Convergence result for the method of steepest descent

Since $\boldsymbol{A}$ is self-adjoint w.r.t. $(\cdot,\cdot)$, the bilinear form

$$(\boldsymbol{y},\,\boldsymbol{z})_{\boldsymbol{A}} \;:=\; (\boldsymbol{A}\,\boldsymbol{y},\,\boldsymbol{z})$$

is an inner product. The corresponding norm $\|\boldsymbol{y}\|_{\boldsymbol{A}} := \sqrt{(\boldsymbol{y},\,\boldsymbol{y})_{\boldsymbol{A}}}$ is called **energy norm**.

Lemma 1.58 The method of steepest descent converges q-linearly with

$$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}\|_{\boldsymbol{A}} \;\le\; \boldsymbol{q}\,\|\boldsymbol{x}_k - \boldsymbol{x}\|_{\boldsymbol{A}}$$

where $\boldsymbol{q} = \frac{\kappa-1}{\kappa+1}$ and $\boldsymbol{\kappa} = \frac{\mu_2}{\mu_1}$ for $\boldsymbol{\mu_1} \le \boldsymbol{\lambda}_{\min}(\boldsymbol{A}),\ \boldsymbol{\mu_2} \ge \boldsymbol{\lambda}_{\max}(\boldsymbol{A})$.

Proof: (1) Auxiliary identity for arbitrary $y \in \mathbb{R}^n$:

$$\begin{aligned}
\|y - x\|_A^2 &= (A\,y,\,y) - 2(A\,x,\,y) + (A\,x,\,x) \;=\; (A\,y,\,y) - 2(\underbrace{A\,x}_{=b},\,x) - (A\,x,\,x) + 2(\underbrace{A\,x}_{=b},\,x) \\
&= (A\,y,\,y) - 2(b,\,y) - [(A\,x,\,x) - 2(b,\,x)] \;=\; 2[J_A(y) - J_A(x)]
\end{aligned}$$

(2) After k steps in the method of steepest descent, we perform one hypothetical step of Richardson's method:

$$\widetilde{x}_{k+1} \;:=\; x_k + \tau\, r_k \;=\; x_k + \tau\, p_k$$

with the optimal parameter τ.
(3) We use Lemma 1.47 to analyze the hypothetical Richardson step.
It is easily seen that A is self-adjoint w.r.t. $(\cdot,\cdot)_A$ and that

$$\begin{aligned}
& \mu_1\,(y,\,y) \le (A\,y,\,y) \le \mu_2\,(y,\,y) \qquad \forall y \in \mathbb{R}^n \\
\overset{\text{Lemma 1.49}}{\iff}\quad & \mu_1 \le \lambda_{\min}(A), \quad \mu_2 \ge \lambda_{\max}(A) \\
\iff\quad & \mu_1\,(y,\,y)_A \le (A\,y,\,y)_A \le \mu_2\,(y,\,y)_A \qquad \forall y \in \mathbb{R}^n
\end{aligned}$$

Hence, Lemma 1.47 (with $\|\cdot\| \mapsto \|\cdot\|_A$) yields:

$$\|\widetilde{x}_{k+1} - x\|_A \;\le\; q\,\|x_k - x\|_A \quad \text{where} \quad q = \frac{\kappa-1}{\kappa+1} \quad \text{and} \quad \kappa = \frac{\mu_2}{\mu_1}\,.$$

(4) Recall that x_{k+1} is the next step in the method of steepest descent. Because of the optimal choice of α_k, we have $J_A(x_{k+1}) \le J_A(\widetilde{x}_{k+1})$. Therefore, with the identity of Part (1):

$$\begin{aligned}
\|x_{k+1} - x\|_A^2 &= 2[J_A(x_{k+1}) - J_A(x)] \\
&\le 2[J_A(\widetilde{x}_{k+1}) - J_A(x)] \;=\; \|\widetilde{x}_{k+1} - x\|_A^2 \;\le\; q^2\,\|x_k - x\|_A^2
\end{aligned}$$

This concludes the proof. □

Remark 1.59

1. The method of steepest descent converges as fast as Richardson's method but does not need a-priori information μ_1, μ_2 or the choice of the damping parameter τ. The (optimal) stepsize α_k is computed automatically using just r_k and A.

2. One can show that the method of steepest descent converges in general not faster than Richardson's method (with $\tau = \tau_{\text{opt}}$).

3. The statement of Lemma 1.58 only holds for the energy norm $\|\cdot\|_A$!