Chapter 9
Gradient Method

Abstract The gradient method is an optimisation method of greedy type. For this purpose, the system of equations has to be rewritten as a minimisation problem (see Section 9.1). The gradient method $\varUpsilon_{\mathrm{grad}}[\varPhi]$ derived in Section 9.2 determines the damping factors of the underlying iteration $\varPhi \in \mathcal{L}$. It turns out that the convergence is not faster than the optimally damped version $\varPhi_{\vartheta_{\mathrm{opt}}}$ of $\varPhi$, but the method can be applied without knowing the spectral values determining ϑ_{opt}. In Section 9.3 we discuss the drawback of the gradient directions and introduce the conjugate directions in preparation for the conjugate gradient method in the next chapter. The final Section 9.4 mentions a variant of the gradient method: the minimal residual iteration which can be applied to any regular matrix A.

9.1 Reformulation as Minimisation Problem

9.1.1 Minimisation Problem

In the following, $A \in \mathbb{R}^{I\times I}$ and $b \in \mathbb{R}^{I}$ are real. We consider a system

$$Ax = b$$

and assume that

$$A \text{ is positive definite.} \tag{9.1}$$

System $Ax = b$ is associated with the function

$$F(x) := \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle . \tag{9.2}$$

The derivative (gradient) of F is $F'(x) = \frac{1}{2}(A + A^{\mathsf{T}})x - b$. Since $A = A^{\mathsf{T}}$ by assumption[1] (9.1), the derivative is equal to

[1] Under the weaker assumption (C.2), the function F can also be minimised, but the minimiser would not be the solution of $Ax = b$; i.e., the method would be inconsistent.

W. Hackbusch, *Iterative Solution of Large Sparse Systems of Equations*,
Applied Mathematical Sciences 95, DOI 10.1007/978-3-319-28483-5_9

$$F'(x) = \operatorname{grad} F(x) = Ax - b.$$

A necessary condition for a *minimum* of F is the vanishing of the gradient: $Ax = b$. Since the Hessian matrix $F''(x) = (F_{x_i x_j})_{i,j \in I} = A$ is positive definite, the solution of $Ax = b$ (in the following denoted by x^*) in fact leads to a minimum. This proves the next lemma.

Lemma 9.1. *Let $A \in \mathbb{R}^{I \times I}$ be positive definite. The solution of the system $Ax = b$ is equivalent to the solution to the minimisation problem*

$$F(x) \overset{!}{=} \min .$$

A second proof of Lemma 9.1 results from the representation

$$F(x) = F(x^*) + \frac{1}{2}\bigl\langle A(x - x^*), x - x^* \bigr\rangle \qquad \text{with } x^* := A^{-1}b. \tag{9.3}$$

This equation proves $F(x) > F(x^*)$ for $x \neq x^*$, i.e., $x^* = A^{-1}b$ is the unique minimiser of F. The representation (9.3) is a particular case of the following Taylor expansion of F around an arbitrary value $\tilde{x} \in \mathbb{R}^I$:

$$F(x) = F(\tilde{x}) + \bigl\langle A\tilde{x} - b, x - \tilde{x} \bigr\rangle + \frac{1}{2}\bigl\langle A(x - \tilde{x}), x - \tilde{x} \bigr\rangle. \tag{9.4}$$

9.1.2 Search Directions

In the following, the minimisation of F with respect to a particular direction $p \in \mathbb{R}^I \backslash \{0\}$ plays a central role. Optimisation over all $x \in \mathbb{R}^I$ is replaced by the one-dimensional minimisation problem (9.5a,b):

$$f(\lambda) \overset{!}{=} \min \quad \text{for the function} \tag{9.5a}$$

$$f(\lambda) := F(x + \lambda p) \qquad (x, p \in \mathbb{R}^I \text{ fixed, } \lambda \in \mathbb{R}). \tag{9.5b}$$

Replacing the variables x and $\tilde{x}$ in (9.4) by $x + \lambda p$ and x, we obtain that

$$f(\lambda) = F(x) + \lambda\bigl\langle Ax - b, p \bigr\rangle + \frac{\lambda^2}{2}\bigl\langle Ap, p \bigr\rangle. \tag{9.5c}$$

$p \neq 0$ implies that $\langle Ap, p \rangle > 0$ (cf. (9.1)); hence, the minimum of the parabola f can be determined from $f'(\lambda) = 0$.

Lemma 9.2. *Assume $p \neq 0$ and (9.1): $A > 0$. The unique minimum of problem (9.5a,b) is attained at*

$$\lambda = \lambda_{\text{opt}}(r, p, A) := \frac{\langle r, p \rangle}{\langle Ap, p \rangle}, \tag{9.6a}$$

where

$$r := b - Ax .$$

In the following, the letter r always denotes the residual (residue) $b - Ax$ of the actual x. It is the negative defect $Ax - b$ and also the negative gradient $F' = Ax - b$.

The *optimal* search direction is evidently $p = x^* - x$ (or a nonvanishing multiple) because $f(\lambda_{\rm opt}) = F(x^*)$ yields the global minimum. However, since $p = x^* - x$ requires knowledge of the solution, another proposal is needed. Let p be normalised by $\|p\|_2 = 1$. The directional derivative $f'(0) = -\langle r, p\rangle = \langle \mathrm{grad}\ F(x), p\rangle$ at $\lambda = 0$ is maximal for the gradient direction $p = -r/\|r\|_2$ and minimal for the reverse direction $p = r/\|r\|_2$. The vector $\mathrm{grad}\ F(x) = -r$ is the direction of the steepest ascent, while the residual r is the direction of the *steepest descent*. This consideration shows the optimality of $p = r$ from a *local* point of view. For $p = r$, the expression (9.6a) becomes

$$\lambda = \lambda_{\rm opt}(r, r, A) = \frac{\|r\|_2^2}{\langle Ar, r\rangle} \qquad \text{for } r := b - Ax \neq 0. \tag{9.6b}$$

The definition

$$\lambda_{\rm opt}(r, 0, A) := 0 \tag{9.6c}$$

is added for formal reasons only: now $\lambda_{\rm opt}(\cdot, \cdot, A)$ is defined for all arguments. As soon as $r = 0$ occurs, x is already the exact solution x^*.

9.1.3 Other Quadratic Functionals

The function F in (9.2) is not the only quadratic function having $x^* := A^{-1}b$ as the minimiser.

Lemma 9.3. *(a) Any quadratic form with a unique minimum at* $x^* = A^{-1}b$ *has the form*

$$F(x) = \tfrac{1}{2}\big\langle HA(x - x^*), A(x - x^*)\big\rangle + c = \tfrac{1}{2}\big\langle H(Ax - b), Ax - b\big\rangle + c \tag{9.7a}$$

with an arbitrary constant c *and*

$$H > 0. \tag{9.7b}$$

Here, in contrast to (9.1), A *may be any regular matrix.*

(b) To ensure that the calculation of $\mathrm{grad}\ F(x) = A^{\mathsf H} HA(x - x^*) = -A^{\mathsf H} Hr$ *from the residual* $r = Ax - b$ *be practical, the matrix* H *must be such that the matrix-vector multiplication* $r \mapsto A^{\mathsf H} Hr$ *is feasible.*

(c) Under assumption (9.1), $H := A^{-1}$ *and* $c := \frac{1}{2}\langle b, x^*\rangle$ *may be chosen. Then* F *in (9.7a) coincides with* F *in (9.2).*

Proof of (c). By (9.1), $H = A^{-1}$ satisfies (9.7b) (cf. Lemma C.4b). A comparison of (9.7a) and (9.3) shows that $c = F(x^*) = \frac{1}{2}\langle Ax^*, x^*\rangle - \langle b, x^*\rangle = -\frac{1}{2}\langle b, x^*\rangle$. □

Conclusion 9.4. *Let A be positive definite. The (energy) scalar product $\langle\cdot,\cdot\rangle_A$ and (energy) norm $\|\cdot\|_A$ are defined by (9.8a):*

$$\langle x,y\rangle_A := \langle Ax,y\rangle\,, \qquad \|x\|_A := \|A^{1/2}x\|_2 = \sqrt{\langle x,x\rangle_A}\,. \tag{9.8a}$$

The minimisation of F in (9.2) is equivalent to the minimisation problem

$$\|x-x^*\|_A \overset{!}{=} \min \qquad \textit{with } x^* := A^{-1}b\,. \tag{9.8b}$$

Proof. Problem (9.8b) may be replaced with $\|x-x^*\|_A^2 \overset{!}{=} \min$. The identity

$$\|x-x^*\|_A^2 = 2\left[F(x)-F(x^*)\right] \qquad \text{(cf. (9.3))} \tag{9.8c}$$

completes the proof. □

Remark 9.5. (a) For the choice $H = I$ and $c = 0$, equation (9.7a) becomes $F(x) = \frac{1}{2}\|Ax-b\|_2^2$ and describes the *least-squares minimisation.*
(b) For $H = A^{-\mathsf{H}}A^{-1} > 0$ and $c = 0$, the identity $F(x) = \frac{1}{2}\|x-x^*\|_2^2$ holds.
(c) For a positive definite K, the minimisation of the norm

$$\|x-x^*\|_K^2 = \|K^{1/2}(x-x^*)\|_2^2$$

corresponds to problem (9.7a) with

$$H = \frac{1}{2}A^{-\mathsf{H}}KA^{-1}, \qquad c = 0\,.$$

According to Lemma 9.3b, multiplying by KA^{-1} must be feasible.

Remark 9.6. Any iteration converging (weakly) monotonically with respect to the norm $\|\cdot\|_A$ leads to a descent sequence

$$F(x^0) \geq F(x^1) \geq \dots\,.$$

9.1.4 Complex Case

In the complex case of $A \in \mathbb{C}^{I\times I}$ and $b \in \mathbb{C}^I$, the function F can again be defined by (9.7a,b), provided that c in (9.7a) is real. Definition (9.2) cannot be generalised without change, since only real functions F can be minimised and, in general, F is not real because of the term $\langle b,x\rangle$. One has to replace F in (9.2) by

$$F(x) := \frac{1}{2}\langle Ax,x\rangle - \mathfrak{Re}\,\langle b,x\rangle \qquad \text{for } x\in\mathbb{C}^I. \tag{9.9a}$$

Exercise 9.7. Assume (9.1) and let F be defined by (9.9a). Prove the following:

(a) F is real, and $\mathfrak{Re}\langle b, x^* \rangle = \langle b, x^* \rangle$ holds for $x^* = A^{-1}b$.

(b) Equations (9.9b,c) hold for $x, y \in \mathbb{K}^I$:

$$F(x) = \tfrac{1}{2}\left[\,\langle A(x - x^*), x - x^* \rangle - \langle b, x^* \rangle\,\right], \tag{9.9b}$$

$$F(x) = F(y) + \mathfrak{Re}\,\langle Ay - b, x - y \rangle + \tfrac{1}{2}\,\langle A(x - y), x - y \rangle\,. \tag{9.9c}$$

(c) The minimum of $f(\lambda) = F(x + \lambda p)$ over $\lambda \in \mathbb{C}$ with F in (9.9a) is attained for the value $\lambda_{\mathrm{opt}}(r, p, A)$ in (9.6a), which in general is complex.

9.2 Gradient Method

Another name for the gradient method is the *method of steepest descent.*

9.2.1 Construction

In general, the gradient method is an algorithm for solving a minimisation problem $F(x) = \min$ with a differentiable function F: $\mathbb{R}^I \to \mathbb{R}$ (cf., e.g., Kosmol [240, §4], Quarteroni–Sacco–Saleri [314, §7.2.2]). We apply the gradient method only to the quadratic function F in (9.2) or (9.7a).

The gradient method minimises F iteratively in the direction of the *steepest descent*:

$$x^0 \in \mathbb{R}^I\text{:}\quad \text{arbitrary starting iterate,} \tag{9.10a}$$

iteration $m = 0, 1, \ldots$:

$$r^m := b - Ax^m, \tag{9.10b}$$

$$x^{m+1} := x^m + \lambda_{\mathrm{opt}}(r^m, r^m, A) r^m. \tag{9.10c}$$

The representation

$$r^{m+1} = b - Ax^{m+1} = b - A(x^m + \lambda_{\mathrm{opt}} r^m) = r^m - \lambda_{\mathrm{opt}} A\, r^m$$

allows the following update of the residual:

$$\boxed{\begin{array}{ll} \textbf{start:} & x^0\text{: arbitrary,}\quad r^0 := b - Ax^0, \\ \textbf{iteration } m = 0, 1, \ldots : & x^{m+1} := x^m + \lambda_{\mathrm{opt}}(r^m, r^m, A)\, r^m, \\ & r^{m+1} := r^m - \lambda_{\mathrm{opt}}(r^m, r^m, A)\, A r^m \end{array}}\quad \begin{array}{r} (9.11\text{a}) \\ (9.11\text{b}) \\ (9.11\text{c}) \end{array}$$

with $\lambda_{\mathrm{opt}}(r^m, r^m, A)$ in (9.6b,c). The advantage of (9.11c) over (9.10b) is the fact that the product Ar^m is already calculated in (9.6b) when λ_{opt} is determined.

The gradient method (9.11a–c) is denoted by $\varUpsilon_{\mathrm{grad}}[\varPhi_1^{\mathrm{Rich}}]$ (cf. §9.2.4).

9.2.2 Properties of the Gradient Method

Remark 9.8. Assume (9.1). (a) In contrast to the previous methods, the iteration $x^m \mapsto \Phi(x^m, b)$ defined in (9.11a–c) is not linear.

(b) $\Phi(\cdot,\cdot)$ is continuous with respect to both of its arguments.

(c) The gradient method is consistent and convergent.

Proof. (a) $\lambda_{\mathrm{opt}}(r^m, r^m) = \lambda_{\mathrm{opt}}(b - Ax^m, b - Ax^m)$ is a nonconstant function of x^m and b. Hence, $\Phi(x, b) = x + \lambda_{\mathrm{opt}}(b - Ax, b - Ax, A)(b - Ax)$ is not linear.

(c) Convergence will be proved in Theorem 9.10. If x^* is a solution of $Ax = b$, the residual r vanishes. Together with (9.6c), we conclude that $\Phi(x^*, b) = x^*$, i.e., F is consistent. □

Although the gradient method is not linear, it can be interpreted as a semi-iterative method applied to a linear basic iteration.

Remark 9.9. The sequence $\{x^m\}$ of the gradient method (9.11a–c) is identical to the sequence $\{y^m\}$ of the semi-iterative Richardson method

$$y^{m+1} = y^m - \Theta_{m+1}\,(Ay^m - b) \;=\; \Phi^{\mathrm{Rich}}_{\Theta_{m+1}}(y^m, b) \tag{9.12}$$

(cf. (8.10b)), if one chooses $y^0 = x^0$ and fixes the factors Θ_{m+1} by

$$\Theta_{m+1} := \lambda_{\mathrm{opt}}(r^m, r^m, A) \qquad \text{with } r^m := b - Ax^m.$$

Theorem 9.10 (convergence). *Let A be positive definite and denote the extreme eigenvalues of A by $\lambda = \lambda_{\min}(A)$ and $\Lambda = \lambda_{\max}(A)$. Let F be defined by (9.2). Then, for any starting iterate x^0, the sequence $\{x^m\}$ of the gradient method converges to the solution $x^* = A^{-1}b$ and satisfies the error estimates*

$$F(x^m) - F(x^*) \le \left(\frac{\Lambda - \lambda}{\Lambda + \lambda}\right)^{2m} \left[F(x^0) - F(x^*)\right], \tag{9.13a}$$

$$\|x^m - x^*\|_A \le \left(\frac{\Lambda - \lambda}{\Lambda + \lambda}\right)^{m} \|x^0 - x^*\|_A\,. \tag{9.13b}$$

Proof. (i) By (9.8c), the estimates (9.13a) and (9.13b) are equivalent.

(ii) For proving (9.13b), it is sufficient to consider the case $m = 1$. The Richardson iteration

$$x^1_{\mathrm{Rich}} = x^0 - \Theta_{\mathrm{Rich}}\left(Ax^0 - b\right) \qquad \text{with } \Theta_{\mathrm{Rich}} = 2/(\Lambda + \lambda)$$

yields the error $e^1_{\mathrm{Rich}} = Me^0$. The iteration matrix $M = M^{\mathrm{Rich}}_{\Theta_{\mathrm{Rich}}} = I - \Theta_{\mathrm{Rich}}A$ has the norm $\|M\|_2 \le \eta$, where

$$\eta = \frac{\Lambda - \lambda}{\Lambda + \lambda} \tag{9.13c}$$

(cf. Theorem 3.23). Since M commutes with A and $A^{1/2}$, we have

$$\tilde{e}^1_{\mathrm{Rich}} = M\,\tilde{e}^0 \qquad \text{for} \quad \tilde{e}^1_{\mathrm{Rich}} := A^{1/2}e^1_{\mathrm{Rich}}, \quad \tilde{e}^0 := A^{1/2}e^0.$$

By $\|\tilde{e}^0\|_2 = \|e^0\|_A$ and $\|\tilde{e}^1_{\mathrm{Rich}}\|_2 = \|e^1_{\mathrm{Rich}}\|_A$, we can estimate e^1_{Rich} by

$$\|e^1_{\mathrm{Rich}}\|_A = \|\tilde{e}^1_{\mathrm{Rich}}\|_2 \le \|M\|_2 \,\|\tilde{e}^0\|_2 \le \eta\,\|e^0\|_A.$$

Both x^1_{Rich} and x^1 are of the form $x^0 + \Theta r^0$. Since the iterate x^1 of the gradient method minimises the error $\|x^1 - x^*\|_A$ (cf. Conclusion 9.4), the assertion follows for $m = 1$: $\|x^1 - x^*\|_A \le \|e^1_{\mathrm{Rich}}\|_A \le \eta\,\|e^0\|_A$. □

Corollary 9.11. (a) The factor η in (9.13c) is the minimal one in (9.13a,b).

(b) The asymptotic convergence rate of the gradient method is η.

(c) η depends only on the condition $\kappa(A) = \mathrm{cond}_2(A) = \Lambda/\lambda$:

$$\eta = \frac{\kappa - 1}{\kappa + 1} \qquad \text{with } \kappa = \kappa(A). \tag{9.14}$$

Proof. Let v_1 and v_2 with $\|v_1\|_2 = \Lambda$ and $\|v_2\|_2 = \lambda$ be the eigenvectors corresponding to Λ and λ. For $x^0 := x^* + e^0$ with $e^0 := v_1 \pm v_2$, one obtains $e^1 = \eta(v_1 \mp v_2)$ and $e^2 = \eta^2(v_1 \pm v_2) = \eta^2 e^0$. $\|e^2\|_A/\|e^0\|_A = \eta^2$ proves part (a). Analogously, $e^{2k} = \eta^{2k}e^0$ shows part (b). □

Usually, the values $\lambda = \lambda_{\min}(A)$ and $\Lambda = \lambda_{\max}(A)$ are not known. Their numerical approximation is discussed below.

Remark 9.12 (approximation of λ and Λ). (a) Let $\langle e^0, v_i\rangle \ne 0$ hold for the eigenvectors v_1 and v_2 of A corresponding to Λ and λ. Then

$$\rho_{m+1,m} := \|x^{m+1} - x^*\|_A / \|x^m - x^*\|_A \qquad (x^m \text{ defined by (9.11a–c)})$$

converges to $\eta = (\kappa - 1)/(\kappa + 1)$ in (9.14).

(b) Using $\rho(M_\Theta^{\mathrm{Rich}}) = 1 - \Theta\lambda$ (e.g., for $\Theta = 1/\|A\|_\infty$), we can approximate λ from the convergence behaviour of the Richardson method. The approximation of η yields an approximation of $\kappa = (1+\eta)/(1-\eta)$ which allows us to determine the other extreme eigenvalue Λ by $\Lambda/\lambda = \kappa$.

Finally, we describe the relation of the gradient method with the Krylov space (cf. §8.1.4).

Proposition 9.13. *The errors $e^m = x^m - x^*$ of the gradient method satisfy*

$$\mathcal{K}_m(A, e^0) = \mathrm{span}\{e^0, e^1, \ldots, e^{m-1}\}$$

for all $m \in \mathbb{N}$. The residuals $r^\mu = -Ae^\mu$ $(0 \le \mu \le m-1)$ span the space $\mathcal{K}_m(A, r^0) = A\mathcal{K}_m(A, e^0)$.

Proof. As long as $r^m \neq 0$, the equivalent semi-iteration corresponds to polynomials p_m of degree $m,$ so that Conclusion 8.13b applies. Otherwise, there is a first m' with $r^{m'} = -Ae^{m'} = 0$. Since $e^{m'} = p_{m'}(A)e^0$, $\deg_A(e^0) \leq m'$ follows (cf. Definition 8.10). Exercise 8.11a states that $\mathcal{K}_m(A, e^0) = \mathcal{K}_{m'}(A, e^0)$ for all $m \geq m'$. Therefore, the statement also holds in the degenerate case of $r^{m'} = 0$. □

9.2.3 Numerical Examples

At first view, the gradient method seems to surpass the semi-iterative method because, in the latter case, the parameters Θ_k have to be chosen *a priori* (cf. (9.12)), whereas the gradient method determines these values *a posteriori* in an optimal way. However, the opposite is the case. While the Chebyshev method leads to an improvement of the order, Corollary 9.11a yields the convergence rate η in (9.13c), which is as slow as the stationary Richardson method with $\Theta = \Theta_{\mathrm{opt}}$ (cf. Theorem 3.23).

In the model case, λ and Λ in (3.1b,c) are known and lead to

$$\eta = \frac{\cos^2(\pi h/2) - \sin^2(\pi h/2)}{\cos^2(\pi h/2) + \sin^2(\pi h/2)} = \cos(\pi h) = 1 - \frac{\pi^2 h^2}{2} + \mathcal{O}(h^4).$$

The low convergence speed of the gradient method is confirmed by the following numerical example (Poisson model problem (2.33a,b)). Table 9.1 contains the results for the step size $h = 1/32$ and the starting iterate $x^0 = 0$. The ratios $\|x^{m+1} - x^*\|_A / \|x^m - x^*\|_A$ in the last column of Table 9.1 clearly approximate the asymptotic convergence rate $h = \cos\frac{\pi}{32} = 0.9951847$. Even after 300 iterations, the value $u_{16,16}$ at the midpoint is wrong by 50%: 0.2778 instead of 0.5. The error measured in the scaled energy norm $h^2\|x^m - x^*\|_A$ deviates very little from the maximum norm $\|e^m\|_\infty$. However, the error with respect to the energy norm $\|\cdot\|_A$ decreases uniformly, whereas the ratios of $\|e^m\|_\infty$ oscillate. Because $\eta = \rho(M^{\mathrm{Jac}})$, the results in Table 9.1 and Table 3.1 prove to be very similar.

m	value in the middle	$\frac{\|e^m\|_A}{\|e^{m-1}\|_A}$	m	value in the middle	$\frac{\|e^m\|_A}{\|e^{m-1}\|_A}$
1	-1.86560_{10}-3		100	-1.89771_{10}-2	0.993444
2	-3.52293_{10}-3	0.844824	110	-5.13520_{10}-3	0.993749
3	-4.84034_{10}-3	0.907804	120	1.01805_{10}-2	0.993990
4	-5.97611_{10}-3	0.935293	200	1.45146_{10}-1	0.994852
5	-7.10198_{10}-3	0.946906	250	2.18301_{10}-1	0.995024
6	-8.16295_{10}-3	0.953838	296	2.73548_{10}-1	0.995102
7	-9.23998_{10}-3	0.958895	297	2.75710_{10}-1	0.995103
8	-1.02699_{10}-2	0.962711	298	2.75702_{10}-1	0.995104
9	-1.13230_{10}-2	0.965778	299	2.77844_{10}-1	0.995105
10	-1.23360_{10}-2	0.968271	300	2.77836_{10}-1	0.995106

Table 9.1 Result of the gradient method $\Upsilon_{\mathrm{grad}}[\Phi_1^{\mathrm{Rich}}]$ for $h = 1/32$ (Poisson model problem).

9.2.4 Gradient Method Based on Other Basic Iterations

Let $\varUpsilon_{\text{grad}} \in \mathcal{N}$ be the notation of the gradient method. Above we applied the gradient method to the Richardson iteration $\varPhi_1^{\text{Rich}}$ resulting in the nonlinear iteration $\varUpsilon_{\text{grad}}[\varPhi_1^{\text{Rich}}]$. Now we discuss $\varUpsilon_{\text{grad}}[\varPhi]$ for other iterations.

9.2.4.1 Standard Version

By Remark 9.9, the gradient method is a particular semi-iterative method with Richardson's iteration as the basic iteration. From the analysis of semi-iterative methods, we know that other basic iterations $\varPhi$ may better suit because of a smaller spectral condition number $\kappa(NA)$ with the matrix $N = N^{\varPhi}[A]$ of the second normal form. This suggests replacing Richardson's iteration by another one (e.g., the SSOR iteration; cf. §8.4.4). For this purpose, the matrix A has to be replaced formally with $\hat{A} := NA$, because the Richardson method applied to the left-transformed (preconditioned) system $\hat{A}x = \hat{b} := Nb$ is equivalent to $\varPhi$ applied to A (cf. Proposition 5.44).

Let A and N be positive definite. Since, in general, the matrix $\hat{A} = NA$ is no longer symmetric, $\hat{A}$ does not satisfy the assumption (9.1), which is necessary for the applicability of the gradient method. A remedy is offered in §5.6.6: the iteration $\check{\varPhi}$ defined by

$$\check{x}^{m+1} = \check{x}^m - N^{1/2}(AN^{1/2}\check{x}^m - b) = \check{x}^m - (\check{A}\check{x}^m - \check{b}), \tag{9.15a}$$

$$\check{A} := N^{1/2}AN^{1/2}, \qquad \check{b} := N^{1/2}b, \tag{9.15b}$$

is equivalent to the basic iteration $\varPhi(x^m, b) = x^m - N(Ax^m - b)$ via the transformation

$$\check{x}^m = N^{-1/2}x^m$$

(multiplying by $N^{\pm 1/2}$ is of course not practically feasible[2]) and represents the Richardson iteration for the system $\check{A}\check{x} = \check{b}$ with the positive definite matrix $\check{A}$. Therefore, the gradient method has to be applied not to F in (9.2) but to

$$\check{F}(\check{x}) := \frac{1}{2}\left\langle \check{A}\check{x}, \check{x}\right\rangle - \left\langle \check{b}, \check{x}\right\rangle. \tag{9.15c}$$

Its negative gradient is the new residual

$$\check{r} := \check{b} - \check{A}\check{x} = N^{1/2}r \qquad (r = b - Ax).$$

The gradient method (9.11b,c) associated with $\check{A}$ yields the iterates

$$\varUpsilon_{\text{grad}}[\check{\varPhi}]: \quad \left.\begin{array}{l} \check{x}^{m+1} := \check{x}^m + \check{\lambda}_{\text{opt}}\,\check{r}^m, \\ \check{r}^{m+1} := \check{r}^m - \check{\lambda}_{\text{opt}}\,\check{A}\,\check{r}^m \end{array}\right\} \quad \text{with} \quad \check{\lambda}_{\text{opt}} := \frac{\|\check{r}^m\|_2^2}{\left\langle \check{A}\check{r}^m, \check{r}^m\right\rangle}.$$

[2] In principle, we may replace the factorisation $N = N^{\frac{1}{2}} \cdot N^{\frac{1}{2}}$ with the Cholesky decomposition $N = VV^{\mathsf{H}}$ and introduce $\check{A} = V^{\mathsf{H}}AV$ (cf. Exercise 5.63).

Inserting $\breve{A} = N^{1/2}AN^{1/2}$, $\breve{x}^m = N^{-1/2}x^m$, $\breve{r}^m = N^{1/2}r^m$, and solving the defining equations for x^{m+1} and r^{m+1}, we obtain the following algorithm for the iterates $\{x^m\}$:

$$\Upsilon_{\mathrm{grad}}[\Phi]:\quad \begin{array}{lll} x^{m+1} := x^m + \check{\lambda}_{\mathrm{opt}} N r^m & \text{with } N = N_\Phi[A], & (9.16\mathrm{a}) \\ r^{m+1} := r^m - \check{\lambda}_{\mathrm{opt}} A N r^m & \text{with} & (9.16\mathrm{b}) \\ \check{\lambda}_{\mathrm{opt}} := \lambda_{\mathrm{opt}}(r^m, Nr^m, A) = \dfrac{\langle Nr^m, r^m\rangle}{\langle ANr^m, Nr^m\rangle}. & & (9.16\mathrm{c}) \end{array}$$

The quantities $N^{\pm\frac{1}{2}}$ do no longer appear, so that $\Upsilon_{\mathrm{grad}}[\Phi]$ defined by (9.16a–c) is a practical algorithm. We call $\Upsilon_{\mathrm{grad}}[\Phi]$ defined in (9.16a–c) the *gradient method applied to the basic iteration* $\Phi(\cdot,\cdot,A)$. The term ‘preconditioned gradient method’ is also used. Note that not the method but the gradient is ‘preconditioned’. While the method (9.11a–c) takes the (negative) gradient r^m as search direction, this is replaced in (9.17a–e) with the ‘preconditioned’ gradient $q = Nr$.

The derivation of (9.16a–c) requires $N > 0$. Nevertheless, $\Upsilon_{\mathrm{grad}}[\Phi]$ is well defined as long as $A > 0$ and N is regular since this guarantees $\langle ANr^m, Nr^m\rangle > 0$ for $r^m \neq 0$ ($r^m = 0$ is a ‘lucky breakdown’ since the exact solution $x = x^m$ is found). However, the convergence statements are restricted to the case $A > 0$ and $N > 0$. Since $A > 0$ implies $N > 0$ for $\Phi \in \mathcal{L}_{\mathrm{sym}}$, symmetric iterations Φ are the natural basic iterations of the gradient method.

In analogy to Remark 9.9, equation (9.16a) proves the next remark.

Remark 9.14. The sequence $\{x^m\}$ of the gradient method (9.16a–c) applied to the positive definite iteration Φ is identical to the sequence $\{y^m\}$ of the semi-iterative method

$$y^{m+1} = y^m - \Theta_{m+1}N(Ay^m - b) = \Theta_{m+1}\Phi(y^m, b, A) + (1 - \Theta_{m+1})y^m$$

with Φ as the basic iteration when the factors Θ_{m+1} are defined by $\check{\lambda}_{\mathrm{opt}}$ in (9.16c).

The amount of work needed by the algorithm (9.16a–c) can be reduced by introducing $q^m := Nr^m$ and $a^m := Aq^m$. Note that q^m and a^m need not be saved for the next iteration step.

$$\begin{array}{|ll|l} \hline \textbf{start:} & x^0 \text{ arbitrary};\quad r^0 := b - Ax^0; & (9.17\mathrm{a}) \\ \textbf{iteration } m = 0, 1, \ldots : & q^m := Nr^m;\quad a^m := Aq^m; & (9.17\mathrm{b}) \\ & \lambda_{\mathrm{opt}} := \lambda_{\mathrm{opt}}(r^m, q^m, A) = \frac{\langle q^m, r^m\rangle}{\langle a^m, q^m\rangle}; & (9.17\mathrm{c}) \\ & x^{m+1} := x^m + \lambda_{\mathrm{opt}} q^m; & (9.17\mathrm{d}) \\ & r^{m+1} := r^m - \lambda_{\mathrm{opt}} a^m; & (9.17\mathrm{e}) \\ \hline \end{array}$$

Remark 9.15. The representation (9.17a–e) shows that for each iteration step only one multiplication by N and one by A are necessary. For $N = I$, we regain the algorithm (9.11a–c).

The convergence of the method (9.17a–e) follows by applying the convergence statement of Theorem 9.10 to the transformed problem (9.15c): $\breve{\Phi}(\breve{x}) = \min$. First, we obtain an error estimate for $\breve{x}^m$ with respect to the corresponding energy norm $\|\cdot\|_{\breve{A}}$. Because of

$$\begin{aligned}\|\breve{x}^m - \breve{x}^*\|_{\breve{A}}^2 &= \left\langle \breve{A}(\breve{x}^m - \breve{x}^*), \breve{x}^m - \breve{x}^* \right\rangle \\ &= \langle A(x^m - x^*), x^m - x^* \rangle \\ &= \|x^m - x^*\|_A^2 \qquad (\,\breve{x}^* = N^{-1/2}x^*\,)\end{aligned}$$

the $\breve{A}$-estimates of $\breve{x}^m - \breve{x}^*$ carry over to the A-norm of the error $x^m - x^*$.

Theorem 9.16 (convergence). *Let A and $N = W^{-1}$ be positive definite. If*

$$\gamma W \le A \le \varGamma W, \tag{9.18a}$$

the iterates in (9.17a–e) satisfy the error estimate

$$\|x^m - x^*\|_A \le \left(\frac{\varGamma - \gamma}{\varGamma + \gamma}\right)^m \|x^0 - x^*\|_A\,. \tag{9.18b}$$

Remark 9.17. Under an assumption analogous to that in Remark 9.12a, we conclude for algorithm (9.17a–e) that the convergence factors converge to $\eta = \frac{\kappa-1}{\kappa+1}$ with $\kappa := \kappa\,(NA) = \varGamma\,/\,\gamma$ (here, γ and $\varGamma$ are the optimal bounds in (9.18a)). Therefore, the gradient method (9.17a–e) can be used to determine the spectral condition number $\varGamma/\gamma$.

We regard the gradient method as a general technique that can be applied to all positive definite iterations Φ and problems with $A > 0$. This is the same situation as the Chebyshev method which also requires specifying the basic iteration.

Theorem 9.18. *Let Φ be a positive definite iteration and assume that $A > 0$. The gradient method applied to Φ converges as fast as the optimally damped iteration $\Phi_{\vartheta_{\mathrm{opt}}}$ for $\vartheta_{\mathrm{opt}} = \frac{2}{\varGamma+\gamma}$. However, the explicit knowledge of the optimal bounds γ and $\varGamma$ in (9.18a) is not necessary.*

Proof. Compare the results of Theorem 6.7 and (9.18b), and use $\gamma = \lambda_{\min}(NA)$ and $\varGamma = \lambda_{\max}(NA)$. □

Now the statement of Proposition 9.13 reads as follows.

Remark 9.19. The errors $e^m = x^m - x^*$ of the gradient method (9.17a–e) satisfy

$$\mathcal{K}_m(NA, e^0) = \operatorname{span}\{e^0, e^1, \ldots, e^{m-1}\}$$

for all $m \in \mathbb{N}$. The residuals $r^\mu = -Ae^\mu$ $(0 \le \mu \le m-1)$ span the space $A\,\mathcal{K}_m(NA, e^0) = \mathcal{K}_m(AN, r^0)$.

9.2.4.2 Residual Oriented Version

In (9.15a) we interpreted the iteration $\varPhi$ as Richardson's iteration applied to the positive definite matrix $\breve{A}$. This is not the only possibility. $\varPhi$ is also equivalent to $\varPhi_1^{\mathrm{Rich}}$ applied to $\bar{A}$:

$$\bar{x}^{m+1} := \bar{x}^m - (\bar{A}\bar{x}^m - \bar{b}) \qquad \text{with}$$
$$\bar{A} := A^{1/2}NA^{1/2} > 0, \quad \bar{b} := A^{1/2}Nb, \quad \bar{x}^m := A^{1/2}x^m. \tag{9.19}$$

Exercise 9.20. Prove the following: (a) The application of the gradient method to the minimisation of $\bar{F}(\bar{x}) := \frac{1}{2}\left\langle \bar{A}\bar{x}, \bar{x}\right\rangle - \left\langle \bar{b}, \bar{x}\right\rangle$ yields (9.20a–c)—denoted by $\varUpsilon_{\mathrm{grad}}^{\mathrm{res}}[\varPhi]$—after a reformulation using the x-quantities:

$$\text{start:} \quad x^0 \text{ arbitrary}, \quad q^0 := N(b - Ax^0), \tag{9.20a}$$
$$x^{m+1} := x^m + \lambda_{\mathrm{opt}}\, q^m \tag{9.20b}$$
$$\text{with } \lambda_{\mathrm{opt}} := \lambda_{\mathrm{opt}}(q^m, Aq^m, N) = \tfrac{\langle q^m, A\,q^m\rangle}{\langle NA\,q^m, A\,q^m\rangle},$$
$$q^{m+1} := q^m - \lambda_{\mathrm{opt}}\, NA\, q^m. \tag{9.20c}$$

(b) The methods $\varUpsilon_{\mathrm{grad}}[\varPhi]$ in (9.17a–e) and $\varUpsilon_{\mathrm{grad}}^{\mathrm{res}}[\varPhi]$ in (9.20a–c) are different. Choosing $N = I$, we do not regain the gradient method (9.11a–c).

(c) Let γ and $\varGamma$ be the bounds in (9.18a). Then the error estimate (9.20d) holds:

$$\|N^{1/2}A(x^m - x^*)\|_2 \le \left(\frac{\varGamma - \gamma}{\varGamma + \gamma}\right)^m \|N^{1/2}A(x^0 - x^*)\|_2. \tag{9.20d}$$

Note that both versions (9.17a–e) and (9.20a–c) lead to the same convergence rate, but the involved norms are different. If $W \sim A$, the norms $\|\cdot\|_A$ and $\|\cdot\|_{ANA}$ are equivalent. On the other hand, for $N = I$ the residual $A(x^m - x^*) = r^m$ is the subject of minimisation and for $N \sim I$ the norms $\|N^{1/2}r^m\|_2$ and $\|r^m\|_2$ are equivalent.

Remark 9.21. The statements of Remark 9.19 also hold for the errors $e^m = x^m - x^*$ of the gradient method (9.20a–c) as well as for the results of the following variant (9.21a–c).

9.2.4.3 Directly Positive Definite Case

Assume $\varPhi \in \mathcal{L}_{>0}$, i.e., the iteration $\varPhi(\cdot, \cdot, A)$ is directly positive definite:

$$N[A]\, A > 0 \qquad \text{(cf. Definition 5.14).}$$

Then the original method (9.10a–c) can be applied with A replaced with the matrix $N[A]A$. Note that in this case the matrix $A \in \mathfrak{D}(\varPhi)$ is only required to be regular.

The quadratic function

$$F(x) := \frac{1}{2}\,\langle NAx, x\rangle - \langle Nb, x\rangle$$

replaces F in (9.2). The corresponding gradient method reads as follows:

$$\text{start:}\quad x^0 \text{ arbitrary},\quad q^0 := N(b - Ax^0), \tag{9.21a}$$

$$x^{m+1} := x^m + \mathring{\lambda}\, q^m \quad \text{with } \mathring{\lambda} := \frac{\|q^m\|_2^2}{\langle NAq^m, q^m\rangle}, \tag{9.21b}$$

$$q^{m+1} := q^m - \mathring{\lambda}\, NA\, q^m. \tag{9.21c}$$

Theorem 9.22. *Let NA be positive definite with*

$$\gamma := \lambda_{\min}(NA) \quad \textit{and} \quad \Gamma := \lambda_{\max}(NA).$$

The iterates in (9.21a–c) satisfy the error estimate

$$\|x^m - x^*\|_{NA} \le \left(\frac{\Gamma - \gamma}{\Gamma + \gamma}\right)^m \|x^0 - x^*\|_{NA}\,.$$

9.2.5 Numerical Examples

The SSOR iteration is used as a basic iteration of the gradient method for the Poisson model case. As in Table 6.1, we choose the relaxation parameter

$$\omega = 1.82126912$$

for the step size $h = 1/32$. The results given in Table 9.2 suggest the convergence rate $\eta \approx 0.769$. From (9.14), we conclude the spectral condition number

$$\Gamma/\gamma = \kappa = (1 + \eta)/(1 - \eta) = 7.66.$$

m	value in the middle	$\frac{\|e^m\|_A}{\|e^{m-1}\|_A}$
1	0.2851075107	0.4576
2	0.9245177570	0.5192
3	0.1780816984	0.5886
4	0.2274720552	0.6454
5	0.2956906889	0.6858
10	0.4381492069	0.7577
20	0.4954559469	0.7672
30	0.4996724015	0.7682
40	0.4999764630	0.7685
50	0.4999983084	0.7687
60	0.4999998782	0.7688
70	0.4999999912	0.7689

Table 9.2 Gradient method $\Upsilon_{\mathrm{grad}}[\Phi^{\mathrm{SSOR}}_{\omega=1.82}]$ applied to the SSOR iteration for $h = 1/32$.

According to Table 6.1, the convergence rate of the SSOR iteration equals 0.8796. From $\rho(M^{\mathrm{SSOR}}) = 1 - \lambda$, we deduce $\lambda = 0.1204$, implying $\Gamma = 7.66$ and $\gamma = 0.922$. Hence,

$$\vartheta_{\mathrm{opt}} = 2/(\Gamma + \gamma) \approx 1.92$$

is the optimal damping or (more precisely) extrapolation factor for $\Phi^{\mathrm{SSOR}}_{\omega=1.82}$ in the Poisson model case with $h = 1/32$.

9.3 Method of the Conjugate Directions

9.3.1 Optimality with Respect to a Direction

The slowness of the gradient method is demonstrated in Theorem 9.10 by the two-dimensional subspace spanned by the two extreme eigenvectors. Therefore, a system of two equations is able illustrate this situation. The matrix

$$A = \operatorname{diag}\{\lambda_1, \lambda_2\} \qquad \text{with} \quad 0 < \lambda_1 \le \lambda_2$$

has the condition $\operatorname{cond}_2(A) = \lambda_2/\lambda_1$. The corresponding function F in (9.2) leads to ellipses as level curves

$$N_c := \{x \in \mathbb{R}^2 : \Phi(x) = c\}, \qquad \text{where} \quad c \in \mathbb{R}.$$

In the two-dimensional case, the gradient method can be illustrated graphically as follows: The point x^m $[x^{m+1}]$ lies on the ellipse $E^{(m)} := N_c$ with $c = F(x^m)$ [or $E^{(m+1)} := N_c$ with $c = F(x^{m+1})$, respectively]. The straight line $x^m x^{m+1}$ is vertical to $E^{(m)}$ and tangential to $E^{(m+1)}$. Therefore, succeeding straight lines (i.e., the corrections $x^{m+1} - x^m$) form right angles. Figure 9.1 shows the case of an elongated ellipse, where the iteration path forms a zigzag line. This illustrates that the approximation to the centre requires many iteration steps. Note that the ellipses are more elongated the larger the condition is. In the case of a circle ($\lambda_1 = \lambda_2$), the first correction would already yield the exact solution x^*.

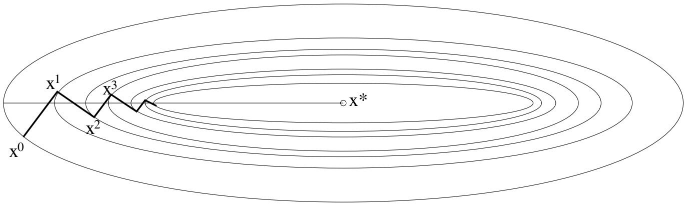

Fig. 9.1 The iterates x^m and the corresponding level lines of the function F.

From the fact that the corrections $x^{m+3} - x^{m+2}$ and $x^{m+1} - x^m$ are parallel, one understands that the iterate x^{m+2} must be corrected in exactly the same direction in which x^m has been corrected previously. Hence, x^{m+2} has lost the property of x^{m+1} being optimal with respect to the direction $x^{m+1} - x^m$. We define:

x is *optimal with respect to a direction* $p \neq 0$, if
$F(x) \le F(x + \lambda p) \qquad$ for all $\lambda \in \mathbb{K}$.

Lemma 9.23. *The optimality of x with respect to p is equivalent to*

$$p \perp r := b - Ax.$$

Proof. A necessary condition for $f(\lambda) = F(x + \lambda p)$ in (9.5c) to be minimal for $\lambda = 0$ is $\langle Ax - b, p\rangle = -\langle r, p\rangle = 0$. As (9.5c) is restricted to the field $\mathbb{R}$, use (9.9c) for the complex case. □

Exercise 9.24. x is called optimal with respect to a subspace $\mathcal{U}$ if $F(x) \leq F(x + \xi)$ for all $\xi \in \mathcal{U}$. Prove that x is optimal with respect to $\mathcal{U}$ if and only if

$$r = b - Ax \perp \mathcal{U}.$$

Remark 9.25. The iterates x^m of the gradient method satisfy (9.22a,b):

$$x^{m+1} \text{ is optimal with respect to } r^m = b - Ax^m, \quad (m \geq 0) \tag{9.22a}$$

$$r^{m+1} \perp r^m. \tag{9.22b}$$

Proof. By Lemma 9.23, (9.22a) and (9.22b) are equivalent. $r^m \perp r^{m+1} = r^m - \lambda_{\text{opt}}(r^m, r^m)\, A\, r^m$ follows from the definition (9.6b,c) of λ_{opt}. □

The principal deficit of the gradient method can be stated as follows. The relation $r^{m+1} \perp r^m$ is not transitive, i.e., $r^m \perp r^{m+1}$ and $r^{m+1} \perp r^{m+2}$ do not imply $r^m \perp r^{m+2}$. Therefore, in general, x^{m+2} has lost its optimality with respect to r^m.

9.3.2 Conjugate Directions

The change of x into $x' := x + q$ $(q \neq 0)$ transforms the residual $r = b - Ax$ of x into the residual

$$r' = b - Ax' = b - A(x + q) = b - Ax - Aq = r - Aq$$

of x'. Let x be optimal with respect to the direction p:

$$r \perp p\,.$$

The new value x' remains optimal with respect to p if and only if $r' \perp p$, i.e., $Aq \perp p$, because the latter property is equivalent to

$$-\langle Aq, p\rangle = \langle r - Aq, p\rangle = \langle r', p\rangle = 0\,.$$

This proves the next statement.

Lemma 9.26. *The optimality of x with respect to $p \neq 0$ implies the optimality of $x' = x + q$ with respect to the same $p \neq 0$ if and only if*

$$Aq \perp p\,. \tag{9.23}$$

Vectors p, q with the property (9.23) are called *conjugate*. The term 'conjugate' can also be replaced with 'A-orthogonal', abbreviated as

$$q \perp_A p\,,$$

where $\perp_A$ denotes orthogonality with respect to the scalar product $\langle\cdot,\cdot\rangle_A$ in (9.8a). Note that the latter definitions only make sense if $A > 0$.

Condition (9.23) leads us to the following method of conjugate directions.

```
        method of conjugate directions                                      (9.24)
start:  $x^0$ arbitrary,   $r^0 := b - Ax^0$;
loop:   for $m = 0, 1, \ldots, n-1$:      $(n := \#I)$
        choose a direction $p^m \neq 0$ which is conjugate to all          (9.24a)
        preceding directions $p^\ell$ $(\ell < m)$;
        $x^{m+1} := x^m + \lambda_{\mathrm{opt}}(r^m, p^m, A), A\,p^m$   with       (9.24b)
        $\lambda_{\mathrm{opt}}(r^m, p^m, A) := \langle r^m, p^m\rangle \,/\, \langle Ap^m, p^m\rangle\,;$        (9.24c)
        $r^{m+1} := r^m - \lambda_{\mathrm{opt}}(r^m, p^m, A)\, Ap^m;$                 (9.24d)
```

The lines (9.24b,c) show that x^{m+1} is optimal with respect to the direction p^m: $F(x^{m+1}) = \min\{F(x^m + \lambda p^m) : \lambda \in \mathbb{K}\}$ or equivalently

$$r^{m+1} \perp p^m. \tag{9.24e}$$

Definition (9.24d) is equivalent to $r^{m+1} := b - Ax^{m+1}$.

The properties of this method are collected below.

Theorem 9.27. *(a) The directions $\{p^m : 0 \leq m \leq n-1\}$ form a basis of pairwise conjugate vectors, i.e., an A-orthogonal basis.*

(b) The algorithm terminates at $m = n{-}1$ with the exact solution $x^{m+1} = x^n = x^*$.

(c) The iterate x^m is optimal with respect to all directions $p^0, \ldots, p^{m-1}$, i.e., it is optimal with respect to the subspace $\mathcal{U}_m := \mathrm{span}\{p^0, p^1, \ldots, p^{m-1}\}$. The residuals r^m satisfy

$$r^m \perp p^\ell \qquad (0 \leq \ell \leq m-1), \tag{9.25a}$$

$$r^m \perp \mathcal{U}_\ell \qquad (1 \leq \ell \leq m). \tag{9.25b}$$

(d) The error $e^m = x^m - x^*$ *fulfils the conditions*

$$e^m \perp_A p^\ell \qquad (0 \le \ell \le m-1). \tag{9.25c}$$

(e) x^m *solves the minimisation problem*

$$F(x^m) = \min_{\lambda_\ell \in \mathbb{K}} \left\{ F(\xi) : \xi = x^0 + \sum_{\ell=0}^{m-1} \lambda_\ell p^\ell \right\} = \min_{\xi - x^0 \in \mathcal{U}_m} F(\xi), \tag{9.25d}$$

where the minimum in (9.25d) is taken at $\lambda_\ell = \lambda_{\text{opt}}(r^\ell, p^\ell, A)$.

Proof. (a) First, we note that the division by $\langle Ap^m, p^m \rangle$ in (9.24c) is well defined because of $p^m \neq 0$, as long as an additional conjugate direction exists, i.e., as long as $m < n$. As soon as $m = n-1$, the vectors $p^0, \ldots, p^{n-1}$ span the whole space $\mathbb{K}^I$ and the process cannot be continued.

(c) The statement (9.25a) is true for $m = 0$ since $\{\ell : 0 \le \ell \le m-1\}$ is the empty set. Suppose that (9.25a) holds for m. By Lemma 9.23, x^m is optimal with respect to all directions p^ℓ $(0 \le \ell \le m-1)$. According to Lemma 9.26, this property is inherited by x^{m+1} because of $p^m \perp_A p^\ell$ $(0 \le \ell \le m-1)$; hence $r^{m+1} \perp p^\ell$ holds for all $0 \le \ell \le m-1$. The missing condition $r^{m+1} \perp p^m$ follows from (9.24e).

(d) (9.25c) follows from (9.25a), as $Ae^m = A(x^m - x^*) = Ax^m - b = -r^m$.

(b) (9.25b) proves that $r^n \perp \mathcal{U}_n$. Since $\mathcal{U}_n = \mathbb{K}^I$ (cf. part (a)), $r_n = 0$ follows, i.e., $x^n = x^*$.

(e) Inserting Eqs. (9.24b) one into another, we obtain

$$x^m = x^0 + \sum_{\ell=0}^{m-1} a_\ell p^\ell \qquad \text{with } a_\ell = \lambda_{\text{opt}}(r^\ell, p^\ell, A).$$

From (9.9c) with $\tilde{x} := x^m$, $x := \xi$, and from $r^m \perp \mathcal{U}_m$, we deduce that

$$\begin{aligned} F(\xi) - F(x^m) &= \mathfrak{Re} \left\langle r^m, \sum_{\ell=0}^{m-1} (\lambda_\ell - a_\ell) p^\ell \right\rangle + \frac{1}{2} \langle A(\xi - x^m), \xi - x^m \rangle \\ &= \frac{1}{2} \|\xi - x^m\|_A^2 \ge 0 \end{aligned}$$

with an equal sign only for $\xi = x^m$, i.e., for $\lambda_\ell = a_\ell$. This proves (9.25d). □

The method of conjugate directions is not interesting in practice, unless the directions p^m in (9.24) are suitably selected. If, for instance, one chooses a fixed conjugate system $\{p^0, \ldots, p^{n-1}\}$, the starting value $x^0 := x^* - p^{n-1}$ with the residual $r^0 = Ap^{n-1}$ leads to a sequence $x^0 = x^1 = \ldots = x^{n-1}$ which only in the last step changes to the exact solution $x^n = x^*$. This explains why, in general, no convergence estimate as in (9.13b) can be given.

9.4 Minimal Residual Iteration

For general matrices A, the function (9.7a) with $H = I$ can be minimised, i.e., the residual $r = b - Ax$ is minimised: $F(x) := \|A(x - x^*)\|_2^2 = \|r\|_2^2 = \min$. Choosing the gradient of F as search direction, we would regain the gradient method in §9.2 applied to the equation $A^{\mathsf{H}}Ax - A^{\mathsf{H}}b$. Instead of this gradient, one can use the residual $r = b - Ax$ of the original system as search direction. This yields the *minimal residual iteration*

$$x^{m+1} = x^m - \frac{\Re\mathfrak{e}\,\langle Ar^m, r^m\rangle}{\langle Ar^m, Ar^m\rangle} r^m, \qquad r^m = b - Ax^m.$$

For general matrices A, the method cannot converge since $r^0 \neq 0$ may lead to $\left\langle Ar^0, r^0\right\rangle = 0$ so that $x^m = x^0 \neq A^{-1}b$ for all m. To avoid this problem, we need the following assumptions.

Theorem 9.28. *Assume* $A + A^{\mathsf{H}} > 0$. *Then the minimal residual iteration converges with the rate*

$$c := \sqrt{\frac{\lambda_{\min}(A + A^{\mathsf{H}})}{2\,\|A\|_2}}. \tag{9.26}$$

The convergence is uniform with respect to the residual: $\|r^{m+1}\|_2 \leq c\,\|r^m\|_2$.

Proof. See Saad [328, Theorem 5.10]. □