Chapter 4
The Richardson Iteration

Remark 4.1. Motivation. The Richardson iteration by itself is not of that much interest in practice. However, it provides the idea for a tool that is then used for the construction of advanced iterative methods, namely the consideration of Krylov subspaces. □

Definition 4.2. Richardson iteration. Let $\underline{x}^{(0)} \in \mathbb{R}^n$ be a given initial iterate. The Richardson[1] iteration for computing a sequence of vectors $\underline{x}^{(k)} \in \mathbb{R}^n, k = 0, 1, 2, \ldots$, has the form

$$\underline{r}^{(k)} = \underline{b} - A\underline{x}^{(k)}, \quad \underline{x}^{(k+1)} = \underline{x}^{(k)} + \alpha_k \underline{r}^{(k)} \tag{4.1}$$

with appropriately chosen numbers $\alpha_k \in \mathbb{R}$. The vector $\underline{r}^{(k)}$ is called residual. A straightforward calculation shows that this method can be written as fixed-point iteration in the following form

$$\underline{x}^{(k+1)} = (I - \alpha_k A)\, \underline{x}^{(k)} + \alpha_k \underline{b}.$$

□

Definition 4.3. Co-domain of a matrix. The set

$$\mathcal{R}(A) = \left\{ \frac{\underline{y}^* A \underline{y}}{\underline{y}^* \underline{y}} \ : \ \underline{y} \in \mathbb{C}^n, \underline{y} \neq \underline{0} \right\} \subset \mathbb{C}$$

is called co-domain of A. □

Remark 4.4. On the co-domain of a matrix. The co-domain of A can be defined by using only the vectors from the unit sphere of $\mathbb{C}^n$, since

[1] Lewis Fry Richardson (1881 – 1953)

$$\frac{\underline{y}^* A \underline{y}}{\underline{y}^* \underline{y}} = \frac{\underline{y}^* A \underline{y}}{\left\|\underline{y}^*\right\|_2 \left\|\underline{y}\right\|_2} = \underbrace{\frac{\underline{y}^*}{\left\|\underline{y}^*\right\|_2}}_{\|\cdot\|_2=1} A \underbrace{\frac{\underline{y}}{\left\|\underline{y}\right\|_2}}_{\|\cdot\|_2=1}.$$

The unit sphere is a compact set (bounded and closed) and the mapping $\underline{y} \mapsto \underline{y}^* A \underline{y} / \underline{y}^* \underline{y}$ is continuous. It follows that $\mathcal{R}(A)$ is also a compact set, see literature. □

Lemma 4.5. Co-domain of the inverse matrix. *Let $A \in \mathbb{R}^{n\times n}$ with $\mathcal{R}(A) \subset \{\lambda \in \mathbb{C} \ : \ Re(\lambda) > 0\}$, i.e., the co-domain of A is a subset of the right half of the complex plane. Then*

$$\mathcal{R}\left(A^{-1}\right) \subset \{\lambda \in \mathbb{C} \ : \ Re(\lambda) > 0\}.$$

Proof. From the assumption, it follows that A is non-singular. Otherwise, there would be a vector $\underline{z} \in \ker(A)$, $\underline{z} \neq \underline{0}$, and

$$\mathrm{Re}\left(\frac{\underline{z}^* \overbrace{A\underline{z}}^{=0}}{\underline{z}^* \underline{z}}\right) = \mathrm{Re}\,(0) = 0.$$

This statement contradicts the assumption on $\mathcal{R}(A)$.

Let $\underline{y} \in \mathbb{C}^n$, $\underline{y} \neq \underline{0}$, be arbitrary and $\underline{z} = A^{-1}\underline{y} \neq \underline{0}$. Hence, $\underline{z}$ is also an arbitrary vector. Using the definition of $\underline{y}$, that the real parts of a complex number and of its conjugate are the same, and $\|A\underline{z}\|_2 \le \|A\|_2 \|\underline{z}\|_2$ yields

$$\begin{aligned}
\mathrm{Re}\left(\frac{\underline{y}^* A^{-1} \underline{y}}{\underbrace{\underline{y}^* \underline{y}}_{\in\mathbb{R}}}\right) &= \frac{1}{\|\underline{y}\|_2^2} \mathrm{Re}\left(\underline{y}^* A^{-1} \underline{y}\right) = \frac{1}{\|A\underline{z}\|_2^2} \mathrm{Re}((A\underline{z})^* \underbrace{A^{-1}A}_{=I}\underline{z}) \\
&= \frac{1}{\|A\underline{z}\|_2^2} \mathrm{Re}(\underbrace{\underline{z}^* A^* \underline{z}}_{\in\mathbb{C}}) = \frac{1}{\|A\underline{z}\|_2^2} \mathrm{Re}\left(\left(\underline{z}^* A^* \underline{z}\right)^*\right) = \frac{1}{\|A\underline{z}\|_2^2} \mathrm{Re}\left(\underline{z}^* A \underline{z}\right) \\
&= \frac{\|\underline{z}\|_2^2}{\|A\underline{z}\|_2^2} \mathrm{Re}\left(\frac{\underline{z}^* A \underline{z}}{\underline{z}^* \underline{z}}\right) \ge \frac{1}{\|A\|_2^2} \mathrm{Re}\left(\frac{\underline{z}^* A \underline{z}}{\underline{z}^* \underline{z}}\right) > 0,
\end{aligned}$$

where the last statement follows from the assumption on $\mathcal{R}(A)$. ■

Theorem 4.6. Convergence of the Richardson iteration. *Let $A \in \mathbb{R}^{n\times n}$ with $\mathcal{R}(A) \subset \{\lambda \in \mathbb{C} \ : \ Re(\lambda) > 0\}$. Then the Richardson iteration (4.1) converges to the solution of the linear system $A\underline{x} = \underline{b}$ for every initial iterate if $\alpha_k = \alpha$, $k = 0, 1, 2, \ldots$, with*

$$0 < \alpha < \min\left\{\beta = Re(\lambda),\ \lambda \in \mathcal{R}\left(A^{-1}\right)\right\}. \tag{4.2}$$

Proof. Note that $\mathcal{R}(A^{-1})$ is a compact set such that the minimum in (4.2) exists. By Lemma 4.5, the minimum is positive such that a positive value α as given in (4.2) exists.

Let $\underline{x}$ be the solution of (1.1). It will be shown that the error $\left\|\underline{x} - \underline{x}^{(k)}\right\|_2$ decreases strongly monotonically and the rate of decrease is strictly lower (uniformly with respect to k) than one. Using (4.1) and $\underline{b} = A\underline{x}$, one has the recursion

$$\begin{aligned}\underline{x} - \underline{x}^{(k+1)} &= \underline{x} - \underline{x}^{(k)} - \alpha \underline{r}^{(k)} = \underline{x} - \underline{x}^{(k)} - \alpha \left(\underline{b} - A\underline{x}^{(k)}\right) \\ &= \underline{x} - \underline{x}^{(k)} - \alpha A \left(\underline{x} - \underline{x}^{(k)}\right).\end{aligned}$$

Hence, it is

$$\begin{aligned}\left\|\underline{x} - \underline{x}^{(k+1)}\right\|_2^2 &= \left(\underline{x} - \underline{x}^{(k)} - \alpha A \left(\underline{x} - \underline{x}^{(k)}\right), \underline{x} - \underline{x}^{(k)} - \alpha A \left(\underline{x} - \underline{x}^{(k)}\right)\right) \\ &= \left\|\underline{x} - \underline{x}^{(k)}\right\|_2^2 - 2\alpha \left(\underline{x} - \underline{x}^{(k)}\right)^T A \left(\underline{x} - \underline{x}^{(k)}\right) + \alpha^2 \left\|A \left(\underline{x} - \underline{x}^{(k)}\right)\right\|_2^2.\end{aligned} \tag{4.3}$$

Denoting $\underline{y} = A\left(\underline{x} - \underline{x}^{(k)}\right)$, using that the transposed of a real number is the same number and (4.2), one obtains

$$\begin{aligned}\frac{\left(\underline{x} - \underline{x}^{(k)}\right)^T A \left(\underline{x} - \underline{x}^{(k)}\right)}{\left\|A\left(\underline{x} - \underline{x}^{(k)}\right)\right\|_2^2} &= \frac{\left(\underline{x} - \underline{x}^{(k)}\right)^T A^T A^{-T} A \left(\underline{x} - \underline{x}^{(k)}\right)}{\left\|A\left(\underline{x} - \underline{x}^{(k)}\right)\right\|_2^2} = \frac{\overbrace{\underline{y}^T A^{-T} \underline{y}}^{\in\mathbb{R}}}{\underline{y}^T \underline{y}} \\ &= \frac{\underline{y}^T A^{-1} \underline{y}}{\underline{y}^T \underline{y}} \ge \min\left\{\mathrm{Re}(\lambda) \ : \ \lambda \in \mathcal{R}\left(A^{-1}\right)\right\} > \alpha, \\ &\Longleftrightarrow \\ \alpha^2 \left\|A\left(\underline{x} - \underline{x}^{(k)}\right)\right\|_2^2 &< \alpha \left(\underline{x} - \underline{x}^{(k)}\right)^T A \left(\underline{x} - \underline{x}^{(k)}\right).\end{aligned}$$

Applying this estimate to the last term of (4.3) yields

$$\begin{aligned}\left\|\underline{x} - \underline{x}^{(k+1)}\right\|_2^2 &< \left\|\underline{x} - \underline{x}^{(k)}\right\|_2^2 - \alpha \left(\underline{x} - \underline{x}^{(k)}\right)^T A \left(\underline{x} - \underline{x}^{(k)}\right) \\ &= \left\|\underline{x} - \underline{x}^{(k)}\right\|_2^2 \left(1 - \alpha \frac{\left(\underline{x} - \underline{x}^{(k)}\right)^T A \left(\underline{x} - \underline{x}^{(k)}\right)}{\left\|\underline{x} - \underline{x}^{(k)}\right\|_2^2}\right).\end{aligned} \tag{4.4}$$

Since $\mathcal{R}(A)$ is compact, there is a $\varepsilon > 0$ such that $\mathrm{Re}(\lambda) \ge \varepsilon$ for all $\lambda \in \mathcal{R}(A)$ (there is no sequence that can converge to the imaginary axis). Hence, it holds that

$$\frac{\left(\underline{x} - \underline{x}^{(k)}\right)^T A \left(\underline{x} - \underline{x}^{(k)}\right)}{\left\|\underline{x} - \underline{x}^{(k)}\right\|_2^2} \ge \varepsilon.$$

Choosing ε such that $\alpha\varepsilon < 1$, then it follows from (4.4) that

$$\left\|\underline{x} - \underline{x}^{(k+1)}\right\|_2^2 < \left\|\underline{x} - \underline{x}^{(k)}\right\|_2^2 (1 - \alpha\varepsilon) =: q \left\|\underline{x} - \underline{x}^{(k)}\right\|_2^2$$

with q independent of k and $q \in (0,1)$. One obtains by induction

$$\left\|\underline{x} - \underline{x}^{(k)}\right\|_2 \le q^{k/2} \left\|\underline{x} - \underline{x}^{(0)}\right\|_2$$

such that $\underline{x}^{(k)} \to \underline{x}$ as $k \to \infty$. ■

Remark 4.7. Choice of α for s.p.d. matrices. Let A be symmetric and positive definite. Using the Rayleigh quotient (2.4) yields for an arbitrary vector $\underline{y} \in \mathbb{C}^n$

$$
\begin{aligned}
&\frac{\mathrm{Re}\left(\underline{y}^* A^{-1}\underline{y}\right)}{\|\underline{y}\|_2^2} \\
&= \frac{1}{\|\underline{y}\|_2^2}\left(\left(\mathrm{Re}\left(\underline{y}\right)\right)^T A^{-1}\mathrm{Re}\left(\underline{y}\right) + \left(\mathrm{Im}\left(\underline{y}\right)\right)^T A^{-1}\mathrm{Im}\left(\underline{y}\right)\right) \\
&= \frac{1}{\|\underline{y}\|_2^2}\left(\left\|\mathrm{Re}\left(\underline{y}\right)\right\|_2^2 \frac{\left(\mathrm{Re}\left(\underline{y}\right)\right)^T A^{-1}\mathrm{Re}\left(\underline{y}\right)}{\left\|\mathrm{Re}\left(\underline{y}\right)\right\|_2^2}\right. \\
&\quad \left. + \left\|\mathrm{Im}\left(\underline{y}\right)\right\|_2^2 \frac{\left(\mathrm{Im}\left(\underline{y}\right)\right)^T A^{-1}\mathrm{Im}\left(\underline{y}\right)}{\left\|\mathrm{Im}\left(\underline{y}\right)\right\|_2^2}\right) \\
&\geq \frac{1}{\|\underline{y}\|_2^2}\left(\left\|\mathrm{Re}\left(\underline{y}\right)\right\|_2^2 \lambda_{\min}\left(A^{-1}\right) + \left\|\mathrm{Im}\left(\underline{y}\right)\right\|_2^2 \lambda_{\min}\left(A^{-1}\right)\right) \\
&= \lambda_{\min}\left(A^{-1}\right) = \frac{1}{\lambda_{\max}\left(A\right)} = \frac{1}{\rho\left(A\right)}.
\end{aligned}
$$

That means, the choice $\alpha < 1/\rho\left(A\right)$ guarantees the convergence of the Richardson method. □

Remark 4.8. Residual minimization for choosing α_k. One possibility to choose α_k in practice consists in the minimization of the norm of the residual

$$
\begin{aligned}
\left\|\underline{r}^{(k+1)}\right\|_2^2 &= \left\|\underline{b} - A\underline{x}^{(k+1)}\right\|_2^2 = \left\|\underline{b} - A\underline{x}^{(k)} - \alpha_k A\underline{r}^{(k)}\right\|_2^2 = \left\|\underline{r}^{(k)} - \alpha_k A\underline{r}^{(k)}\right\|_2^2 \\
&= \left\|\underline{r}^{(k)}\right\|_2^2 - \alpha_k \left(\underline{r}^{(k)}\right)^T A\underline{r}^{(k)} - \alpha_k \underbrace{\left(A\underline{r}^{(k)}\right)^T \underline{r}^{(k)}}_{\in\mathbb{R}} + \alpha_k^2 \left\|A\underline{r}^{(k)}\right\|_2^2 \\
&= \left\|\underline{r}^{(k)}\right\|_2^2 - 2\alpha_k \left(\underline{r}^{(k)}\right)^T A\underline{r}^{(k)} + \alpha_k^2 \left\|A\underline{r}^{(k)}\right\|_2^2.
\end{aligned}
$$

The necessary condition for a minimum

$$
\frac{d}{d\alpha_k}\left\|\underline{r}^{(k+1)}\right\|_2^2 = -2\left(\underline{r}^{(k)}\right)^T A\underline{r}^{(k)} + 2\alpha_k \left\|A\underline{r}^{(k)}\right\|_2^2 = 0
$$

gives

$$
\alpha_k = \frac{\left(\underline{r}^{(k)}\right)^T A\underline{r}^{(k)}}{\left\|A\underline{r}^{(k)}\right\|_2^2}. \tag{4.5}
$$

Since

$$\frac{d^2}{d\alpha_k^2}\left\|\underline{r}^{(k+1)}\right\|_2^2 = 2\left\|A\underline{r}^{(k)}\right\|_2^2 > 0,$$

if $\underline{r}^{(k)} \neq \underline{0}$, one obtains in fact a minimum. $\Box$

Remark 4.9. Spaces spanned by the iterates. It is by (4.1)

$$\begin{aligned}\underline{x}^{(1)} &\in \underline{x}^{(0)} + \operatorname{span}\left\{\underline{r}^{(0)}\right\},\\ \underline{x}^{(2)} &\in \underline{x}^{(1)} + \operatorname{span}\left\{\underline{r}^{(1)}\right\} = \underline{x}^{(0)} + \operatorname{span}\left\{\underline{r}^{(0)}, \underline{r}^{(1)}\right\}.\end{aligned}$$

It holds

$$\underline{r}^{(1)} = \underline{b} - A\underline{x}^{(1)} = \underline{b} - A\underline{x}^{(0)} - \alpha_0 A\underline{r}^{(0)} = \underline{r}^{(0)} - \alpha_0 A\underline{r}^{(0)}$$

and consequently

$$\underline{x}^{(2)} \in \underline{x}^{(0)} + \operatorname{span}\left\{\underline{r}^{(0)}, A\underline{r}^{(0)}\right\}.$$

One obtains by induction

$$\underline{x}^{(k)} \in \underline{x}^{(0)} + \operatorname{span}\left\{\underline{r}^{(0)}, A\underline{r}^{(0)}, \ldots, A^{k-1}\underline{r}^{(0)}\right\}.$$

$\Box$

Definition 4.10. Krylov subspace. Let $\underline{q} \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$. Then, the space

$$K_m(\underline{q}, A) := \operatorname{span}\left\{\underline{q}, A\underline{q}, \ldots, A^{m-1}\underline{q}\right\}$$

is called the Krylov[2] subspace of order m that is spanned by $\underline{q}$ and A. $\Box$

Remark 4.11. Next goal. It holds that $\underline{x}^{(k)} \in \underline{x}^{(0)} + K_k\left(\underline{r}^{(0)}, A\right)$. In the following, Richardson's method will be improved by constructing the iterates $\underline{x}^{(k)}$ in this manifold with respect to certain optimality criteria. $\Box$

[2] Aleksey Nikolaevich Krylov (1863 – 1945)