MATH60005/70005: Optimization (Autumn 22-23)

Week 4: Exercises

Dr Dante Kalise
Department of Mathematics
Imperial College London, United Kingdom
dkaliseb@imperial.ac.uk

Sara Bicego (GTA)
Department of Mathematics
Imperial College London, United Kingdom
s.bicego21@imperial.ac.uk

1. Find the exact linesearch stepsize when $f(\mathbf{x})$ is a quadratic function $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} + \mathbf{c}$ where $\mathbf{A}$ is an $n \times n$ positive definite matrix, $\mathbf{b} \in \mathbb{R}^n$ and $\mathbf{c} \in \mathbb{R}$.

2. Let $\mathbf{A}$ be a symmetric $n \times n$ matrix, $\mathbf{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Then the function $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} + \mathbf{c}$ is a $C^{1,1}$ function. The smallest Lipschitz constant of f is $2\|\mathbf{A}\|_2$

3. Show that $f(\mathbf{x}) = \sqrt{1 + \mathbf{x}^2} \in C_L^{1,1}$.

4. Give an example of a function $f \in C_L^{1,1}(\mathbb{R})$ and a starting point $x_0 \in \mathbb{R}$ such that the problem $\min f(x)$ has an optimal solution and the gradient method with constant stepsize $t = \frac{2}{L}$ diverges.

5. Consider the localization problem where we are given m locations of sensors $\mathscr{A} := \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$, with each sensor in $\mathbb{R}^n$, and approximate distances between the sensors and an unknown source located at $\mathbf{x} \in \mathbb{R}^n$: $d_i \approx \|\mathbf{x} - \mathbf{a}_i\|$. We try to find the source location $\mathbf{x}$ given the sensor locations $\mathscr{A}$ and the approximate distances $d_1, d_2, \ldots, d_m$. For this, we write the optimization problem:

$$\min_{\mathbf{x}} \left\{ f(\mathbf{x}) \equiv \sum_{i=1}^{m} \left(\|\mathbf{x} - \mathbf{a}_i\| - d_i \right)^2 \right\}.$$

   a) State the first-order optimality condition for this problem, and show that for $\mathbf{x} \notin \mathscr{A}$ it is equivalent to

$$\mathbf{x} = \frac{1}{m} \left\{ \sum_{i=1}^{m} \mathbf{a}_i + \sum_{i=1}^{m} d_i \frac{\mathbf{x} - \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|} \right\}$$

b) Show that the iteration:

$$\mathbf{x}^{k+1} = \frac{1}{m}\left\{\sum_{i=1}^{m}\mathbf{a}_i + \sum_{i=1}^{m} d_i \frac{\mathbf{x}^k - \mathbf{a}_i}{\left\|\mathbf{x}^k - \mathbf{a}_i\right\|}\right\}$$

is a gradient method, assuming that $\mathbf{x}^k \notin \mathscr{A}$ for all $k \geq 0$. What is the stepsize?

c) Write an explicit Gauss-Newton iteration of the form

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \mathbf{d}^k ,$$

giving an expression for $\mathbf{d}^k$ in terms of the Jacobian and vectorized cost for this problem, without computing the inverse.

6. Consider the quadratic function $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top Q\mathbf{x}$$

where Q is a symmetric matrix of size 2×2 with eigenvalues $0 < \lambda_{\text{min}} < \lambda_{\text{max}}$. Suppose we apply the gradient descent method to the problem of minimizing f, with exact line search and initial point

$$\mathbf{x}_0 = \frac{1}{\lambda_{\text{min}}}\mathbf{u}_{\text{min}} + \frac{1}{\lambda_{\text{max}}}\mathbf{u}_{\text{max}}$$

where $\mathbf{u}_{\text{min}}$ and $\mathbf{u}_{\text{max}}$ are the norm one eigenvectors associated with λ_{min} and λ_{max}, respectively.

a) Show that after 1 iteration

$$\mathbf{x}_1 = \left(\frac{\lambda_{\text{max}} - \lambda_{\text{min}}}{\lambda_{\text{max}} + \lambda_{\text{min}}}\right)\left(\frac{1}{\lambda_{\text{min}}}\mathbf{u}_{\text{min}} - \frac{1}{\lambda_{\text{max}}}\mathbf{u}_{\text{max}}\right) .$$

b) Assuming that

$$\mathbf{x}_k = \left(\frac{\lambda_{\text{max}} - \lambda_{\text{min}}}{\lambda_{\text{max}} + \lambda_{\text{min}}}\right)^k \left(\frac{1}{\lambda_{\text{min}}}\mathbf{u}_{\text{min}} + \frac{(-1)^k}{\lambda_{\text{max}}}\mathbf{u}_{\text{max}}\right) \quad \text{for } \ k = 0, 1, \ldots ,$$

show that

$$\frac{f(\mathbf{x}_{k+1})}{f(\mathbf{x}_k)} = \left(\frac{\lambda_{\text{max}} - \lambda_{\text{min}}}{\lambda_{\text{max}} + \lambda_{\text{min}}}\right)^2 .$$

Using this, what can be said about the convergence of this method based on the ratio $\kappa = \frac{\lambda_{\text{max}}}{\lambda_{\text{min}}}$?

Quadratic Optimization Benchmark

Consider the quadratic minimization problem

$$\min_{\mathbf{x}} \left\{ \mathbf{x}^\top \mathbf{A}\mathbf{x} : \mathbf{x} \in \mathbb{R}^5 \right\}$$

where $\mathbf{A}$ is the 5×5 Hilbert matrix defined by

$$\mathbf{A}_{i,j} = \frac{1}{i+j-1}, \qquad i, j = 1, 2, 3, 4, 5$$

The matrix can be constructed via the MATLAB command `A=hilb(5)`. Run the following methods and compare the number of iterations required by each of the methods when the initial vector is $\mathbf{x}^0 = (1, 2, 3, 4, 5)^\top$ to obtain a solution $\mathbf{x}^*$ with $\|\nabla f(\mathbf{x})\| \leq 10^{-4}$:

- Gradient method with backtracking stepsize rule and parameters $\alpha = 0.5, \beta = 0.5, s = 1$
- Gradient method with backtracking stepsize rule and parameters $\alpha = 0.1, \beta = 0.5, s = 1$
- Diagonally scaled gradient method with diagonal elements $\mathbf{D}_{i,i} = \frac{1}{\mathbf{A}_{i,i}}, i = 1{,}2{,}3{,}4{,}5$ and exact line search;
- Diagonally scaled gradient method with diagonal elements $\mathbf{D}_{i,i} = \frac{1}{\mathbf{A}_{i,i}}, i = 1{,}2{,}3{,}4{,}5$ and backtracking line search with parameters $\alpha = 0.1, \beta = 0.5\ s = 1$.

Solutions

1) For the quadratic function $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} + \mathbf{c}$, the gradient reads $\nabla f(\mathbf{x}) = 2\big(\mathbf{A}\mathbf{x} + \mathbf{b}\big)$, and the gradient descend iteration is

$$\mathbf{x}^{k+1} = \mathbf{x}^k - t^k \nabla f(\mathbf{x}^k),$$

where we tune the stepsize t^k using linesearch. This amounts to solve

$$\min_{t \geq 0} \left\{ g(t) := f\big(\mathbf{y} + t\mathbf{d}\big) \right\}, \quad \text{with } \mathbf{d} = -\nabla f(\mathbf{x}^k), \quad \mathbf{y} = \mathbf{x}^k.$$

Substituting the definition of $f(\cdot)$ and $\nabla f(\cdot)$ into $g(t)$, we obtain

$$\begin{aligned} g(t) &= (\mathbf{y} + t\mathbf{d})^\top \mathrm{A}(\mathbf{y} + t\mathbf{d}) + 2\mathbf{b}^\top(\mathbf{y} + t\mathbf{d}) + \mathbf{c} \\ &= t^2(\mathbf{d}^\top \mathbf{A}\mathbf{d}) + 2(\mathbf{d}^\top \mathbf{A}\mathbf{y} + \mathbf{d}^\top \mathbf{b})t + \mathbf{x}^\top \mathbf{A}\mathbf{y} + 2\mathbf{b}^\top \mathbf{y} + \mathbf{c} \\ &= t^2(\mathbf{d}^\top \mathbf{A}\mathbf{d}) + 2(\mathbf{d}^\top \mathbf{A}\mathbf{y} + \mathbf{d}^\top \mathbf{b})t + f(\mathbf{y})\,. \end{aligned}$$

To find the minimizer of $g(t)$, we impose the first order optimality condition for

$$g'(t) := 2t(\mathbf{d}^\top \mathbf{A}\mathbf{d}) + 2(\mathbf{d}^\top \mathbf{A}\mathbf{y} + \mathbf{d}^\top \mathbf{b})\,,$$

i.e. we are looking for $t \geq 0$ such that $g'(t) = 0$. This leads to

$$t = -\frac{\mathbf{d}^\top(2\mathbf{A}\mathbf{y})}{2(\mathbf{d}^\top\mathbf{A}\mathbf{d})} = -\frac{\mathbf{d}^\top(\nabla f(\mathbf{y}))}{2(\mathbf{d}^\top\mathbf{A}\mathbf{d})}$$

and substituting back $\mathbf{d} = -\nabla f(\mathbf{x}^k)$, $\mathbf{y} = \mathbf{x}^k$, we have

$$t^k = +\frac{\|\nabla f(\mathbf{x}^k)\|^2}{2\nabla f(\mathbf{x}^k)^\top\mathbf{A}\nabla f(\mathbf{x}^k)}.$$

To conclude, we need check whether the computed stepsize is positive. Under the assumption $\nabla f(\mathbf{x}^k) \neq 0$, we have that both the numerator and the denominator (remember that $\mathbf{A} \succ 0$) are strictly positive, hence $t^k > 0$, and finally

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \frac{\|\nabla f(\mathbf{x}^k)\|^2}{2\nabla f(\mathbf{x}^k)^\top\mathbf{A}\nabla f(\mathbf{x}^k)}\nabla f(\mathbf{x}^k).$$

2) We want to show that – for f and ∇f as before – we have

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\|, \quad \text{for } L = 2\|\mathbf{A}\|_2.$$

Substituting the expression of the gradient on the left-hand side, we have

$$\|2(\mathbf{A}\mathbf{x} + b) - 2(\mathbf{A}\mathbf{y} + b)\| = 2\|\mathbf{A}(\mathbf{x} - \mathbf{y})\|$$

for which the Lipschitz condition becomes

$$\|\mathbf{A}(\mathbf{x} - \mathbf{y})\| \leq \|A\|_2\|\mathbf{x} - \mathbf{y}\|.$$

Thus, we aim at showing $\|\mathbf{A}(\mathbf{z})\| \leq \|A\|_2\|\mathbf{z}\|$ by using the definition of norm

$$\|\mathbf{A}\|_2 = \|\mathbf{A}\|_{2,2} = \max_{\|z\|_2 \leq 1} \|\mathbf{A}\mathbf{z}\|.$$

We precede by contradiction: assume that $\|\mathbf{A}\mathbf{z}\| > \|A\|_2\|\mathbf{z}\|$. Dividing both sides by $\|z\|$, we obtain

$$\left\|\mathbf{A}\frac{\mathbf{z}}{\|\mathbf{z}\|}\right\| > \|A\|_2,$$

which is equivalent to $\|\mathbf{A}\mathbf{v}\| > \|\mathbf{A}\|_2$ for all $\|\mathbf{v}\| = 1$. In particular, this holds for the maximum

$$\max_{\|\mathbf{v} \leq 1\|} \|\mathbf{A}\mathbf{v}\| > \|\mathbf{A}\|_2$$

which contradicts the definition of norm. Thus, we have the required inequality

$$\|\mathbf{A}\mathbf{z}\| \leq \|\mathbf{A}\|_2\|\mathbf{z}\|,$$

for $\mathbf{z} = \mathbf{x} - \mathbf{y}$.

3) We start by dealing with the one-dimensional case. If we define the function f and its derivative as

$$f(x) = \sqrt{1+x^2}\,, \qquad f'(x) = \frac{x}{\sqrt{1+x^2}},$$

the Lipschitz condition reads

$$\left\| \frac{x}{\sqrt{1+x^2}} - \frac{y}{\sqrt{1+y^2}} \right\| \le L\|x-y\|.$$

Since the above inequality is difficult to prove, we rely on the link between Lipschitz continuity and the norm of the Hessian: for f convex and twice differentiable (as in this case), we have

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \le L\|\mathbf{x}-\mathbf{y}\| \iff \|\nabla^2 f(\mathbf{x})\| \le L.$$

Since we are considering $f : \mathbb{R} \to \mathbb{R}$, we want to bound the absolute value of the second derivative

$$f''(x) = \frac{\sqrt{1+x^2} - \dfrac{\sqrt{1+x^2}}{x^2}}{1+x^2} = \frac{1}{(1+x^2)^{\frac{3}{2}}}, \qquad \|f''(x)\| \le 1 \iff f \in C_1^{1,1}.$$

Moving to the multi-dimensional case, we consider

$$f(\mathbf{x}) = \sqrt{1+\|\mathbf{x}\|^2}\,, \qquad \nabla f(\mathbf{x}) = \frac{\mathbf{x}}{\sqrt{1+\|\mathbf{x}\|^2}}, \qquad \mathbf{x} = [x_1, \dots, x_n]^\top \in \mathbb{R}^n$$

for which the partial derivatives read

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} = \frac{\delta_{i,j}}{\sqrt{1+\|\mathbf{x}\|^2}} - \frac{x_i x_j}{\left(1+\|x\|^2\right)^{\frac{3}{2}}}, \qquad i,j = 1,\dots,n$$

where $\delta_{i,j}$ are the Dirac deltas defined as $\delta_{i,j} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$

Moreover, we obtain the Hessian

$$\mathcal{H} := \nabla^2 f(\mathbf{x}) = a\mathbb{1} - a^3 \mathbf{x}\mathbf{x}^\top\,, \qquad a = \frac{1}{\sqrt{1+\|x\|^2}}\,,$$

whose norm can be computed as

$$\|\nabla^2 f(\mathbf{x})\| = \sqrt{\lambda_{max}(\mathcal{H}^\top\mathcal{H})} = \sqrt{\lambda_{max}(\mathcal{H}^2)} = \sqrt{\lambda_{max}(\mathcal{H})^2} = \left|\lambda_{max}(\mathcal{H})\right|.$$

The Hessian $\mathcal{H}$ has eigenvectors $\mathbf{x}$ and $\mathbf{x}^\perp$, associated to eigenvalues $\lambda_1 = (a - a^3\|\mathbf{x}\|^2)$ and $\lambda_2 = a$ respectively. We conclude by noticing that $a \ge 0$, hence $\lambda_{max} = \lambda_2 = a$, and finally

$$a = \frac{1}{\sqrt{1+\|\mathbf{x}\|^2}} \le 1 \iff f \in C_1^{1,1}.$$

4) We consider the function $f(x) = x^2$, with first derivative $f'(x) = 2x$. Then, f is L-Lipschitz continuous with $L = 2$, since we have that

$$\|f'(x) - f'(y)\| \leq 2\|x - y\| .$$

The gradient descend iteration for f with constant stepsize $t = \dfrac{2}{L} = 1$ reads

$$x^{k+1} = x^k - t2x^k = x^k - 2x^k = -x^k.$$

Hence, the method diverges for every $x_0 \neq 0$, as its iterations oscillate repeatedly between x_0 and $-x_0$. It would be enough to consider $t = \dfrac{2}{L} - \varepsilon$, with $\varepsilon > 0$ to have convergence of the gradient method for f.

5a) The first order optimality condition reads $\nabla f(\mathbf{x}) = 0$. Recalling that for $g(\mathbf{x}) = \|\mathbf{x}\|$ we write (for $\mathbf{x} \neq \mathbf{0}$) its gradient $\nabla g(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|$, a direct calculation shows that

$$\nabla f(\mathbf{x}) = 2\sum_{i=1}^{m}(\|\mathbf{x} - \mathbf{a}_i\| - d_i)\frac{\mathbf{x} - \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|} = 2\left(\sum_{i=1}^{m}(\mathbf{x} - \mathbf{a}_i) - \sum_{i=1}^{m} d_i \frac{\mathbf{x} - \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|}\right).$$

Then, setting $\nabla f(\mathbf{x}) = \mathbf{0}$ leads to

$$\begin{aligned}
\nabla f(\mathbf{x}) &= \mathbf{0} \\
2\left(\sum_{i=1}^{m}(\mathbf{x} - \mathbf{a}_i) - \sum_{i=1}^{m} d_i \frac{\mathbf{x} - \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|}\right) &= \mathbf{0} \\
\sum_{i=1}^{m} \mathbf{x} &= \sum_{i=1}^{m} \mathbf{a}_i + \sum_{i=1}^{m} d_i \frac{\mathbf{x} - \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|} \\
\mathbf{x} &= \frac{1}{m}\left(\sum_{i=1}^{m} \mathbf{a}_i + \sum_{i=1}^{m} d_i \frac{\mathbf{x} - \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|}\right).
\end{aligned}$$

5b) If the iteration is a gradient method, it can be expressed as

$$\mathbf{x}^{k+1} = \frac{1}{m}\left\{\sum_{i=1}^{m} \mathbf{a}_i + \sum_{i=1}^{m} d_i \frac{\mathbf{x}^k - \mathbf{a}_i}{\|\mathbf{x}^k - \mathbf{a}_i\|}\right\} = \mathbf{x}^k + t^k \mathbf{d}^k$$

From part a) we now that

$$\mathbf{x}^k - \frac{1}{m}\left\{\sum_{i=1}^{m} \mathbf{a}_i + \sum_{i=1}^{m} d_i \frac{\mathbf{x}^k - \mathbf{a}_i}{\|\mathbf{x}^k - \mathbf{a}_i\|}\right\} = \frac{1}{2m}\nabla f(\mathbf{x}^k),$$

or rearranging

$$\mathbf{x}^k - \frac{1}{2m}\nabla f(\mathbf{x}^k) = \frac{1}{m}\left\{\sum_{i=1}^{m} \mathbf{a}_i + \sum_{i=1}^{m} d_i \frac{\mathbf{x}_k - \mathbf{a}_i}{\|\mathbf{x}_k - \mathbf{a}_i\|}\right\},$$

that is, the iteration corresponds to gradient descent with constant stepsize $t = \frac{1}{2m}$.

5c) In the Gauss-Newton method the direction $\mathbf{d}^k$ is given by

$$\mathbf{d}^k = (J(\mathbf{x}^k)^\top J(\mathbf{x}^k))^{-1} J(\mathbf{x}^k)^\top F(\mathbf{x}^k)\,,$$

where $F(\mathbf{x})$ in $\mathbb{R}^m$ corresponds to the vector function associated to the cost

$$F(\mathbf{x}) = \begin{bmatrix} \|\mathbf{x}-\mathbf{a}_1\| - d_1 \\ \vdots \\ \|\mathbf{x}-\mathbf{a}_m\| - d_m \end{bmatrix},$$

and $J(\mathbf{x})$ in $\mathbb{R}^{m\times n}$ is the Jacobian matrix given by

$$J(\mathbf{x}) = \begin{bmatrix} \frac{(\mathbf{x}-\mathbf{a}_1)^\top}{\|\mathbf{x}-\mathbf{a}_1\|} \\ \vdots \\ \frac{(\mathbf{x}-\mathbf{a}_m)^\top}{\|\mathbf{x}-\mathbf{a}_m\|} \end{bmatrix}.$$

6a) For a quadratic function, one has that the stepsize when performing an exact line search at the point $\mathbf{x}_k$ in the direction $-\mathbf{d}_k \equiv -\nabla f(\mathbf{x}_k) = -Q\mathbf{x}_k$ is

$$\alpha_k = \frac{\mathbf{d}_k^\top \mathbf{d}_k}{\mathbf{d}_k^\top Q \mathbf{d}_k}$$

Thus, we obtain

$$\begin{aligned}
\mathbf{d}_0 &= Q\left(\frac{1}{\lambda_{\min}}\mathbf{u}_{\min} + \frac{1}{\lambda_{\max}}\mathbf{u}_{\max}\right) = \mathbf{u}_{\min} + \mathbf{u}_{\max} \\
\mathbf{d}_0^\top \mathbf{d}_0 &= (\mathbf{u}_{\min} + \mathbf{u}_{\max})^\top (\mathbf{u}_{\min} + \mathbf{u}_{\max}) = \|\mathbf{u}_{\min}\|^2 + \|\mathbf{u}_{\max}\|^2 = 2 \\
\mathbf{d}_0^\top Q \mathbf{d}_0 &= (\mathbf{u}_{\min} + \mathbf{u}_{\max})^\top (\lambda_{\min}\mathbf{u}_{\min} + \lambda_{\max}\mathbf{u}_{\max}) = \lambda_{\min} + \lambda_{\max}
\end{aligned}$$

Therefore,

$$\alpha_0 = \frac{2}{\lambda_{\min} + \lambda_{\max}}$$

and

$$\begin{aligned}
\mathbf{x}_1 = \mathbf{x}_0 - \alpha_0 \mathbf{d}_0 &= \frac{1}{\lambda_{\min}}\mathbf{u}_{\min} + \frac{1}{\lambda_{\max}}\mathbf{u}_{\max} - \frac{2}{\lambda_{\min} + \lambda_{\max}}(\mathbf{u}_{\min} + \mathbf{u}_{\max}) \\
&= \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}\left(\frac{1}{\lambda_{\min}}\mathbf{u}_{\min} - \frac{1}{\lambda_{\max}}\mathbf{u}_{\max}\right)
\end{aligned}$$

6b) Using the expression for $\mathbf{x}_k$ we obtain

$$\begin{aligned}
\mathbf{x}_k^\top Q \mathbf{x} &= \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}\right)^{2k} \left(\frac{1}{\lambda_{\min}}\mathbf{u}_{\min} + \frac{(-1)^k}{\lambda_{\max}}\mathbf{u}_{\max}\right)^\top Q \left(\frac{1}{\lambda_{\min}}\mathbf{u}_{\min} + \frac{(-1)^k}{\lambda_{\max}}\mathbf{u}_{\max}\right) \\
&= \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}\right)^{2k} \left(\frac{1}{\lambda_{\min}}\mathbf{u}_{\min} + \frac{(-1)^k}{\lambda_{\max}}\mathbf{u}_{\max}\right)^\top (\mathbf{u}_{\min} + (-1)^k \mathbf{u}_{\max}) \\
&= \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}\right)^{2k} \left(\frac{1}{\lambda_{\min}} + \frac{(-1)^{2k}}{\lambda_{\max}}\right).
\end{aligned}$$

The expression for $f(\mathbf{x}_{k+1})$ follows analogously evaluating at $k+1$, and noting that $(-1)^{2k} = (-1)^{2k+2}$, we conclude

$$\frac{f(\mathbf{x}_{k+1})}{f(\mathbf{x}_k)} = \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}\right)^2 .$$

This indicates the value of the function decreases by a factor of

$$\left(\frac{\kappa - 1}{\kappa + 1}\right)^2 ,$$

where $\kappa > 1$. The closer κ gets to 1, the faster the method. As κ increases, the method becomes slower.