17 SOLUTION OF NONLINEAR EQUILIBRIUM EQUATIONS

We have seen that the finite element method for general nonlinear problems leads to the solution of nonlinear equations. The present chapter is therefore devoted to a discussion of various methods to solve nonlinear equations. Many of these methods have their origin not only in solid and structural mechanics, but also in nonlinear optimization theory and as general references, the reader may consult Bathe (1996), Belytschko *et al.* (2000), Crisfield (1991), Fletcher (1980), Luenberger (1984), Papadrakakis (1993) and Zienkiewicz and Taylor (1991) for relevant information.

The nonlinear equations of interest here are the equilibrium equations given by (16.10) and (16.11), i.e.

$$\boxed{\boldsymbol{\psi} = \mathbf{0}} \tag{17.1}$$

where

$$\boxed{\boldsymbol{\psi} = \boldsymbol{f}_{int} - \boldsymbol{f}} \tag{17.2}$$

and $\boldsymbol{f}$ denotes the *external forces*, i.e. the load on the body, defined by

$$\boldsymbol{f} = \int_S \boldsymbol{N}^T \boldsymbol{t} dS + \int_V \boldsymbol{N}^T \boldsymbol{b} dV$$

whereas the *internal forces* $\boldsymbol{f}_{int}$ are defined by

$$\boldsymbol{f}_{int} = \int_V \boldsymbol{B}^T \boldsymbol{\sigma} dV \tag{17.3}$$

Expression (17.1) holds for any body in equilibrium and our problem is to satisfy this equation.

In order to solve the boundary value problem in question, we have to consider the response of the actual material. Here, we will for illustration purposes assume elasto-plasticity i.e.

$$\boxed{\dot{\boldsymbol{\sigma}} = \boldsymbol{D}_t \dot{\boldsymbol{\varepsilon}}} \tag{17.4}$$

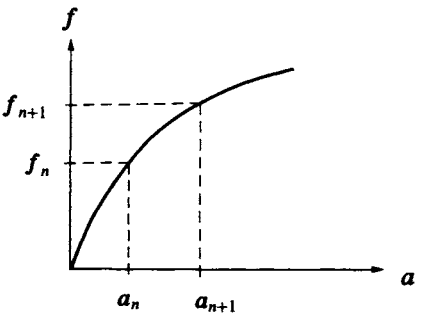

Figure 17.1: State n is known; we want to determinate state $n + 1$.

where $\boldsymbol{D}_t$ is the tangential stiffness equal to the elastic stiffness if elastic loading occurs and equal to the elasto-plastic tangential stiffness if plastic loading is present. Apart from this constitutive relation that is specific for elasto-plastic problems, all aspects and results that are related to the solution of the nonlinear equilibrium equations are general and they therefore also hold for viscoplasticity and creep, for instance. The equations of equilibrium are also referred to as the *global equations*, as they hold for the entire body, whereas the constitutive relations (17.4) are referred to as the *local equations*, as they hold for each material point.

The external loading $\boldsymbol{f}$ is assumed to be known and both (17.1) and (17.4) are nonlinear equation systems. However, since the elasto-plastic response of the material depends on the entire load history, we cannot simply solve (17.1) and (17.4) by imposing the entire external load directly. Instead, we have to adopt a *stepwise*, i.e. an *incremental solution procedure*, where the external loading is increased in small steps.

With reference to Fig. 17.1, we assume that we have reached state n where everything is known. That means the nodal displacements $\boldsymbol{a}_n$, the external forces $\boldsymbol{f}_n$, the strains $\boldsymbol{\varepsilon}_n$ and the stresses $\boldsymbol{\sigma}_n$ are all known quantities. To start the process, the state n may be taken to be the state where the body is completely unloaded. The external load is now increased to $\boldsymbol{f}_{n+1}$ and we want to determine $\boldsymbol{a}_{n+1}$, $\boldsymbol{\varepsilon}_{n+1}$ and $\boldsymbol{\sigma}_{n+1}$ at state $n + 1$.

Before a general approach to the solution of (17.1) is presented, it is of interest to discuss the simplest possible approach, the *Euler forward scheme*.

17.1 Euler forward scheme

In order to illustrate some important issues, we will start with the simple Euler forward scheme. Historically, this method was the first to be used for the solution of nonlinear finite element problems within solid mechanics. However,

it will turn out that this method possesses a fundamental drawback, but the illustration of this drawback may serve as a prelude to a more formal and correct manner of approaching our problem.

Since the constitutive relation (17.4) is given in an incremental fashion, it is tempting to differentiate (17.1) with respect to time to obtain

$$\int_V \boldsymbol{B}^T \dot{\boldsymbol{\sigma}} dV = \dot{\boldsymbol{f}} \tag{17.5}$$

where

$$\dot{\boldsymbol{f}} = \int_S \boldsymbol{N}^T \dot{\boldsymbol{t}} dS + \int_V \boldsymbol{N}^T \dot{\boldsymbol{b}} dV \tag{17.6}$$

According to (17.4) and our finite element approximation, we have

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{D}_t \dot{\boldsymbol{\varepsilon}} = \boldsymbol{D}_t \boldsymbol{B} \dot{\boldsymbol{a}}$$

Insertion of this expression into (17.5) provides

$$\boldsymbol{K}_t \dot{\boldsymbol{a}} = \dot{\boldsymbol{f}} \tag{17.7}$$

where the *tangential stiffness matrix* $\boldsymbol{K}_t$ is defined by

$$\boxed{\boldsymbol{K}_t = \int_V \boldsymbol{B}^T \boldsymbol{D}_t \boldsymbol{B} dV} \tag{17.8}$$

It is evident that $\boldsymbol{K}_t$ represents the current, i.e. the tangential stiffness of the body.

Referring to Fig. 17.1, everything is known at state n. The external load is increased to the known quantity $\boldsymbol{f}_{n+1}$ and we want to determine $\boldsymbol{a}_{n+1}$, $\boldsymbol{\varepsilon}_{n+1}$ and $\boldsymbol{\sigma}_{n+1}$ at state $n+1$. Therefore, multiplying (17.7) by dt and integrating from state n to state $n+1$, we obtain

$$\int_{\boldsymbol{a}_n}^{\boldsymbol{a}_{n+1}} \boldsymbol{K}_t d\boldsymbol{a} = \boldsymbol{f}_{n+1} - \boldsymbol{f}_n \tag{17.9}$$

The fundamental problem that we are faced with is that $\boldsymbol{K}_t$ is not a constant matrix; in fact, $\boldsymbol{K}_t$ depends on the displacements, cf. (17.8), since $\boldsymbol{D}_t$ does so. In addition, it is not even known how $\boldsymbol{K}_t$ varies from state n to state $n+1$. However, at state n everything is known, i.e. also the constitutive matrix $\boldsymbol{D}_t$ is known and the simplest thing we can do is to approximate $\boldsymbol{K}_t$ in (17.9) by $\boldsymbol{K}_t$ evaluated at state n. With the obvious notation

$$(\boldsymbol{K}_t)_n = \int_V \boldsymbol{B}^T (\boldsymbol{D}_t)_n \boldsymbol{B} dV$$

we then obtain approximately from (17.9) that

$$\boxed{(\boldsymbol{K}_t)_n (\boldsymbol{a}_{n+1} - \boldsymbol{a}_n) = \boldsymbol{f}_{n+1} - \boldsymbol{f}_n \qquad \textit{Euler forward}} \tag{17.10}$$

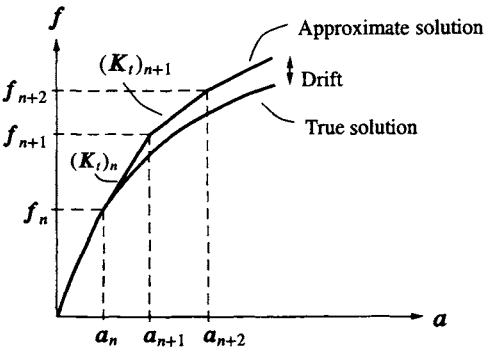

Figure 17.2: Euler forward scheme.

Except for a_{n+1}, everything is known and we can therefore solve this linear equation system to provide a_{n+1}. The approach obtained is termed the *Euler forward scheme*. The word 'forward' refers to the fact that we use our information about state n to determine state $n+1$ by a direct extrapolation.

When a_{n+1} has been determined from (17.10), the corresponding strain is given by $\varepsilon_{n+1} = Ba_{n+1}$. Moreover, the stress state σ_{n+1} is obtained by integration of the constitutive relation (17.4) from state n to state $n+1$. In a symbolic manner, we write

$$\sigma_{n+1} = \sigma_n + \int_{\varepsilon_n}^{\varepsilon_{n+1}} D_t d\varepsilon \tag{17.11}$$

For points that respond elastically this integration is trivial. However, for points that respond in a elasto-plastic manner, the integration is far from being trivial. We will deal with this problem in detail in the next chapter; at the present stage, we simply accept that this integration can be performed.

Since (17.10) is certainly an approximate solution to our original problem, the repeated use of (17.10) to trace the response of the body is bound to introduce some errors. An illustration of this effect is shown in Fig. 17.2.

In this figure, we assume that the exact solution is known at state n. Increasing the external load from f_n to f_{n+1}, (17.10) then determines the nodal displacements a_{n+1}. Since $(K_t)_n$ is the tangential stiffness matrix at state n, we obtain the over-shooting effect shown in Fig. 17.2. If this process is now repeated with the state $n+1$ considered to be the known state, we obtain the response to the next step also shown in Fig. 17.2. It is apparent that use of the Euler forward scheme gives rise to a certain drift from the true solution.

This immediately raises the question whether it is possible to correct this drift and thereby improve our solution. In order to do so, we must find a quantitative expression for the drift and this problem is complicated by the fact that, in general, we do not know the true solution. The key point in this discussion

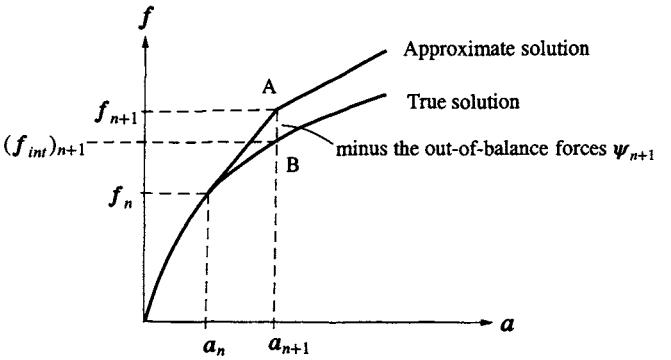

Figure 17.3: Discussion of drift and equilibrium; $(f_{int})_{n+1}$ denotes the internal forces corresponding to the stresses σ_{n+1}. These stresses are calculated from the nodal displacements a_{n+1}.

turns out to be that of equilibrium.

Referring to (17.2), it is recalled that f_{int} represents the internal forces that the stresses give rise to. Since f denotes the known external forces, (17.1) states that equilibrium of the body requires that the external forces be equal to the internal forces. With the stresses σ_{n+1} determined by (17.11), we can calculate the internal forces from (17.3) according to

$$(f_{int})_{n+1} = \int_V B^T \sigma_{n+1} dV$$

The internal forces $(f_{int})_{n+1}$ correspond to the stresses σ_{n+1} and these stresses are calculated from the nodal displacements a_{n+1}. Since the integration (17.11) may for the time being be assumed to be exact, the internal forces $(f_{int})_{n+1}$ take the value illustrated in Fig. 17.3. It is evident that $(f_{int})_{n+1} - f_{n+1}$, i.e. the drift indicated in Fig. 17.2, is expressed by the fact that the Euler forward scheme does not fulfill equilibrium of the body. The drift $(f_{int})_{n+1} - f_{n+1}$ is also called the *out-of-balance* or *residual forces* and the equilibrium condition (17.1) states that the out-of-balance forces, i.e. the residual forces, must be zero.

We have seen that the Euler forward scheme results in a certain drift from the true solution and this drift manifests itself in the form that the residual forces are different from zero, i.e. the equilibrium conditions for the body are not fulfilled.

Against this background, it seems natural to evaluate the nonlinear equilibrium equations (17.1) in more detail and investigate how we can ensure that our solution fulfills these conditions. Before we turn to this subject, we summarize the Euler forward algorithm as shown in Box 17.1. In this box, the means to enforce the boundary conditions has not been indicated; this aspect is dealt with in Section 17.6.

Box 17.1 Euler forward algorithm

- *Initiation of quantities*

  $\boldsymbol{a}_0 = \boldsymbol{0}\,; \quad \boldsymbol{\varepsilon}_0 = \boldsymbol{0} \quad \boldsymbol{\sigma}_0 = \boldsymbol{0}\,; \quad \boldsymbol{f}_0 = \boldsymbol{0}$

- *For load step* $n = 0, 1, 2, \ldots . N_{max}$

  - *Determine new load level* $\boldsymbol{f}_{n+1}$
  - *Calculate* $\boldsymbol{K}_t = \int_V \boldsymbol{B}^T (\boldsymbol{D}_t)_n \boldsymbol{B} dV$
  - *Calculate* $\boldsymbol{a}_{n+1}$ *from* $\boldsymbol{K}_t(\boldsymbol{a}_{n+1} - \boldsymbol{a}_n) = \boldsymbol{f}_{n+1} - \boldsymbol{f}_n$
  - *Calculate* $\boldsymbol{\varepsilon}_{n+1} := \boldsymbol{B}\boldsymbol{a}_{n+1}$
  - *Determine* $\boldsymbol{\sigma}_{n+1}$ *by integration of the constitutive equations* (see next chapter)
  - *Accept quantities*

    $\boldsymbol{a}_{n+1}\,; \quad \boldsymbol{\varepsilon}_{n+1}\,; \quad \boldsymbol{\sigma}_{n+1}$

- *End load step loop*

17.2 General iteration format

We have argued for the requirement that our solution fulfills the equilibrium equations (17.1). Since these equations are nonlinear, this section is devoted to a discussion of some general aspects relating to the solution of nonlinear equations and in the next sections, we will then adopt these viewpoints to the nonlinear FE equations.

Suppose that we have some unknowns collected in the column matrix $\boldsymbol{a}$. These unknowns are determined by the following nonlinear equation system

$$\boldsymbol{a} = \boldsymbol{F}(\boldsymbol{a}) \tag{17.12}$$

To solve this equation system, a common *iteration scheme* is

$$\boldsymbol{a}^i = \boldsymbol{F}(\boldsymbol{a}^{i-1})\,; \qquad i = 1, 2, \ldots \tag{17.13}$$

where i denotes the number of iterations and where $\boldsymbol{a}^0$ is the so-called *start vector* that must be specified by us ($\boldsymbol{a}^{i-1} = \boldsymbol{a}^0$ for $i = 1$).

Let the true solution of (17.12) be denoted by $\boldsymbol{z}$ and the question then arises whether the iteration scheme (17.13) converges towards the true solution $\boldsymbol{z}$. A strict mathematical proof that defines the conditions for which the iteration scheme (17.13) converges towards the true solution, is given, for instance, by Dahlquist and Björk (1974). However, it turns out to be difficult to make practical use of these mathematical conditions and we will therefore simply state the

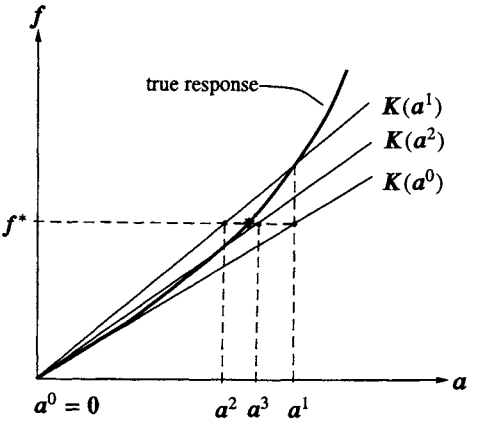

Figure 17.4: Convergent iteration scheme. True solution for external force f^* is indicated by $(*)$.

convergence theorem in the following verbal form

> *If the start vector a^0 is sufficiently close to the true solution of (17.12), then the iteration scheme (17.13) will converge towards this true solution*

What in this formulation is meant by 'sufficiently close' cannot be given an explicit formulation that is applicable in practice. However, it emphasizes an important point that also arises in nonlinear finite element calculations, namely that for a given elasto-plastic problem, say, we cannot, in general, be sure that a given solution scheme will provide the solution. Instead, we may have to modify the load steps, the number of iterations and even be forced to adopt another solution scheme in order to obtain a solution. In general therefore, nonlinear finite elements calculations are far from being trivial, but a solid knowledge of the underlying theory and experience in the solution of similar problems greatly enhance the possibility of achieving a solution.

To illustrate the procedure, we will consider the following equation system

$$\boldsymbol{K}(\boldsymbol{a})\boldsymbol{a} = \boldsymbol{f}$$

where $\boldsymbol{K}$ is a nonlinear global stiffness matrix, $\boldsymbol{f}$ the external load and $\boldsymbol{a}$ the nodal displacements; the stiffness matrix $\boldsymbol{K}(\boldsymbol{a})$ can be viewed as the secant stiffness. This expression is typical for elastic problems involving geometrical nonlinearities and in this case one may often have the response shown in Fig. 17.4. Here, the structure stiffens with increasing loading and an example is the response of a laterally loaded plate with fixed supports, which stiffens due to the membrane forces created by large deformations. According to (17.12)

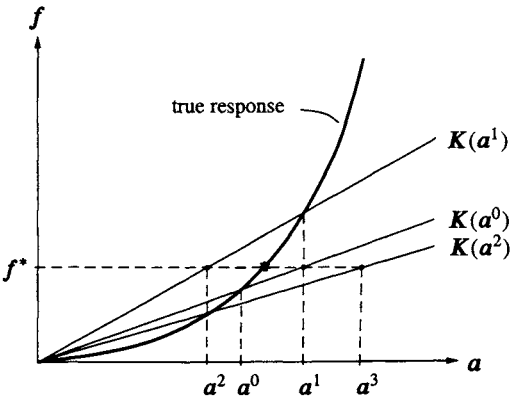

Figure 17.5: Divergent iteration scheme. True solution for external force f^* is indicated by $(*)$.

and (17.13), we adopt the following iteration scheme after the external load has been increased from zero to $f = f^*$

$$a^i = [K(a^{i-1})]^{-1} f^* ; \qquad i = 1, 2, \ldots$$

Choosing the start vector as $a^0 = 0$, then $K(a^0) = K(0)$ becomes equal to the initial elastic stiffness, and we then obtain the convergent iteration scheme shown in Fig. 17.4.

To illustrate that even small changes in the response may create a divergent iteration scheme, consider the response shown in Fig. 17.5. This response is quite close to that shown in Fig. 17.4, except that the stiffening effect is larger. The starting vector a^0 is now taken to be different from zero; in fact the start vector is rather close to the exact solution. Nevertheless, Fig. 17.5 shows a divergent iteration scheme and it is somewhat surprising that the qualitatively similar responses shown in Figs. 17.4 and 17.5 can give rise to completely different iteration courses.

17.3 Standard iteration format for equilibrium iterations - Iteration matrix

Having discussed the general iteration format which started with the nonlinear equation system (17.12) and resulted in the iteration scheme (17.13), we will now see how the nonlinear equilibrium equations arising in nonlinear finite element calculations can be treated in the same manner.

We have seen that the backbone in nonlinear FE analysis is the fulfillment of the nonlinear equilibrium equations (17.1). Since the stresses depend on the

nodal displacements $\boldsymbol{a}$, and since we want to fulfill equilibrium at a given fixed external loading, we have from (17.1)-(17.3)

$$\boldsymbol{\psi}(\boldsymbol{a}) = \mathbf{0} \tag{17.14}$$

This relation may also be expressed in the form of the homogeneous equation system

$$\mathbf{0} = -(\boldsymbol{A}(\boldsymbol{a}))^{-1}\boldsymbol{\psi}(\boldsymbol{a}) \tag{17.15}$$

where $\boldsymbol{A}^{-1}$ is a square matrix. In the most general case, $\boldsymbol{A}^{-1}$ may also depend on the unknown $\boldsymbol{a}$-values. In order that (17.15) should provide the trivial solution $\boldsymbol{\psi} = \mathbf{0}$ only, we must require

$$\boxed{\det \boldsymbol{A}^{-1} \neq 0} \tag{17.16}$$

This requirement is certainly fulfilled if $\boldsymbol{A}^{-1}$ is positive definite, but even if $\boldsymbol{A}^{-1}$ is not positive definite, its use in (17.15) is acceptable as long as (17.16) holds.

Expression (17.15) may even be formulated as

$$\boldsymbol{a} = \boldsymbol{a} - (\boldsymbol{A}(\boldsymbol{a}))^{-1}\boldsymbol{\psi}(\boldsymbol{a})$$

or

$$\boldsymbol{a} = \boldsymbol{F}(\boldsymbol{a}) \quad \text{where} \quad \boldsymbol{F}(\boldsymbol{a}) = \boldsymbol{a} - (\boldsymbol{A}(\boldsymbol{a}))^{-1}\boldsymbol{\psi}(\boldsymbol{a}) \tag{17.17}$$

By these manipulations, we have retrieved formulation (17.12).

In accordance with the general iteration scheme (17.13), we then obtain from (17.17) that

$$\boldsymbol{a}^i = \boldsymbol{a}^{i-1} - (\boldsymbol{A}(\boldsymbol{a}^{i-1}))^{-1}\boldsymbol{\psi}(\boldsymbol{a}^{i-1})$$

which may be written as

$$\boxed{\begin{array}{l}\textit{Standard iteration format} \\ \boldsymbol{A}(\boldsymbol{a}^{i-1})(\boldsymbol{a}^i - \boldsymbol{a}^{i-1}) = -\boldsymbol{\psi}(\boldsymbol{a}^{i-1})\,; \qquad i = 1, 2, \cdots \\ \textit{where } \boldsymbol{A} \textit{ is the iteration matrix}\end{array}} \tag{17.18}$$

All quantities denoted by $i - 1$ are known quantities and (17.18) therefore determines the new nodal displacement estimate $\boldsymbol{a}_i$. This iteration scheme is the vehicle by which we solve (17.14) and it will be referred to as *the standard iteration format*; the matrix $\boldsymbol{A}$ is called the *iteration matrix*.

The iteration matrix $\boldsymbol{A}$ must be chosen by us and we will later see how the format (17.18) enables us to derive a very large group of iteration methods applicable to nonlinear FE calculations in a very simple and convenient manner. Before we probe further into this topic, it is worthwhile to scrutinize the restrictions that we must place on the iteration matrix $\boldsymbol{A}$ in more detail. We have

already indicated restriction (17.16). Suppose that equilibrium is satisfied, i.e. $\boldsymbol{\psi}(\boldsymbol{a}^{i-1}) = \mathbf{0}$. Then the correct solution has been achieved and the iteration scheme (17.18) should then imply no further changes of the nodal displacements, i.e. $\boldsymbol{a}^i = \boldsymbol{a}^{i-1}$. However, the only possibility that $\boldsymbol{A}(\boldsymbol{a}^i - \boldsymbol{a}^{i-1}) = \mathbf{0}$ provides this solution is

$$\boxed{\det \boldsymbol{A} \neq 0} \tag{17.19}$$

Apart from the restrictions given by (17.16) and (17.19), the iteration matrix may be chosen arbitrarily, but we will experience that especially restriction (17.19) has important consequences for nonlinear finite element calculations.

Referring to our fundamental problem illustrated in Fig. 17.1, we start from state n where equilibrium is fulfilled and where the nodal displacements $\boldsymbol{a}_n$, the strains $\boldsymbol{\varepsilon}_n$, the stresses $\boldsymbol{\sigma}_n$ and the external loading $\boldsymbol{f}_n$ are all known. The external loading is then increased to $\boldsymbol{f}_{n+1}$ and we then want to determine the corresponding displacements $\boldsymbol{a}_{n+1}$, the strains $\boldsymbol{\varepsilon}_{n+1}$ and thereby also the stresses $\boldsymbol{\sigma}_{n+1}$ (obtained by an integration of the constitutive equations). The purpose of the iteration scheme (17.18) is to fulfill equilibrium at state $n+1$. Since the external loading is given by $\boldsymbol{f}_{n+1}$, the out-of-balance forces $\boldsymbol{\psi}(\boldsymbol{a}^{i-1})$ defined by (17.2)-(17.3) become

$$\boxed{\boldsymbol{\psi}(\boldsymbol{a}^{i-1}) = \int_V \boldsymbol{B}^T \boldsymbol{\sigma}^{i-1} dV - \boldsymbol{f}_{n+1}} \tag{17.20}$$

Whereas the external loading $\boldsymbol{f}_{n+1}$ is specified by us, the stress $\boldsymbol{\sigma}^{i-1}$ must be obtained by integration of the constitutive relations (17.4). The last state where the stresses were known and where the equilibrium equations were satisfied was state n. For each point in the body, $\boldsymbol{\sigma}^{i-1}$ is therefore obtained by the following symbolic integration

$$\begin{aligned} \boldsymbol{\sigma}^{i-1} &= \boldsymbol{\sigma}_n + \boldsymbol{D}(\boldsymbol{\varepsilon}^{i-1} - \boldsymbol{\varepsilon}_n) \quad \text{if the point behaves elastically} \\ \boldsymbol{\sigma}^{i-1} &= \boldsymbol{\sigma}_n + \int_{\boldsymbol{\varepsilon}_n}^{\boldsymbol{\varepsilon}^{i-1}} \boldsymbol{D}^{ep} d\boldsymbol{\varepsilon} \quad \text{if the point behaves plastically} \end{aligned} \tag{17.21}$$

The specific manner by which the integration is performed when plastic behavior occurs is dealt with in the next chapter; at the present stage, we simply accept that this integration can be performed. In practice, the determination of the stresses $\boldsymbol{\sigma}^{i-1}$ by means of (17.21) is not performed for all points in the body, but only for the *Gauss points*.

To start the general iteration scheme (17.18) for $i = 1$, we need to specify a starting value of $\boldsymbol{a}$, i.e. we have to specify $\boldsymbol{a}^0$. The most recent known value of $\boldsymbol{a}$ is $\boldsymbol{a}_n$ and it is therefore appropriate to choose $\boldsymbol{a}^0 = \boldsymbol{a}_n$ and thereby $\boldsymbol{\varepsilon}^0 = \boldsymbol{\varepsilon}_n$ which, according to (17.21) implies $\boldsymbol{\sigma}^0 = \boldsymbol{\sigma}_n$. Likewise, as the iteration matrix

$\boldsymbol{A}^0$ we take the one given by $\boldsymbol{A}_n$. The starting conditions for the iteration scheme therefore become

$$\boldsymbol{a}^0 = \boldsymbol{a}_n \; ; \qquad \boldsymbol{\sigma}^0 = \boldsymbol{\sigma}_n \; ; \qquad \boldsymbol{A}^0 = \boldsymbol{A}_n \tag{17.22}$$

From (17.20), we then obtain

$$\boldsymbol{\psi}^0 = \int_V \boldsymbol{B}^T \boldsymbol{\sigma}_n dV - \boldsymbol{f}_{n+1} \tag{17.23}$$

By means of (17.18), we have within each load step created a series of iterations which gradually improve the solution. The quantity that controls the iterations is the out-of-balance forces $\boldsymbol{\psi}(\boldsymbol{a}^{i-1})$, which measure the difference between the internal forces $\int_V \boldsymbol{B}^T \boldsymbol{\sigma}^{i-1} dV$ and the external forces $\boldsymbol{f}_{n+1}$. When the out-of-balance forces approach zero, the correction to the nodal displacements also approaches zero. In practice therefore, the iteration scheme (17.18) is stopped when the out-of-balance forces become smaller than a certain amount specified by the user of the FE program. We shall return to this subject in Section 17.7. When $\boldsymbol{\psi}(\boldsymbol{a}^{i-1})$ is small enough, we accept the solution $\boldsymbol{a}^{i-1}$, i.e

When convergence is accepted

$$\boldsymbol{a}_{n+1} = \boldsymbol{a}^{i-1} \; ; \quad \boldsymbol{\varepsilon}_{n+1} = \boldsymbol{\varepsilon}^{i-1} \; ; \quad \boldsymbol{\sigma}_{n+1} = \boldsymbol{\varepsilon}^{i-1} \tag{17.24}$$

Since the objective of the iterations indicated in (17.18) is to fulfill the equilibrium equations, these iterations are called *equilibrium iterations*. Moreover, since the external load is increased incrementally and as iterations are performed within each load step, this approach is often called an *incremental-iterative approach.*

Let us return to the integration of the constitutive equations given by (17.21). The integration limits turn out to be very important and we observe that they are given by $\boldsymbol{\varepsilon}_n$, which is the last state where the equilibrium conditions were satisfied, and $\boldsymbol{\varepsilon}^{i-1}$ which is the current value of the strains. That is, with the formulation (17.21), we deliberately ignore the intermediate states between $\boldsymbol{\varepsilon}_n$ and $\boldsymbol{\varepsilon}^{i-1}$ since these intermediate states may, for instance, involve false elastic unloading excursions. If we performed the integration of the constitutive equations by first integrating from $\boldsymbol{\varepsilon}_n$ to $\boldsymbol{\varepsilon}^1$, then from $\boldsymbol{\varepsilon}^1$ to $\boldsymbol{\varepsilon}^2$ and so on, we would force the stress state $\boldsymbol{\sigma}^{i-1}$ to depend on some possibly false intermediate states, i.e. false load history and this pitfall is avoided by formulation (17.21). Apart from that in the next chapter, we shall see how integration of (17.21) is performed in practice.

It may come as some surprise that the only point where the constitutive equations are involved is in the integration scheme (17.21). Moreover, the iteration strategy (17.18) has the very important property that possible non-zero out-of-balance forces in one load step are automatically transferred to the next load step. That is, there is no possibility for an accumulation of non-zero out-of-balance forces.

It is of significant interest that all iteration strategies proposed in the literature and used to solve nonlinear FE equations are, in principle, embraced by the scheme (17.18) and (17.20)-(17.24). The key point is the selection of the iteration matrix $\boldsymbol{A}$ where

The choice of a specific iteration matrix $\boldsymbol{A}$ *represents the establishment of a specific solution strategy* (17.25)

We recall that the only restrictions on the iteration matrix $\boldsymbol{A}$ are given by (17.16) and (17.19) and if (17.18) is to be an expression that is dimensionally correct, $\boldsymbol{A}$ must be a stiffness matrix; except for that, each choice of $\boldsymbol{A}$ represents a valid solution strategy. Certainly, this is not to say that all choices of $\boldsymbol{A}$ are equally promising, but the statement (17.25) is the vehicle that produces an arsenal of different methods where specific advantages and shortcomings are related to each choice of the iteration matrix $\boldsymbol{A}$. With this general framework, we will now discuss some classical and some more recent iteration strategies.

17.4 Newton-Raphson scheme

It was mentioned that the choice of different iteration matrices $\boldsymbol{A}$ in the standard iteration scheme (17.18) creates different solution strategies. The choice of $\boldsymbol{A}$ is subject to the constraints (17.16) and (17.19) and that it should represent some stiffness matrix, but apart from that we can make any choice. Evidently, not all choices create an efficient solution scheme and to obtain some kind of feeling for what is an appropriate choice, we will now identify the iteration matrix $\boldsymbol{A}$ that corresponds to the well-known *Newton-Raphson scheme*. Often this method is simply called *Newton's method*, but it was derived simultaneously by Raphson; Bićanić and Johnson (1979) give the relevant historical background.

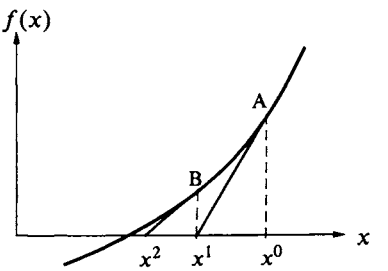

Figure 17.6: Newton-Raphson strategy for a one-dimensional problem.

For a one-dimensional problem, the essence of the Newton-Raphson strategy is illustrated in Fig. 17.6. The problem is to identify the solution to the

nonlinear equation $f(x) = 0$. A starting value x^0 is guessed by us. At the corresponding point A on the curve $f(x)$, the tangent is determined and this tangent is extrapolated to obtain the next estimate x^1 for the solution. This process is then repeated so that the tangent at point B provides the next estimate x^2, etc.

It appears that the fundamental idea of the Newton-Raphson approach is to linearize the nonlinear function about a given point. This means that the nonlinear function is approximated by a Taylor expansion about the point in question and in this Taylor expansion, terms higher than the linear ones are ignored.

In our case, the nonlinear multi-dimensional function is given by the equilibrium equations, i.e.

$$\boldsymbol{\psi}(\boldsymbol{a}) = \mathbf{0}$$

where

$$\boldsymbol{\psi}(\boldsymbol{a}) = \int_V \boldsymbol{B}^T \boldsymbol{\sigma} dV - \boldsymbol{f} \tag{17.26}$$

and the external forces $\boldsymbol{f}$ are known and fixed whereas the stresses $\boldsymbol{\sigma}$ depend on the nodal displacements $\boldsymbol{a}$. Assume now that the approximation $\boldsymbol{a}^{i-1}$ to the true solution $\boldsymbol{a}$ has been established. Ignoring higher-order terms, a Taylor expansion of $\boldsymbol{\psi}$ about $\boldsymbol{a}^{i-1}$ yields

$$\boldsymbol{\psi}(\boldsymbol{a}^i) = \boldsymbol{\psi}(\boldsymbol{a}^{i-1}) + (\frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{a}})^{i-1}(\boldsymbol{a}^i - \boldsymbol{a}^{i-1}) \tag{17.27}$$

This expression provides the linearized approximation to the true expression for $\boldsymbol{\psi}(\boldsymbol{a}^i)$; it therefore represents the tangent to the curve at point $\boldsymbol{a}^{i-1}$. Similar to the one-dimensional case, we require $\boldsymbol{\psi}(\boldsymbol{a}^i) = \mathbf{0}$ and it then follows from (17.27) that

$$\mathbf{0} = \boldsymbol{\psi}(\boldsymbol{a}^{i-1}) + (\frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{a}})^{i-1}(\boldsymbol{a}^i - \boldsymbol{a}^{i-1}) \tag{17.28}$$

To proceed further, we have to identify the derivative $\partial \boldsymbol{\psi}/\partial \boldsymbol{a}$. Since the external loading is fixed, (17.26) implies

$$\frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{a}} = \int_V \boldsymbol{B}^T \frac{d\boldsymbol{\sigma}}{d\boldsymbol{a}} dV \tag{17.29}$$

From the constitutive relation $\dot{\boldsymbol{\sigma}} = \boldsymbol{D}_t \dot{\boldsymbol{\varepsilon}}$, cf. (17.4), and noting that the variable now is $\boldsymbol{a}$, we obtain

$$d\boldsymbol{\sigma} = \boldsymbol{D}_t d\boldsymbol{\varepsilon} = \boldsymbol{D}_t \boldsymbol{B} d\boldsymbol{a}$$

i.e.

$$\frac{d\boldsymbol{\sigma}}{d\boldsymbol{a}} = \boldsymbol{D}_t \boldsymbol{B} \tag{17.30}$$

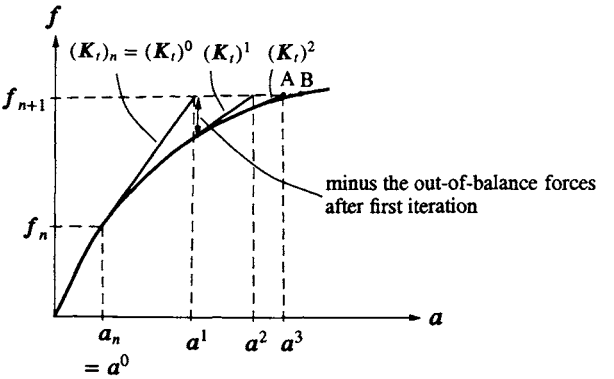

Figure 17.7: Newton-Raphson scheme showing the equilibrium iterations. Point B is the true solution and point A is the solution obtained after three iterations.

Insertion of (17.30) into (17.29) yields

$$\frac{\partial \psi}{\partial a} = K_t \quad \text{where} \quad K_t = \int_V B^T D_t B dV \tag{17.31}$$

where K_t is the tangent stiffness matrix of the body, cf. (17.8).

With this result, (17.28) takes the form

$$(K_t)^{i-1}(a^i - a^{i-1}) = -\psi(a^{i-1}) \tag{17.32}$$

and a comparison with the standard iteration scheme (17.18) shows that

$$\boxed{A = K_t \;\Rightarrow\; \textit{Newton-Raphson scheme}} \tag{17.33}$$

Therefore, this choice of the iteration matrix A results in the well-known Newton-Raphson approach.

For the first iteration $i = 1$, we find with (17.23) and the starting conditions (17.22) that

$$(K_t)_n(a^1 - a_n) = f_{n+1} - \int_V B^T \sigma_n dV \qquad \text{first iteration}$$

If equilibrium is fulfilled at state n then $f_n = \int_V B^T \sigma_n dV$ and we immediately observe that the first iteration is, in fact, identical to the Euler forward scheme, cf. (17.10). From (17.32) follows that in each equilibrium iteration, the Newton-Raphson approach makes use of the current tangential stiffness matrix and this feature is illustrated in Fig. 17.7. The Newton-Raphson approach is one of the most often used solution schemes and the reason is the fast convergence as illustrated in Fig. 17.7. Since the Newton-Raphson method starts with a Euler forward prediction and then continues with successive corrections

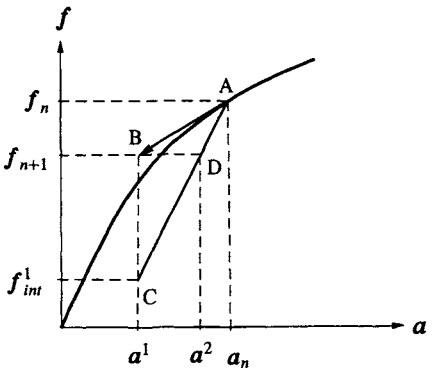

Figure 17.8: Newton-Raphson scheme during unloading of elasto-plastic body. Correct solution is obtained after two iterations.

to fulfill equilibrium, the terminology of a *predictor-corrector scheme* is often used.

In Fig. 17.7, we assumed that plastic loading occurs and it may be of interest to investigate the Newton-Raphson approach when elastic unloading occurs and Fig. 17.8 illustrates this issue. It appears that the correct response will be predicted after two iterations. First, unloading occurs along AB using the tangential stiffness at point A. Observing that, in reality, elastic unloading occurs along AC, the internal forces corresponding to the displacements $\boldsymbol{a}^1$ are given by point C. That is, at that stage the out-of-balance forces $\boldsymbol{\psi}(\boldsymbol{a}^1)$ become with (17.20) $\boldsymbol{\psi}(\boldsymbol{a}^1) = \boldsymbol{f}^1_{int} - \boldsymbol{f}_{n+1} = -|BC|$. When these residual forces are applied in (17.32) in the next iteration where $i = 2$ and observing that the current tangential stiffness now amounts to the elastic stiffness at point C, we arrive at the correct point D after the second iteration.

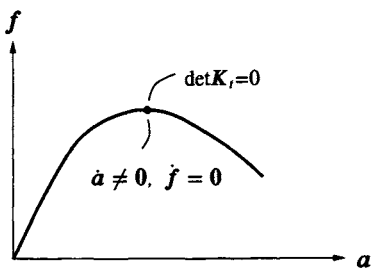

Figure 17.9: The Newton-Raphson scheme will have difficulties close to a peak load.

Since the tangential matrix $\boldsymbol{K}_t$ varies with the loading, it is of interest to investigate whether at some state in the Newton-Raphson scheme we may violate

Box 17.2 Newton-Raphson algorithm

- *Initiation of quantities*

  $\boldsymbol{a}_0 = \boldsymbol{0}\,; \quad \boldsymbol{\varepsilon}_0 = \boldsymbol{0}\,; \quad \boldsymbol{\sigma}_0 = \boldsymbol{0}\,; \quad \boldsymbol{f}_0 = \boldsymbol{0}\,; \quad \boldsymbol{f}_{int} = \boldsymbol{0}$

- *For load step* $n = 0, 1, 2, \ldots . N_{max}$

  - *Determine new load level* $\boldsymbol{f}_{n+1}$
  - *Initiation of iteration quantities*

    $\boldsymbol{a}^0 := \boldsymbol{a}_n$
  - *Iterate* $i = 1, 2, \ldots$ *until* $|\boldsymbol{\psi}|_{norm} = |\boldsymbol{f}_{int} - \boldsymbol{f}_{n+1}|_{norm} < epsilon$

    - *Calculate* $\boldsymbol{K}_t = \int_V \boldsymbol{B}^T \boldsymbol{D}_t^{i-1} \boldsymbol{B} dV$
    - *Calculate* $\boldsymbol{a}^i$ *from* $\boldsymbol{K}_t(\boldsymbol{a}^i - \boldsymbol{a}^{i-1}) = \boldsymbol{f}_{n+1} - \boldsymbol{f}_{int}$
    - *Calculate* $\boldsymbol{\varepsilon}^i := \boldsymbol{B}\boldsymbol{a}^i$
    - *Determine* $\boldsymbol{\sigma}^i$ *by integration of the constitutive equations* (see next chapter)
    - *Calculate internal forces* $\boldsymbol{f}_{int} = \int_V \boldsymbol{B}^T \boldsymbol{\sigma}^i dV$
  - *End iteration loop*
  - *Accept quantities*

    $\boldsymbol{a}_{n+1} := \boldsymbol{a}^i \; ; \quad \boldsymbol{\varepsilon}_{n+1} := \boldsymbol{\varepsilon}^i \; ; \quad \boldsymbol{\sigma}_{n+1} := \boldsymbol{\sigma}^i \; ; \quad \boldsymbol{f}_{int}$
- *End load step loop*

the important restriction given by (17.19), which with (17.33) becomes

$$\det \boldsymbol{K}_t \neq 0$$

This restriction is violated if $\det \boldsymbol{K}_t = 0$ which implies that the homogeneous equation system $\boldsymbol{K}_t \dot{\boldsymbol{a}} = \boldsymbol{0}$ possesses a non-trivial $\dot{\boldsymbol{a}}$-solution. As illustrated in Fig. 17.9, this situation corresponds to the peak load where we have $\dot{\boldsymbol{a}} \neq \boldsymbol{0}$, $\dot{\boldsymbol{f}} = \boldsymbol{0}$ which satisfy the equation $\boldsymbol{K}_t \dot{\boldsymbol{a}} = \dot{\boldsymbol{f}}$ from (16.17) when $\det \boldsymbol{K}_t = 0$. Use of the Newton-Raphson scheme will therefore imply increasing difficulties when a peak load is approached. If the response of the body, as illustrated in Fig. 17.9, exhibits a softening branch, great difficulties are encountered when we try to trace the response over the peak load and into the softening region. This latter observation is relevant, not only for the Newton-Raphson scheme, but also for other methods and we shall return to this subject later on.

We finally observe that in every Newton-Raphson iteration, a new $\boldsymbol{K}_t$-matrix needs to be established and that, in principle, the inverse matrix $(\boldsymbol{K}_t)^{-1}$ also

needs to be identified, cf. (17.32), i.e. each such iteration is costly. With this remark, we may summarize the properties of the Newton-Raphson scheme as follows:

* *Newton-Raphson works well both in loading and unloading* * *Newton-Raphson provides a fast convergence* * *problems may occur close to peak points* * *every Newton-Raphson iteration is costly*

(17.34)

The Newton-Raphson algorithm is summarized in Box. 17.2; here the means to impose the boundary conditions have not been indicated, see Section 17.6.

17.5 Initial stiffness and modified Newton-Raphson schemes

It was observed that the choice $\boldsymbol{A} = \boldsymbol{K}_t$ leads to the well-known and efficient Newton-Raphson scheme. However, we have previously stressed the freedom that we have when choosing the iteration matrix $\boldsymbol{A}$ in the standard iteration scheme (17.18). Against this background, it is not surprising that a number of choices exists for the iteration matrix which are modifications of the Newton-Raphson approach. Since such modifications exist, the true Newton-Raphson method is occasionally referred to as the *full* Newton-Raphson scheme. Scrutinizing the standard iteration scheme (17.18), it follows that any $\boldsymbol{A}$-matrix must have the same dimension as the stiffness matrix of the body.

Recognizing that every Newton-Raphson iteration is costly, it is tempting to choose a constant $\boldsymbol{A}$-matrix in all iterations. The most evident choice would then be $\boldsymbol{A} = \boldsymbol{K}$ where $\boldsymbol{K}$ is the linear elastic stiffness matrix. Since $\boldsymbol{K}$ is the initial stiffness of the body, this choice leads to the *initial stiffness method*, i.e.

$\boldsymbol{A} = \boldsymbol{K} \quad \Rightarrow \quad$ *Initial stiffness scheme*

Since the elastic stiffness matrix $\boldsymbol{K}$ is positive definite, we have $det\ \boldsymbol{K} \neq 0$ and none of the restrictions given by (17.16) and (17.19) are violated.

The performance of this approach is illustrated in Fig. 17.10 and a comparison with Fig. 17.7 clearly illustrates that the initial stiffness method simplifies the calculations considerably, but at the expense of a much slower convergence. This is especially true when the structure approaches its peak load, where the force-displacement curve is flat. For elastic unloading of the body, it is evident that the initial stiffness method will provide the correct response in just one

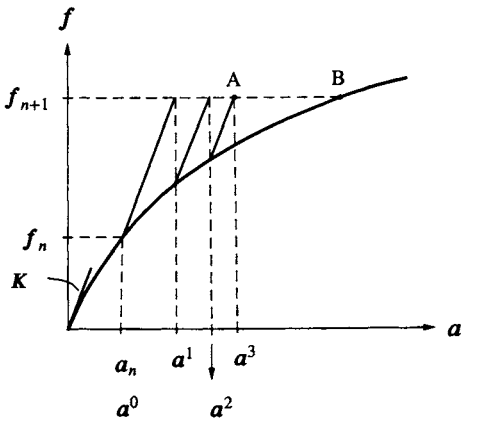

Figure 17.10: Initial stiffness method. Point B is the true solution and point A is the solution obtained after three iterations.

iteration. We are then led to the following conclusions

* *Initial stiffness approach works well both in loading and unloading*
* *every iteration is cheap*
* *Initial stiffness approach converges slowly*

The initial stiffness method may be recast into other equivalent forms that have been proposed in the literature. With Hooke's law $\boldsymbol{\sigma} = \boldsymbol{D}(\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^p) = \boldsymbol{DBa} - \boldsymbol{D\varepsilon}^p$, the internal forces become

$$\boldsymbol{f}_{int}^{i-1} = \int_V \boldsymbol{B}^T \boldsymbol{\sigma}^{i-1} dV = \boldsymbol{K}\boldsymbol{a}^{i-1} - \int_V \boldsymbol{B}^T \boldsymbol{D}(\boldsymbol{\varepsilon}^p)^{i-1} dV$$

where $\boldsymbol{K} = \int_V \boldsymbol{B}^T \boldsymbol{DB} dV$ is the elastic stiffness matrix. Insertion of the expression above into the general iteration scheme (17.18) and (17.20) with $\boldsymbol{A} = \boldsymbol{K}$ gives

$$\boldsymbol{K}\boldsymbol{a}^i = \boldsymbol{f}_{n+1} + \int_V \boldsymbol{B}^T \boldsymbol{D}(\boldsymbol{\varepsilon}^p)^{i-1} dV \tag{17.35}$$

This formulation was suggested by Argyris (1965b) under the name *initial strain method.* The reason for this terminology is that linear elasticity with initial strains is given by $\boldsymbol{\sigma} = \boldsymbol{D}(\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^o)$ where $\boldsymbol{\varepsilon}^o$ are the initial strains, cf. (4.62). It appears that the plastic strains $\boldsymbol{\varepsilon}^p$ may be interpreted as initial strains. On the other hand, if we define the quantity $\boldsymbol{\sigma}^o$ by $\boldsymbol{\sigma}^o = -\boldsymbol{D\varepsilon}^p$ then Hooke's law can be written as $\boldsymbol{\sigma} = \boldsymbol{D\varepsilon} + \boldsymbol{\sigma}^o$, i.e. for zero strains, we have $\boldsymbol{\sigma} = \boldsymbol{\sigma}^o$ and $\boldsymbol{\sigma}^o$ may therefore be viewed as initial stresses. Replacing $\boldsymbol{D\varepsilon}^p$ by $-\boldsymbol{\sigma}^o$, the

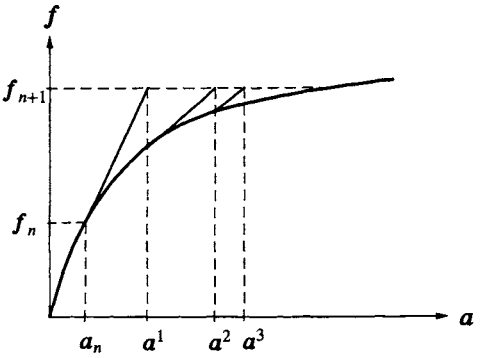

Figure 17.11: Modified Newton-Raphson. Updating of the tangential stiffness matrix after the first iteration in each load step.

scheme (17.35) was proposed by Zienkiewicz *et al.* (1969) under the name: *initial stress method.* However, as we have seen, both the initial strain and the initial stress method are identical to the initial stiffness method; it is only the physical interpretation that differs.

A compromise between full Newton-Raphson, where updating of the tangential stiffness matrix $\boldsymbol{K}_t$ occurs in every iteration, and the initial stiffness method, where no updating occurs at all, is the modification where updating is performed only once in each load step. This approach is called *modified Newton-Raphson.* Intuitively, it seems reasonable to perform this updating at the beginning of the load step. However, we experienced in Fig. 17.8 that full Newton-Raphson predicts the correct response during elastic unloading after two iterations. Therefore, the approach suggested above would cause problems during unloading. Instead, the updating of the tangential stiffness matrix is chosen to occur after the first equilibrium iteration in each load step. This approach is illustrated in Fig. 17.11.

Finally, we may mention the so-called *self-correcting procedure* of Stricklin *et al.* (1971) and further elaborated on by Stricklin and Haisler (1977). Essentially, it consists of using the full Newton-Raphson approach, but make just one equilibrium iteration in every load step. That is, the general iteration scheme (17.18) is used with $i = 1$ and with $\boldsymbol{A} = \boldsymbol{K}_t$. We have previously emphasized that the scheme (17.18) has the advantage that possible non-zero out-of-balance forces in one load step are automatically transferred to the next load step and the self-correcting procedure takes full advantage of this property.

To illustrate this issue, assume that the numerical procedure has determined point A in Fig. 17.12 to be the response at state n. For the nodal displacements $\boldsymbol{a}_n$, the corresponding internal forces are given by $(\boldsymbol{f}_{int})_n = \int_V \boldsymbol{B}^T \boldsymbol{\sigma}_n dV$, i.e. point P. The out-of-balance forces are therefore given by $(\boldsymbol{f}_{int})_n - \boldsymbol{f}_n = -|\mathrm{AP}|$. When the external loading is increased to $\boldsymbol{f}_{n+1}$, we obtain from (17.18) with

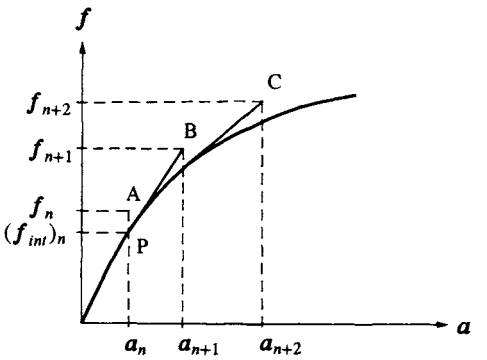

Figure 17.12: Self-correcting procedure of Stricklin *et al.* (1971); non-zero out-of-balance forces are automatically transferred to the next load step.

$\boldsymbol{A} = \boldsymbol{K}_t$ and $i = 1$

$$(\boldsymbol{K}_t)^0(\boldsymbol{a}^1 - \boldsymbol{a}^0) = -\boldsymbol{\psi}^0 \tag{17.36}$$

Since, we take just one iteration, the quantity $\boldsymbol{a}^1$ are taken as $\boldsymbol{a}_{n+1}$. With (17.22) and (17.23), (17.36) then becomes

$$(\boldsymbol{K}_t)_n(\boldsymbol{a}_{n+1} - \boldsymbol{a}_n) = -((\boldsymbol{f}_{int})_n - \boldsymbol{f}_{n+1}) = \boldsymbol{f}_{n+1} - \boldsymbol{f}_n - ((\boldsymbol{f}_{int})_n - \boldsymbol{f}_n)$$

It appears that the out-of-balance forces $(\boldsymbol{f}_{int})_n - \boldsymbol{f}_n$ at state n are automatically transferred to the next step. This feature is illustrated in Fig. 17.12 where the predicted response then is given by points A, B and C.

If elastic unloading occurs in a load step, the response in this load step will follow the tangential elasto-plastic stiffness instead of the elastic stiffness. It is only if elastic unloading continues in the next load step that the procedure will recognize the elastic unloading effects and use the elastic stiffness. Therefore the procedure will give an inferior prediction during elastic unloading and this can only be accepted as a reasonable approximation if the load steps are very small. The method is frequently used in, for instance, large strain plasticity problems where the load steps of other reasons are kept small and where the method often is referred to as the *incremental linear method.*

17.6 Consideration of boundary conditions

The essence in any equilibrium iteration is to produce a linear equation system that can be solved to provide the new nodal displacement estimate $\boldsymbol{a}^i$. In reality, this solution cannot be established before the boundary conditions of the problem have been introduced into the equation system and we have not, in fact, been precise on that point previously.

As an example, we may refer to the Newton-Raphson scheme given by (17.32) where the tangential stiffness matrix $\boldsymbol{K}_t$ defined by (17.31) enters the equation system. When establishing $\boldsymbol{K}_t$ in this manner, no considerations were taken of the kinematic boundary conditions; therefore, (17.32) also allows rigid-body motions. If $\dot{\boldsymbol{a}} \neq \boldsymbol{0}$ corresponds to a rigid-body motion, this would create no strains in the body, i.e. $\dot{\boldsymbol{\varepsilon}} = \boldsymbol{B}\dot{\boldsymbol{a}} = \boldsymbol{0}$. It follows that the homogeneous equation system $\boldsymbol{K}_t\dot{\boldsymbol{a}} = \boldsymbol{0}$ possesses a non-trivial solution $\dot{\boldsymbol{a}} \neq \boldsymbol{0}$ and $\boldsymbol{K}_t$ is therefore singular, i.e. $det\ \boldsymbol{K}_t = 0$. This means, in fact, that we cannot solve the equation system (17.32).

To obtain a solution of our boundary value problem, we therefore have to introduce the pertinent boundary conditions that, among other things, ensure that rigid-body motions are prevented.

Let us write the standard iteration scheme (17.18) and (17.20) in a simplified form as

$$\boldsymbol{A}\delta\boldsymbol{a} = \boldsymbol{f} - \boldsymbol{f}_{int} \tag{17.37}$$

where

$$\boldsymbol{A} = \boldsymbol{A}^{i-1} \; ; \qquad \delta\boldsymbol{a} = \boldsymbol{a}^{i} - \boldsymbol{a}^{i-1} \tag{17.38}$$

Moreover

$$\boldsymbol{f} = \int_S \boldsymbol{N}^T \boldsymbol{t}_{n+1} dS + \int_V \boldsymbol{N}^T \boldsymbol{b}_{n+1} dV \; ; \quad \boldsymbol{f}_{int} = \int_V \boldsymbol{B}^T \boldsymbol{\sigma}^{i-1} dV \tag{17.39}$$

where we notice that the stresses $\boldsymbol{\sigma}^{i-1}$ are known, i.e. the internal forces $\boldsymbol{f}_{int}$ are also known.

Since $\boldsymbol{A}$ may always be interpreted as a stiffness matrix of the body, it follows in accordance with the discussion above that

$$\det \boldsymbol{A} = 0$$

i.e., equation system (17.37) can first be solved after the boundary conditions have been introduced.

The boundary conditions are prescribed as

$$\boxed{\begin{aligned} \boldsymbol{u} &= \textit{ is given along } S_u \\ \boldsymbol{t} &= \textit{ is given along } S_t \end{aligned}}$$

That is, the displacement vector $\boldsymbol{u}$ is prescribed along the boundary surface S_u and the traction vector $\boldsymbol{t}$ is prescribed along the boundary surface S_t. This is illustrated in Fig. 17.13a).

We now split the finite element nodes into two groups: one group along the boundary S_u where the displacements are prescribed and another group which contains the remaining nodes. Correspondingly, the nodal displacements $\boldsymbol{a}$ can

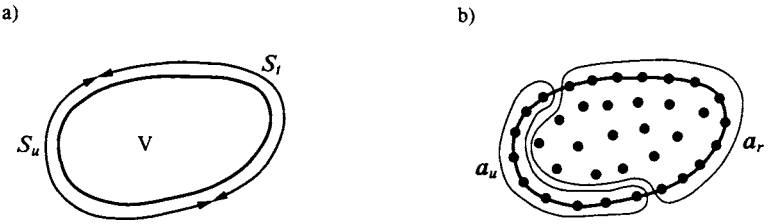

Figure 17.13: a) Along S_u the displacements are given and along S_t the traction vector is known; b) split of nodal displacements into prescribed values $\boldsymbol{a}_u$ along S_u and the remaining, i.e. unknown, values $\boldsymbol{a}_r$ in the remaining part of the body.

then be partitioned into $\boldsymbol{a}_u$ and $\boldsymbol{a}_r$, respectively, i.e. $\boldsymbol{a}_u$ are known whereas the remaining nodal displacements $\boldsymbol{a}_r$ are to be determined; this partitioning is illustrated in Fig. 17.13b). In accordance with this procedure, we may, in principle, partition (17.37) and with evident notation, we obtain

$$\begin{bmatrix} \boldsymbol{A}_{uu} & \boldsymbol{A}_{ur} \\ \boldsymbol{A}_{ru} & \boldsymbol{A}_{rr} \end{bmatrix} \begin{bmatrix} \delta \boldsymbol{a}_u \\ \delta \boldsymbol{a}_r \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}_u - (\boldsymbol{f}_{int})_u \\ \boldsymbol{f}_r - (\boldsymbol{f}_{int})_r \end{bmatrix} \tag{17.40}$$

Since all the internal forces $\boldsymbol{f}_{int}$ are known, both $(\boldsymbol{f}_{int})_u$ and $(\boldsymbol{f}_{int})_r$ are known forces. However, of the external forces $\boldsymbol{f}$, only the part $\boldsymbol{f}_r$ is, in fact, known whereas the part $\boldsymbol{f}_u$ is unknown, since it corresponds to the reactions where the displacements are prescribed.

From the second row of (17.40), we find

$$\boldsymbol{A}_{rr}\delta \boldsymbol{a}_r = \boldsymbol{f}_r - (\boldsymbol{f}_{int})_r - \boldsymbol{A}_{ru}\delta \boldsymbol{a}_u$$

where the right-hand side is known. Written explicitly, using (17.38) and (17.39), we have

$$\boxed{\boldsymbol{A}_{rr}^{i-1}(\boldsymbol{a}_r^i - \boldsymbol{a}_r^{i-1}) = (\boldsymbol{f}_{n+1})_r - (\boldsymbol{f}_{int}^{i-1})_r - \boldsymbol{p}_r^{i-1}} \tag{17.41}$$

where $\boldsymbol{p}_r^{i-1}$ represents the quantity

$$\boldsymbol{p}_r^{i-1} = \boldsymbol{A}_{ru}^{i-1}(\boldsymbol{a}_u^i - \boldsymbol{a}_u^{i-1})$$

The known displacement changes along the boundary S_u are imposed in the very first integration, i.e.

$$\boldsymbol{p}_r^0 = (\boldsymbol{A}_{ru})_n[(\boldsymbol{a}_u)_{n+1} - (\boldsymbol{a}_u)_n]$$

In the next iterations, no change occurs of the displacements $\boldsymbol{a}_u$ along S_u, i.e.

$$\boldsymbol{p}_r^{i-1} = \boldsymbol{0} \quad \text{when} \quad i = 2, 3, \ldots \tag{17.42}$$

In reality, it is on the equation system (17.41) we perform the equilibrium equations and all the previous as well as the following sections should be interpreted in this manner.

The iteration scheme will always take at least one iteration and it is only if the out-of-balance forces after this first iteration are too large that the iterations will continue. Therefore, we obtain with (17.41) and (17.42) that the part of the out-of-balance forces on which we measure equilibrium is $\boldsymbol{\psi}_r^{i-1} = (\boldsymbol{f}_{int}^{i-1})_r - (\boldsymbol{f}_{n+1})_r$. Consequently

$$\boxed{\begin{array}{l}\textit{Fulfillment of equilibrium is checked for the} \\ \textit{out-of-balance forces } \boldsymbol{\psi}_r^{i-1} = (\boldsymbol{f}_{int}^{i-1})_r - (\boldsymbol{f}_{n+1})_r\end{array}} \tag{17.43}$$

The iterations are stopped when equilibrium has been fulfilled. In that case $\delta \boldsymbol{a}_r = \boldsymbol{a}_r^i - \boldsymbol{a}_r^{i-1} = \boldsymbol{0}$ and as also $\delta \boldsymbol{a}_u = \boldsymbol{0}$ (after the first iteration), (17.40) shows that the reaction forces along S_u, i.e. $(\boldsymbol{f}_{n+1})_u$ are given by

$$(\boldsymbol{f}_{n+1})_u = (\boldsymbol{f}_{int}^{i-1})_u \tag{17.44}$$

Since the stresses $\boldsymbol{\sigma}^{i-1}$ are known, the internal force $\boldsymbol{f}_{int}^{i-1}$ and thereby also the part $(\boldsymbol{f}_{int}^{i-1})_u$ are known. After equilibrium has been obtained, the unknown reactions $(\boldsymbol{f}_{n+1})_u$ along S_u are therefore determined by (17.44).

On the other hand, if we always determine the reaction forces by (17.44) then equilibrium along the boundary S_u is always fulfilled, i.e. $\boldsymbol{\psi}_u^{i-1} = (\boldsymbol{f}_{int}^{i-1})_u - (\boldsymbol{f}_{n+1})_u = \boldsymbol{0}$, and as $(\boldsymbol{\psi}^{i-1})^T = [(\boldsymbol{\psi}_r^{i-1})^T, (\boldsymbol{\psi}_u^{i-1})^T]$, we may instead of (17.43) check the equilibrium conditions on the total out-of-balance force $\boldsymbol{\psi}^{i-1}$.

17.7 Convergence criteria

Let us return to the general iteration scheme (17.18)-(17.24) which starts with the iteration counter $i = 1$ and then continues for increasing i-values. Hopefully, the iteration scheme converges in the sense that the out–of–balance forces $\boldsymbol{\psi}$ approach zero, which implies that the new solution $\boldsymbol{a}^i$ differs only insignificantly from the previous solution $\boldsymbol{a}^{i-1}$. In practice however, the out–of–balance forces $\boldsymbol{\psi}$ will never be exactly zero, so we have to specify some threshold for $\boldsymbol{\psi}$ that terminates the iterations. Such a threshold value is called a *convergence criterion.*

Convergence criteria may be formulated in a number of different manners and for a more detailed discussion, we refer to Bathe and Cimento (1980), Bathe (1996), Bergan and Clough (1972) and Crisfield (1991). Instead of a threshold value for $\boldsymbol{\psi}$ we may measure the improvement of our solution in each iteration in terms of $\boldsymbol{a}^i - \boldsymbol{a}^{i-1}$ and apply a convergence criterion for this quantity.

To enforce a threshold value on the out–of–balance forces $\boldsymbol{\psi}^{i-1}$, we have to measure $\boldsymbol{\psi}^{i-1}$ against something and for this purpose it seems logical to choose

the external forces $\boldsymbol{f}_{n+1}$. The quantities $\boldsymbol{\psi}^{i-1}$ and $\boldsymbol{f}_{n+1}$ contain many components and in order to compare these quantities, we may use their lengths measured by the scalar product. Therefore, an often used force convergence criterion is given by

$$[(\boldsymbol{\psi}^{i-1})^T \boldsymbol{\psi}^{i-1}]^{1/2} \leq \varepsilon_{F_1} (\boldsymbol{f}_{n+1}^T \boldsymbol{f}_{n+1})^{1/2} \tag{17.45}$$

where ε_{F_1} expresses the threshold. Typical values for this threshold are $\varepsilon_{F_1} = 10^{-3} - 10^{-2}$, cf. Zienkiewicz and Taylor (1991) and Crisfield (1991), and the smaller the value ε_{F_1}, the less is the acceptable error in the out–of–balance forces. If ε_{F_1} is chosen too small, many (costly) iterations are performed without improving the solution significantly and if ε_{F_1} is chosen too large, the solution becomes inaccurate.

It is a general experience in the literature that it is difficult to recommend a specific tolerance that is always of relevance; one tolerance may work well in some case and may be inferior in other situations. For a given problem, one often starts with a rather crude tolerance and then performs calculations with narrower tolerances in an effort to judge whether a converged solution has been obtained.

In (17.45), the out-of-balance forces are compared with the total external force. When the load steps are small, the out-of-balance forces are often compared with the load step itself, i.e.

$$[(\boldsymbol{\psi}^{i-1})^T \boldsymbol{\psi}^{i-1}]^{1/2} \leq \varepsilon_{F_2} [\Delta \boldsymbol{f}^T \Delta \boldsymbol{f}]^{1/2}$$

where $\Delta \boldsymbol{f} = \boldsymbol{f}_{n+1} - \boldsymbol{f}_n$; Bathe and Cimento (1980) suggested $\varepsilon_{F_2} \approx 0.1$.

For plastic structures with a small plastic modulus H, force criteria may be misleading; the out-of-balance force may be small, but the structure is far from its correct displacement pattern. Therefore, in addition to the force convergence criteria, use is often made of some displacement convergence criteria, for instance

$$[(\delta \boldsymbol{u}^i)^T \delta \boldsymbol{u}^i]^{1/2} \leq \varepsilon_D [\boldsymbol{a}_n^T \boldsymbol{a}_n]^{1/2}$$

where $\delta \boldsymbol{u}^i = \boldsymbol{a}^i - \boldsymbol{a}^{i-1}$ and $\varepsilon_D \approx 10^{-3}$, cf. Bathe and Cimento (1980).

The convergence criteria above possess the drawback that quantities having one dimension may be added to quantities having another dimension; the components of $\boldsymbol{a}$ may involve both displacements and rotations and, likewise, the components of $\boldsymbol{\psi}$ may involve forces as well as moments. A convergence criterion in terms of an energy criterion avoids this problem since force components are here multiplied by displacements whereas moments are multiplied by rotations, i.e. it involves so-called *conjugated quantities*. Therefore an energy convergence criterion in the form

$$[(\boldsymbol{\psi}^{i-1})^T \boldsymbol{a}_n]^{1/2} \leq \varepsilon_E [\boldsymbol{f}_n^T \boldsymbol{a}_n]^{1/2}$$

is often adopted; the tolerance $\varepsilon_E = 10^{-3} - 10^{-2}$ is often used, cf. Zienkiewicz and Taylor (1991).

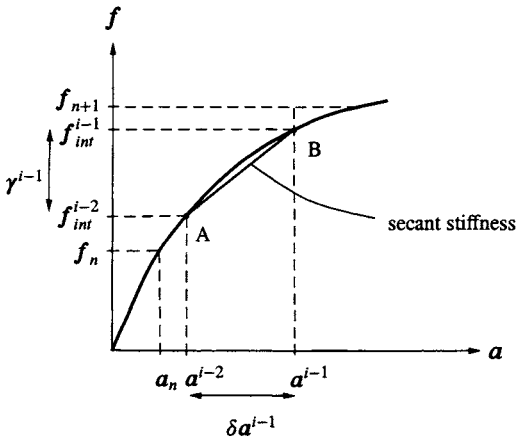

Figure 17.14: Current state given by point B and previous state given by point A; secant stiffness between these points.

17.8 Quasi-Newton methods

The Newton-Raphson scheme possesses a number of advantages among which its fast convergence is the major reason for its popularity, cf. (17.34). However, it was mentioned that every Newton-Raphson iteration requires the formulation of a new tangent stiffness matrix $\boldsymbol{K}_t$ and that, in principle, the inverse matrix $\boldsymbol{K}_t^{-1}$ also needs to be established; this means that every Newton-Raphson iteration is costly. This drawback can be obviated by the initial stiffness method where the elastic stiffness matrix is used in every iteration or in the modified Newton-Raphson method where the stiffness matrix is occasionally updated. However, these approaches have a slower convergence rate than the full Newton-Raphson scheme. It is of considerable interest that there exists an entirely different approach by which we can maintain a fast convergence without performing a costly matrix inversion in every iteration.

In the previous iteration schemes, we only took advantage of the information of the current state defined by $\boldsymbol{a}^{i-1}$ and $\boldsymbol{\psi}^{i-1}$ and with this information, we tried to establish a new estimate $\boldsymbol{a}^i$ and thereby $\boldsymbol{\psi}^i$. Essentially, this means that all previous information of the response is ignored. To illustrate this viewpoint, consider Fig. 17.14 and suppose that we increase the external loading from $\boldsymbol{f}_n$ to $\boldsymbol{f}_{n+1}$. Equilibrium iterations are then performed with, say, a Newton-Raphson procedure. At the present stage, we have then obtained the nodal displacements $\boldsymbol{a}^{i-1}$ at point B whereas the previous nodal displacements are given by $\boldsymbol{a}^{i-2}$ at point A.

In a Newton-Raphson procedure, we then determine the tangential stiffness matrix $\boldsymbol{K}_t$ at point B to obtain the new nodal displacement estimate $\boldsymbol{a}^i$. Our objective is now to establish a procedure that as closely as possible retains the

fast convergence of the Newton-Raphson approach. Instead of the tangential stiffness at point B, it is then tempting to use the secant stiffness between point A and point B. This secant stiffness is a close approximation to the tangential stiffness at point B. To define this secant stiffness matrix, we observe that the internal forces at point B are given by $\boldsymbol{f}_{int}^{i-1} = \int_V \boldsymbol{B}^T \boldsymbol{\sigma}^{i-1} dV$ whereas the internal forces at point A are given by $\boldsymbol{f}_{int}^{i-2} = \int_V \boldsymbol{B}^T \boldsymbol{\sigma}^{i-2} dV$. These internal forces are illustrated in Fig. 17.14. Define the quantities $\delta \boldsymbol{a}^{i-1}$ and $\boldsymbol{\gamma}^{i-1}$ by

$$\delta \boldsymbol{a}^{i-1} = \boldsymbol{a}^{i-1} - \boldsymbol{a}^{i-2}\,; \qquad \boldsymbol{\gamma}^{i-1} = \boldsymbol{f}_{int}^{i-1} - \boldsymbol{f}_{int}^{i-2} \tag{17.46}$$

These quantities are illustrated in Fig. 17.14. The secant stiffness $\boldsymbol{K}_s^{i-1}$ between point A and B is then defined by

$$\boldsymbol{K}_s^{i-1} \delta \boldsymbol{a}^{i-1} = \boldsymbol{\gamma}^{i-1}$$

which implies

$$\boxed{\delta \boldsymbol{a}^{i-1} = \boldsymbol{H}^{i-1} \boldsymbol{\gamma}^{i-1} \quad \textit{Quasi-Newton relation}} \tag{17.47}$$

where $\boldsymbol{H}^{i-1}$ is the inverse of $\boldsymbol{K}_s^{i-1}$. This relation is called the *quasi-Newton relation* since it defines a matrix $\boldsymbol{H}^{i-1}$ that is almost equal to the inverse of the tangential stiffness at point B that arises in the Newton-Raphson scheme. It appears that if in the general iteration scheme (17.18) we make the choice

$$(\boldsymbol{A}^{i-1})^{-1} = \boldsymbol{H}^{i-1} \tag{17.48}$$

then we will obtain a convergence that is almost as fast as the Newton-Raphson scheme.

Before we proceed further, assume that the total number of degrees of freedom in the finite element discretization is N, then the matrix $\boldsymbol{H}^{i-1}$ has the dimension $N \times N$. However, the quasi-Newton equation (17.47) - i.e. the secant relation - only comprises N equations and this implies that $\boldsymbol{H}^{i-1}$ is not defined uniquely by (17.47) unless in the trivial case where $N = 1$. To establish $\boldsymbol{H}^{i-1}$ so that it satisfies (17.47) therefore leaves us with considerable freedom.

That apart, our goal was to establish a method by which costly matrix inversions are avoided. Therefore, suppose that in the previous iteration we have obtained the matrix $\boldsymbol{H}^{i-2}$ we then want to be able to establish $\boldsymbol{H}^{i-1}$ by means of the simple updating scheme

$$\boxed{\boldsymbol{H}^{i-1} = \boldsymbol{H}^{i-2} + \boldsymbol{H}_{corr}^{i-1}} \tag{17.49}$$

where the *correction matrix* $\boldsymbol{H}_{corr}^{i-1}$ should be easy to determine. If this can be achieved, we have fulfilled our objective of devising a method with almost as good convergence properties as the Newton-Raphson method without having to perform costly inversions of some stiffness matrix.

Methods that satisfy (17.47)-(17.49) are called *quasi-Newton methods* - occasionally also called *variable matrix methods* - and they were originally derived within nonlinear optimization theory. The literature on quasi-Newton methods is very extensive and reviews are given by Dennis and Moré (1977), Fletcher (1980) and Luenberger (1984).

17.8.1 Rank one correction

To provide a simple illustration of a quasi-Newton method, suppose that we take the updating scheme (17.49) in the form

$$\boldsymbol{H}^{i-1} = \boldsymbol{H}^{i-2} + \boldsymbol{v}^{i-1}(\boldsymbol{v}^{i-1})^T \tag{17.50}$$

where $\boldsymbol{v}^{i-1}$ is a column matrix that is to be determined so that the quasi-Newton equation (17.47) is fulfilled. Insertion of (17.50) in (17.47) gives

$$\delta \boldsymbol{a}^{i-1} = \boldsymbol{H}^{i-2}\boldsymbol{\gamma}^{i-1} + \boldsymbol{v}^{i-1}((\boldsymbol{v}^{i-1})^T\boldsymbol{\gamma}^{i-1})$$

which provides

$$\boldsymbol{v}^{i-1} = \frac{1}{(\boldsymbol{v}^{i-1})^T\boldsymbol{\gamma}^{i-1}}(\delta \boldsymbol{a}^{i-1} - \boldsymbol{H}^{i-2}\boldsymbol{\gamma}^{i-1}) \tag{17.51}$$

and thereby

$$((\boldsymbol{v}^{i-1})^T\boldsymbol{\gamma}^{i-1})^2 = (\delta \boldsymbol{a}^{i-1} - \boldsymbol{H}^{i-2}\boldsymbol{\gamma}^{i-1})^T\boldsymbol{\gamma}^{i-1}$$

Insertion of (17.51) in (17.50) and use of the expression above give the result

$$\boldsymbol{H}^{i-1} = \boldsymbol{H}^{i-2} + \frac{(\delta \boldsymbol{a}^{i-1} - \boldsymbol{H}^{i-2}\boldsymbol{\gamma}^{i-1})(\delta \boldsymbol{a}^{i-1} - \boldsymbol{H}^{i-2}\boldsymbol{\gamma}^{i-1})^T}{(\delta \boldsymbol{a}^{i-1} - \boldsymbol{H}^{i-2}\boldsymbol{\gamma}^{i-1})^T\boldsymbol{\gamma}^{i-1}} \tag{17.52}$$

Since $\delta \boldsymbol{a}^{i-1}$ and $\boldsymbol{\gamma}^{i-1}$ are known, it appears that if $\boldsymbol{H}^{i-2}$ is also known then the new value of $\boldsymbol{H}^{i-1}$, i.e. the new value of the inverse of the secant stiffness matrix, is obtained by performing the simple operations in (17.52) and no costly inversion technique is necessary.

To further illustrate this approach, let us recall some results from matrix algebra, cf. for instance Strang (1980). For any square matrix $\boldsymbol{P}$, the eigenvalue problem is given by $\boldsymbol{P}\boldsymbol{c} = \lambda \boldsymbol{c}$, where $\boldsymbol{c}$ is the eigenvector and λ is the corresponding eigenvalue. The eigenvalue problem can also be written as the homogeneous equation system $(\boldsymbol{P} - \lambda \boldsymbol{I})\boldsymbol{c} = \boldsymbol{0}$ and a non-trivial $\boldsymbol{c}$-solution requires that $\det(\boldsymbol{P} - \lambda \boldsymbol{I}) = 0$; this condition is the *characteristic equation*. If the matrix $\boldsymbol{P}$ possesses an eigenvalue that is zero, then we must have $\det \boldsymbol{P} = 0$. Let the matrix $\boldsymbol{P}$ be of dimension $N \times N$, then the characteristic equation $\det(\boldsymbol{P} - \lambda \boldsymbol{I}) = 0$ becomes a polynomial of order N and we therefore have N eigenvalues. The number of non-zero eigenvalues of the matrix $\boldsymbol{P}$ is called the *rank of the matrix* and it is denoted by r. If $\det(\boldsymbol{P} - \lambda \boldsymbol{I}) = 0$ possesses no zero eigenvalues, the

rank is $r = N$. If it possesses one zero eigenvalue, the rank is $r = N - 1$ and if it possesses two zero eigenvalues, the rank is $r = N - 2$ and it is then said that the matrix $\boldsymbol{P}$ has the eigenvalue zero with a *multiplicity* of two. Therefore, if the homogeneous equation system $\boldsymbol{Pc} = \boldsymbol{0}$ has q non-trivial $\boldsymbol{c}$ - solutions that are linearly independent then the rank of $\boldsymbol{P}$ is $r = N - q$ and the matrix $\boldsymbol{P}$ possesses the zero eigenvalue with a multiplicity of q.

In the present case, consider the correction matrix $\boldsymbol{H}_{corr}^{i-1}$ which according to (17.49) and (17.50) is given by $\boldsymbol{H}_{corr}^{i-1} = \boldsymbol{v}^{i-1}(\boldsymbol{v}^{i-1})^T$ and the dimension of $\boldsymbol{H}_{corr}^{i-1}$ is $N \times N$. Investigate the possibility of a non-trivial solution of the homogeneous equation system $\boldsymbol{H}_{corr}^{i-1}\boldsymbol{c} = \boldsymbol{0}$, i.e. $\boldsymbol{v}^{i-1}((\boldsymbol{v}^{i-1})^T\boldsymbol{c}) = \boldsymbol{0}$. It appears that there are $q = N - 1$ linear independent $\boldsymbol{c}$-vectors that fulfill $(\boldsymbol{v}^{i-1})^T\boldsymbol{c} = 0$, i.e. $\boldsymbol{H}_{corr}^{i-1}$ has a zero eigenvalue with a multiplicity of $q = N - 1$. The rank of $\boldsymbol{H}_{corr}^{i-1}$ is therefore $r = N - q = N - (N - 1) = 1$. Therefore, the quasi-Newton method defined by (17.50) is called a *rank one correction.*

The rank one correction quasi-Newton method defined by (17.50) was suggested by Broyden (1967) and later independently by others. It appears that if $\boldsymbol{H}^{i-2}$ is symmetric then this method has the neat property that the new update $\boldsymbol{H}^{i-1}$ is also symmetric. However, if the denominator in (17.52) is also small this may lead to numerical difficulties. Therefore, we shall not dwell on this approach any further and here merely take it as a simple illustration of a quasi-Newton method.

17.8.2 Rank two correction. BFGS-method

To establish a quasi-Newton method that is of relevance for our purpose, consider the following form

$$\boldsymbol{H}^{i-1} = (\boldsymbol{I} + \boldsymbol{v}\boldsymbol{w}^T)^T\boldsymbol{H}^{i-2}(\boldsymbol{I} + \boldsymbol{v}\boldsymbol{w}^T) \tag{17.53}$$

where the matrix $\boldsymbol{v}\boldsymbol{w}^T$ must be chosen so that (17.53) fulfills the quasi-Newton equation (17.47). Carrying out the multiplications and assuming $\boldsymbol{H}^{i-2}$ to be symmetric we obtain

$$\boldsymbol{H}^{i-1} = \boldsymbol{H}^{i-2} + (\boldsymbol{v}^T\boldsymbol{H}^{i-2}\boldsymbol{v})\boldsymbol{w}\boldsymbol{w}^T + \boldsymbol{w}(\boldsymbol{H}^{i-2}\boldsymbol{v})^T + \boldsymbol{H}^{i-2}\boldsymbol{v}\boldsymbol{w}^T \tag{17.54}$$

which is evidently of the format given by (17.49). In the quasi-Newton approach defined by (17.50), the correction matrix $\boldsymbol{H}_{corr}^{i-1}$ was found to be of rank one. Let us now determine the rank of the correction matrix in the approach defined by (17.54). Therefore, we shall investigate non-trivial $\boldsymbol{c}$-solutions of the homogeneous equation system $\boldsymbol{H}_{corr}^{i-1}\boldsymbol{c} = \boldsymbol{0}$. With (17.54) and (17.49) and defining the vector $\boldsymbol{b}$ by $\boldsymbol{b} = \boldsymbol{H}^{i-2}\boldsymbol{v}$, the following equation system is therefore considered

$$\boldsymbol{H}_{corr}^{i-1}\boldsymbol{c} = (\boldsymbol{v}^T\boldsymbol{b})\boldsymbol{w}(\boldsymbol{w}^T\boldsymbol{c}) + \boldsymbol{w}(\boldsymbol{b}^T\boldsymbol{c}) + \boldsymbol{b}(\boldsymbol{w}^T\boldsymbol{c}) = \boldsymbol{0}$$

This equation system is fulfilled if both $\boldsymbol{w}^T\boldsymbol{c} = 0$ and $\boldsymbol{b}^T\boldsymbol{c} = 0$. With $\boldsymbol{H}_{corr}^{i-1}$ having the dimension $N \times N$, these conditions can be fulfilled for $N - 2$ linear

independent c-vectors. Referring to our previous discussion, $\boldsymbol{H}_{corr}^{i-1}$ has therefore the eigenvalue zero with a multiplicity of $q = N - 2$ and the rank r of $\boldsymbol{H}_{corr}^{i-1}$ is then $r = N - q = N - (N - 2) = 2$. Consequently, the quasi-Newton approach defined by (17.54) is a *rank two correction.*

The next topic is to identify the vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ in (17.54). Since the vector $\delta\boldsymbol{a}^{i-1}$ is one of the vectors we know, it seems natural to choose $\boldsymbol{w}$ to be proportional to $\delta\boldsymbol{a}^{i-1}$. According to (17.53), $\boldsymbol{v}$ and $\boldsymbol{w}$ only occur in the combination $\boldsymbol{v}\boldsymbol{w}^T$ or $\boldsymbol{w}\boldsymbol{v}^T$ and the length of one of these vectors can therefore be chosen arbitrarily; this will only affect the length of the other vector. Let us therefore choose $\boldsymbol{w}$ according to

$$\boldsymbol{w} = \frac{1}{b}\delta\boldsymbol{a}^{i-1} \qquad \text{where} \quad b = (\boldsymbol{\gamma}^{i-1})^T\delta\boldsymbol{a}^{i-1} \tag{17.55}$$

Since both $\delta\boldsymbol{a}^{i-1}$ and $\boldsymbol{\gamma}^{i-1}$ are known, also $\boldsymbol{w}$ is known.

With (17.55), (17.54) becomes

$$\boldsymbol{H}^{i-1} = \boldsymbol{H}^{i-2} + (\boldsymbol{v}^T\boldsymbol{H}^{i-2}\boldsymbol{v})\frac{1}{b^2}\delta\boldsymbol{a}^{i-1}(\delta\boldsymbol{a}^{i-1})^T + \frac{1}{b}\delta\boldsymbol{a}^{i-1}(\boldsymbol{H}^{i-2}\boldsymbol{v})^T + \frac{1}{b}\boldsymbol{H}^{i-2}\boldsymbol{v}(\delta\boldsymbol{a}^{i-1})^T \tag{17.56}$$

Insertion into the quasi-Newton equation (17.47) results in

$$\delta\boldsymbol{a}^{i-1} = \boldsymbol{H}^{i-2}\boldsymbol{\gamma}^{i-1} + (\boldsymbol{v}^T\boldsymbol{H}^{i-2}\boldsymbol{v})\frac{1}{b}\delta\boldsymbol{a}^{i-1} + \frac{1}{b}\delta\boldsymbol{a}^{i-1}(\boldsymbol{v}^T\boldsymbol{H}^{i-2}\boldsymbol{\gamma}^{i-1}) + \boldsymbol{H}^{i-2}\boldsymbol{v}$$

where it was assumed that $\boldsymbol{H}^{i-2}$ is symmetric. Determination of $\boldsymbol{v}$ from the expression gives

$$\boldsymbol{v} = m(\boldsymbol{H}^{i-2})^{-1}\delta\boldsymbol{a}^{i-1} - \boldsymbol{\gamma}^{i-1} \tag{17.57}$$

where

$$m = 1 - \frac{1}{b}\boldsymbol{v}^T\boldsymbol{H}^{i-2}(\boldsymbol{v} + \boldsymbol{\gamma}^{i-1}) \tag{17.58}$$

Expression (17.57) leads to

$$\boldsymbol{v}^T\boldsymbol{H}^{i-2}(\boldsymbol{v} + \boldsymbol{\gamma}^{i-1}) = m^2(\delta\boldsymbol{a}^{i-1})^T(\boldsymbol{H}^{i-2})^{-1}\delta\boldsymbol{a}^{i-1} - m\,b \tag{17.59}$$

where advantage was taken of (17.55). Insertion of (17.59) in (17.58) results in

$$m = \pm\left[\frac{b}{(\delta\boldsymbol{a}^{i-1})^T(\boldsymbol{H}^{i-2})^{-1}\delta\boldsymbol{a}^{i-1}}\right]^{1/2} \tag{17.60}$$

The vector $\boldsymbol{v}$ given by (17.57) and (17.60) has now been expressed in terms of known quantities. Moreover, this expression for $\boldsymbol{v}$ ensures that $\boldsymbol{H}^{i-1}$ given

by (17.56) fulfills the quasi-Newton equation (17.47). It is easy to be convinced that irrespective of the sign we choose for m in (17.60), (17.56) provides the same updated matrix $\boldsymbol{H}^{i-1}$. In (17.60), we shall for convenience choose the positive sign.

In view of this comment, the results (17.53), (17.55), (17.57) and (17.60) can be summarized as

BFGS-method

$$\boldsymbol{H}^{i-1} = (\boldsymbol{I} + \boldsymbol{v}\boldsymbol{w}^T)^T \boldsymbol{H}^{i-2}(\boldsymbol{I} + \boldsymbol{v}\boldsymbol{w}^T)$$

where

$$\boldsymbol{v} = \left[\frac{(\boldsymbol{\gamma}^{i-1})^T \delta\boldsymbol{a}^{i-1}}{(\delta\boldsymbol{a}^{i-1})^T (\boldsymbol{H}^{i-2})^{-1} \delta\boldsymbol{a}^{i-1}}\right]^{1/2} (\boldsymbol{H}^{i-2})^{-1} \delta\boldsymbol{a}^{i-1} - \boldsymbol{\gamma}^{i-1} \tag{17.61}$$

and

$$\boldsymbol{w} = \frac{\delta\boldsymbol{a}^{i-1}}{(\boldsymbol{\gamma}^{i-1})^T \delta\boldsymbol{a}^{i-1}}$$

This comprises the so-called *BFGS-method*, suggested by Broyden (1970), Fletcher (1970), Goldfarb (1970) and Shannon (1970). The BFGS-method can be expressed in various identical forms and the product form adopted in (17.61) was proposed by Matthies and Strang (1979). Indeed, Matthies and Strang (1979) were the first to apply the BFGS-method in a finite element context.

In the BFGS-scheme (17.61), we apparently need to establish the inverse matrix $(\boldsymbol{H}^{i-2})^{-1}$. This inverse matrix always occurs in the form $(\boldsymbol{H}^{i-2})^{-1}\delta\boldsymbol{a}^{i-1}$ and we will now show that this term is already known. From (17.46a) and the general iteration scheme (17.18), we have

$$\boldsymbol{A}^{i-2}\delta\boldsymbol{a}^{i-1} = \boldsymbol{A}^{i-2}(\boldsymbol{a}^{i-1} - \boldsymbol{a}^{i-2}) = -\boldsymbol{\psi}^{i-2}$$

and since (17.48) shows that $\boldsymbol{A}^{i-2} = (\boldsymbol{H}^{i-2})^{-1}$, we obtain

$$(\boldsymbol{H}^{i-2})^{-1}\delta\boldsymbol{a}^{i-1} = -\boldsymbol{\psi}^{i-2}$$

The out-of-balance force $\boldsymbol{\psi}^{i-2}$ is already known and this facilitates the use of the BFGS-scheme (17.61) considerably.

The product form for the updated matrix $\boldsymbol{H}^{i-1}$ given by (17.61) is of advantage for several reasons. Defining the matrix $\boldsymbol{M}$ by

$$\boldsymbol{M} = \boldsymbol{I} + \boldsymbol{v}\boldsymbol{w}^T \tag{17.62}$$

we obtain

$$\boldsymbol{H}^{i-1} = \boldsymbol{M}^T \boldsymbol{H}^{i-2} \boldsymbol{M} \tag{17.63}$$

The stiffness matrix in the finite element formulation is often symmetric and so is therefore its inverse. In the previous derivations, we assumed $\boldsymbol{H}^{i-2}$ to be

symmetric and (17.63) then shows the interesting result that the updated matrix $\boldsymbol{H}^{i-1}$ is also symmetric. Let us next assume that $\boldsymbol{H}^{i-2}$ is positive definite. To investigate whether the updated matrix $\boldsymbol{H}^{i-1}$ is also positive definite, we choose an arbitrary vector $\boldsymbol{c}$ and calculate the quadratic form of $\boldsymbol{H}^{i-1}$. Denoting the value of the quadratic form by a, we obtain with (17.63)

$$a = \boldsymbol{c}^T \boldsymbol{H}^{i-1} \boldsymbol{c} = \boldsymbol{y}^T \boldsymbol{H}^{i-2} \boldsymbol{y} \tag{17.64}$$

where $\boldsymbol{y}$ is defined by

$$\boldsymbol{M}\boldsymbol{c} = \boldsymbol{y}$$

It appears that if $\boldsymbol{c} \neq \boldsymbol{0}$ implies $\boldsymbol{y} \neq \boldsymbol{0}$ then, since $\boldsymbol{H}^{i-2}$ was assumed to be positive definite, $a > 0$ and $\boldsymbol{H}^{i-1}$ is therefore also positive definite.

The only situation where $\boldsymbol{c} \neq \boldsymbol{0}$ implies $\boldsymbol{y} = \boldsymbol{0}$ is if $\boldsymbol{M}\boldsymbol{c} = \boldsymbol{0}$ possesses a non-trivial solution. In turn this requires that the matrix $\boldsymbol{M}$ has an eigenvalue that is zero. Let us therefore determine the eigenvalues λ of the matrix $\boldsymbol{M}$, i.e. $\boldsymbol{M}\boldsymbol{c} = \lambda \boldsymbol{c}$. With (17.62), we then obtain

$$(1 - \lambda)\boldsymbol{c} + \boldsymbol{v}(\boldsymbol{w}^T \boldsymbol{c}) = \boldsymbol{0} \tag{17.65}$$

Let us assume that $\boldsymbol{w}^T \boldsymbol{c} = 0$; with the dimension of $\boldsymbol{M}$ being $N \times N$, this condition is fulfilled by $N - 1$ linearly independent $\boldsymbol{c}$-vectors and for each of these $\boldsymbol{c}$-vectors, (17.65) is fulfilled for $\lambda = 1$. We conclude

$$\lambda = 1 \qquad \text{with multiplicity } N - 1$$

Assume next that $\boldsymbol{w}^T \boldsymbol{c} \neq 0$. Expression (17.65) then shows that the eigenvector $\boldsymbol{c}$ must be proportional to $\boldsymbol{v}$, i.e. $\boldsymbol{c} = \alpha \boldsymbol{v}$ where α is a constant. Use of $\boldsymbol{c} = \alpha \boldsymbol{v}$ in (17.65) gives $[1 - \lambda + (\boldsymbol{w}^T \boldsymbol{v})]\boldsymbol{v} = \boldsymbol{0}$, i.e. we obtain the following remaining eigenvalue

$$\lambda = 1 + (\boldsymbol{w}^T \boldsymbol{v})$$

which with $\boldsymbol{v}$ and $\boldsymbol{w}$ given by (17.61) becomes

$$\lambda = \left[\frac{(\delta \boldsymbol{a}^{i-1})^T (\boldsymbol{H}^{i-2})^{-1} \delta \boldsymbol{a}^{i-1}}{(\boldsymbol{\gamma}^{i-1})^T \delta \boldsymbol{a}^{i-1}} \right]^{1/2} \tag{17.66}$$

Since $\boldsymbol{H}^{i-2}$ and thereby $(\boldsymbol{H}^{i-2})^{-1}$ was assumed to be positive definite, the eigenvalue given by (17.66) can never be zero. In conclusion, we have proved that the matrix $\boldsymbol{M}$ has no zero eigenvalues and the discussion relating to (17.64) then shows that the updated matrix $\boldsymbol{H}^{i-1}$ is positive definite, if $\boldsymbol{H}^{i-2}$ is so. In conclusion

If $\boldsymbol{H}^{i-2}$ is symmetric and positive definite then the BFGS-method implies that the updated matrix $\boldsymbol{H}^{i-1}$ is symmetric and positive definite

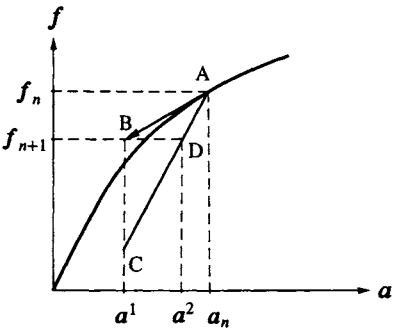

Figure 17.15: BFGS-method during unloading of elasto-plastic body. Correct solution is obtained after two iterations.

This interesting property implies, for instance, that if the very first $\boldsymbol{H}^{i-2}$-matrix is chosen as the inverse of the elastic stiffness matrix, which certainly is symmetric and positive definite, then all the following updates $\boldsymbol{H}^{i-1}$ generated by (17.61) are symmetric and positive definite.

We will now show that the BFGS-method for unloading of a elasto-plastic body provides the correct elastic unloading after two iterations; this important property is a result of the secant relation expressed by the quasi-Newton equation (17.47). In the first iteration and referring to Fig. 17.15, unloading takes place along some secant direction AB giving the first displacement estimate $\boldsymbol{a}^1$. The internal forces corresponding to $\boldsymbol{a}^1$ are given by point C. In the next iteration, the secant stiffness is determined by the last two sets of displacements and the last two sets of internal forces. Since we can assume that equilibrium is fulfilled at point A, this new secant stiffness is given by CA equal to the elastic stiffness. With this stiffness together with the residual force vector given by CB we reach, in the second iteration, exactly the correct response given by point D.

Efficient implementation of the BFGS-method evidently hinges on keeping the number of computer operations as small as possible and this requires a number of considerations. As we are here only concerned with the concepts behind the BFGS-method, the reader is referred to Bathe and Cimento (1980), Crisfield (1991), Fletcher (1980), Luenberger (1984) and Matthies and Strang (1979) for information on implementation techniques. Here we merely observe that the BFGS-method provides an attractive approach, which gives a convergence rate that is almost as fast as the Newton-Raphson without performing costly matrix inversions. The BFGS-method therefore provides an interesting compromise between the Newton-Raphson scheme and the modified Newton-Raphson scheme. The advantages within a finite element context are well documented, see Bathe (1996), Bathe and Cimento (1980), Crisfield (1991) and Matthies and Strang (1979). The use of a nonsymmetric quasi-Newton approach in finite element calculations of fluid mechanics has been investigated by Engelman *et al.* (1981).

17.9 Line search

Irrespective of the choice of iteration matrix $\boldsymbol{A}$, it turns out to be possible to reduce the number of equilibrium iterations significantly by introducing the concept of *line search*. Like many other issues relating to solution of nonlinear equations, this concept has its origin in optimization theory, see for instance Luenberger (1984).

Let us define the quantity $\boldsymbol{s}^{i-1}$ by

$$\boxed{\boldsymbol{s}^{i-1} = -(\boldsymbol{A}^{i-1})^{-1}\boldsymbol{\psi}^{i-1} \qquad \textit{search direction}} \tag{17.67}$$

Then the general iteration scheme (17.18) can be written as

$$\boldsymbol{a}^i = \boldsymbol{a}^{i-1} + \boldsymbol{s}^{i-1} \tag{17.68}$$

It appears that the new solution $\boldsymbol{a}^i$ is obtained by correcting the old solution $\boldsymbol{a}^{i-1}$ by the quantity $\boldsymbol{s}^{i-1}$. This quantity is called the *search direction* since it is the direction in which we search for the new solution. With this interpretation, it is tempting to accept the search direction $\boldsymbol{s}^{i-1}$, but leave open how far we shall go in this direction. Therefore, instead of (17.68), we now adopt the scheme

$$\boxed{\boldsymbol{a}^i = \boldsymbol{a}^{i-1} + \beta^{i-1}\boldsymbol{s}^{i-1}} \tag{17.69}$$

where the parameter β^{i-1}, the *acceleration factor*, is to be determined. With this viewpoint, we accept the search direction $\boldsymbol{s}^{i-1}$ and adjust the acceleration factor β^{i-1} in some way so that (17.69) gives us the best possible new estimate. The scheme is illustrated in Fig. 17.16

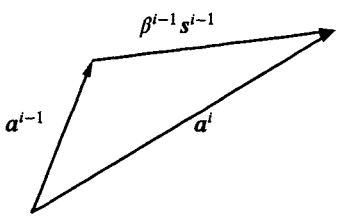

Figure 17.16: Scheme when line search is adopted

A comparison of (17.69) and (17.18) shows that instead of working with the iteration matrix $\boldsymbol{A}^{i-1}$ we now work with the iteration matrix $\frac{1}{\beta^{i-1}}\boldsymbol{A}^{i-1}$. The new scheme is therefore fully acceptable as long as $\beta^{i-1} \neq 0$ and it appears that line search can be applied in combination with any of the methods discussed previously.

However, the question remains of how to determine β^{i-1} - occasionally called the *step length* - so that the scheme (17.69) is optimal in some sense. The concept of *line search* means that we go from an old solution $\boldsymbol{a}^{i-1}$ corresponding to $\beta^{i-1} = 0$ in the search direction $\boldsymbol{s}^{i-1}$ until we obtain the best possible new solution $\boldsymbol{a}^i$. Since the iteration scheme without line search corresponds to $\beta^{i-1} = 1$, it is evident that we must have $\beta^{i-1} > 0$. The best estimate for $\boldsymbol{a}^i$ is the one which makes the out-of-balance forces $\boldsymbol{\psi}(\boldsymbol{a}^i) = \boldsymbol{0}$. In general, this condition cannot be fulfilled since it requires that the search direction be the correct one. This is just to say that we cannot expect to obtain that all components of $\boldsymbol{\psi}(\boldsymbol{a}^i)$ become zero just by adjusting one parameter β^{i-1}. Since β^{i-1} is one parameter, we must establish one condition from which β^{i-1} can be determined. Such a condition can be obtained by requiring that the new out-of-balance force $\boldsymbol{\psi}(\boldsymbol{a}^i)$ is orthogonal to a vector, i.e. the component of $\boldsymbol{\psi}(\boldsymbol{a}^i)$ in the direction of this vector is required to be zero. The question then arises which vector we shall choose when enforcing this orthogonality requirement. Referring to Fig. 17.16, three vectors are possible candidates: $\boldsymbol{a}^i$, $\boldsymbol{a}^{i-1}$ and $\boldsymbol{s}^{i-1}$. It turns out that the search direction $\boldsymbol{s}^{i-1}$ is the most natural choice, i.e. we adopt the orthogonality condition $(\boldsymbol{s}^{i-1})^T\boldsymbol{\psi}(\boldsymbol{a}^i) = 0$, which with (17.69) reads

$$\boxed{(\boldsymbol{s}^{i-1})^T\boldsymbol{\psi}(\boldsymbol{a}^{i-1} + \beta^{i-1}\boldsymbol{s}^{i-1}) = 0 \qquad \textit{line search}} \tag{17.70}$$

Since $\boldsymbol{s}^{i-1}$ and $\boldsymbol{a}^{i-1}$ are known quantities, this orthogonality condition provides one (nonlinear) relation by which β^{i-1} can be determined.

To motivate the use of the search direction $\boldsymbol{s}^{i-1}$ in the orthogonality condition (17.70), it is instructive to consider linear elasticity. In that case $\boldsymbol{\sigma} = \boldsymbol{D\varepsilon} = \boldsymbol{DBa}$ and the out-of-balance forces $\boldsymbol{\psi}$ defined by (17.2) then becomes

$$\boldsymbol{\psi} = \boldsymbol{Ka} - \boldsymbol{f} \tag{17.71}$$

where $\boldsymbol{K} = \int_V \boldsymbol{B}^T\boldsymbol{DB}dV$ is the elastic stiffness matrix. The equilibrium condition $\boldsymbol{\psi} = \boldsymbol{0}$ is then fulfilled by solving the linear equation system $\boldsymbol{Ka} - \boldsymbol{f} = \boldsymbol{0}$.

Further insight into the linear problem can be obtained by introducing the *potential energy* Π of the linear elastic body. If the body is discretized by means of finite elements, the potential energy Π is defined by

$$\Pi = \frac{1}{2}\boldsymbol{a}^T\boldsymbol{Ka} - \boldsymbol{a}^T\boldsymbol{f} \tag{17.72}$$

where the external force $\boldsymbol{f}$ is viewed as constant. It appears that $\Pi = \Pi(\boldsymbol{a})$ and therefore

$$\frac{\partial\Pi}{\partial\boldsymbol{a}} = \boldsymbol{Ka} - \boldsymbol{f}$$

A comparison with (17.71) shows that equilibrium is expressed as $\partial\Pi/\partial\boldsymbol{a} = \boldsymbol{0}$, i.e. at equilibrium the potential energy Π takes an extremum value.

Evidently, it is possible to solve the linear equation system $\boldsymbol{K}\boldsymbol{a} - \boldsymbol{f} = \boldsymbol{0}$ directly, but it is also possible to solve this linear equation system in an iterative manner using the scheme (17.69). With $\boldsymbol{a}^i$ given by (17.69), the potential energy defined by (17.72) becomes.

$$\Pi(\boldsymbol{a}^i) = \frac{1}{2}(\boldsymbol{a}^i)^T \boldsymbol{K}\boldsymbol{a}^i - (\boldsymbol{a}^i)^T \boldsymbol{f} \tag{17.73}$$

Since $\boldsymbol{a}^{i-1}$ and $\boldsymbol{s}^{i-1}$ are given quantities, $\boldsymbol{a}^i$ given by (17.69) depends only on β^{i-1} and we therefore have $\Pi = \Pi(\beta^{i-1})$. From (17.73) it then follows that

$$\frac{d\Pi}{d\beta^{i-1}} = (\boldsymbol{s}^{i-1})^T(\boldsymbol{K}\boldsymbol{a}^i - \boldsymbol{f}) \tag{17.74}$$

and

$$\frac{d^2\Pi}{d(\beta^{i-1})^2} = (\boldsymbol{s}^{i-1})^T \boldsymbol{K}\boldsymbol{s}^{i-1} > 0 \tag{17.75}$$

where it was used that $\boldsymbol{K}$ is positive definite. In accordance with (17.71), we have for linear elasticity that

$$\boldsymbol{\psi}^i = \boldsymbol{K}\boldsymbol{a}^i - \boldsymbol{f} \tag{17.76}$$

We observed above that the potential energy takes an extremum value at equilibrium. As we are performing equilibrium iterations, the condition of equilibrium has not been achieved, but with the quantities $\boldsymbol{a}^{i-1}$ and $\boldsymbol{s}^{i-1}$ given, the best possible new solution is obtained by requiring Π to be extremum, i.e. $d\Pi/d\beta^{i-1} = 0$. Therefore, (17.74) gives with (17.76)

$$(\boldsymbol{s}^{i-1})^T \boldsymbol{\psi}^i = 0 \tag{17.77}$$

According to (17.75), $d^2\Pi/d(\beta^{i-1})^2 > 0$ and the extremum property of the potential energy Π is therefore that of a minimum. Similar to (17.76), we have

$$\boldsymbol{\psi}^{i-1} = \boldsymbol{K}\boldsymbol{a}^{i-1} - \boldsymbol{f}$$

With (17.76) and (17.69), (17.77) therefore provides the following solution

$$\beta^{i-1} = -\frac{(\boldsymbol{s}^{i-1})^T \boldsymbol{\psi}^{i-1}}{(\boldsymbol{s}^{i-1})^T \boldsymbol{K}\boldsymbol{s}^{i-1}} \tag{17.78}$$

which is the best possible choice of β^{i-1}. If $\boldsymbol{\psi}^{i-1} = \boldsymbol{0}$, then the estimate $\boldsymbol{a}^{i-1}$ fulfills the equilibrium equations and it is therefore the solution sought for. Accordingly, (17.78) provides $\beta^{i-1} = 0$ when $\boldsymbol{\psi}^{i-1} = \boldsymbol{0}$. However, if $\boldsymbol{\psi}^{i-1} \neq \boldsymbol{0}$ then (17.78) gives a solution for β^{i-1} which minimizes the potential energy Π as much as possible.

It appears that for linear problems, use of the search direction $\boldsymbol{s}^{i-1}$ in the orthogonality condition is the best possible choice, cf. (17.77). It therefore seems

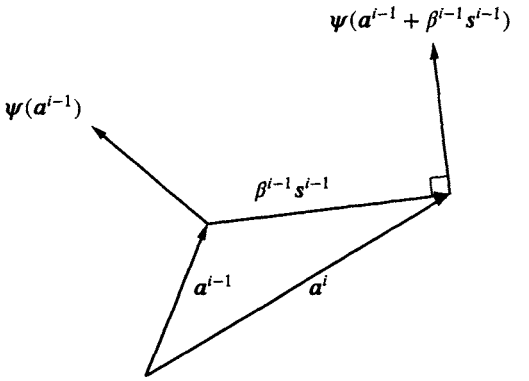

Figure 17.17: Line search for orthogonality between s^{i-1} and $\psi(a^{i-1} + \beta^{i-1}s^{i-1})$.

reasonable to adopt the same orthogonality condition also for other problems and this was already anticipated by the choice (17.70).

To further evaluate (17.70), we consider the scalar product $(s^{i-1})^T\psi(a^{i-1})$ which with (17.67) can be written as

$$(s^{i-1})^T\psi(a^{i-1}) = (\psi(a^{i-1}))^T s^{i-1} = -(\psi(a^{i-1}))^T(A^{i-1})^{-1}\psi(a^{i-1}) \quad (17.79)$$

Since the iteration matrix A^{i-1} is usually a positive definite matrix, (17.79) shows that $(s^{i-1})^T\psi(a^{i-1}) < 0$ usually holds. We then obtain the illustration shown in Fig. 17.17, where it is observed that $\psi(a^{i-1})$ corresponds to $\beta^{i-1} = 0$. When β^{i-1} is changed, the out-of-balance forces change and β^{i-1} is changed such that s^{i-1} and $\psi(a^{i-1} + \beta^{i-1}s^{i-1})$ eventually become orthogonal. Therefore, when β^{i-1} is determined from (17.70) this implies that the out-of-balance force has no components in the search direction s^{i-1}.

As already mentioned, (17.70) provides for, say, plasticity problems a nonlinear relation for determination of β^{i-1}. Therefore, in principle, solution of (17.70) requires an iterative procedure and in each of these iterations, a new out-of-balance force vector needs to be determined. This in turn requires determination of the stress field, which means integration of the constitutive equations for all (Gauss) points in the body. In the next chapter, we shall treat integration of the nonlinear elasto-plastic constitutive equations in detail, but already at this stage it is evident that this integration is far from being trivial. Therefore, using a number of iterations to solve β^{i-1} from (17.70) in an accurate manner requires a computer effort that is not insignificant. Moreover, since the search direction s^{i-1} is only an approximation to the correct one, this also questions the meaningfulness of achieving an accurate β^{i-1}-solution. Finally, whereas the orthogonality expressed of (17.70) provides the best possible β^{i-1}-solution for

linear problems, it can only be expected to provide a fair β^{i-1}-solution for non-linear problems.

All these arguments suggest that there is no sense in trying to solve (17.70) in an accurate manner and, instead, an approximate value for β^{i-1} is sufficient. Indeed, this is supported by numerical experience both within structural mechanics, see Crisfield (1991), and optimization theory, see Luenberger (1984). This numerical experience also shows that also use of an approximate β^{i-1}-value, in general, reduces the number of equilibrium iterations significantly and reduces the total computer effort considerably. Due to its effectiveness, line search is available in most general purpose finite element programs.

With DOF being the number of degrees of freedom for the structure and B the bandwidth of the equation system, the computational cost for solving the equilibrium equation system by means of Gauss elimination is proportional to DOF·B^2, cf. for instance Ottosen and Petersson (1992). The stresses σ^{i-1} are evaluated at each Gauss point in the structure and the number of Gauss points is roughly proportional to DOF=degrees of freedom. It is concluded that the advantage with the use of line search increases with the number of degrees of freedom.

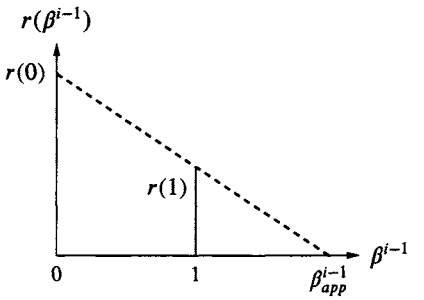

Figure 17.18: Approximate determination of β^{i-1}.

To arrive at an approximate β^{i-1}-value, we first observe that $\beta^{i-1} = 1$ corresponds to the iteration scheme without line search, cf. (17.69) and (17.18). We therefore certainly expect that $\beta^{i-1} > 0$. Define the quantity $r(\beta^{i-1})$ by

$$r(\beta^{i-1}) = -(\boldsymbol{s}^{i-1})^T \boldsymbol{\psi}(\boldsymbol{a}^{i-1} + \beta^{i-1}\boldsymbol{s}^{i-1})$$

where the minus sign is motivated by the discussion following (17.79). The quantity $\boldsymbol{\psi}(\boldsymbol{a}^{i-1})$ has already been calculated, cf. (17.67), and it is therefore easy to calculate $r(\beta^{i-1} = 0)$; calculate also r for $\beta^{i-1} = 1$. We then obtain the situation sketched in Fig. 17.18. According to (17.70) we want $r(\beta^{i-1}) = 0$ and making the linear extrapolation shown in Fig. 17.18, we obtain the following approximate value for β^{i-1}

$$\beta^{i-1}_{app} = \frac{r(0)}{r(0) - r(1)} \tag{17.80}$$

If the difference $r(0) - r(1)$ is small, β_{app}^{i-1} becomes large and, in principle, we may even encounter that β_{app}^{i-1} as predicted by (17.80) becomes negative. Since the entire motivation for the orthogonality condition (17.70) holds only strictly for linear problems, very small or very large β^{i-1}-values should be considered with caution. In practice therefore, acceptable values for β^{i-1} are restricted to a certain interval, say $0.3 \leq \beta^{i-1} \leq 3$. If (17.80) provides $\beta_{app}^{i-1} < 0.3$ then $\beta^{i-1} = 0.3$ is used and if (17.80) predicts $\beta^{i-1} > 3$ then $\beta^{i-1} = 3$ is used.

17.10 Limit points

In the previous discussion of incremental-iterative schemes, we have assumed that, for each load step, the load is increased by a certain amount and then held constant during the equilibrium iterations. This approach works well when the displacements increase with increasing load level.

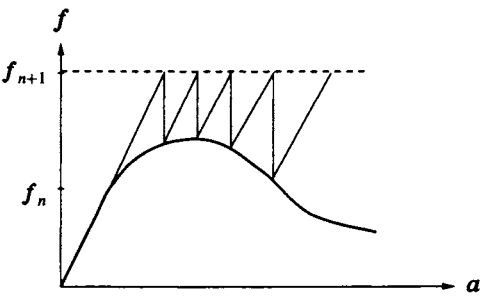

Figure 17.19: Structure displaying a softening response after the peak load. With the load f_{n+1} held fixed, equilibrium iterations will continue for ever (initial stiffness method used in the figure).

Consider now the situation where the structure displays a softening response, i.e. the structure has a maximum load capacity with a descending branch after the peak load, cf. Fig. 17.19. Evidently, the approach where the load is increased from f_n to f_{n+1} and then held constant will not be able to trace the softening branch; as soon as the load f_{n+1} exceeds the maximum load capacity of the structure, equilibrium iterations will simply continue for ever. This is illustrated in Fig. 17.19 using the initial stiffness method. Even if the response after the peak load does not display a descending branch, but exhibits a horizontal plateau (e.g. ideal plasticity), equilibrium iterations will continue for ever, if f_{n+1} exceeds the peak load.

From an engineering point of view, the peak load, i.e. the *limit load*, is often of fundamental interest. Occasionally, it is argued that once the equilibrium iterations do not converge then the limit load has been reached. However, this

is a dangerous approach since equilibrium iterations may diverge for a number of other reasons. Moreover, it is often interesting to trace the descending branch after the peak load since it contains important information of the ductility or brittleness of the structural response, cf. Fig 17.20. In Fig. 17.20c) the phenomenon of *snap-back* is illustrated.

Figure 17.20: a) Ductile response; b) brittle response; c) snap-back.

The situation of *snap-through* often encountered in buckling problems is shown in Fig. 17.21. With the iterative approach discussed so far, one would only be able to trace the response OABC whereas the part ADB of the response cannot be traced.

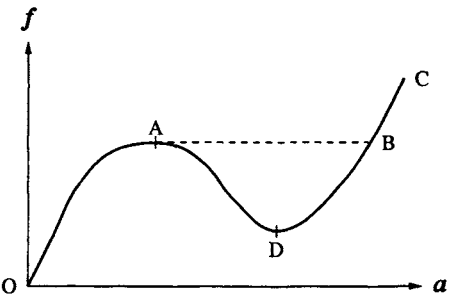

Figure 17.21: Snap-through of structure

It is evidently of importance to derive iterative schemes that are able to trace the various post-peak responses discussed above. Strategies that possess this property are called *path-following methods* and borrowing from the terminology of buckling analysis they are also often called *continuation methods*.

Whereas we previously accepted the load level and then modified the displacements until equilibrium is fulfilled, it is evident that we must now devise iterative schemes where both the displacements and the load level are adjusted, i.e.

Determination of the post-peak response requires simultaneous adjustment of the displacements and of the load level

Before entering this discussion, it is of interest to be able to characterize the different kinds of response discussed in relation to Figs. 17.20 and 17.21. Assume for simplicity that the external loading is proportional, i.e.

$$\boldsymbol{f} = \lambda \boldsymbol{f}_{ref} \tag{17.81}$$

where $\boldsymbol{f}_{ref}$ is a constant *reference load* and the proportionality factor $\lambda = \lambda(t)$, where t is the time, is called the *load parameter*. According to (17.2), the out-of-balance forces $\boldsymbol{\psi}$ are then given by

$$\boldsymbol{\psi} = \boldsymbol{f}_{int} - \lambda \boldsymbol{f}_{ref}$$

where the internal forces $\boldsymbol{f}_{int}$ depend on the stresses and thereby on the nodal displacements $\boldsymbol{a}$. The equilibrium condition is therefore given by

$$\boldsymbol{\psi} = \boldsymbol{0}\,; \quad \boldsymbol{\psi} = \boldsymbol{\psi}(\boldsymbol{a}, \lambda)$$

During equilibrium we have $\dot{\boldsymbol{\psi}} = \boldsymbol{0}$, which leads to

$$\frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{a}} \dot{\boldsymbol{a}} + \frac{\partial \boldsymbol{\psi}}{\partial \lambda} \dot{\lambda} = \boldsymbol{0}$$

i.e.

$$\boxed{\boldsymbol{K}_t \dot{\boldsymbol{a}} - \dot{\lambda} \boldsymbol{f}_{ref} = \boldsymbol{0}} \tag{17.82}$$

where use was made of (17.31).

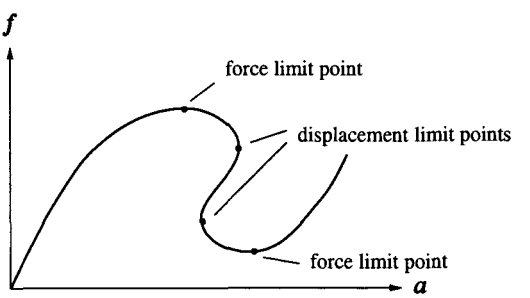

Figure 17.22: Illustration of force and displacement limit points

Let us make the following definitions:

$$\boxed{\begin{array}{l} \textit{Force limit point: } \dot{\lambda} = 0 \textit{ and } \dot{\boldsymbol{a}} \neq \boldsymbol{0} \\ \textit{Displacement limit point: } \dot{\lambda} \neq 0 \textit{ and } \dot{\boldsymbol{a}} = \boldsymbol{0} \end{array}} \tag{17.83}$$

These *limit points* are illustrated in Fig. 17.22 and we may note that in the literature such limit points are often called *turning points*.

To determine a force limit point, (17.82) and (17.83a) lead to $\boldsymbol{K}_t\dot{\boldsymbol{a}} = \boldsymbol{0}$ and a non-trivial solution requires $\det \boldsymbol{K}_t = 0$, i.e.

$$\boxed{\det \boldsymbol{K}_t = 0 \quad \Rightarrow \quad \textit{force limit point}}$$

Let us next assume that $\det \boldsymbol{K}_t \neq 0$; then (17.82) implies

$$\dot{\boldsymbol{a}} = \dot{\lambda}\boldsymbol{K}_t^{-1}\boldsymbol{f}_{ref}$$

A comparison with (17.83b) shows that the only possibility for a displacement limit point is that $\boldsymbol{f}_{ref}$ fulfills $\boldsymbol{K}_t^{-1}\boldsymbol{f}_{ref} = \boldsymbol{0}$, which requires $\det \boldsymbol{K}_t^{-1} = 0$, i.e.

$$\boxed{\det \boldsymbol{K}_t^{-1} = 0 \text{ and } \boldsymbol{K}_t^{-1}\boldsymbol{f}_{ref} = \boldsymbol{0} \quad \Rightarrow \quad \textit{displacement limit point}}$$

17.11 Bergan's minimum residual force method

In the previous iteration schemes, the load was increased step-wise and after each load increase, the load was held constant and, via equilibrium iterations, the nodal displacements were then adjusted in an iterative fashion until equilibrium was satisfied.

An elegant and entirely different approach, which allows force limit points to be passed as well as tracing of the post-peak response, was proposed by Bergan (1979, 1980). For a certain load level, the standard iteration scheme (17.18) is adopted, i.e.

$$\boldsymbol{A}^{i-1}(\boldsymbol{a}^i - \boldsymbol{a}^{i-1}) = -\boldsymbol{\psi}^{i-1} \tag{17.84}$$

where

$$\boldsymbol{\psi}^{i-1} = \boldsymbol{f}_{int}^{i-1} - \lambda^{i-1}\boldsymbol{f}_{ref}$$

and proportional loading according to (17.81) was assumed. It follows that

$$\boldsymbol{\psi}^0 = (\boldsymbol{f}_{int})_n - \lambda^0\boldsymbol{f}_{ref} \tag{17.85}$$

where it was used that $\boldsymbol{f}_{int}^0 = \int_V \boldsymbol{B}^T\boldsymbol{\sigma}^0 dV = \int_V \boldsymbol{B}^T\boldsymbol{\sigma}_n dV = (\boldsymbol{f}_{int})_n$, cf. (17.22). Moreover, in (17.85) we set

$$\lambda^0 = \lambda_n + \Delta\lambda^* \tag{17.86}$$

where $\boldsymbol{f}_n = \lambda_n\boldsymbol{f}_{ref}$ and the quantity $\Delta\lambda^*$ is specified by us.

From (17.84), the displacements $\boldsymbol{a}^i$ can now be determined; the corresponding stresses $\boldsymbol{\sigma}^i$ can now also be obtained by integration of the constitutive equations. With the stress field $\boldsymbol{\sigma}^i$ known throughout the body, the corresponding internal forces can be determined in the usual manner by means of

$$\boldsymbol{f}_{int}^i = \int_V \boldsymbol{B}^T\boldsymbol{\sigma}^i dV \tag{17.87}$$

The new residual forces then become

$$\boldsymbol{\psi}^i = \boldsymbol{f}_{int}^i - \lambda^i \boldsymbol{f}_{ref} \tag{17.88}$$

where the new load parameter λ^i is viewed as unknown. The key point in *Bergan's minimum residual force method* is then to adjust the load parameter λ^i so that the residual force $\boldsymbol{\psi}^i$ becomes minimum.

Define the quantity b by

$$b = (\boldsymbol{\psi}^i)^T \boldsymbol{\psi}^i$$

With both $\boldsymbol{f}_{ref}$ and $\boldsymbol{f}_{int}^i$ being known and fixed and λ^i the only unknown, we obtain

$$\frac{db}{d\lambda^i} = -2(\boldsymbol{\psi}^i)^T \boldsymbol{f}_{ref}$$

as well as $d^2b/(d\lambda^i)^2 = 2\boldsymbol{f}_{ref}^T \boldsymbol{f}_{ref} > 0$. Therefore, the minimum of the quantity b with respect to λ^i is given by $db/d\lambda^i = 0$, i.e.

$$\boxed{(\boldsymbol{\psi}^i)^T \boldsymbol{f}_{ref} = 0} \tag{17.89}$$

Insertion of (17.88) gives the new load parameter λ^i according to

$$\boxed{\lambda^i = \frac{(\boldsymbol{f}_{int}^i)^T \boldsymbol{f}_{ref}}{\boldsymbol{f}_{ref}^T \boldsymbol{f}_{ref}}} \tag{17.90}$$

In view of the orthogonality expressed by (17.89) and the definitions of residual forces given by (17.88), we obtain the graphical illustration of the method shown in Fig. 17.23. After $\boldsymbol{a}^i$ has been determined from (17.84), these nodal displacements are accepted and the corresponding fixed internal forces $\boldsymbol{f}_{int}^i$ are then given by (17.87). The new load parameter λ^i is now adjusted according to (17.90) until the new external force $\lambda^i \boldsymbol{f}_{ref}$ and the new residual force $\boldsymbol{\psi}^i$ become orthogonal. Each iteration can therefore be viewed as a *two-step procedure*.

With this new load parameter, we perform a new iteration, i.e. the iteration number i is increased by one and the procedure is repeated until convergence – according to the procedures discussed in Section 17.6 – has been obtained. In that case, we put $\boldsymbol{a}_{n+1} = \boldsymbol{a}^i$ and $\lambda_{n+1} = \lambda^i$.

Bergan's minimum residual method is simple and very effective. In (17.84), we can take any iteration matrix; for instance, $\boldsymbol{A}^{i-1}$ may be taken as the tangential stiffness matrix $\boldsymbol{K}_t^{i-1}$ or the constant initial stiffness matrix $\boldsymbol{K}$. This latter choice is often adopted since it is cheap and since it allows force limit points to be passed and post-peak response to be traced. These features are illustrated in the one-dimensional case shown in Fig. 17.24.

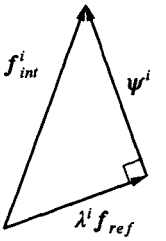

Figure 17.23: Orthogonal property obtained by Bergan's minimum residual method.

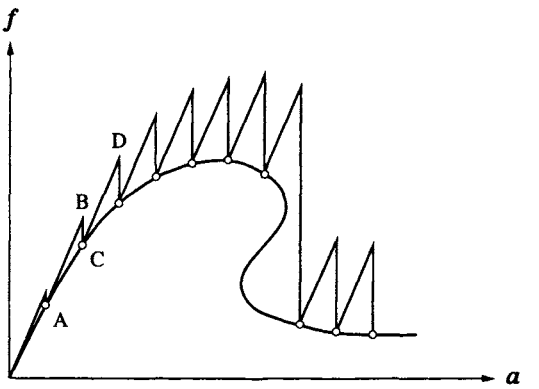

Figure 17.24: One-dimensional system. Bergan's method used with the initial stiffness matrix; after equilibrium is fulfilled, the imposed load increment is the same in all load steps. The predictions are indicated by (o).

Assume that point A has been reached. The load is then increased by some amount specified by us ($\Delta\lambda^*$ in (17.86)) to point B and the corresponding displacement is determined by (17.84) using the initial stiffness. These displacements are accepted and the corresponding internal force is determined by (17.87); this internal force corresponds to point C. The load is now adjusted so that the residual force, i.e. BC, becomes minimum. In the present one-dimensional case, this means reducing the load from B to C, i.e. the new residual force becomes in the one-dimensional case exactly equal to zero. Therefore in one iteration, exact equilibrium has been obtained. The loading is now increased by an amount specified by us ($\Delta\lambda^*$ in (17.86)) which brings us to point D and the procedure is repeated; in the case shown in Fig. 17.24 all specified load steps $\Delta\lambda^*$ are taken as equal.

While exact fulfillment of equilibrium in just one iteration is not a general property, but holds only for the one-dimensional problems, Fig. 17.24 illustrates that Bergan's method allows us to pass the force limit point and to trace the post-peak response. However, the figure also illustrates that Bergan's method cannot trace a snap-back response.

Box 17.3 Bergan's minimum residual force method

- *Initiation of quantities*

  $\boldsymbol{a}_0 = \boldsymbol{0}\,; \quad \boldsymbol{\varepsilon}_0 = \boldsymbol{0}\,; \quad \boldsymbol{\sigma}_0 = \boldsymbol{0}\,; \quad \boldsymbol{f}_0 = \boldsymbol{0}\,; \quad \boldsymbol{f}_{int} = \boldsymbol{0}$

- *For load step* $n = 0, 1, 2, \ldots . N_{max}$
  - *Determine new load level* $\lambda \boldsymbol{f}_{ref}$
  - *Initiation of iteration quantities*

    $\boldsymbol{a}^0 := \boldsymbol{a}_n$
  - *Iterate* $i = 1, 2, \ldots$ *until* $|\boldsymbol{\psi}|_{norm} = |\boldsymbol{f}_{int} - \lambda \boldsymbol{f}_{ref}|_{norm} <$ *epsilon*
    - *Calculate* $\boldsymbol{K} = \int_V \boldsymbol{B}^T \boldsymbol{D} \boldsymbol{B} dV$
    - *Calculate* $\boldsymbol{a}^i$ *from* $\boldsymbol{K}(\boldsymbol{a}^i - \boldsymbol{a}^{i-1}) = \lambda \boldsymbol{f}_{ref} - \boldsymbol{f}_{int}$
    - *Calculate* $\boldsymbol{\varepsilon}^i := \boldsymbol{B}\boldsymbol{a}^i$
    - *Determine* $\boldsymbol{\sigma}^i$ *by integration of the constitutive equations* (see next chapter)
    - *Calculate internal forces* $\boldsymbol{f}_{int}$
    - *Calculate load parameter* $\lambda = \dfrac{\boldsymbol{f}_{int}^T \boldsymbol{f}_{ref}}{\boldsymbol{f}_{ref}^T \boldsymbol{f}_{ref}}$
  - *End iteration loop*
  - *Accept quantities*

    $\boldsymbol{a}_{n+1} := \boldsymbol{a}^i\,; \quad \boldsymbol{\varepsilon}_{n+1} := \boldsymbol{\varepsilon}^i\,; \quad \boldsymbol{\sigma}_{n+1} := \boldsymbol{\sigma}^i\,; \quad \boldsymbol{f}_{int}$
- *End load step loop*

With the use of the initial stiffness matrix $\boldsymbol{K}$, each iteration using Bergan's method is in practice as cheap as iterations using the standard initial stiffness method (Bergan's method only requires the additional computation of λ^i given by (17.90), which is inexpensive). However, the standard initial stiffness method is not able to pass the force limit point and to trace the post-peak response. In addition to these features Bergan's method improves the convergence speed very significantly, as can be seen by comparing Figs. 17.10 and 17.24.

Here, we have formulated Bergan's method for proportional loading expressed by $\boldsymbol{f} = \lambda \boldsymbol{f}_{ref}$, but it is straightforward to generalize the method to non-proportional loading. In that case, the loading is considered as piece-wise linear and the above approach only requires modest modifications.

The algorithm for Bergan's method in combination with the initial stiffness matrix $\boldsymbol{K}$ is summarized in Box. 17.3; here the means to enforce the kinematic

boundary conditions have not been indicated.

It is also noticed that a closely related method that improves the possibilities for passing displacement limit points was proposed by Krenk (1995) and Krenk and Hededal (1995). Instead of the orthogonality relation (17.89), this method makes use of the orthogonality relation $(\boldsymbol{\psi}^i)^T(\boldsymbol{a}^i - \boldsymbol{a}^{i-1}) = 0$ to determine the new load parameter λ^i.

Finally, we mention the very successful *arc-length* method where a combination of load and displacement control is adopted. This method facilitates the passage of limit points and a detailed discussion is provided, for instance by Crisfield (1991).