalue of $G$. Knowing that any eigenvalue of $G$ lies between $-1$ and 1, the method utilizes the Chebyshev polynomials $c_k(x)$ defined recursively by $c_0(x) = 1$, $c_1(x) = x$, and $c_{k+1}(x) = 2x c_k(x) - c_{k-1}(x)$, and define:

$$p_k(x) = \frac{1}{c_k(\mu)} c_k\left(-1 + 2\frac{x - \lambda_{\min}}{\lambda_{\max} - \lambda_{\min}}\right) ,$$

where $\mu = -1 + 2\frac{1 - \lambda_{\min}}{\lambda_{\max} - \lambda_{\min}}$, $-1 < \lambda_{\min} < \lambda_{\max} < 1$ are the smallest and largest eigenvalues of $G$, respectively.

There are two benefits of using the Chebyshev polynomials. First, $c_k(x)$ satisfies $|c_k(x)| \leq 1$ on $[-1,1]$ and it grow rapidly off this interval; thus the following estimate expects to be small:

$$||\boldsymbol{y}_k - \boldsymbol{x}||_2 \leq \frac{1}{|c_k(\mu)|} ||\boldsymbol{\varepsilon}_0||_2 \,.$$

Secondly, the recursive relation in the Chebyshev polynomials enables the following algorithm that completely removes the need to compute the iterates $\boldsymbol{x}_k$ but calculate $\boldsymbol{y}_k$ directly:

$$\begin{aligned} &\boldsymbol{y}_{k+1} = \omega_{k+1}(\boldsymbol{y}_k - \boldsymbol{y}_{k-1} + \gamma \boldsymbol{z}_k) + \boldsymbol{y}_{k-1} \,, \\ &M \boldsymbol{z}_k = \boldsymbol{b} - A \boldsymbol{y}_k \,, \end{aligned}$$

where $\gamma = \frac{2}{2-\lambda_{\min}-\lambda_{\max}}$ , $\omega_{k+1} = 2\frac{2-\lambda_{\min}-\lambda_{\max}}{\lambda_{\max}-\lambda_{\min}}\frac{c_k(\mu)}{c_{k+1}(\mu)}$.

Another use of the iterative methods is to construct preconditioners. A (left) preconditioner $P$ modifies the original equation $A\boldsymbol{x} = \boldsymbol{b}$ to:

$$PA\boldsymbol{x} = P\boldsymbol{b}\,. \tag{5.3}$$

In general, the preconditioner depends highly on the problems to be solved, such as the low-Mach preconditioner for low-speed aerodynamic problems.

From a pure linear algebra point of view, though, the iterative method provides a class of preconditioners given by $P = M^{-1}$. In this case, we still need to solve a non-trivial system with $M^{-1}A$, but the hope is that $M^{-1}A$ will be better conditioned than $A$ itself. Corresponding to the previous methods, we have the following preconditioners:

$$P_{\text{jac}} = D^{-1}\,, \quad P_{\text{gs}} = (L+D)^{-1}\,, \quad P_{\text{sor}} = (L+\omega^{-1}D)^{-1}\,.$$

## Exercises

**Exercise 1.** *Use mathematical induction to show (1.6) in the case $n_l = 3$.*

**Exercise 2.** *Let $A$ be diagonally dominant and we want to complete the proof that the Gauss-Seidel method converges. Particularly let $G = -(L+D)^{-1}U$, show that $||G||_\infty \le \rho_{gs}$, which is given by (3.4).*
***Hint**: Use the definition of induced matrix norms, we just need to show that for all $||\boldsymbol{x}||_\infty = 1$, there is $||\boldsymbol{y}||_\infty \le \rho_{gs}$, where $\boldsymbol{y} = G\boldsymbol{x}$. And for this purpose, use $D\boldsymbol{y} = -L\boldsymbol{y} - U\boldsymbol{x}$.*

**Exercise 3.** *Show that if $A$ is symmetric, diagonally dominant, and all its diagonal elements are positive, then $A$ is positive definite.*
***Hint**: Show that $\boldsymbol{x}^t A\boldsymbol{x} \ge 0$ and derive the condition for the equality to hold. For this purpose, use the inequality $a_{ij}x_iy_j \ge -\left(\frac{1}{2}|a_{ij}|x_i^2 + \frac{1}{2}|a_{ji}|x_j^2\right)$.*

**Exercise 4.** *Prove that if $A$ is symmetric positive definite, then the SOR method with $0 < \omega < 2$ converges.*

**Exercise 5.** *Let us consider the SOR method with $\omega > 0$, show that:*

$$|\det G_\omega| = |1-\omega|^n\,.$$

*Then deduce that $\rho(G_\omega) \ge |1-\omega|$, the Kahan theorem.*
***Hint**: The determinant of a matrix $A$ is the product of all the eigenvalues of $A$.*

## References


[1] William Morton Kahan. Gauss-Seidel methods of solving large systems of linear equations. PhD thesis, University of Toronto, 1958.