An Infinitely Large Solid Mechanics Napkin

Andrew Chen

Last updated: 2025-02-19

0 Preface

0.1 Introduction: what's this about a napkin?

Inspiration for visual and content styles, including the title, comes from the work of Evan Chen[1], whose *Infinitely Large Napkin* is designed to be a complete and essential set of notes from pure mathematics.

Here, I have adapted the style of Evan's *Napkin* for a typeset version of notes[2] from a semester-long solid mechanics course, for example **2.002** as taught by Professor Ken Kamrin, or **2.071** as taught by Professor Lallit Anand. These notes roughly follow the content of 2.002, with additional topics in finite elasticity developed in section 4 and finite plasticity developed in section 6.6.

This set of notes is far from perfect. If you spot errors or inconsistencies, please let me know at `achen314@mit.edu`.

0.2 Additional Resources

The following collection of texts and notes offer a varied and thorough treatment of continuum mechanics and materials science. The reader is encouraged to consult them when working through these notes.

- Mechanics and Materials II (2.002) on OpenCourseWare
- Anand, L., and Govindjee, S. (2020). *Continuum Mechanics of Solids.* Oxford University Press. ISBN: 978-0-19-886472-1.
- Anand, L., Kamrin, K., and Govindjee, S. (2022). *Introduction to Mechanics of Solid Materials.* Oxford University Press. ISBN: 978-0-19-286607-3. In these notes this textbook is referred to as *AKG*. (Numbered exercises refer to particular problems in the accompanying exercise book.)
- Gurtin, M., Fried, E., and Anand, L. (2010). *The Mechanics and Thermodynamics of Continua.* Cambridge University Press. ISBN: 978-0-521-40598-0.
- Abeyaratne, R. (2024). *Lecture Notes on the Mechanics of Elastic Solids*, Volume 3 (*An Introduction to Finite Elasticity*), accessible at this link
- Ashby, Michael F., and Jones, David R. H. (2012). *Engineering Materials 1: An Introduction to Properties, Applications, and Design*, fourth edition. ISBN: 978-0-08-096665-6.
- Callister, William D., and Rethwisch, David G. (2010). *Materials Science and Engineering: An Introduction*, eighth edition. ISBN: 978-0-470-41997-7.
- Course notes from ME185 at UC Berkeley, *Introduction to Continuum Mechanics*, as taught by Professor Panayiotis Papadopoulos, accessible at this link
- Course notes from Professor Piaras Kelly, University of Auckland, accessible at this link

[1] No relation (!)

[2] This set of notes is not meant to be a complete or standalone text. Please also refer to the more complete sources in section 0.2.

0.3 Notation

In general, boldface lowercase Roman letters, e.g. $\mathbf{u}$, represent vector quantities. Boldface uppercase Roman letters, e.g. $\mathbf{S}$, represent tensor quantities of second order. Blackboard characters, e.g. $\mathbb{C}$, represent tensor quantities of fourth order.

The major *exceptions* to these rules are for vector quantities which refer to vectors within the reference configuration, which will be denoted in uppercase boldface letters, e.g. $\mathbf{X}$, and for the small strain tensor $\boldsymbol{\varepsilon}$ and the small stress tensor $\boldsymbol{\sigma}$.

I will denote the Cauchy stress tensor by $\mathbf{T}$, the first Piola stress tensor by $\mathbf{P}$, and the second Piola stress tensor by $\mathbf{S}$.

Subscripts will nearly always refer to a component of a vector or tensor[3], cf. $\mathbf{u} = (u_x, u_y, u_z)$ versus $[\mathbf{u}] = u_i$. Derivatives will always be indicated using e.g. $d(\cdot)/dx$, $\partial(\cdot)/\partial x$, or $f'(x)$ only when it is explicitly clear that $f = f(x)$ alone.

"Vocabulary" words which have a particular definition will be given in **blue boldface text** and when necessary, explicit definitions will be preceded by the boldface word **Definition**.

Text in a dark red box represents a *Key Fact* which is usually an important relation, consequence, or conceptual summary. (Acknowledgment for this notation goes to Professor Alexander Paulin.)

Here is an example of what a *Key Fact* looks like.

[3]The summation convention (see section 1.2) for subscript indices will *only* ever be used with the subscripts i, j, and k in three dimensions and with the subscripts α, β, and γ in two dimensions; that is, subscripts like x, y, r, etc. will never be summed over.

Contents

1 Mathematical Preliminaries

In this chapter we summarize the math concepts that underlie solid mechanics, namely the mathematics of vectors and tensors in finite-dimensional vector spaces. Unless otherwise noted we will restrict our attention to three-dimensional vectors and tensors, motivated by the desire to describe deformations in our spatially three-dimensional universe. To this end, let us agree that by default we will consider a three-dimensional vector space E^3 called the "Euclidean space" with a (non-unique) right-hand[1] orthonormal basis defined as $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and a fixed origin $\mathcal{O}$. To make things a bit easier to write down, we will adopt a convention whereby Latin subscripts i, j, etc. refer to each one of these three dimensions in turn, so that the abbreviated notation $\{\mathbf{e}_i\}$ is equivalent to the tuple $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$.

1.1 Vectors and Tensors

Definition 1.1.1. A **vector** $\mathbf{u}$ is a directed element that extends from the origin $\mathcal{O}$ to another point $U = (U_1, U_2, U_3)$ in E^3. In this way, as long as the origin is specified, the vector is *associated* with this point in a one-to-one manner. Given the orthonormal basis $\{\mathbf{e}_i\}$, the components of $\mathbf{u}$ can then be associated with the coordinates of the point U by

$$u_i = \mathbf{u} \cdot \mathbf{e_i} = U_i.$$

The vector components relative to a basis u_i can be organized into an array, for example $\mathbf{u} = (3, 1, 4)$, or into a componentwise addition

$$\mathbf{u} = \sum_{i=1}^{3} u_i \mathbf{e}_i,$$

for example $\mathbf{u} = 3\mathbf{e}_1 + 1\mathbf{e}_2 + 4\mathbf{e}_3$.

Importantly, any other set of orthonormal basis vectors could be chosen to describe E^3. The components U_i of the point U, and thus the components u_i of the vector $\mathbf{u}$, would then change with the basis, even though both $\mathcal{O}$ and the point U remain fixed in space. The moral here is that the representation of a vector in a basis is unique to that basis.

Definition 1.1.2. The magnitude of the **cross product** between two vectors $\mathbf{u}$ and $\mathbf{v}$, written $\mathbf{u} \times \mathbf{v}$, quantifies the area of the parallelogram defined as having $\mathbf{u}$ and $\mathbf{v}$ as its edges. Analagously, given three non-coplanar vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$, the **scalar triple product** is defined to be $[\mathbf{u}, \mathbf{v}, \mathbf{w}] \equiv \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$. As the name implies, the result is a scalar; this scalar quantifies the volume of the parallepiped defined as having $\mathbf{u}, \mathbf{v}, \mathbf{w}$ as a set of edges.

Definition 1.1.3. A **tensor** $\mathbf{A}$ is a linear transformation[2] which maps a given vector in E^3 to another vector in E^3. The "transformation" part of this definition means that the

[1] Here, *right-handedness* is the requirement that $\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3$.

[2] This type of linear transformation which operates on vectors is specifically a *second-order* tensor. It is no coincidence that the components of this type of tensor can be specified in an array of order two, and consequently that the component representation, i.e., A_{ij}, requires two subscripts. Generalizing, vectors are sometimes called *first-order* tensors, although we will avoid that term here.

tensor *operates* on an "input" vector (e.g. $\mathbf{u}$) and results in an "output" vector (e.g. $\mathbf{v}$), which is written

$$\mathbf{v} = \mathbf{A}\mathbf{u}.$$

The "linear" part of the definition means that $\mathbf{A}$ has the following properties, for all vectors $\mathbf{u}$, $\mathbf{v} \in E^3$ and for all scalars c:

$$\mathbf{A}(\mathbf{u} + \mathbf{v}) = \mathbf{A}\mathbf{u} + \mathbf{A}\mathbf{v}$$
$$\mathbf{A}(c\mathbf{u}) = c\mathbf{A}\mathbf{u}$$

As with vectors, once a basis is specified, the components A_{ij} of a tensor can be written down in terms of that basis. Analogously to the formulation $u_i = \mathbf{u} \cdot \mathbf{e}_i$, we can "extract" the components of a tensor $\mathbf{A}$ relative to a basis $\{\mathbf{e}_i\}$ as

$$A_{ij} = \mathbf{e}_i \cdot \mathbf{A}\mathbf{e}_j.$$

Commonly the components A_{ij} relative to a basis are organized into tabular form inside a *matrix*, just as the components u_i of a vector can be organized into an array. Two special tensors are the **identity tensor** $\mathbf{I}$ (in some texts, also written $\mathbf{1}$), which maps all vectors to themselves, and the **zero tensor** $\mathbf{0}$, which maps all vectors to the zero vector.

Example 1.1.4 (Special tensors)
For all vectors $\mathbf{u} \in E^3$,

$$\mathbf{I}\mathbf{u} = \mathbf{u}$$
$$\mathbf{0}\mathbf{u} = \mathbf{0}$$

Definition 1.1.5. The **tensor product** is a special operation that creates a linear mapping (i.e., a tensor) using two vectors. Given two vectors $\mathbf{v}$, $\mathbf{w} \in E^3$, the tensor product between them is written $\mathbf{v} \otimes \mathbf{w}$ and defines a mapping such that

$$(\mathbf{v} \otimes \mathbf{w})\mathbf{u} = (\mathbf{u} \cdot \mathbf{w})\mathbf{v}$$

for all $\mathbf{u} \in E^3$. The ij-th component of the tensor product is given by $(\mathbf{v} \otimes \mathbf{w})_{ij} = v_i w_j$.

This definition allows us to construct a *basis of tensors* in the same way that the three vectors $\{\mathbf{e}_i\}$ served as a basis for all other vectors in E^3. Specifically, it can be shown that the *nine* tensor products

$$\mathbf{e}_i \otimes \mathbf{e}_j$$

(recall that both i and j iterate over the values 1, 2, and 3) define a (non-unique) basis of all linear transformations within E^3. Thus, analagously to the componentwise addition for vectors, we can write

$$\mathbf{A} = \sum_{i=1}^{3} \sum_{j=1}^{3} A_{ij} \mathbf{e}_i \otimes \mathbf{e}_j$$

for a tensor $\mathbf{A}$.

1.2 Summation Convention

One can imagine that it gets annoying to write sums like $\sum_{i=1}^{3}$ every time we want to describe the components of a vector or tensor relative to a basis. Fortunately most physicists, going back to Einstein, are equally notationally lazy, and have developed a shorthand — the *summation convention* — to get around this. The generally agreed-upon rules of the convention are that:

1. Latin subscripts i, j, etc. are assumed to take on the exact values 1, 2, 3.
2. Greek subscripts α, β, etc. are assumed to take on the exact values 1, 2 (for two-dimensional problems).
3. If a subscript appears exactly twice in a single term, there is an implicit sum from 1 to 3 over that subscript for that term.
4. A subscript is not allowed to appear more than twice in a single term, *unless it is explicitly specified that the summation convention is suspended for that term.* If this is the case usually something like "(no sum on i)" is written.

Sometimes, twice-repeating subscripts are referred to as "dummy" because the actual character used to write the subscript does not matter; it is simply being summed over. Likewise, subscripts that are not summed over are called "free". The following examples should help clarify the rules:

Example 1.2.1 (Summation convention)

Some examples of the summation convention shortcut:

1. $\mathbf{u} = u_i \mathbf{e}_i \iff \mathbf{u} = \sum_{i=1}^{3} u_i \mathbf{e}_i = u_1 \mathbf{e}_1 + u_2 \mathbf{e}_2 + u_3 \mathbf{e}_3$
2. $\mathbf{u} \cdot \mathbf{v} = u_i v_i \iff \mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{3} u_i v_i = u_1 v_1 + u_2 v_2 + u_3 v_3$
3. $\mathbf{A} = A_{ij} \mathbf{e}_i \otimes \mathbf{e}_j \iff \mathbf{A} = \sum_{i=1}^{3} \sum_{j=1}^{3} A_{ij} \mathbf{e}_i \otimes \mathbf{e}_j$ (nine terms!)
4. $\mathbf{A}\mathbf{u} = A_{ij} u_j \iff \mathbf{A}\mathbf{u} = \sum_{j=1}^{3} A_{ij} u_j = S_{i1} u_1 + S_{i2} u_2 + S_{i3} u_3$. Notice that this actually represents three different values, one of which is $S_{11} u_1 + S_{12} u_2 + S_{13} u_3$.
5. In the previous example, we could have equivalently written $\mathbf{A}\mathbf{u} = A_{ik} u_k$ or $\mathbf{A}\mathbf{u} = A_{iq} u_q$ or $\mathbf{A}\mathbf{u} = A_{iz} u_z$; here k (or q or z) is a dummy variable, whereas i is free.
6. The expression $u_i v_j$ has no summation, but represents nine individual terms, one of which is $u_2 v_3$.
7. The expression $A_{pq} B_{qr} C_{qs}$ contains a syntax error, because the subscript q is repeated three times.

8. The equation $A_{ij} = B_{ik}$ contains a syntax error, because the free indices on each side of the equation are mismatched.

9. The equations $v_i = A_{ij}u_j$ and $v_i = A_{ji}u_j$ are both syntactically correct but describe two different linear transformations. However, the equations $v_i = A_{ij}u_j$ and $v_j = A_{ji}u_i$ describe the same linear transformation.

10. In Cartesian coordinates, the vector divergence is given by $\operatorname{div} \mathbf{u} = \dfrac{\partial u_i}{\partial x_i} \iff \operatorname{div} \mathbf{u} = \displaystyle\sum_{i=1}^{3} \frac{\partial u_i}{\partial x_i} = \frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3}$

Exercise 1.2.2. Working in components, **show that** for all vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$, that

$$(\mathbf{a} \otimes \mathbf{b})(\mathbf{c} \otimes \mathbf{d}) = (\mathbf{b} \cdot \mathbf{c})(\mathbf{a} \otimes \mathbf{d}).$$

Moreover if $\mathbf{A}$ is a tensor, **show that**

$$\mathbf{A}(\mathbf{a} \otimes \mathbf{b}) = (\mathbf{A}\mathbf{a}) \otimes \mathbf{b}.$$

1.3 Kronecker Delta and Levi-Civita Symbol

Definition 1.3.1. The **Kronecker delta** symbol δ_{ij} has a value of either 0 or 1, according to the rule

$$\delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j. \end{cases}$$

This definition is motivated by the dot product of two basis vectors, which should return 1 if the basis vectors are identical or 0 otherwise. Written out, this statement is simply

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}.$$

Example 1.3.2

The Kronecker delta can be used under summation:

$$\delta_{ii} = 3.$$

Recall that this notation actually means

$$\delta_{ii} = \sum_{i=1}^{3} \delta_{ii} = \delta_{11} + \delta_{22} + \delta_{33},$$

and each of the terms on the right hand side evaluates to 1 by definition. This is equivalent to organizing the nine possible values of δ_{ij} in a 3×3 matrix and taking its trace.

Definition 1.3.3. The **Levi-Civita** (also called the **alternating**) symbol e_{ijk} has a

value of -1, 0, or 1, according to the rule

$$e_{ijk} = \begin{cases} +1, & \text{if } \{i,j,k\} = \{1,2,3\},\ \{2,3,1\},\ \text{or } \{3,1,2\} \\ -1, & \text{if } \{i,j,k\} = \{2,1,3\},\ \{1,3,2\},\ \text{or } \{3,2,1\} \\ 0, & \text{if an index is repeated.} \end{cases}$$

The ordering of subscripts corresponding to $e_{ijk} = 1$ is referred to as an *even permutation* of $\{1,2,3\}$, and the ordering of subscripts corresponding to $e_{ijk} = -1$ is referred to as an *odd permutation*.

This definition is motivated by the cross product of basis vectors in our right-hand orthonormal basis. A right-hand cross-product should return $+1$, a left-hand cross-product should return -1, and a cross-product with any vector and itself should return 0. This is captured by the statement

$$\mathbf{e}_i \cdot (\mathbf{e}_j \times \mathbf{e}_k) = e_{ijk}.$$

The symbols δ_{ij} and e_{ijk} are very powerful when used with other terms in index notation. For example, the vector cross product can be written compactly as $\mathbf{u} \times \mathbf{v} = e_{ijk} u_j v_k \mathbf{e_i}$, or equivalently the i-th *component* of the cross-product can be written as

$$(\mathbf{u} \times \mathbf{v})_i = e_{ijk} u_j v_k.$$

Notice that in both of these expressions sums are implicit on j and k (they are dummy variables), but not on i (it is a free variable), and as such each expression actually represents three individual equations, corresponding to each component of the cross-product vector.

1.4 Properties of Tensors

Let $\mathbf{A}$ and $\mathbf{B}$ be second-order tensors in E^3, and let $\mathbf{u}$ and $\mathbf{v}$ be vectors in E^3.

Definition 1.4.1. The **transpose** of $\mathbf{A}$, denoted $\mathbf{A}^T$, is the unique tensor such that

$$\mathbf{u} \cdot \mathbf{A}\mathbf{v} = \mathbf{v} \cdot \mathbf{A}^T \mathbf{u}.$$

In any basis, the components of $\mathbf{A}$ and $\mathbf{A}^T$ are related by

$$[A^T]_{ij} = A_{ji}.$$

If a tensor is equal to its transpose, the tensor is called **symmetric**. If a tensor is equal to the *negative* of its transpose, the tensor is called **skew**.

Definition 1.4.2. The **product of two tensors** between $\mathbf{A}$ and $\mathbf{B}$ represents a composite mapping when applied to a vector, for example

$$\mathbf{A}\mathbf{B}\mathbf{u} = \mathbf{A}(\mathbf{B}\mathbf{u})$$

describes first the action of the linear transformation $\mathbf{B}$ on the vector $\mathbf{u}$, then the action of the linear transformation $\mathbf{A}$ on the result of the first mapping. When a basis is specified, each component of the composite map is related to the components of each tensor by

$$[AB]_{ij} = A_{ik} B_{kj}.$$

In general, $\mathbf{A}\mathbf{B} \neq \mathbf{B}\mathbf{A}$ — the order matters!

Definition 1.4.3. The **contraction of two tensors** $\mathbf{A}$ and $\mathbf{B}$, also called the *inner product*, is the function which yields the scalar denoted $\mathbf{A} \cdot \mathbf{B}$ and defined to be

$$\mathbf{A} \cdot \mathbf{B} = \text{tr}\,(\mathbf{A}\mathbf{B}^T)$$

or in component form

$$\mathbf{A} \cdot \mathbf{B} = A_{ij}B_{ij}.$$

This computation is analogous to the vector dot product and results in similar corollaries. For example, we define the **magnitude of a tensor** as $|\mathbf{A}| \equiv \sqrt{\mathbf{A} \cdot \mathbf{A}}$ and we say that two tensors $\mathbf{A}$ and $\mathbf{B}$ are *orthogonal* if $\mathbf{A} \cdot \mathbf{B} = 0$. Moreover, the contraction between a symmetric and a skew tensor is exactly zero.

Definition 1.4.4. The **trace** of a tensor $\mathbf{A}$, denoted tr $\mathbf{A}$, is formally defined as the scalar which satisfies

$$\text{tr}\,(\mathbf{u} \otimes \mathbf{v}) = \mathbf{u} \cdot \mathbf{v}.$$

This is equal to the sum of the diagonal components of $\mathbf{A}$ expressed in *any* basis. A tensor is called *traceless* or **deviatoric** if its trace is exactly zero.

Exercise 1.4.5. True or false? Every skew tensor is deviatoric.

Definition 1.4.6. The **determinant** of a tensor $\mathbf{A}$ is formally defined as the scalar det $\mathbf{A}$ such that for any non-coplanar vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$,

$$\det \mathbf{A} = \frac{\mathbf{A}\mathbf{u} \cdot (\mathbf{A}\mathbf{v} \times \mathbf{A}\mathbf{w})}{\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})}.$$

This is equal to the usual matrix determinant of $\mathbf{A}$ expressed in *any* basis, which is easier to compute. The determinant represents *how much* the volume of the parallelepiped defined by the vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ changes after each of the vectors undergoes the linear transformation $\mathbf{A}$.

Definition 1.4.7. If the linear transformation represented by a tensor is both (a) injective (one-to-one) between elements in its domain and co-domain, and (b) surjective (onto) its entire co-domain, the mapping is called *bijective* and the tensor is said to be **invertible**. If $\mathbf{A}$ is an invertible tensor, then we write its inverse $\mathbf{A}^{-1}$ and

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}.$$

Thus, $\mathbf{A}^{-1}$ represents the inverse linear transformation from $\mathbf{A}$. There are a number of equivalent conditions for invertability, some of which require a basis to be specified. One useful condition is that $\mathbf{A}$ is invertible if and only if $\det \mathbf{A} \neq 0$.

Definition 1.4.8. If the linear transformation represented by a tensor preserves the lengths of vectors and preserves the angle between vectors, it is said to be **orthogonal**. If $\mathbf{A}$ is an orthogonal tensor, it satisfies the relation

$$\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} = \mathbf{I},$$

and the inverse of the linear transformation $\mathbf{A}$ is exactly its *transpose*. Orthogonal linear transformations represent *rotations* (if $\det \mathbf{A} = 1$, also called *proper orthogonal*) and *reflections* ($\det \mathbf{A} = -1$, also called *improper orthogonal*).

Definition 1.4.9. A tensor $\mathbf{A}$ is **positive definite** if and only if for all nonzero vectors $\mathbf{u}$,

$$\mathbf{u} \cdot \mathbf{A}\mathbf{u} > 0.$$

Exercise 1.4.10. Let $\mathbf{F}$ be an invertible tensor. **Show that** the tensor $\mathbf{C} \equiv \mathbf{F}^T\mathbf{F}$ is symmetric and positive definite.

Definition 1.4.11. A vector $\mathbf{m}$ is an **eigenvector** of $\mathbf{A}$ with corresponding **eigenvalue** λ if and only if

$$\mathbf{A}\mathbf{m} = \lambda\mathbf{m},$$

i.e., the action of the linear transformation $\mathbf{A}$ on $\mathbf{m}$ changes its magnitude by a factor of λ but not its direction. We say two tensors are *collinear* if they share the same eigenvectors. Eigenvalues of a tensor $\mathbf{A}$ can be found by solving the **characteristic equation**

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0,$$

which in our case will be a third-order polynomial equation in λ. The three roots of this equation are the three eigenvalues. For each eigenvalue λ_i, the corresponding eigenvector can be found by computing the kernel of $(\mathbf{A} - \lambda_i\mathbf{I})$, i.e. the vector $\mathbf{m_i}$ for which

$$(\mathbf{A} - \lambda_i\mathbf{I})\mathbf{m_i} = \mathbf{0}.$$

Furthermore, the characteristic equation can be written in the form

$$\lambda^3 - I_1\lambda^2 + I_2\lambda - I_3 = 0$$

where the I_i represent **invariants** of $\mathbf{A}$, scalar quantities which are functions of the tensor but independent of any basis representation. The invariants in the characteristic equation are related to the trace and determinant of the tensor and to the eigenvalues:

$$\begin{aligned} I_1 &= \operatorname{tr}\mathbf{A} = \lambda_1 + \lambda_2 + \lambda_3 \\ I_2 &= \frac{1}{2}[(\operatorname{tr}\mathbf{A})^2 - \operatorname{tr}\mathbf{A}^2] = \lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1 \\ I_3 &= \det\mathbf{A} = \lambda_1\lambda_2\lambda_3 \end{aligned}$$

(Note that other invariants exist; for example, the quantity I_2^2 is always invariant of basis.) Finally, if the tensor $\mathbf{A}$ is symmetric, the eigenvalues are guaranteed to all be real, and a basis of eigenvectors can be established (i.e., there exist three mutually orthogonal eigenvectors). In this situation the tensor can be written as a diagonal matrix with respect to this basis of eigenvectors. If the tensor is symmetric *and* positive-definite, the above is true, plus the fact that all eigenvalues will be positive.

Exercise 1.4.12. Let $\mathbf{A}$ be a symmetric, positive-definite tensor. **Show that** any eigenvalue of $\mathbf{A}$ must be positive.

1.4.1 Basis Transformation

The usual representation of a vector or tensor is intimately tied to the basis used to represent it. However, there is no single basis that must always be used. To this end, the transformation equations in this section can be used to write down the components of a vector or tensor in a new basis, provided the new basis vectors $\mathbf{e}'_i$ are known in terms of the original basis vectors $\mathbf{e}_i$.

Definition 1.4.13. The **direction cosine** between two basis vectors, denoted $\cos(\mathbf{e}'_i, \mathbf{e}_i)$, is the cosine of the angle between the two vectors. Provided these vectors are of unit magnitude, the direction cosine is equal to their dot product, $\cos(\mathbf{e}'_i, \mathbf{e}_i) = \mathbf{e}'_i \cdot \mathbf{e}_i$.

Key Equation 1.4.14

A vector $\mathbf{v}$ and a tensor $\mathbf{T}$ in a given basis $\{\mathbf{e}_i\}$ may be rewritten as $\mathbf{v}'$ and $\mathbf{T}'$, respectively, in terms of a new basis $\{\mathbf{e}'_i\}$ by

$$\mathbf{v}' = \mathbf{Q}\mathbf{v}$$
$$\mathbf{T}' = \mathbf{Q}\mathbf{T}\mathbf{Q}^T$$

where $\mathbf{Q}$ is the orthogonal tensor of direction cosines,

$$[\mathbf{Q}] = \begin{bmatrix} \cos(\mathbf{e}'_1, \mathbf{e}_1) & \cos(\mathbf{e}'_1, \mathbf{e}_2) & \cos(\mathbf{e}'_1, \mathbf{e}_3) \\ \cos(\mathbf{e}'_2, \mathbf{e}_1) & \cos(\mathbf{e}'_2, \mathbf{e}_2) & \cos(\mathbf{e}'_2, \mathbf{e}_3) \\ \cos(\mathbf{e}'_3, \mathbf{e}_1) & \cos(\mathbf{e}'_3, \mathbf{e}_2) & \cos(\mathbf{e}'_3, \mathbf{e}_3) \end{bmatrix}$$

Because all the $\mathbf{e}_i$ and all the $\mathbf{e}'_i$ are unit vectors, it follows that

$$\cos(\mathbf{e}'_i, \mathbf{e}_j) = \mathbf{e}'_i \cdot \mathbf{e}_j.$$

Note that because $\mathbf{Q}$ is orthogonal, the inverse transformation (i.e. from $\{\mathbf{e}'_i\}$ back to $\{\mathbf{e}_i\}$) can be computed using $\mathbf{Q}^T$ in place of $\mathbf{Q}$ and vice versa.

Example 1.4.15

A basis $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ obtained by rotating the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ through an angle θ about $\mathbf{e}_3$ has the relations

$$\begin{cases} \mathbf{e}'_1 &= \cos(\theta)\mathbf{e}_1 + \sin(\theta)\mathbf{e}_2 \\ \mathbf{e}'_2 &= -\sin(\theta)\mathbf{e}_2 + \cos(\theta)\mathbf{e}_1 \\ \mathbf{e}'_3 &= \mathbf{e}_3 \end{cases}$$

and thus the rotation tensor $\mathbf{Q}$ between these bases has components

$$[\mathbf{Q}] = \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

1.4.2 Decomposition of Tensors

A tensor $\mathbf{A}$ can be decomposed a number of useful ways:

1. Symmetric-skew decomposition:

$$\mathbf{A} = \underbrace{\frac{1}{2}(\mathbf{A} + \mathbf{A}^T)}_{\text{symmetric}} + \underbrace{\frac{1}{2}(\mathbf{A} - \mathbf{A}^T)}_{\text{skew}}$$

2. Spherical-deviatoric decomposition:

$$\mathbf{A} = \underbrace{\mathbf{A} - \frac{1}{3}(\text{tr } \mathbf{A})\mathbf{I}}_{\text{deviatoric}} + \underbrace{\frac{1}{3}(\text{tr } \mathbf{A})\mathbf{I}}_{\text{spherical}}$$

Note that the spherical component is diagonal. Sometimes, the deviatoric part of $\mathbf{A}$ is denoted $\mathbf{A}'$.

3. Spectral decomposition[3]: for a *symmetric* tensor $\mathbf{A}$ the eigenvectors $\{\mathbf{m}_i\}$ form an orthonormal basis.

   a) *Three distinct eigenvalues* λ_i:

   $$\mathbf{A} = \sum_{i=1}^{3} \lambda_i \mathbf{m}_i \otimes \mathbf{m}_i \qquad [\mathbf{A}] = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

   The matrix representation (right) gives $\mathbf{A}$ expressed in the basis of eigenvectors.

   b) *Two distinct eigenvalues,* $\lambda_1 \neq \lambda_2 = \lambda_3$:

   $$\mathbf{A} = \lambda_1 \mathbf{m}_1 \otimes \mathbf{m}_1 + \lambda_2(\mathbf{I} - \mathbf{m}_1 \otimes \mathbf{m}_1)$$

   The eigenvectors are m_1 and any vector orthogonal to m_1.

   c) *No distinct eigenvalues,* $\lambda_1 = \lambda_2 = \lambda_3 \equiv \lambda$

   $$\mathbf{A} = \lambda \mathbf{I}$$

   This happens if and only if $\mathbf{A}$ is spherical. Any vector is then an eigenvector of $\mathbf{A}$.

   Note that the spectral decomposition of a tensor allows us to define the *tensor square root*:

   Definition 1.4.16. Let $\mathbf{A}$ be a tensor with the spectral decomposition

   $$\mathbf{A} = \sum_{i=1}^{3} \lambda_i \mathbf{m}_i \otimes \mathbf{m}_i.$$

   Then its **tensor square root** is expressed in the same basis of eigenvectors as

   $$\sqrt{\mathbf{A}} \equiv \sum_{i=1}^{3} \sqrt{\lambda_i} \mathbf{m}_i \otimes \mathbf{m}_i.$$

   Remark 1.4.17: Actually, the tensor square root is a special case of the general rule that if $\mathbf{A}$ has the spectral decomposition given, then for all $n > 0$, the n-th power of $\mathbf{A}$ can be expressed as

   $$\mathbf{A}^n \equiv \sum_{i=1}^{3} (\lambda_i)^n \mathbf{m}_i \otimes \mathbf{m}_i.$$

4. Polar decomposition: if a tensor $\mathbf{A}$ has positive determinant, $\det \mathbf{A} > 0$, there exist unique symmetric, positive definite tensors $\mathbf{U}$ and $\mathbf{V}$ as well as a proper orthogonal (rotation) tensor $\mathbf{R}$ such that

   $$\mathbf{A} = \mathbf{R}\mathbf{U} = \mathbf{V}\mathbf{R}.$$

[3] Not all tensors admit a spectral decomposition. Of practical use to us are tensors for which the eigenvalues are *real*.

Exercise 1.4.18. Let $\mathbf{A}$ and $\mathbf{B}$ be tensors and let $\mathbf{B}'$ denote the deviatoric part of $\mathbf{B}$. **Show that**

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{A} \cdot \mathbf{B}'.$$

Exercise 1.4.19. Let $\mathbf{C}$ be a symmetric tensor and let $\mathbf{D}$ be a skew tensor. **Show that**

$$\mathbf{C} \cdot \mathbf{D} = 0.$$

1.4.3 Fourth-order Tensors

Definition 1.4.20. A **fourth-order tensor** is a linear operator that maps a second-order tensor to another second-order tensor. For a fourth-order tensor $\mathbb{C}$, this operation is written

$$\mathbf{B} = \mathbb{C}\mathbf{A}.$$

The linearity properties (addition and scalar multiplication) introduced previously apply to fourth-order tensors, because the fourth-order tensor produces a linear transformation. Moreover, once a basis is specified, the components of a fourth-order tensor can be written down using four subscripts. The basis may be specified directly in terms of four linearly independent vectors, or by means of second-order "basis tensors":

$$C_{ijkl} = \mathbf{E}_{ij}(\mathbb{C}\mathbf{E}_{kl}) = (\mathbf{e}_i \otimes \mathbf{e}_j)\mathbb{C}(\mathbf{e}_k \otimes \mathbf{e}_l).$$

Example 1.4.21

For all linear transformations $\mathbf{A}$,

- the identity fourth-order tensor $\mathbb{I}$ having components $\mathbb{I}_{ijkl} = \delta_{ik}\delta_{jl}$ maps $\mathbf{A}$ to itself,
$$\mathbb{I}\mathbf{A} = \mathbf{A};$$
- the fourth-order tensor $\bar{\mathbb{I}}$ having components $\bar{\mathbb{I}}_{ijkl} = \delta_{il}\delta_{jk}$ maps $\mathbf{A}$ to its *transpose*,
$$\bar{\mathbb{I}}\mathbf{A} = \mathbf{A}^T;$$
- the fourth-order tensor $\mathbb{I}^{\text{sym}} = \frac{1}{2}(\mathbb{I} + \bar{\mathbb{I}})$ having components $\mathbb{I}^{\text{sym}}_{ijkl} = \frac{1}{2}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})$ maps $\mathbf{A}$ to its *symmetric part*,
$$\mathbb{I}^{\text{sym}}\mathbf{A} = \text{sym } \mathbf{A};$$
- the fourth-order tensor $\mathbf{I} \otimes \mathbf{I}$ having components $\delta_{ij}\delta_{kl}$ maps the tensor $\mathbf{A}$ to the tensor $(\text{tr}\mathbf{A})\mathbf{I}$.

Exercise 1.4.22. Write the components of the fourth-order tensor which maps any second-order tensor to its deviatoric part.

1.5 Vector Calculus

In this section we will briefly review the calculus of scalar, vector, and tensor fields in E^3. In the context of solid mechanics, a scalar (or vector or tensor) *field* refers to a scalar (or

vector or tensor) quantity, e.g. temperature (or velocity or stress), which is defined at every point in a relevant subset of E^3 we call the *domain*. Usually these fields are taken to be "smooth" enough such that they can be differentiated as necessary[4].

To formulate these fields, let $\mathbf{x}$ represent a vector variable, for example one that represents position in E^3. Let $\phi = \phi(\mathbf{x})$ be a scalar function, $\mathbf{u} = \mathbf{u}(\mathbf{x})$ be a vector function, and $\mathbf{T} = \mathbf{T}(\mathbf{x})$ be a tensor function of $\mathbf{x}$ all defined everywhere on E^3. Accordingly, we define an orthonormal basis such that the components of $\mathbf{u}(\mathbf{x})$ may be written as $[u_i]$, and the components of $\mathbf{T}(\mathbf{x})$ may be written as $[T_{ij}]$. It will be simplest to specify the expressions for vector calculus operations in terms of their components with respect to this basis.

Definition 1.5.1. The **gradient** of a function quantifies its rate of change at the position $\mathbf{x}$ with respect to a given direction.

- The gradient of a scalar field is a vector field,

$$[\text{grad } \phi(\mathbf{x})]_i = \frac{\partial \phi(\mathbf{x})}{\partial x_i}$$

- The gradient of a vector field is a tensor field,

$$[\text{grad } \mathbf{v}(\mathbf{x})]_{ij} = \frac{\partial \mathbf{v}_i(\mathbf{x})}{\partial x_j}$$

Accordingly, the directional derivative in a given direction $\mathbf{b}$ (i.e. rate of change in the direction $\mathbf{b}$) is given by

$$(\text{grad } \phi) \cdot \mathbf{b} \qquad \text{or} \qquad (\text{grad } \mathbf{v})\mathbf{b}.$$

Definition 1.5.2. The **divergence** of a function at a point quantifies the extent to which that point is a *source* or *sink* of the relevant quantity.

- The divergence of a vector field is a scalar field,

$$[\text{div } \mathbf{v}(\mathbf{x})] = \frac{\partial \mathbf{v}_i(\mathbf{x})}{\partial x_i} = \text{tr}(\text{grad } \mathbf{v}(\mathbf{x}))$$

- The divergence of a tensor field is a vector field,

$$[\text{div } \mathbf{T}(\mathbf{x})]_i = \frac{\partial \mathbf{T}_{ij}(\mathbf{x})}{\partial x_j}$$

Definition 1.5.3. The **curl** of a function at a point quantifies the extent to which the relevant quantity experiences *local* rotation there.

- The curl of a vector field is a vector field,

$$[\text{curl } \mathbf{v}(\mathbf{x})]_i = e_{ijk} \frac{\partial \mathbf{v}_k(\mathbf{x})}{\partial x_j}$$

- The curl of a tensor field is a tensor field,

$$[\text{curl } \mathbf{T}(\mathbf{x})]_{ji} = e_{ipq} \frac{\partial \mathbf{T}_{jq}(\mathbf{x})}{\partial x_p}$$

[4] Apologies to the mathematicians for this extremely non-rigorous statement. For the most part, we simply request that *enough* continuous derivatives exist in order for the relevant fields of interest to play nice.

Notice that the gradient *increases* the order of a field, the divergence *decreases* the order of a field, and the curl *preserves* the order of a field. It is possible to combine the gradient and divergence operations into a second-order derivative which preserves the order of a field; this composite operation is called the **Laplacian** and denoted Δ.

- The Laplacian of a scalar field is a scalar field,

$$\Delta\phi = \text{div grad}\phi = \frac{\partial^2\phi}{\partial x_i \partial x_i}$$

- The Laplacian of a vector field is a vector field,

$$[\Delta\mathbf{v}(\mathbf{x})]_i = [\text{div grad}\mathbf{v}]_i = \frac{\partial^2 v_i}{\partial x_j \partial x_j}$$

- The Laplacian of a tensor field is a tensor field,

$$[\Delta T]_{ij} = \frac{\partial T_{ij}}{\partial x_k \partial x_k}$$

Exercise 1.5.4. Let $\mathbf{A}$ be a tensor with deviatoric part $\mathbf{A}'$. **Compute** the following derivatives:

1. $\dfrac{\partial(\text{tr }\mathbf{A})}{\partial\mathbf{A}}$
2. $\dfrac{\partial\mathbf{A}'}{\partial\mathbf{A}}$
3. $\dfrac{\partial}{\partial\mathbf{A}}\left(\sqrt{\dfrac{3}{2}}|\mathbf{A}'|\right)$

Of use later on will be the principle of *localization*.

Key Equation 1.5.5 (Localization theorem)

Let $g(\mathbf{x})$ be a continuous scalar, vector, or tensor field valid on a subset of E^3 called $\mathcal{R}$. If and only if

$$\int_{\mathcal{P}} g(\mathbf{x})\, dV = 0$$

for all $\mathcal{P} \subset \mathcal{R}$, then

$$g(\mathbf{x}) = 0$$

everywhere in $\mathcal{R}$.

To sketch an outline of the proof, assume for contradiction that $g(\mathbf{x}) \neq 0$ somewhere in $\mathcal{R}$, and call the region for which this is true $\mathcal{F}$. Without loss of generality we can take $g(\mathbf{x}) > 0$ throughout $\mathcal{F}$. Then we could pick a small and very carefully crafted $\mathcal{P}$ with size $\mathcal{F}/2$. Evaluating the integral $\int_{\mathcal{P}} g(\mathbf{x})\, dV$ would then result in a positive result, which is a contradiction. Therefore our assumption must be wrong and $g(\mathbf{x})$ must be zero throughout $\mathcal{R}$.

Finally we will make use of the *divergence theorem*, which gives us a recipe to convert between volume integrals taken over an enclosed region and surface integrals taken over its corresponding boundary.

Key Equation 1.5.6 (Divergence theorem)

Let $\mathcal{R}$ be a domain with boundary $\partial\mathcal{R}$ with unit outward normal $\mathbf{n}$. Then, for a scalar field ϕ, vector field $\mathbf{v}$, and tensor field $\mathbf{T}$,

$$\begin{aligned}\int_{\partial\mathcal{R}} \phi\mathbf{n}\, dA &= \int_{\mathcal{R}} \text{grad}\ \phi\, dV \\ \int_{\partial\mathcal{R}} \mathbf{v}\cdot\mathbf{n}\, dA &= \int_{\mathcal{R}} \text{div}\ \mathbf{v}\, dV \\ \int_{\partial\mathcal{R}} \mathbf{T}\mathbf{n}\, dA &= \int_{\mathcal{R}} \text{div}\ \mathbf{T}\, dV\end{aligned}$$

In words, the divergence theorem says that the *flux* of a particular field quantity through the boundary of a volume is equivalent to the net gradient or divergence of that field within the same volume.

Exercise 1.5.7. For all vector fields $\mathbf{v}$ defined on a surface $\partial\mathcal{R}$, **show that**

$$\int_{\partial\mathcal{R}} (\text{curl}\ \mathbf{v})\cdot\mathbf{n}\, dA = 0.$$

Finally we note that a common shorthand used to notate spatial derivatives (i.e., derivatives of some quantity $A_{ijk\ldots p}$ with respect to some direction x_q) within index notation is to write

$$\frac{\partial A_{ijk\ldots p}}{\partial x_q} \equiv A_{ijk\ldots p,q},$$

replacing the derivative term with a comma *after* which comes the index corresponding to the direction of differentiation. As a concrete example, it is commonly written that

$$[\text{div}\ \mathbf{T}(\mathbf{x})]_i = \frac{\partial \mathbf{T}_{ij}(\mathbf{x})}{\partial x_j} \equiv T_{ij,j}.$$

2 Kinematics

In this chapter we discuss the *kinematics* of deformation, *i.e.* the mathematical description of the deformation without any reference to material properties or other material-dependent mechanical behavior. Consequently, the results we develop in this section apply equally to *any* material we might consider. We will see later on that the constitutive theories that *do* describe mechanical behavior — for example, the stresses that are created as a result of deformation — necessarily rely on the kinematical description of the deformation we develop here.

2.1 Motion and the Deformation Gradient

We consider a body occupying the "reference" configuration $\mathcal{R}_R$ at a given reference time t_0.[1] We endow the reference system with a triad of basis vectors and an origin, such that the material element at a location in this *reference configuration* is defined by a vector $\mathbf{X}$ which references this origin.[2]

At a later time $t > t_0$, suppose *something* has happened to the body such that it now occupies a different configuration, $\mathcal{R}$. We call this *something* a **deformation**. To distinguish material elements in this *deformed configuration* we use a vector $\mathbf{x}$. Now, the mapping of elements $\boldsymbol{\chi} : \mathbf{X} \mapsto \mathbf{x}$ must be injective, because it cannot physically happen that material elements collapse onto themselves or split apart as a result of the deformation, and it must be surjective, because by construction the co-domain $\mathcal{R}$ contains exactly the deformed material elements and nothing more.[3]

Therefore, there exists a bijective map formally called the **motion function**,

$$\mathbf{x} = \boldsymbol{\chi}(\mathbf{X}, t),$$

which connects points in the reference configuration to their image in the deformed configuration. The inverse map, sometimes called the **reference map**, is written as $\boldsymbol{\chi}^{-1}$.

Definition 2.1.1. The **displacement** $\mathbf{u}(\mathbf{X}, t)$ is the difference between a material point's image at a particular time t and its original location,

$$\mathbf{u}(\mathbf{X}, t) \equiv \mathbf{x} - \mathbf{X}.$$

Its first and second time derivatives, denoted $\dot{\mathbf{u}}$ and $\ddot{\mathbf{u}}$, are called **velocity** and **acceleration**, respectively.

[1] The reference configuration does not *need* to be a physical configuration taken up by the body at any real time; the motivation is simply to construct a mathematically straightforward reference state which then can be used to describe the body's subsequent motion. On the other hand, the reference configuration does not to be *unique* from any other configuration taken up by the body. For example, we will encounter deformations that move material points relative to each other, but do not change the overall three-dimensional volume occupied by the body.

[2] The use of $\mathbf{X}$ to indicate position in the reference configuration is standard, which unfortunately conflicts with the convention of reserving uppercase Latin letters for tensors.

[3] Note here that we consider only a *continuum* description of the body which cannot describe events like crack growth, void nucleation, and the like. We defer the description of these failure processes to a later chapter.

Definition 2.1.2. The **deformation gradient tensor** (or just *deformation gradient*) $\mathbf{F}$ describes the deformation in terms of what happens to each basis vector,

$$\mathbf{F} \equiv \frac{\partial \mathbf{x}(\mathbf{X}, t)}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial x_1}{\partial X_1} & \frac{\partial x_1}{\partial X_2} & \frac{\partial x_1}{\partial X_3} \\ \frac{\partial x_2}{\partial X_1} & \frac{\partial x_2}{\partial X_2} & \frac{\partial x_2}{\partial X_3} \\ \frac{\partial x_3}{\partial X_1} & \frac{\partial x_3}{\partial X_2} & \frac{\partial x_3}{\partial X_3} \end{bmatrix}.$$

The determinant of $\mathbf{F}$, which is one of its invariants, is denoted J. For valid deformations, $J > 0$.

Definition 2.1.3. Similarly, the **displacement gradient tensor** (or *displacement gradient*) is the tensor corresponding to the gradient of the displacement vector $\mathbf{u}$,

$$\mathbf{H} \equiv \nabla \mathbf{u} = \frac{\partial \mathbf{x}(\mathbf{X}, t)}{\partial \mathbf{X}} - \mathbf{I}.$$

The displacment gradient and deformation gradient are not independent. Rather:

Key Equation 2.1.4

The displacement gradient and deformation gradient tensors are related by

$$\mathbf{H} = \mathbf{F} - \mathbf{I}.$$

If they are (both) independent of $\mathbf{X}$, the deformation is called *homogeneous*.

The tensors $\mathbf{F}$ (and $\mathbf{H}$ by relation) are so important because they *completely define* what happens to one-, two-, and three-dimensional infinitesimal material elements, which are fundamentally all defined by "material fibers", or infinitesimal one-dimensional material line elements. Specifically,

- The infinitesimal material line element $d\mathbf{X}$ becomes $d\mathbf{x} = \mathbf{F}d\mathbf{X}$ (also, $d\mathbf{X} = \mathbf{F}^{-1}d\mathbf{x}$).
- The infinitesimal area element $\mathbf{n}_R dA_R$ becomes $\mathbf{n}dA = J\mathbf{F}^{-T}\mathbf{n}_R dA_R$ (**Nanson's formula**).
  - This infinitesimal "patch" of area $\mathbf{n}_R dA_R$ is actually defined by its *edge vectors* $\mathbf{a}$ and $\mathbf{b}$ (via $\mathbf{n}_R dA = \mathbf{a} \times \mathbf{b}$), which each get mapped to $\mathbf{Fa}$ and $\mathbf{Fb}$ respectively, so that $\mathbf{n}dA = \mathbf{Fa} \times \mathbf{Fb}$.
- The infinitesimal volume element dV_R becomes $dV = JdV_R$.
  - Similarly, the infinitesimal volume element dV_R is defined by three vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ by the paralellepiped formula.

Exercise 2.1.5. Recalling that $J \equiv \det \mathbf{F}$, **show that**

$$\mathbf{F}^T(\mathbf{Fa} \times \mathbf{Fb}) = J(\mathbf{a} \times \mathbf{b}),$$

from which Nanson's formula follows.

2.2 Polar Decomposition of F

We have seen that $\mathbf{F}$ acts by transforming material elements like $d\mathbf{X}$ to their deformed configuration $d\mathbf{x}$. To break down the action of the deformation, the polar decomposition theorem is useful. (We are guaranteed a polar decomposition because we required the determinant of $\mathbf{F}$ to be strictly positive.) Applied to $\mathbf{F}$, we get the following important result:

Key Equation 2.2.1

Given any deformation gradient tensor $\mathbf{F}$, we can write

$$\mathbf{VR} = \mathbf{F} = \mathbf{RU}$$

where the **rotation tensor** $\mathbf{R}$ is orthogonal, and the **left stretch tensor** $\mathbf{V}$ and **right stretch tensor** $\mathbf{U}$ are both symmetric and positive definite.

The stretch and rotation tensors can be computed by

$$\begin{aligned}\mathbf{U} &= \sqrt{\mathbf{F}^T\mathbf{F}},\\ \mathbf{V} &= \sqrt{\mathbf{F}\mathbf{F}^T},\\ \mathbf{R} &= \mathbf{F}\mathbf{U}^{-1} = \mathbf{V}^{-1}\mathbf{F}\end{aligned}$$

and consequently $\mathbf{V} = \mathbf{RUR}^T$, i.e. $\mathbf{V}$ is just $\mathbf{U}$ having been transformed using the rotation tensor $\mathbf{R}$ in the basis transformation formula.

It turns out that computing the square root of a tensor gets annoying, and we lose no information in keeping the square, so we define the **left and right Cauchy-Green tensors** as

$$\begin{aligned}\mathbf{B} &\equiv \mathbf{V}^2 = \mathbf{F}\mathbf{F}^T\\ \mathbf{C} &\equiv \mathbf{U}^2 = \mathbf{F}^T\mathbf{F}\end{aligned}$$

Notice that $\mathbf{U}$ and $\mathbf{V}$, being the same tensor under a basis transformation, contain the same eigenvalues, which we call λ_i. Let $\{\mathbf{r}_i\}$ denote the eigenvectors of $\mathbf{U}$. Note that the vectors $\mathbf{r}_i$ are associated with the reference configuration, because $\mathbf{U}$ acts *first* on a vector $d\mathbf{X}$ *before* it is rotated by $\mathbf{R}$ to the deformed configuration. Then, again because $\mathbf{U}$ and $\mathbf{V}$ are related by a basis transformation through $\mathbf{R}$, the eigenvectors of $\mathbf{V}$, $\{\mathbf{l}_i\}$, must be related by $\{\mathbf{l}_i\} = \mathbf{R}\{\mathbf{r}_i\}$. The vectors $\mathbf{l}_i$ are, therefore, associated with the *deformed* configuration, acting on the vector $\mathbf{R}d\mathbf{X}$.

It then follows that

$$\mathbf{V} = \sum_i \lambda_i\, \mathbf{l}_i \otimes \mathbf{l}_i, \qquad \mathbf{U} = \sum_i \lambda_i\, \mathbf{r}_i \otimes \mathbf{r}_i$$

$$\mathbf{B} = \sum_i \lambda_i^2\, \mathbf{l}_i \otimes \mathbf{l}_i, \qquad \mathbf{C} = \sum_i \lambda_i^2\, \mathbf{r}_i \otimes \mathbf{r}_i$$

$$\mathbf{F} = \sum_i \lambda_i\, \mathbf{l}_i \otimes \mathbf{r}_i$$

$$\mathbf{R} = \mathbf{l}_i \otimes \mathbf{r}_i$$

Remark 2.2.2: The spectral representation of the tensors above is useful to illustrate the fact that $\mathbf{F}$ and $\mathbf{R}$ are called "two-point" tensors, as in the most general form they transform vectors corresponding to the basis in the reference configuration, $\{\mathbf{r}_i\}$, to vectors corresponding to the basis in the deformed configuration, $\{\mathbf{l}_i\}$. The rotation vector $\mathbf{R}$ thus exactly specifies how the two bases may be related.

Correspondingly, the right stretch tensor $\mathbf{U}$ and the right Cauchy-Green tensor $\mathbf{C}$ transform vectors within the basis of the *reference* configuration, and the left stretch tensor $\mathbf{V}$ and the left Cauchy-Green tensor $\mathbf{B}$ transform vectors within the basis of the *deformed* configuration.

Physically, $\mathbf{R}$ is always responsible for a rigid rotation of the material element. As their names imply, the stretch tensors $\mathbf{U}$ and $\mathbf{V}$ are responsible for changing the length of the material element, and possibly further changing its orientation (note that in general, $\mathbf{U}d\mathbf{X}$ is not parallel to $d\mathbf{X}$). It is only material elements *aligned with the eigenvectors of the stretch tensors* that undergo a pure stretch (i.e., change in length without rotation). These material elements stretch precisely by λ_i, which are called *principal stretches*.

Recall that in the world of linear transformations, the order of their application matters. To that end, while the *total* effect of applying the sequence $\mathbf{VR}$ is exactly the same as the effect of applying the sequence $\mathbf{RU}$ (namely, $\mathbf{F}$), the "route" which a material element takes is different in the intermediate steps. In the application of $\mathbf{VR}d\mathbf{X}$, the undeformed element $d\mathbf{X}$ is first rigidly rotated to $\mathbf{R}d\mathbf{X}$. Then, it undergoes a stretch (and perhaps more rotation as discussed above) to $\mathbf{VR}d\mathbf{X}$. In the case of the right stretch sequence, the stretching part is accomplished first, followed by a rigid rotation of the now-stretched $\mathbf{U}d\mathbf{X}$.

Finally, we note that because $\mathbf{R}$ simply rigidly rotates material elements, it is $\mathbf{U}$ (or $\mathbf{V}$) that contributes entirely to any elongation or contraction of material elements. Thus, the **fiber stretch** $\lambda(\mathbf{m})$ in an arbitrary direction $\mathbf{m}$ is given by

$$\lambda(\mathbf{m}) = |\mathbf{U}\mathbf{m}| \qquad \text{or} \qquad \lambda^2 = \mathbf{m} \cdot \mathbf{C}\mathbf{m},$$

the latter being typically easier to compute. Note that by the definition of an eigenvalue, we recover the principal stretches when $\mathbf{m}$ is an eigenvector.

Similarly the **angle change** between material elements is likewise entirely attributed to $\mathbf{U}$ or $\mathbf{V}$, as the rotation tensor accomplishes only a solid-body (rigid) rotation. The angle change is defined in terms of two material elements $d\mathbf{X}^{(1)}$ and $d\mathbf{X}^{(2)}$ that are mutually orthogonal in the reference configuration (i.e., they are oriented in directions $\mathbf{m}_1$ and $\mathbf{m}_2$ such that $\mathbf{m}_1 \cdot \mathbf{m}_2 = 0$). After the deformation, the *decrease* in the angle between them is

$$\gamma \equiv \frac{\pi}{2} - (\text{new angle}) = \sin^{-1}\left(\frac{\mathbf{m}_1 \cdot \mathbf{C}\mathbf{m}_2}{\lambda(\mathbf{m}_1)\,\lambda(\mathbf{m}_2)}\right).$$

If the $\mathbf{m}_i$ are chosen to be any two eigenvectors of $\mathbf{U}$ (i.e., any of the $\mathbf{r}_i$), then $\gamma = 0$, i.e. the triad of principal basis vectors never experiences shear.

Exercise 2.2.3. Consider the simple shear deformation given by

$$\begin{cases} x_1 &= X_1 + \gamma X_2 \\ x_2 &= X_2 \\ x_3 &= X_3, \end{cases}$$

where $\gamma > 0$ is a constant. **Sketch the deformation** of a unit cube (one corner at the origin, and the other corner at $(1, 1, 1)$) under simple shear. **Compute** the tensors $\mathbf{F}$, $\mathbf{B}$,

and $\mathbf{V}$ for simple shear. **Compute** J and show that simple shear is a *volume-preserving* deformation.

Exercise 2.2.4. The simple shear deformation is volume-preserving, but the unit cube occupies a different region of space before and after the deformation. **Describe in words or with a sketch** a deformation which is *not* the identity, but where material elements occupy the *same volume* in the undeformed and deformed configurations.

2.3 Strain: Finite Deformations

In general, **strain** is a measure of how much a particular deformation is different from a rigid-body rotation. Of course, this information is already entirely contained in $\mathbf{F}$ or equivalently $\mathbf{H}$, which takes on the value $\mathbf{I}$ (or $\mathbf{0}$, respectively) when the deformation is exactly a rigid-body rotation. Nevertheless a number of strain measures exist to quantify this difference, all used in slightly different contexts, but all generally functions of "descendants" of $\mathbf{F}$ like $\mathbf{U}$.

Definition 2.3.1. The **Biot strain tensor** is defined as

$$\mathbf{E} \equiv \mathbf{U} - \mathbf{I}.$$

Definition 2.3.2. The **Green-St. Venant strain tensor** is defined as

$$\mathbf{E} \equiv \frac{1}{2}(\mathbf{C} - \mathbf{I}),$$

recalling that knowing $\mathbf{C}$ is enough to quantify the stretch in any fiber direction. This definition is "nice" because the computation of $\mathbf{C}$ is easy, compared to $\mathbf{U}$ (eww, square roots).

Definition 2.3.3. There are two (sometimes called "right" and "left", respectively) **Hencky logarithmic strain tensors**, defined to be

$$\ln \mathbf{U} = \sum_i (\ln \lambda_i)\, \mathbf{r}_i \otimes \mathbf{r}_i$$
$$\ln \mathbf{V} = \sum_i (\ln \lambda_i)\, \mathbf{l}_i \otimes \mathbf{l}_i.$$

2.4 Strain: Infinitesimal Deformations

When $\mathbf{F}$ is *close to* $\mathbf{I}$, the resulting deformation is called *infinitesimal* (or, less precisely, *small*). When this is the case, the preceding results can be *linearized*, i.e., the governing equations can be expanded using something similar to a Taylor expansion, with terms having quadratic or higher order then discarded.

A consequence of this is that the *multiplicative* stretch-rotation decomposition of deformation we have seen before, e.g. $\mathbf{F} = \mathbf{RU}$, becomes an *additive* one. Specifically, in the infinitesimal case, we work with the symmetric-skew decomposition of the displacement gradient tensor $\mathbf{H}$. In this limit, all of the previous strain measures collapse onto one, which we call the *infinitesimal strain tensor*.

Definition 2.4.1. When $|\mathbf{H}| \equiv |\nabla \mathbf{u}| \ll 1$, the **infinitesimal strain tensor** is defined to be the symmetric part of $\mathbf{H}$,

$$\boldsymbol{\varepsilon} = \frac{1}{2}\left(\mathbf{H} + \mathbf{H}^T\right), \qquad \varepsilon_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i})$$

with the following components in terms of the displacement $\mathbf{u}(\mathbf{X}) = (u_1, u_2, u_3)$:

$$[\boldsymbol{\varepsilon}] = \begin{bmatrix} \frac{\partial u_1}{\partial X_1} & \frac{1}{2}\left(\frac{\partial u_1}{\partial X_2} + \frac{\partial u_2}{\partial X_1}\right) & \frac{1}{2}\left(\frac{\partial u_1}{\partial X_3} + \frac{\partial u_3}{\partial X_1}\right) \\ & \frac{\partial u_2}{\partial X_2} & \frac{1}{2}\left(\frac{\partial u_2}{\partial X_3} + \frac{\partial u_3}{\partial X_2}\right) \\ \text{symm.} & & \frac{\partial u_3}{\partial X_3} \end{bmatrix}$$

With respect to the dimensions of the reference body, the entries in the infinitesimal strain tensor have a nice physical description. The diagonal entries of the strain tensor, for example ε_{11}, represent the *change in length per unit original length* of a line element in that direction (for example, $\mathbf{e}_1$). The off-diagonal entries, for example ε_{13}, represent *one-half* the change (decrease, if positive) in angle between line elements originally oriented in those orthogonal directions (for example, $\mathbf{e}_1$ and $\mathbf{e}_3$). Most generally, for any two vectors $\mathbf{a}$ and $\mathbf{b}$, the operation $\mathbf{a} \cdot \boldsymbol{\varepsilon}\mathbf{b}$ computes *half the change in dot product* between $\mathbf{a}$ and $\mathbf{b}$.

Exercise 2.4.2. AKG 1.15.

Remark 2.4.3: The off-diagonal components ε_{ij}, $i \neq j$ in the strain tensor are called the *tensorial shear strain* components (cf. the diagonal terms which are called the *normal strain* components). Confusingly, the *engineering shear strain components* are defined to be twice the magnitude of the corresponding tensorial shear strain components, $\gamma_{ij} = 2\varepsilon_{ij}$, $i \neq j$. For example, in terms of $\mathbf{u}(\mathbf{X})$,

$$\gamma_{12} = \left(\frac{\partial u_1}{\partial X_2} + \frac{\partial u_2}{\partial X_1}\right).$$

Derivation 2.4.4 (Infinitesimal strain tensor as a linearization of a large-strain tensor)
Consider the Green-St. Venant strain tensor

$$\mathbf{E} \equiv \frac{1}{2}(\mathbf{C} - \mathbf{I}),$$

where $\mathbf{C} = \mathbf{F}^T\mathbf{F}$ and $\mathbf{H} = \mathbf{F} - \mathbf{I}$. Then

$$\begin{aligned} \mathbf{E} &= \frac{1}{2}\left[\mathbf{F}^T\mathbf{F} - \mathbf{I}\right] \\ &= \frac{1}{2}\left[(\mathbf{H} + \mathbf{I})^T(\mathbf{H} + \mathbf{I}) - \mathbf{I}\right] \\ &= \frac{1}{2}\left[(\mathbf{H}^T\mathbf{H} + \mathbf{H} + \mathbf{H}^T + \mathbf{I}) - \mathbf{I}\right] \\ &= \frac{1}{2}(\mathbf{H}^T\mathbf{H} + \mathbf{H} + \mathbf{H}^T) \end{aligned}$$

and when $|\mathbf{H}|$ is small, the term $\mathbf{H}^T\mathbf{H}$ can be neglected, leaving us with the usual infinitesimal strain tensor $\boldsymbol{\varepsilon}$. (A similar computation can be undertaken starting with $\mathbf{B}$, whereby the second-order term becomes $\mathbf{H}\mathbf{H}^T$ instead.)

The infinitesimal strain tensor admits a spherical-deviatoric decomposition,

$$\boldsymbol{\varepsilon} = \underbrace{\boldsymbol{\varepsilon} - \frac{1}{3}(\operatorname{tr}\boldsymbol{\varepsilon})\mathbf{I}}_{\text{strain deviator } \boldsymbol{\varepsilon}'} + \underbrace{\frac{1}{3}(\operatorname{tr}\boldsymbol{\varepsilon})\mathbf{I}}_{\text{volumetric strain}}$$

where the **strain deviator** $\boldsymbol{\varepsilon}'$ records the part of strain responsible for shape change, and the spherical **volumetric strain** records the amount of volume change. Specifically,

for infinitesimal deformations the *local* volume change per unit volume is given by

$$\operatorname{tr}\boldsymbol{\varepsilon} = \varepsilon_{kk}.$$

This is exactly three times the number that appears in the diagonal elements of the volumetric strain tensor.

(Actually, it turns out that the "linearized" version of $J \equiv \det \mathbf{F}$ is exactly $1 + \operatorname{tr}\boldsymbol{\varepsilon}$, so in the limit of infinitesimal deformations these two measures of volume change are identical in nature.)

It should be emphasized that the strain tensor can vary in space and hence describes the *local* state of shape and/or volume change. If an infinitesimal spherical element undergoes a deformation whose strain deviator is nonzero, it will no longer be spherical (i.e., it will have changed shape). Conversely if the strain deviator is zero but the volumetric strain is nonzero, the element simply dilates or contracts according to the value of ε_{kk}.

Exercise 2.4.5. The matrix of strain components in a body, with respect to a basis, is

$$[\boldsymbol{\varepsilon}] = \begin{pmatrix} 2 & -0.3 & 0 \\ -0.3 & 1 & 2 \\ 0 & 2 & 1 \end{pmatrix} \times 10^{-3}$$

Consider a tetrahedron defined by points $O(0,0,0)$, $A(1,0,0)$, $B(0,1,0)$, and $C(0,0,1)$ in the given basis, as well as point $D(1/2,1/2,0)$ at the midpoint of edge AC.

Calculate:

1. the normal strain (i.e., change in length relative to original length) of fibers OA, AC and DB;
2. the change in angle between fibers BD and AC; and
3. the volumetric strain (i.e., relative change in volume) of the tetrahedron $OABC$.

Remember that all vectors should be normalized to unit vectors.

Finally, observing that the infinitesimal strain tensor $\boldsymbol{\varepsilon}$ corresponds to the symmetric part of the skew-symmetric decomposition of $\mathbf{H}$, we define the *infinitesimal rotation tensor* to be the skew part. Namely,

Definition 2.4.6. When $|\mathbf{H}| \equiv |\nabla \mathbf{u}| \ll 1$, the **infinitesimal rotation tensor** is defined to be

$$\boldsymbol{\omega} = \frac{1}{2}\left(\mathbf{H} - \mathbf{H}^T\right), \qquad \omega_{ij} = \frac{1}{2}(u_{i,j} - u_{j,i})$$

The tensor $\boldsymbol{\omega}$ is skew-symmetric ($\omega_{ij} = -\omega_{ji}$), and thus its diagonal components are zero. The three independent components ω_{12}, ω_{13}, and ω_{23} can be coordinated with a three-component vector[4] $\mathbf{w} \equiv \frac{1}{2}\operatorname{curl}\mathbf{u}$ such that $\boldsymbol{\omega}\mathbf{m} = \mathbf{w} \times \mathbf{m}$ for all $\mathbf{m}$.

[4] In particular, $\omega_{12} = -w_3$, $\omega_{13} = w_2$, and $\omega_{23} = -w_1$.

In general the tensor $\boldsymbol{\omega}$ quantifies the amount of local rotation at a given infinitesimally small point. The magnitude $|\boldsymbol{\omega}| = |\mathbf{w}|$ corresponds to the angle of rotation, and the unit vector in the direction of $\mathbf{w}$ points out the axis of rotation.

Taken together, the strain and rotation tensors "break down" how a body changes under infinitesimal deformation: $\mathbf{H} = \boldsymbol{\varepsilon} + \boldsymbol{\omega}$. For example, in the special case where $\boldsymbol{\varepsilon} = \mathbf{0}$ at a given point, the neighborhood of that point rotates like a rigid body.

Finally, it should be restated that a *rigid-body translation*, where $\mathbf{u}$ is a constant vector (independent of position), does *not* contribute to $\nabla\mathbf{u}$ and thus does not contribute to $\boldsymbol{\varepsilon}$ or to $\boldsymbol{\omega}$.

Exercise 2.4.7. AKG 1.1, 1.2, 1.5.

Exercise 2.4.8. Describe in words a motion that produces a non-zero displacement gradient tensor, but has no strain.

2.4.1 Compatibility

Given a displacement field $\mathbf{u} = \mathbf{u}(\mathbf{X})$, it is straightforward to compute the strain field: the components of $[\boldsymbol{\varepsilon}]$ are given by $\varepsilon_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i})$.

However, a problem arises when we go in reverse. That is, given a strain field $[\boldsymbol{\varepsilon}]$ with the intent to solve for the original displacement field $\mathbf{u}$, we are given a system with six independent components (the ε_{ij}) but only three unknowns (the u_i). Because there are more constraints than unknowns, *it is possible that a solution may not exist.* Thus, to guarantee that a given strain field "comes from" a displacement field, we need enough additional conditions to guarantee the existence of a solution. These additional conditions are called the **compatibility conditions**.

Remark 2.4.9: This situation is analagous to the observation that a given vector field $\mathbf{v}(\mathbf{x}, t)$ may or may not have a scalar *potential function* $\varphi(\mathbf{x}, t)$, for which grad $\varphi = \mathbf{v}$. Given a vector field, a valid potential function only exists when the three *compatibility equations* curl $\mathbf{v} = \mathbf{0}$ must be satisfied.

A non-rigorous physical interpretation of the compatibility requirement is that a valid displacement field must not "break up" or "cause overlap between" elements in the deformed configuration.

In two dimensions, where three strain components $\{\varepsilon_{11}, \varepsilon_{22}, \varepsilon_{12}\}$ are given and two displacement components $\{u_1, u_2\}$ are requested, the (one) compatibility equation is

$$\varepsilon_{11,22} + \varepsilon_{22,11} - 2\varepsilon_{12,12} = 0.$$

Equivalently, if a given set of strain components $\{\varepsilon_{11}, \varepsilon_{22}, \varepsilon_{12}\}$ is compatible, than *any* two out of the three equations

$$\begin{aligned} \frac{\partial u_1}{\partial x_1} &= \varepsilon_{11} \\ \frac{\partial u_2}{\partial x_2} &= \varepsilon_{22} \\ \frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1} &= 2\varepsilon_{12} \end{aligned}$$

must yield the same values for u_1 and u_2.

The requirement that the three strain components satisfy the compatibility condition is *necessary and sufficient* for the *existence* of a valid displacement field.

Example 2.4.10

The matrix

$$[\mathbf{e}] = \frac{1}{L^2}\begin{pmatrix} x_2^2 & x_1^2 \\ x_1^2 & x_2^2 \end{pmatrix}$$

does not satisfy the compatibility equation in two dimensions, so it cannot be a strain field. In other words, there is no displacement field $\mathbf{u}(x_1, x_2) = (u_1(x_1, x_2), u_2(x_1, x_2))$ which could be differentiated to produces this "strain" matrix. However, the matrix

$$[\boldsymbol{\varepsilon}] = A\begin{pmatrix} x_1^2 + x_2^2 & x_1 x_2 \\ x_1 x_2 & x_2^2 \end{pmatrix}$$

does satisfy the compatibility equations. Therefore, a displacement field $\mathbf{u}(x_1, x_2)$ can be found, at least down to a constant of integration which represents a rigid translation.

In three dimensions there are six compatibility conditions that must be satisfied. They can be summarized as

$$\operatorname{curl}(\operatorname{curl} \boldsymbol{\varepsilon}) = \mathbf{0}, \qquad \varepsilon_{ij,kl} + \varepsilon_{kl,ij} - \varepsilon_{ik,jl} - \varepsilon_{jl,ik} = 0$$

for $i, j, k, l = \{1, 2, 3\}$ (no sums).

Exercise 2.4.11. AKG 1.13, 1.14.

3 Balance Laws

In this section we discuss the global and local forms of the balance laws for mass, forces, moments, and energy, which are again applicable to deformations without considering their material or mechanical properties. These balance laws follow from physical principles as applied to continua. For the majority of this section we will work in terms of the deformed coordinate $\mathbf{x}$, because the reference configuration is typically assumed to be free of forces. Accordingly we will define an origin $\mathcal{O}$ from which we can take the vector $\mathbf{r}$ to represent the directed distance between $\mathcal{O}$ and $\mathbf{x}$.

3.1 Balance of Mass

The **balance of mass** requires that mass is neither created nor destroyed as a result of deformation.

Key Equation 3.1.1 (Balance of mass)

If the mass density of the reference configuration is $\rho_R(\mathbf{X})$ and the mass density of the deformed configuration is $\rho(\mathbf{x})$,

$$\int_{\mathcal{R}_R} \rho_R(\mathbf{X})\, dV_R = \int_{\mathcal{R}} \rho(\mathbf{x})\, dV$$

from which it follows by localization that

$$\rho(\mathbf{x}) = \frac{1}{J}\rho_R(\mathbf{X}).$$

3.2 Balance of Forces and Moments

In continuum mechanics, two classes of forces are assumed to exist at any point within a body. The first class consists of forces that act on *surfaces* and are therefore defined *per unit area*. These **tractions** are either due to internal contact forces, if the point in question is on the interior of a body, or due to contact forces exerted by the environment on the boundary of the body, if the point in question is on the exterior of a body. The second class consists of forces that act on *volumes* and are therefore defined *per unit volume*. These **body forces** are assumed to be from an external source, gravity being the most common example.

Definition 3.2.1. The **traction vector** $\mathbf{t}(\mathbf{x}, \mathbf{n})$ at a point $\mathbf{x}$ is the total surface force per unit area acting on a section plane passing through $\mathbf{x}$ with outward-pointing normal vector $\mathbf{n}$. By convention, we take the normal vector to be pointing away from the region of interest, so that this force per unit area is taken to be the force per unit area exerted *on the inside* section *from the outside* section, where $\mathbf{n}$ points from the inside to the outside.

Note that the value of $\mathbf{t}$ depends not only on the location $\mathbf{x}$, but also on the choice of section plane, as defined by $\mathbf{n}$.

Definition 3.2.2. The **body force vector** $\mathbf{b}_0(\mathbf{x})$ quantifies the total force per unit volume acting at a point $\mathbf{x}$ from external (environmental) sources.

Remark 3.2.3: Some textbooks generalize the body force vector as

$$\mathbf{b}(\mathbf{x}) = \mathbf{b}_0(\mathbf{x}) - \rho\ddot{\mathbf{u}},$$

where the term $\rho\ddot{\mathbf{u}}$ is the *inertial body force* associated with acceleration in the body.

With this formulation, the balance of forces (or balance of linear momentum) for a continuum is simply the requirement that $\mathbf{F}_{\text{net}} = m\mathbf{a}$, i.e. Newton's second law. The balance of moments (or balance of angular momentum) is analogously Newton's second law for angular acceleration, obtained by cross-product multiplying both sides of $\mathbf{F}_{\text{net}} = m\mathbf{a}$ by the moment arm $\mathbf{r}$. We commonly write these statements on a *per-unit-volume* basis, so that the density appears instead of the mass.

Key Equation 3.2.4 (Balance of linear and angular momentum, global version)
The balance of linear momentum requires that

$$\int_{\partial\mathcal{R}} \mathbf{t}(\mathbf{n})\, dA + \int_{\mathcal{R}} \mathbf{b}_0\, dV = \int_{\mathcal{R}} \rho\ddot{\mathbf{u}}\, dV$$

and the balance of angular momentum follows,

$$\int_{\partial\mathcal{R}} \mathbf{r} \times \mathbf{t}(\mathbf{n})\, dA + \int_{\mathcal{R}} \mathbf{r} \times \mathbf{b}_0\, dV = \int_{\mathcal{R}} \rho(\mathbf{r} \times \ddot{\mathbf{u}})\, dV$$

Remark 3.2.5: In the *special case* where the acceleration term $\ddot{\mathbf{u}}$ is zero or negligible compared to the external force and moment terms, the right hand side of (3.2.4) vanishes and the resulting formulation is called the *equations of equilibrium* (and the continuum is said to be *in equilibrium*).

3.2.1 Cauchy's Result and the Stress Tensor

It can be shown that Newton's third law, the law of action and reaction, applies to continua in the following form, for all $\mathbf{n}$:

$$\mathbf{t}(\mathbf{x}, -\mathbf{n}) = -\mathbf{t}(\mathbf{x}, \mathbf{n})$$

Using this, the traction at a point in a direction $\mathbf{n}$ can be determined solely with $\mathbf{n}$ and a tensor independent of $\mathbf{n}$ called the *Cauchy stress tensor*.

Key Equation 3.2.6
The traction $\mathbf{t}(\mathbf{n})$ at a point $\mathbf{x}$ is a *linear* function of $\mathbf{n}$ through the **Cauchy stress tensor** $\mathbf{T} = \mathbf{T}(\mathbf{x})$ evaluated at $\mathbf{x}$:

$$\mathbf{t}(\mathbf{x}, \mathbf{n}) = \mathbf{T}\mathbf{n}$$

Formally, the tensor $\mathbf{T}$ is defined as $\mathbf{T} \equiv \mathbf{t}_i \otimes \mathbf{e}_i$ (yes, there is a sum over i here), where $\mathbf{t}_i$ stands for $\mathbf{t}(\mathbf{n} = \mathbf{e}_i, \mathbf{x})$, i.e. the traction *in the coordinate direction* $\mathbf{e}_i$. Said another way,

At a given location x, knowing the traction in the coordinate directions $\{\mathbf{e}_i\}$ is necessary and sufficient to determine the traction in any arbitrary direction n.

The components T_{ij} of the stress tensor can thus be read off as "the j-direction component[1] of the traction on the i-face[2] of an infinitesimally small section-cut cube at the location of interest, with the cube edges cut parallel to the basis vectors". The *local* versions of linear and angular momentum can now be stated in terms of $\mathbf{T}$:

Key Equation 3.2.7 (Balance of linear and angular momentum, local version)

At every point $\mathbf{x}$ in the deformed body, the balance of linear momentum requires that

$$\operatorname{div} \mathbf{T} + \mathbf{b} = \rho \ddot{\mathbf{u}} \qquad T_{ij,j} + b_i = \rho \ddot{u}_i$$

and the balance of angular momentum requires that

$$\mathbf{T} = \mathbf{T}^T \qquad T_{ij} = T_{ji}.$$

Derivation 3.2.8

From the global balance of linear momentum,

$$\int_{\partial \mathcal{R}} \mathbf{t}(\mathbf{n})\, dA + \int_{\mathcal{R}} \mathbf{b}_0\, dV = \int_{\mathcal{R}} \rho \ddot{\mathbf{u}}\, dV,$$

we can use the definition of the stress tensor to replace the traction $\mathbf{t}(\mathbf{n})$ with $\mathbf{Tn}$, then use the divergence theorem to convert the area integral into a volume integral:

$$\begin{aligned} \int_{\partial \mathcal{R}} \mathbf{Tn}\, dA + \int_{\mathcal{R}} \mathbf{b}_0\, dV &= \int_{\mathcal{R}} \rho \ddot{\mathbf{u}}\, dV \\ \int_{\mathcal{R}} \operatorname{div} \mathbf{T}\, dV + \int_{\mathcal{R}} \mathbf{b}_0\, dV &= \int_{\mathcal{R}} \rho \ddot{\mathbf{u}}\, dV \\ \int_{\mathcal{R}} (\operatorname{div} \mathbf{T} + \mathbf{b}_0 - \rho \ddot{\mathbf{u}})\, dV &= 0 \end{aligned}$$

Since this is true for *any* $\mathcal{R}$, it follows that, by localization, $\operatorname{div} \mathbf{T} + \mathbf{b}_0 - \rho \ddot{\mathbf{u}} = 0$.

Derivation 3.2.9

A formal proof of the symmetry of the stress tensor can be carried out in a similar method as 3.2.8, by cross-multiplying both sides by an arbitrary vector $\mathbf{r}$. This proof can be found in most standard textbooks or is left as an exercise to the reader.

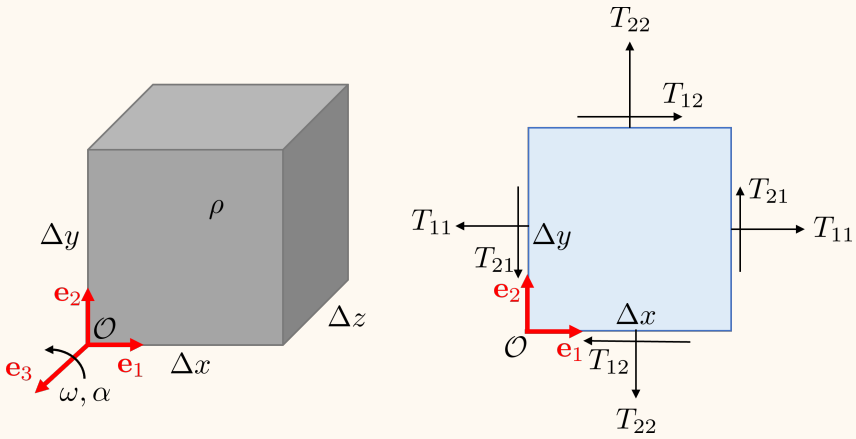
For more practical purposes, here we give a proof of the same result but specalized to an infinitesimal rectangular volume in a fixed rectangular basis. Specifically, we aim to show that even if the rectangular volume is given arbitrary angular velocity

[1] Actually, because $\mathbf{T}$ is symmetric, the roles of i and j may be switched here.
[2] The i-face is the face with outward normal vector $\mathbf{e}_i$.

and arbitrary angular acceleration, the tensor describing its stress state is always symmetric. (In *equilibrium*, both of these values are specialized to zero, but here we prove the more general result.)

To this end, consider the rectangular element shown in the figure; it has sides oriented parallel to the coordinate system $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, side lengths Δx, Δy, Δz, and uniform density $\rho > 0$. One corner of the element is taken to coincide with the origin $\mathcal{O}$ of the coordinate system.

It follows that the tractions on the edges are the components T_{ij}, as shown in the figure. (Thus far we have not assumed any properties of the stress tensor; only that it exists.)



We can apply the classical "angular Newton's second law" (the principle of conservation of angular momentum),

$$\mathbf{M}_{\text{net}} = I\boldsymbol{\alpha},$$

where $\mathbf{M}_{\text{net}}$ is the net moment about any point, and $\boldsymbol{\alpha}$ is the resulting angular acceleration about that point. We will compute the moment about $\mathcal{O}$ in the $\mathbf{e}_1 - \mathbf{e}_2$ plane; the other coordinate axes follow an identical result with a simple permutation of the indices.

The net moment around the origin, counterclockwise positive, is

$$\mathbf{M}_{\text{net}} = (T_{21}\Delta y\Delta z)\Delta x - (T_{12}\Delta x\Delta z)\Delta y = (T_{21} - T_{12})\Delta x\Delta y\Delta z.$$

Meanwhile, the mass moment of inertia is

$$I_z = \frac{1}{12}(\underbrace{\rho\Delta x\Delta y\Delta z}_{\text{mass}})(\Delta x^2 + \Delta y^2) + \frac{1}{2}(\rho\Delta x\Delta y\Delta z)(\Delta x^2 + \Delta y^2),$$

where the second term comes from the parallel axis theorem. Hence the angular Newton's second law result is

$$(T_{21} - T_{12})\Delta x\Delta y\Delta z = \frac{7}{2}\rho\alpha(\Delta x\Delta y\Delta z)(\Delta x^2 + \Delta y^2)$$
$$(T_{21} - T_{12}) = \frac{7}{2}\rho\alpha(\Delta x^2 + \Delta y^2)$$

Finally, in the infinitesimal case, where $\Delta x \to 0$ and $\Delta y \to 0$, the preceding result yields

$$T_{12} = T_{21},$$

for any ρ and any α, the result we were seeking. Notice that the value of the instantaneous angular velocity (say ω) does not appear and hence the above result holds for any ω.

Taking moments in a similar way about the other (x- and y-) axes demonstrates that $T_{13} = T_{31}$ and $T_{23} = T_{32}$, which completes the proof.

Again, the special case of *equilibrium* (zero or negligible acceleration terms) results in the simplified form

$$\operatorname{div} \mathbf{T} + \mathbf{b} = \mathbf{0}$$

of the linear momentum balance. Interestingly, the angular momentum balance is unchanged. Recall that the statement of angular momentum balance can be satisfied with or without external moments, and with or without any sort of rotation of the body. Indeed,

***As long as the angular momentum balance is satisfied*, the stress tensor $\mathbf{T}$ is symmetric, independent of any rotational velocity or acceleration. The angular momentum balance is typically taken to be satisfied in most realistic solid mechanics problems unless an exotic material model is assumed.**

Exercise 3.2.10. Prove each statement or provide a counterexample.

1. If the traction at a given point in a body (with well-defined normal vector) is zero, then the stress tensor evalulated at that point is zero.
2. If the stress tensor evaluated at a point in a body (with well-defined normal vector) is zero, then the traction at that point is zero.

3.2.2 Principal Stresses

Being symmetric, $\mathbf{T}$ admits a basis of eigenvectors $\{\mathbf{m}_i\}$ and eigenvalues τ_i for which

$$\mathbf{t}(\mathbf{n} = \mathbf{m}_i) = \mathbf{T}\mathbf{m}_i = \tau_i \mathbf{m}_i,$$

i.e. the stresses in these special directions are entirely normal, with zero shear component. These eigenvectors are called the **principal stress directions**, and the corresponding eigenvalues τ_i represent the **principal stresses**, or the normal stresses in these directions. (Sometimes the principal stresses will be written τ_i^P for clarity.) The principal stress with the largest magnitude is usually called the *maximum principal stress* and has important applications in failure criteria, for example.

3.2.3 Spherical-Deviatoric Split of $\mathbf{T}$

As with the small strain tensor, the stress tensor $\mathbf{T}$ admits a physically meaningful spherical-deviatoric decomposition. Namely,

$$\mathbf{T} = \underbrace{\mathbf{T} - \frac{1}{3}(\operatorname{tr} \mathbf{T})\mathbf{I}}_{\text{stress deviator } \mathbf{T}'} + \underbrace{\frac{1}{3}(\operatorname{tr} \mathbf{T})\mathbf{I}}_{\text{volumetric stress}}$$

where the **stress deviator** $\mathbf{T}'$ quantifies the part of the stress state that *acts to* change the shape of the body. The volumetric part of the stress is the tensor which has zeros in the off-diagonal components and $(1/3)$ tr $\mathbf{T}$ in the diagonal components; this scalar is called the *negative* of the **mean normal stress**. Note that the mean normal stress is defined this way so that a *compressive* mean normal stress is positive. In the special case where $\mathbf{T}' = \mathbf{0}$, the stress state is said to be **hydrostatic**. For example, a fluid at rest experiences a hydrostatic stress state.

3.2.4 Referential Stress Measures: Piola Stresses

Sometimes it is useful to record the stress state in the deformed configuration with regard to the body's *reference* configuration. This is analagous to the concept of *engineering stress*, in contrast to *true stress*. From Nanson's relation, $\mathbf{n} = J\mathbf{F}^T\mathbf{n}_R$ and we can define the **first Piola stress** as

$$\mathbf{P} \equiv J\mathbf{T}\mathbf{F}^{-T}.$$

Thus, $\mathbf{P}$ catalogues the stress per unit *reference area*[3], and

$$\int_{\partial\mathcal{R}_R} \mathbf{P}\mathbf{n}_R \, dA_R = \int_{\partial\mathcal{R}} \mathbf{T}\mathbf{n} \, dA.$$

The balance of linear and angular momentum can then be expressed in terms of the referential variable $\mathbf{X} = (X_1, X_2, X_3)$ as

$$\begin{aligned} \text{Div } \mathbf{P} + \mathbf{b}_R &= \rho_R \ddot{\chi} \\ \mathbf{P}\mathbf{F}^T &= \mathbf{F}\mathbf{P}^T \end{aligned}$$

where $\mathbf{b}_R = J\mathbf{b}$ is the body force per unit *reference volume*. The divergence operator is written with a capital letter here to indicate that it is taken with respect to the referential variable $\mathbf{X}$, not $\mathbf{x}$ as before.

Observe that the first Piola stress is *not* symmetric. To "fix" this problem we define a *symmetric* **second Piola stress** as[4]

$$\mathbf{S} \equiv \mathbf{F}^{-1}\mathbf{P} = J\mathbf{F}^{-1}\mathbf{T}\mathbf{F}^{-T}.$$

Remark 3.2.11: Recall that $\mathbf{F}$ is a two-point tensor which transforms vectors from the reference configuration basis to the deformed configuration basis. The first Piola stress $\mathbf{P}$ is likewise a mixed-basis tensor which transforms vectors from the reference configuration basis to the deformed configuration basis. Conversely, the second Piola stress $\mathbf{S}$ stays within the reference configuration basis, and the Cauchy stress $\mathbf{T}$ stays within the deformed configuration basis.

Specifically, suppose we use the basis $\{\mathbf{e}_i\}$ in the reference configuration and the

[3] Observe that $\mathbf{P}$ has a *mixed basis*, since it maps normal vectors in the reference configuration to traction vectors which exist only in the deformed configuration.

[4] In contrast to $\mathbf{P}$, the symmetric $\mathbf{S}$ maps tensors in the reference configuration to tensors in the reference configuration.

basis $\{\mathbf{e}'_i\}$ in the deformed configuration. Then we may write

$$\begin{aligned}\mathbf{F} &= F_{ij}\mathbf{e}'_i \otimes \mathbf{e}_j,\\ \mathbf{P} &= P_{ij}\mathbf{e}'_i \otimes \mathbf{e}_j,\\ \mathbf{S} &= S_{ij}\mathbf{e}_i \otimes \mathbf{e}_j,\\ \mathbf{T} &= T_{ij}\mathbf{e}'_i \otimes \mathbf{e}'_j.\end{aligned}$$

Note that $F_{ij} = \mathbf{e}'_i \cdot \mathbf{F}\mathbf{e}_j$, and so on. This emphasizes that $\mathbf{F}$ *acts on* vectors in the basis $\{\mathbf{e}_i\}$ and maps them to vectors in the basis $\{\mathbf{e}'_i\}$. The principal basis of the reference configuration is represented by $\{\mathbf{r}_i\}$, and the principal basis of the deformed configuration is represented by $\{\mathbf{i}_i\}$.

Exercise 3.2.12. AKG 1.16, 1.19, 1.22, 1.26 through 1.30.

3.3 Balance of Energy (First Law)

The first law of thermodynamics expresses the statement of conservation of energy. Specifically, it equates the rate of change of internal and kinetic energies within the body to the rate of heat transfer to the body, and the mechanical power (from tractions and body forces) expended on the body. We define the following energy variables:

- The specific[5] internal energy is $\varepsilon_m(\mathbf{x}, t)$, so the net internal energy is
$$\mathcal{E}_\mathcal{R} = \int_\mathcal{R} \rho\varepsilon_m \, dV$$
- The kinetic energy is
$$\mathcal{K}_\mathcal{R} = \int_\mathcal{R} \frac{1}{2}\rho|\dot{\mathbf{u}}| \, dV$$
- The total heat flow *into* $\mathcal{R}$ is
$$\mathcal{Q}_\mathcal{R} = \int_{\partial\mathcal{R}} -\mathbf{q}\cdot\mathbf{n} \, dA + \int_\mathcal{R} r \, dV,$$
where $\mathbf{q}(\mathbf{x}, t)$ is the heat flux into $\mathcal{R}$ across the boundary $\partial\mathcal{R}$ (the negative sign is used here because $\mathbf{n}$ is the *outward-pointing* normal), and $r(\mathbf{x}, t)$ is a scalar field representing the volumetric heat supply from sources far from $\mathcal{R}$, such as radiative sources.
- The external power *expended on* $\mathcal{R}$ due to tractions and body forces is
$$\mathcal{W}_{\text{ext},\mathcal{R}} = \int_{\partial\mathcal{R}} \mathbf{T}\mathbf{n}\cdot\dot{\mathbf{u}} \, dA + \int_\mathcal{R} \mathbf{b}\cdot\dot{\mathbf{u}} \, dV$$

Key Equation 3.3.1

With the definitions above, the first law of thermodynamics says that globally,

$$\frac{\partial}{\partial t}\mathcal{E}_\mathcal{R} + \frac{\partial}{\partial t}\mathcal{K}_\mathcal{R} = \mathcal{Q}_\mathcal{R} + \mathcal{W}_{\text{ext},\mathcal{R}},$$

[5] meaning *per-unit-mass*

with the local form

$$\rho \frac{\partial \varepsilon_m}{\partial t} = \mathbf{T} \cdot \mathbf{D} - \operatorname{div} \mathbf{q} + r,$$

for

$$\mathbf{D} \equiv \operatorname{sym} \mathbf{L} = \operatorname{sym} \operatorname{grad} \dot{\mathbf{u}}, \qquad D_{ij} = \frac{1}{2}(\dot{u}_{i,j} + \dot{u}_{j,i}),$$

i.e. the *rate of deformation* tensor $\mathbf{D}$ is the symmetric part of the velocity gradient tensor $\mathbf{L} \equiv \nabla \dot{\mathbf{u}}$.

With regard to the referential formulation, it can be shown that

$$\int_{\mathcal{R}} \mathbf{T} \cdot \mathbf{D} \, dV = \int_{\mathcal{R}_R} \mathbf{P} \cdot \dot{\mathbf{F}} \, dV_R = \int_{\mathcal{R}_R} \frac{1}{2} \mathbf{S} \cdot \dot{\mathbf{C}} \, dV_R. \qquad (*)$$

In the reference configuration, we call the pairs $(\mathbf{P}, \mathbf{F})$ and $(\mathbf{S}, (1/2)\mathbf{C})$ **work-conjugate**. Moreover, the quantity expressed by the equation $(*)$ is called the **stress power**.

3.4 Imbalance of Entropy (Second Law)

The second law of thermodynamics says that the entropy in a system is never decreasing; that is, it is either constant or increasing (when a part of the system is *creating* entropy). We define the following entropy variables:

- The specific entropy is $\eta_m(\mathbf{x}, t)$, so the net entropy is

$$\mathcal{S}_{\mathcal{R}} = \int_{\mathcal{R}} \rho \eta_m \, dV$$

- The total entropy flow *into* $\mathcal{R}$ is

$$\mathcal{J}_{\mathcal{R}} = \int_{\partial \mathcal{R}} -\mathbf{j} \cdot \mathbf{n} \, dA + \int_{\mathcal{R}} j \, dV,$$

  where $\mathbf{j}(\mathbf{x}, t)$ is the entropy flux into $\mathcal{R}$ across the boundary $\partial \mathcal{R}$ (the negative sign is used here because $\mathbf{n}$ is the *outward-pointing* normal), and $j(\mathbf{x}, t)$ is a scalar field representing the volumetric entropy supply from sources far from $\mathcal{R}$.

- The absolute temperature $\theta(\mathbf{x}, t) > 0$ is a strictly positive scalar field which relates entropy and heat flow by

$$\mathbf{j} = \frac{\mathbf{q}}{\theta} \qquad \text{and} \qquad j = \frac{r}{\theta}$$

Key Equation 3.4.1

One statement of the second law is that the rate of change of the net entropy $\partial \mathcal{S}_{\mathcal{R}} / \partial t$ must be equal to or greater than the total entropy flow $\mathcal{J}_{\mathcal{R}}$, which leads to the global statement

$$\frac{\partial}{\partial t} \int_{\partial \mathcal{R}} -\mathbf{j} \cdot \mathbf{n} \, dA + \int_{\mathcal{R}} j \, dV \geq \int_{\partial \mathcal{R}} -\frac{\mathbf{q}}{\theta} \cdot \mathbf{n} \, dA + \int_{\mathcal{R}} \frac{r}{\theta} \, dV,$$

with the local form

$$\rho \frac{\partial \eta_m}{\partial t} \geq -\operatorname{div} \left(\frac{\mathbf{q}}{\theta} \right) + \left(\frac{r}{\theta} \right).$$

3.5 Dissipation (Free-Energy Imbalance)

The **specific Helmholtz free energy** $\psi_m(\mathbf{x}, t)$ is a measure of the useful work in a system at a constant temperature θ. It is defined in terms of the specific internal energy ε_m and specific entropy η_m as

$$\psi_m(\mathbf{x}, t) = \varepsilon_m(\mathbf{x}, t) - \theta\eta_m(\mathbf{x}, t).$$

Using this definition, the first and second laws can be combined into a statement of *free-energy imbalance*, which allows us to define the **dissipation**.

Key Equation 3.5.1

On a per-unit-mass basis, an expression of the free-energy imbalance is that

$$\mathcal{D} \equiv \mathbf{T} \cdot \mathbf{D} - \rho\eta_m \frac{\partial\theta}{\partial t} - \frac{1}{\theta}\mathbf{q} \cdot \operatorname{grad}\theta - \rho\frac{\partial\psi_m}{\partial t} \geq 0.$$

In the special case where the temperature field is constant in space and time, i.e. $\theta(\mathbf{x}, t) = \text{const.}$, the influence of thermal factors is negligible and the theory is said to be *mechanical*. In this situation the free-energy imbalance simplifies to

$$\mathcal{D} \equiv \mathbf{T} \cdot \mathbf{D} - \rho\frac{\partial\psi_m}{\partial t} \geq 0.$$

In the referential formulation, the mechanical free-energy imbalance can be rewritten as

$$\mathcal{D}_R = \mathbf{P} \cdot \dot{\mathbf{F}} - \rho_R\frac{\partial\tilde{\psi}_R}{\partial t} = \mathbf{S} \cdot \frac{1}{2}\dot{\mathbf{C}} - \rho_R\frac{\partial\tilde{\psi}_R}{\partial t} \geq 0,$$

where $\tilde{\psi}_R$ is the specific free energy in the reference configuration, i.e. the free energy per unit *reference mass*. The quantity $\rho_R\tilde{\psi}_R$ can be taken together to define ψ_R, the free energy per unit reference volume.

Remark 3.5.2: In the case of *small deformations*, we may replace the rate-of-deformation tensor $\mathbf{D}$ with $\dot{\varepsilon}$, the time derivative of the small strain tensor. Moreover, we may assume that the density ρ is independent of time, so that it may be combined with the per-unit-mass quantities ε_m, η_m, and ψ_m to arrive at *per-unit-volume* quantities $\varepsilon \equiv \rho\varepsilon_m$, $\eta \equiv \rho\eta_m$, and $\psi \equiv \rho\psi_m$.

4 Elasticity

An **elastic** material is taken to have the following properties:

- The stress $\mathbf{P}$ at a reference position[1] $\mathbf{X}$ in the material depends only on the deformation of the material elements in the immediate neighborhood of that position. Namely, the stress only depends on $\mathbf{F}$, which entirely characterizes the local deformation at every position.
- The stress does not depend on the *rate* of deformation. That is, it does not depend on any time derivatives of $\mathbf{F}$.
- The stress at a given time t depends only on the value of $\mathbf{F}$ at that time, and not on the *history* of the deformation.
- The material does not dissipate energy. Specifically, the rate at which external work is done onto the body (i.e., the stress power) exactly equals the rate of increase of stored energy in the material (plus the rate of increase of kinetic energy if the accelerations are non-negligible).

4.1 Finite Elasticity

In this section we are concerned with elasticity for *finite* (cf. *infinitesimal*) deformations as described by the deformation gradient tensor $\mathbf{F}$. From the last of the properties of an elastic material described above, we can express the dissipation inequality instead as an equality,

$$\mathcal{D}_R = \mathbf{P} \cdot \dot{\mathbf{F}} - \frac{\partial \psi_R}{\partial t} = 0.$$

We further assume[2] that the free energy function depends only on $\mathbf{F}$, so that

$$\psi_R = \psi_R(\mathbf{F}).$$

It can be shown that these two results yield the following important relationship between ψ_R, $\mathbf{F}$, and $\mathbf{P}$.

Key Equation 4.1.1

In an elastic material, the stress response is fully determined by the free-energy function and the current value of the deformation gradient,

$$\mathbf{P} = \frac{\partial \psi_R}{\partial \mathbf{F}} \implies \mathbf{T} = \frac{1}{J} \frac{\partial \psi_R}{\partial \mathbf{F}} \mathbf{F}^T$$

Materials that obey this property are also said to be **hyperelastic**.

[1] It will be convenient for us to work in terms of a *referential* description; however, the "conversion rules" between $\mathbf{P}$ and $\mathbf{T}$, and between $\mathbf{X}$ and $\mathbf{x}$, can always be used as long as $\mathbf{F}$ is known.

[2] This assumption is justified by the requirement that the material response can only depend on the current kinematic state of the body, which is fully defined by by $\mathbf{F}$.

Consequently, characterization of an elastic material simply boils down to finding an appropriate free-energy function. We can further restrict possible forms for this free-energy function using physical arguments.

By the principle of *material frame indifference*, the free energy function is a physical quality of a material's deformation and hence cannot change with a rigid motion. Mathematically, $\psi_R(\mathbf{F}) = \psi_R(\mathbf{QF})$ for all orthogonal $\mathbf{Q}$. This turns out to be equivalent to the requirement that $\psi_R(\mathbf{F}) = \hat{\psi}_R(\mathbf{C})$, where $\mathbf{C} = \mathbf{U}^2 = \mathbf{F}^T\mathbf{F}$. In words, this additional requirement says that the free-energy function can only depend on the stretching component of the stretch-rotation decomposition of deformation. Therefore, we can write

$$\mathbf{P} = 2\mathbf{F}\frac{\partial\hat{\psi}_R}{\partial\mathbf{C}} \implies \mathbf{T} = \frac{2}{J}\mathbf{F}\frac{\partial\hat{\psi}_R}{\partial\mathbf{C}}\mathbf{F}^T.$$

This is the most general form of the elastic constitutive relation that we can write which holds for all hyperelastic materials. Any further specialization depends on *material symmetry*, which is a physical attribute of a specific type of material. Fortunately, many everyday materials exhibit some amount of material symmetry, which simplifies the formulation of their free-energy function. Let us first make precise what we mean by material symmetry.

Definition 4.1.2. A **material symmetry transformation** is a rotation of the reference configuration that leaves the pointwise (free-energy, and therefore stress) response to deformation unaltered.

4.1.1 Isotropy

Many engineering materials are *isotropic*, which fortunately simplifies their constitutive treatment.

Definition 4.1.3. An **isotropic** material is a material for which *every* rotation is a material symmetry transformation,

$$\hat{\psi}_R(\mathbf{Q}^T\mathbf{CQ}) = \hat{\psi}_R(\mathbf{C})$$

for all orthogonal $\mathbf{Q}$. A special case of this is the rotation which transforms $\mathbf{C}$ to $\mathbf{B}$, namely $\mathbf{R}$. For isotropic materials, it follows that

$$\hat{\psi}_R(\mathbf{C}) = \hat{\psi}_R(\mathbf{B}).$$

Therefore, for isotropic materials it can be shown that

$$\mathbf{P} = 2\frac{\partial\hat{\psi}_R}{\partial\mathbf{B}}\mathbf{F} \implies \mathbf{T} = \frac{2}{J}\frac{\partial\hat{\psi}_R}{\partial\mathbf{B}}\mathbf{B}.$$

Moreover, $\mathbf{T}$ and $\mathbf{B}$ have the same principal directions $\mathbf{l}_i$.

Next, for isotropic materials, the free-energy function can be shown to only depend on the *invariants* of $\mathbf{B}$ (or equivalently, only on the invariants of $\mathbf{C}$). For the ease of notation we define

$$\psi_i \equiv \frac{\partial\hat{\psi}_R}{\partial I_i}$$

for the invariants[3] I_1, I_2, I_3 of $\mathbf{B}$ or of $\mathbf{C}$. Then, computing $(\partial\hat{\psi}_R/\partial\mathbf{B})$ in terms of invariants gives

$$\mathbf{P} = 2\mathbf{F}\left[I_3\psi_3\mathbf{C}^{-1} + [\psi_1 + I_1\psi_2]\mathbf{I} - \psi_2\mathbf{C}\right]$$
$$\mathbf{T} = 2J\psi_3\mathbf{I} + \frac{2}{J}[\psi_1 + I_1\psi_2]\mathbf{B} - \frac{2}{J}\psi_2\mathbf{B}^2.$$

[3] Recall that $J = \sqrt{I_3}$.

Finally, we can express the free-energy function $\hat{\psi}_R$ in terms of the principal stretches λ_i of $\mathbf{U}$ and $\mathbf{V}$ (recall that the eigenvalues of $\mathbf{B}$ and of $\mathbf{C}$ are λ_i^2),

$$\hat{\psi}_R = \tilde{\psi}_R(\lambda_i).$$

This is useful because it allows us to write the spectral decomposition of the resulting stresses in terms of these principal stretch values. Namely (dropping the decorators above ψ),

Key Equation 4.1.4 (Isotropic elastic material)

For an isotropic elastic material, the stresses are related to the principal stretches by

$$\mathbf{P} = \sum_{i=1}^{3} \left(\frac{\partial \psi_R}{\partial \lambda_i} \right) \mathbf{l}_i \otimes \mathbf{r}_i \qquad \text{(no sum on } i\text{)}$$

$$\mathbf{T} = \sum_{i=1}^{3} \left(\frac{\lambda_i}{J} \frac{\partial \psi_R}{\partial \lambda_i} \right) \mathbf{l}_i \otimes \mathbf{l}_i \qquad \text{(no sum on } i\text{)}$$

It is thus possible to read off the principal Cauchy stresses (the principal values of $\mathbf{T}$) as

$$\tau_i = \left(\frac{\lambda_i}{J} \frac{\partial \psi_R}{\partial \lambda_i} \right) \qquad \text{(no sum on } i\text{)}.$$

If the deformation is pure homogeneous strain, $\mathbf{R} = \mathbf{I}$ and the bases $\{\mathbf{l}_i\}$ and $\{\mathbf{r}_i\}$ coincide, making $\mathbf{P}$ symmetric. In this case the principal values of $\mathbf{P}$ can be read off as well to be

$$\sigma_i = \left(\frac{\partial \psi_R}{\partial \lambda_i} \right).$$

4.1.2 Incompressibility

Definition 4.1.5. An **incompressible** material is one for which *every* deformation preserves volume. That is, for all admissible deformation gradients $\mathbf{F}$,

$$J = \det \mathbf{F} = 1.$$

It follows that for incompressible materials, $\det \mathbf{B} = \det \mathbf{C} = \lambda_1 \lambda_2 \lambda_3 = 1$. The following observation about incompressible materials underlies the formulation of an appropriate free-energy function:

If a material is isotropic and incompressible, the addition of a purely hydrostatic stress to any existing stress state does not alter the internal energy of the material. Equivalently, the stress components associated with a hydrostatic pressure do not contribute to the stress power.

It is important to note that the stress components associated with a hydrostatic pressure do not contribute to the stress power *because they do not cause deformation.* The hydrostatic pressure components *do* alter the material's state of stress, and because of this the expressions for the stress tensors must include an extra hydrostatic term.

Remark 4.1.6: As a simple thought experiment, consider a sphere made of an isotropic, incompressible material. The addition of a hydrostatic state of pressure cannot result in a deviatoric deformation (i.e., for the sphere to deform into a non-spherical shape) by symmetry, and it cannot result in a dilatation or compaction by incompressibility. Therefore the deformation must be identically zero.

For any incompressible material, let $p(\mathbf{x}, t)$ be a scalar field corresponding to an arbitrary hydrostatic pressure. Then

$$\mathbf{P} = -p\mathbf{F}^{-T} + 2\mathbf{F}\frac{\partial \psi_R}{\partial \mathbf{C}} \implies \mathbf{T} = -p\mathbf{I} + \frac{2}{J}\mathbf{F}\frac{\partial \psi_R}{\partial \mathbf{C}}\mathbf{F}^T,$$

where the negative sign is conventional to indicate a compressible pressure. In terms of the invariants of $\mathbf{B}$, we also have

$$\begin{aligned}\mathbf{T} &= -p\mathbf{I} + (2\psi_1 + 2I_1\psi_2)\mathbf{B} - 2\psi_2\mathbf{B}^2 \\ &= -p\mathbf{I} + 2\psi_1\mathbf{B} - 2\psi_2\mathbf{B}^{-1},\end{aligned}$$

where we use the notation $\psi_i \equiv \partial\psi_R/\partial I_i$ as defined above.

Key Equation 4.1.7 (Isotropic, incompressble elastic material)

For an *isotropic, incompressible* material, the stresses are related to the principal stretches by

$$\mathbf{P} = \sum_{i=1}^{3}\left(\frac{\partial \psi_R}{\partial \lambda_i} - \frac{p}{\lambda_i}\right)\mathbf{l}_i \otimes \mathbf{r}_i \qquad \text{(no sum on } i\text{)}$$

$$\mathbf{T} = \sum_{i=1}^{3}\left(\lambda_i\frac{\partial \psi_R}{\partial \lambda_i} - p\right)\mathbf{l}_i \otimes \mathbf{l}_i \qquad \text{(no sum on } i\text{)}$$

It is thus possible to read off the principal Cauchy stresses (the principal values of $\mathbf{T}$) as

$$\tau_i = \left(\lambda_i\frac{\partial \psi_R}{\partial \lambda_i} - p\right) \qquad \text{(no sum on } i\text{)}.$$

If the deformation is pure homogeneous strain, $\mathbf{R} = \mathbf{I}$ and the bases $\{\mathbf{l}_i\}$ and $\{\mathbf{r}_i\}$ coincide, making $\mathbf{P}$ symmetric. In this case the principal engineering stresses (the principal values of $\mathbf{P}$) can be read off as well to be

$$\sigma_i = \left(\frac{\partial \psi_R}{\partial \lambda_i} - \frac{p}{\lambda_i}\right).$$

Moreover, in this case the principal Cauchy stresses τ_i are related to the principal engineering stresses σ_i by

$$\tau_i = \lambda_i\sigma_i.$$

The existence of this "arbitrary pressure field" has two important physical interpretations.

In an incompressible, isotropic material:

1. **An undeformed body may still be experiencing a hydrostatic state of stress.**

2. **In a deformed body, the stress is permitted to have an arbitrary extra pressure whose value is independent of the deformation.**

In all conditions, *traction boundary conditions* are required to determine the pressure p.

4.2 Free-Energy Functions for Incompressible Materials with Finite Strain

This section compiles a few typical examples of specialized free-energy functions for incompressible materials capable of attaining large, finite strains, typically elastomers. The free energy functions are most often specialized in terms of the principal stretches $(\lambda_1, \lambda_2, \lambda_3)$, with the incompressibility constraint $\lambda_1\lambda_2\lambda_3 = 1$ for all deformations. Alternatively, they may be specified in terms of the principal invariants I_1, I_2, I_3 of $\mathbf{B}$ or equivalently of $\mathbf{C}$,

$$\begin{aligned} I_1 &= \lambda_1^2 + \lambda_2^2 + \lambda_3^2 \\ I_2 &= \lambda_2^2\lambda_3^2 + \lambda_3^2\lambda_1^2 + \lambda_1^2\lambda_2^2 = \lambda_1^{-2} + \lambda_2^{-2} + \lambda_3^{-2} \\ I_3 &= \lambda_1^2\lambda_2^2\lambda_3^2 = 1. \end{aligned}$$

1. The **Neo-Hookean** free-energy function has one parameter, the shear modulus $\mu_0 > 0$:
$$\psi_R = \frac{\mu_0}{2}(I_1 - 3),$$
and the principal values of the Cauchy stress are
$$\tau_i = \mu_0\lambda_i^2 - p.$$

2. The **Mooney-Rivlin** free-energy function has two parameters, $C_1 > 0$ and $C_2 > 0$:
$$\psi_R = \frac{C_1}{2}(I_1 - 3) + \frac{C_2}{2}(I_2 - 3),$$
and the principal values of the Cauchy stress are
$$\tau_i = C_1\lambda_i^2 - C_2\lambda_i^{-2} - p.$$

3. The **Ogden** free-energy function generalizes the two-term Mooney-Rivlin model to contain M terms. It takes $2M$ parameters of the form μ_r and α_r, for $r = 1, 2, 3, ..., M$:
$$\psi_R = \sum_{r=1}^{M} \frac{\mu_r}{\alpha_r}(\lambda_1^{\alpha_r} + \lambda_2^{\alpha_r} + \lambda_2^{\alpha_r} - 3)$$
with the constraints that (1) for each r, $\mu_r\alpha_r > 0$ (no sum over r), and (2)
$$\sum_{r=1}^{M} \mu_r\alpha_r = 2\mu_0$$
where μ_0 is the "ground state" shear modulus (i.e., the shear modulus of the material in the undeformed stress-free configuration). The principal values of the Cauchy stress are
$$\tau_i = \sum_{r=1}^{M} \mu_r\lambda_i^{\alpha_r} - p.$$

4. The **Gent** free-energy function incorporates the physical concept of a *locking stretch*, which represents a finite value of I_1 (and consequently a finite value for the λ_i) where the polymer chains are all extended to their limit, and consequently a further stretch is accompanied by a dramatic increase in stress. (The previous models fail to account for the locking stretch phenomenon and as such do not fit well to experimental data at large stretch values.) The Gent model takes two parameters, the shear modulus $\mu_0 > 0$ and the stiffening parameter $I_m > 0$, which limits[4] the maximum possible value of I_1:

$$\psi_R = -\frac{\mu_0}{2} I_m \ln\left(1 - \frac{I_1 - 3}{I_m}\right)$$

Under the Gent model, the Cauchy stress is given by

$$\mathbf{T} = -p\mathbf{I} + \left[\mu_0 \left(\frac{I_m}{I_m - (I_1 - 3)}\right)\right] \mathbf{B}.$$

5. The **Arruda-Boyce** free-energy function also models the locking stretch phenomenon as motivated from statistical mechanics of polymer chain entropy. This model takes two parameters, μ_0 and λ_L (the *network locking stretch*):

$$\psi_R = \mu_0 \lambda_L^2 \left[\left(\frac{\bar{\lambda}}{\lambda_L}\right)\beta + \ln\left(\frac{\beta}{\sinh\beta}\right)\right]$$

with

$$\beta = \mathcal{L}^{-1}\left(\frac{\bar{\lambda}}{\lambda_L}\right), \qquad \mathcal{L}(u) \equiv \coth(u) - \frac{1}{u} \approx \frac{3u - u^3}{1 - u^2}$$

and $\bar{\lambda}$ the *root-mean-square average stretch*:

$$\bar{\lambda} \equiv \sqrt{\frac{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}{3}} = \sqrt{\frac{I_1}{3}}.$$

Under the Arruda-Boyce model, the Cauchy stress is given by

$$\mathbf{T} = -p\mathbf{I} + \left[\mu_0 \left(\frac{\lambda_L}{3\bar{\lambda}}\right) \mathcal{L}^{-1}\left(\frac{\bar{\lambda}}{\lambda_L}\right)\right] \mathbf{B}.$$

Remark 4.2.1: In the Gent and Arruda-Boyce models, the Cauchy stress may be rewritten as

$$\mathbf{T} = -p\mathbf{I} + \mu\mathbf{B},$$

where μ represents a *generalized shear modulus*, which in each case is a function of the principal stretches in such a way that the stress grows without bound as the stretch increases, reflecting the chain locking phenomenon. If we take $\mu = \mu_0 = \text{const.}$, we recover the Neo-Hookean model, which does not account for chain-locking.

Many other free-energy functions have been formulated on the basis of physical arguments and empirical data; the reader is encouraged to consult the vast body of literature in this regard.

[4] Specifically, $I_1 < I_m + 3$. As $I_1 \to I_m + 3$, the stress increases without bound, which models the dramatic stiffening seen in experiments.

Exercise 4.2.2. AKG 6.1 through 6.5, 6.8 through 6.11.

Exercise 4.2.3. Consider again the *simple shear* deformation; recall the deformation gradient is

$$\mathbf{F} = \mathbf{I} + \gamma \mathbf{e}_1 \otimes \mathbf{e}_2,$$

where $\gamma > 0$ is a constant. **Compute** the tensor $\mathbf{B}$ for this deformation and **list** the principal invariants of $\mathbf{B}$. Given the form of $\mathbf{B}$, one would likely expect that a nonzero shear stress T_{12} is required to sustain this deformation. If the material is made of an incompressible Mooney-Rivlin material, **show that** two nonzero normal stress components are required *in addition to the* T_{12} *shear stress component* in order to sustain the deformation. If the material is made of an incompressible Neo-Hookean material, **show that** only one nonzero normal stress is required. **What is** the difference between the *mathematical* formulation of the two material models?

4.3 Linear Elasticity

In the special case where the deformation is close to the identity,

$$\mathbf{F} \approx \mathbf{I} \implies |\mathbf{H}| \ll 1,$$

the form of the free-energy function may be simplified. In this situation, the tensors governing deformation and stress may be *linearized*, i.e., they may be Taylor expanded and terms with order higher than one may be neglected. Following linearization,

- the reference and deformed configurations are sufficiently similar that we no longer need to refer to them separately, and specifically $\partial(\cdot)/\partial x_i = \partial(\cdot)/\partial X_i$;
- the first Piola stress $\mathbf{P}$ and the Cauchy stress $\mathbf{T}$ are *identical*, and we can denote this common *stress* by the symmetric tensor $\boldsymbol{\sigma}$; and
- the rate of deformation tensor $\mathbf{D}$ is equal to the time derivative of the small strain tensor, $\dot{\boldsymbol{\varepsilon}}$.

Thus, the dissipation equation, which is the combined statement of the first and second laws of thermodynamics in an isothermal case, can be written as

$$\mathcal{D} \equiv \boldsymbol{\sigma} \cdot \dot{\boldsymbol{\varepsilon}} - \frac{\partial \psi}{\partial t} = 0,$$

so for a scalar free energy function of the assumed form $\psi = \psi(\boldsymbol{\varepsilon})$ we have

$$\boldsymbol{\sigma} = \frac{\partial \psi(\boldsymbol{\varepsilon})}{\partial \boldsymbol{\varepsilon}}.$$

To write down a particular form for $\psi(\boldsymbol{\varepsilon})$, we again turn to the Taylor expansion of ψ about $\boldsymbol{\varepsilon} = \mathbf{0}$. In the linear theory, it is sufficient to keep the first non-zero term of this expansion. We assume that when $\boldsymbol{\varepsilon} = \mathbf{0}$, the body is stress-free, and without loss of generality we set the scale for ψ to be zero in this state as well. Then, the first non-zero term in the expansion is the *quadratic* term. We arrive at the following form of the free-energy function:

Key Equation 4.3.1

When the deformation is infinitesimal,

$$\sigma = \frac{\partial \psi(\boldsymbol{\varepsilon})}{\partial \boldsymbol{\varepsilon}},$$

with the first non-zero Taylor expansion term yielding

$$\psi = \frac{1}{2} \frac{\partial^2 \psi}{\partial \varepsilon_{pq} \partial \varepsilon_{rs}}\bigg|_{\boldsymbol{\varepsilon}=\mathbf{0}} \varepsilon_{pq}\varepsilon_{rs} = \frac{1}{2} C_{pqrs}\varepsilon_{pq}\varepsilon_{rs} = \frac{1}{2}\boldsymbol{\varepsilon} \cdot \mathbb{C}\boldsymbol{\varepsilon}$$

where $\mathbb{C}$ is the symmetric, positive definite fourth-order tensor that maps the second-order tensor $\boldsymbol{\varepsilon}$ to the second-order tensor $\boldsymbol{\sigma}$, i.e.

$$\boldsymbol{\sigma} = \mathbb{C}\boldsymbol{\varepsilon} \qquad \sigma_{ij} = C_{ijkl}\varepsilon_{kl}$$

The **elasticity tensor** $\mathbb{C}$ is sufficient to characterize the stress field given the infinitesimal strain field associated with a small deformation.

Remark 4.3.2: It appears that there are $3^4 = 81$ entries in the tensor C_{ijkl}. However, by the symmetry of the second partial derivative from which it is defined, $\mathbb{C}$ has the *major symmetry* $C_{pqrs} = C_{rspq}$. Moreover, because σ_{ij} and ε_{kl} are both symmetric themselves, $\mathbb{C}$ has the *minor symmetries* $C_{pqrs} = C_{qprs}$ and $C_{pqrs} = C_{pqsr}$. Taken together, there are at most 21 independent components of $\mathbb{C}$. This corresponds to an arbitrarily anisotropic linear elastic material.

The linear mapping $\mathbb{C}$ is a bijection, so it has an inverse, called the **compliance tensor** which is denoted by $\mathbb{S}$, for which

$$\boldsymbol{\varepsilon} = \mathbb{S}\boldsymbol{\sigma}; \qquad \psi = \frac{1}{2}\boldsymbol{\sigma} \cdot \mathbb{S}\boldsymbol{\sigma}.$$

The compliance tensor is also positive definite and possesses the same symmetries as $\mathbb{C}$.

4.3.1 Elasticity Tensor with Material Symmetry

In the most general case of complete anisotropy, there are 21 independent components of $\mathbb{C}$ (see Remark 4.3.2). However, many real materials exhibit some amount of material symmetry, which allow the number of independent constants to be reduced. For linearly elastic materials, the definition of a material symmetry transformation *specializes* in the following manner:

Definition 4.3.3. For *linearly elastic materials* governed by the elasticity tensor $\mathbb{C}$, an orthogonal tensor $\mathbf{Q}$ represents a **material symmetry transformation** if and only if

$$C_{ijkl} = Q_{ip}Q_{jp}Q_{kr}Q_{ls}C_{pqrs}.$$

In words, this says that if the components of $\mathbb{C}$ do not change after the basis change associated with a rotation $\mathbf{Q}$, then $\mathbf{Q}$ represents a material symmetry transformation. Different types of symmetry are characterized by the tensors $\mathbf{Q}$ (which may represent reflections or rotations) which are material symmetry transformations for a given unit cell.

Example 4.3.4 (Cubic material symmetry)

The following $\mathbf{Q}$ are material symmetry transformations for a material with *cubic symmetry*: the reflections

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and the rotations

$$\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}.$$

In words, a material has cubic symmetry if it has three orthogonal planes of reflection symmetry, and three axes of 90°-rotation symmetry. As a result, the number of independent constants falls to just three,

$$[\mathbb{C}] = \begin{bmatrix} C_{1111} & C_{1122} & C_{1122} & 0 & 0 & 0 \\ C_{1122} & C_{1111} & C_{1122} & 0 & 0 & 0 \\ C_{1122} & C_{1122} & C_{1111} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{1212} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{1212} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{1212} \end{bmatrix}$$

If the material is **isotropic**, then *any* orthogonal $\mathbf{Q}$ is a material symmetry transformation and the constitutive relation has the following special form:

Key Equation 4.3.5

For an isotropic linear elastic material,

$$\boldsymbol{\sigma} = \mathbb{C}\boldsymbol{\varepsilon} = 2\mu\boldsymbol{\varepsilon} + \lambda(\operatorname{tr}\boldsymbol{\varepsilon})\mathbf{I}$$

with μ and λ the only two independent material constants, which are generally called the *elastic moduli*. The positive-definiteness of $\mathbb{C}$ requires that

$$\mu > 0, \qquad \kappa \equiv \lambda + \frac{2}{3}\mu > 0.$$

The following alternative form of the constitutive relation in terms of the deviatoric strain $\boldsymbol{\varepsilon}' \equiv \boldsymbol{\varepsilon} - (1/3)(\operatorname{tr}\boldsymbol{\varepsilon})\mathbf{I}$ is also useful:

$$\boldsymbol{\sigma} = 2\mu\boldsymbol{\varepsilon}' + \kappa(\operatorname{tr}\boldsymbol{\varepsilon})\mathbf{I}$$

Finally, the following inverted form gives the strain in terms of a known stress state:

$$\boldsymbol{\varepsilon} = \frac{1}{2\mu}\boldsymbol{\sigma}' + \frac{1}{9\kappa}(\operatorname{tr}\boldsymbol{\sigma})\mathbf{I}$$

In particular, μ is called the **shear modulus** and κ is called the **bulk modulus**.

Exercise 4.3.6. Starting with the isotropic linear elastic constitutive relation

$$\boldsymbol{\sigma} = 2\mu\boldsymbol{\varepsilon}' + \kappa(\operatorname{tr}\boldsymbol{\varepsilon})\mathbf{I},$$

derive the inverted form

$$\boldsymbol{\varepsilon} = \frac{1}{2\mu}\boldsymbol{\sigma}' + \frac{1}{9\kappa}(\operatorname{tr}\boldsymbol{\sigma})\mathbf{I}.$$

Remark 4.3.7: The definition of the shear modulus is motivated by the *simple shear test*, for which $\boldsymbol{\varepsilon} = \tau(\mathbf{e}_1 \otimes \mathbf{e}_2 + \mathbf{e}_2 \otimes \mathbf{e}_1)$ is prescribed and correspondingly $\boldsymbol{\sigma} = (\gamma/2)(\mathbf{e}_1 \otimes \mathbf{e}_2 + \mathbf{e}_2 \otimes \mathbf{e}_1)$ is measured. In this case $\mu \equiv \tau/\gamma$.

Similarly the definition of the bulk modulus is motivated by the (slightly-harder-to-perform) *uniform compaction test*, for which $\boldsymbol{\varepsilon} = (-\Delta/3)\mathbf{I}$ is prescribed and correspondingly $\boldsymbol{\sigma} = -p\mathbf{I}$ is measured. Then $\kappa \equiv p/\Delta$.

The stress state $\boldsymbol{\sigma} = \sigma\mathbf{e}_1 \otimes \mathbf{e}_1$ in a *simple tension test* for which the corresponding strain state is $\boldsymbol{\varepsilon} = \varepsilon\mathbf{e}_1 \otimes \mathbf{e}_1 + l(\mathbf{e}_2 \otimes \mathbf{e}_2 + \mathbf{e}_3 \otimes \mathbf{e}_3)$ motivates the definition of the **elastic modulus** E and the **Poisson ratio** ν such that for this test,

$$E = \frac{\sigma}{\varepsilon} \qquad \text{and} \qquad \nu = -\frac{l}{\varepsilon}.$$

Namely, in terms of the shear modulus and bulk modulus,

$$E \equiv \frac{9\kappa\mu}{3\kappa + \mu} > 0, \qquad -1 < \nu \equiv \frac{3\kappa - 2\mu}{6\kappa + 2\mu} < \frac{1}{2}$$

or inversely

$$\mu = \frac{E}{2(1+\nu)}, \qquad \kappa = \frac{E}{3(1-2\nu)}.$$

With these definitions we can rewrite the isotropic linear elastic constitutive relation in another common form:

Key Equation 4.3.8

For an isotropic linear elastic material,

$$\boldsymbol{\sigma} = \frac{E}{1+\nu}\left[\boldsymbol{\varepsilon} + \frac{\nu}{1-2\nu}(\operatorname{tr}\boldsymbol{\varepsilon})\mathbf{I}\right]$$

with the inverse relationship

$$\boldsymbol{\varepsilon} = \frac{1}{E}\left[(1+\nu)\boldsymbol{\sigma} - \nu(\operatorname{tr}\boldsymbol{\sigma})\mathbf{I}\right].$$

Remark 4.3.9: The special case of an *incompressible* linearly elastic material corresponds to

$$\kappa \to \infty, \qquad \nu \to \frac{1}{2}.$$

The practical significance of this formulation is that materials can be *modeled* as *nearly-incompressible* whenever they are *relatively* easy to distort compared to shape-

change; mathematically, this occurs when the ratio of material properties $\mu/\kappa \ll 1$. For example, rubber has $\mu/\kappa \approx 10^{-4}$.

Exercise 4.3.10. AKG 2.1, 2.5, 2.19 through 2.31.

4.3.2 Remarks on the Physical Basis of Rubber Elasticity

As discussed above, rubber is commonly modeled as an incompressible material due to the ratio of its shear modulus to its bulk modulus being much smaller than unity. Physically, rubber consists of a *network* of polymer chains: individual chains of covalent bonds linking individual backbone units, called monomers, together. These polymer chains are both crosslinked (meaning that occasional covalent bonds form *across* chains) and physically engangled; moreover, nearby monomers experience van der Waals interactions. This complex network of polymer chains is commonly described by the analogy of a bowl of cooked spaghetti.

In the case of rubber, these polymer chains are indeed so tightly volumetrically packed by nature of their flexibility that it is quite difficult to further compact them. However, *relative* to the effort required to compact them, it is very easy to force chains to glide against one another, which is exactly what happens in shear. For this reason, we generally observe that in rubbers, $\mu \ll \kappa$. This is thus the physical argument for the incompressibility assumption.

Another curious phenomenon is that the elasticity in rubber chains is due to *entropy*, in stark contrast to the elasticity in a metal which is due to bond stretching[5]. Specifically, in the absence of external forces, the natural equilibrium state of a rubber specimen is that which aims to maximize the total entropy of a system, and perturbing this specimen away from this state (e.g., by pushing, pulling, or applying another external force) causes a decrease in entropy. The desire of the system to "re-maximize" its entropy gives rise to an elastic restoring force, which we observe as an apparent stiffness.

Being due to an entropic effect, the shear modulus of a rubber is therefore also dependent on the ambient temperature. In general,

$$\mu = \hat{\mu}(T) \propto N k_B T,$$

where N is the average number of chains per reference volume, k_B is Boltzmann's constant, and T is the temperature in absolute units. This is why rubber components are often *stiffer* on a hot day compared to a cold day! (Recall that the modulus of metallic or ceramic materials typically *decreases* with increasing temperature.)

Example 4.3.11

Consider the following simple experiment: a 1 kg weight is placed atop a rubber pad at an ambient temperature of 25°C. Suppose the temperature of the setup is uniformly increased to 60°C. We would then observe the weight moving *upwards* from its initial position, because the rubber would *stiffen* and therefore displace less from its original amount.

[5] See chapter 22 of AKG for a more detailed treatment of the entropic elasticity phenomenon.

4.3.3 Two-Dimensional Problems

Long Cylindrical Bodies: Plane Strain

In the case of a long cylindrical[6] body for which one length dimension (without loss of generality, in $\mathbf{e}_3$) is much much larger than the other two length dimensions (in $\mathbf{e}_1$ and $\mathbf{e}_2$), commonly admits a **plane strain** deformation in which the displacement field in the long dimension is *negligible* compared to the other two. Thus,

$$\mathbf{u}(\mathbf{x}) = (u_1, u_2, 0) \implies \varepsilon_{13} = \varepsilon_{23} = \varepsilon_{33} = 0$$

and if the material is isotropic,

$$\sigma_{\alpha\beta} = \frac{E}{1+\nu}\left(\varepsilon_{\alpha\beta} + \frac{\nu}{1-2\nu}\varepsilon_{\gamma\gamma}\delta_{\alpha\beta}\right), \qquad \varepsilon_{\alpha\beta} = \frac{1}{2}(u_{\alpha,\beta} + u_{\beta,\alpha}) = \frac{1+\nu}{E}(\sigma_{\alpha\beta} - \nu\sigma_{\gamma\gamma}\delta_{\alpha\beta})$$

where the Greek subscripts are only allowed to be 1 and 2 (and hence $\varepsilon_{\gamma\gamma} = \varepsilon_{11} + \varepsilon_{22}$). The stress components in the long direction is

$$\sigma_{13} = \sigma_{23} = 0, \sigma_{33} = \nu(\sigma_{11} + \sigma_{22}).$$

The stress and strain tensors have at most the following nonzero components:

$$[\boldsymbol{\sigma}] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & 0 \\ \sigma_{12} & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix}, \qquad [\boldsymbol{\varepsilon}] = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & 0 \\ \varepsilon_{12} & \varepsilon_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thin Cylindrical Bodies: Plane Stress

In the case of a short cylindrical body, with one length dimension (again assume this to be $\mathbf{e}_3$) much much *shorter* than the other two, it may happen that the traction components pointing in the lateral directions are negligible compared to the in-plane tractions. This situation is called **plane stress**. In this special case, the stress tensor has zero components in the direction 3, i.e. $\sigma_{13} = \sigma_{23} = \sigma_{33} = 0$.

If the material is isotropic,

$$\sigma_{\alpha\beta} = \frac{E}{1+\nu}\left(\varepsilon_{\alpha\beta} + \frac{\nu}{1-\nu}\varepsilon_{\gamma\gamma}\delta_{\alpha\beta}\right), \qquad \varepsilon_{\alpha\beta} = \frac{1+\nu}{E}\left(\sigma_{\alpha\beta} - \frac{\nu}{1+\nu}\sigma_{\gamma\gamma}\delta_{\alpha\beta}\right).$$

The strain components in the long direction are

$$\varepsilon_{13} = \varepsilon_{23} = 0, \varepsilon_{33} = \frac{-\nu}{1-\nu}(\varepsilon_{11} + \varepsilon_{22}).$$

The stress and strain tensors have at most the following nonzero components:

$$[\boldsymbol{\sigma}] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & 0 \\ \sigma_{12} & \sigma_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad [\boldsymbol{\varepsilon}] = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & 0 \\ \varepsilon_{12} & \varepsilon_{22} & 0 \\ 0 & 0 & \varepsilon_{33} \end{bmatrix}.$$

In both plane strain and in plane stress, the (two) equations of equilibrium read

$$\sigma_{\alpha\beta,\beta} + b_\alpha = 0,$$

for a two-dimensional body force vector $\mathbf{b} = (b_1, b_2, 0)$.

[6] A shape is cylindrical if it can be created by the extrusion of a 2D profile into the third dimension. The profile, which becomes the cross-section of the body, need not be circular.

Exercise 4.3.12. AKG 2.6 through 2.8.

4.4 Mixed Problem of Elastostatics

When the accelerations of a body are negligible, the balance of forces and moments reduces to the equilibrium equations. In this case, the **mixed problem of elastostatics** can be stated in the following manner. Essentially, this is the broadest form that will generalize all specific problems in elasticity.

Given a body filling a region $\mathcal{R}$ with boundary $\partial\mathcal{R}$ having:

- complementary subsurfaces $\mathcal{S}_1$ and $\mathcal{S}_2$ such that $\mathcal{S}_1 \cap \mathcal{S}_2 = 0$ and $\mathcal{S}_1 \cup \mathcal{S}_2 = \partial\mathcal{R}$;
- a free-energy function ψ_R, which can describe linear or non-linear behavior;
- a body force distribution $\mathbf{b}(\mathbf{X})$;
- and boundary conditions

$$\begin{cases} \mathbf{u} & = \hat{\mathbf{u}} \text{ on } \mathcal{S}_1, \\ \mathbf{T}\mathbf{n} & = \hat{\mathbf{t}} \text{ on } \mathcal{S}_2, \end{cases}$$

  where the "hat" vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{t}}$ represent *prescribed* displacements and tractions, respectively;

we wish to find:

- a displacement field $\mathbf{u}$;
- a deformation field $\mathbf{F}$, which may be specialized to $\boldsymbol{\varepsilon}$, or expressed in terms of another tensor with which $\mathbf{F}$ has a one-to-one relationship;
- and a stress field $\mathbf{T}$ (or $\mathbf{S}$, etc.), which may be specialized to $\boldsymbol{\sigma}$, over $\mathcal{R}$;

that satisfy the field equations. In the case of the linear theory, the field equations have the form

$$\begin{cases} \boldsymbol{\varepsilon} & = \frac{1}{2}\left(\nabla\mathbf{u} + (\nabla\mathbf{u})^T\right) \\ \boldsymbol{\sigma} & = \mathbb{C}\boldsymbol{\varepsilon} \\ \operatorname{div}\boldsymbol{\sigma} + \mathbf{b} & = \mathbf{0} \end{cases} .$$

It can be shown that the solution, i.e. the set $\{\mathbf{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}\}$, is *unique*, except in the case where no displacement boundary conditions are specified. In that special case any two solutions can differ at most by a rigid displacement.

Moreover, in the linear theory, linearity can be exploited in the form of *superposition*. This means that

for linear problems, the solutions to some simple problems can be combined to generate solutions to more complicated problems.

The linear superposition principle is stated as follows:

Key Equation 4.4.1 (Linear superposition principle)

Suppose the set $\{\mathbf{u}_1, \boldsymbol{\varepsilon}_1, \boldsymbol{\sigma}_1\}$ is a solution to the mixed problem of elastostatics corresponding to a body force $\mathbf{b}_1$, prescribed displacements $\hat{\mathbf{u}}_1$ on $\mathcal{S}_1$, and prescribed tractions $\hat{\mathbf{t}}_1$ on $\mathcal{S}_2$, and suppose the set $\{\mathbf{u}_2, \boldsymbol{\varepsilon}_2, \boldsymbol{\sigma}_2\}$ is a solution to the mixed problem of elastostatics corresponding to a body force $\mathbf{b}_2$, prescribed displacements

$\hat{\mathbf{u}}_2$ on $\mathcal{S}_1$, and prescribed tractions $\hat{\mathbf{t}}_2$ on $\mathcal{S}_2$.

Then, the set $\{\mathbf{u} \equiv \mathbf{u}_1 + \mathbf{u}_2, \boldsymbol{\varepsilon} \equiv \boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2, \boldsymbol{\sigma} \equiv \boldsymbol{\sigma}_1 + \boldsymbol{\sigma}_2\}$ is a solution of the problem with body force $\mathbf{b} \equiv \mathbf{b}_1 + \mathbf{b}_2$ and boundary conditions

$$\begin{cases} \mathbf{u} & = \hat{\mathbf{u}}_1 + \hat{\mathbf{u}}_2 \text{ on } \mathcal{S}_1, \\ \mathbf{Tn} & = \hat{\mathbf{t}}_1 + \hat{\mathbf{t}}_2 \text{ on } \mathcal{S}_2. \end{cases}$$

4.5 Elastic Wave Propagation

In this special section we will look at one of very few cases in solid mechanics where the acceleration is *non-zero* (or non-negligible)[7]. Specifically we are concerned with the propagation of *elastic* waves, which are often *pressure waves* in elastic bodies. In all cases we will assume that the amplitudes of displacement are *small*, so that the linear elastic relations apply. The vector equation of momentum balance with a non-zero acceleration term reads

$$\operatorname{div} \boldsymbol{\sigma} = \rho \ddot{\mathbf{u}},$$

and in particular we are interested in displacement solutions of the form

$$\mathbf{u}(\mathbf{x}, t) = \mathbf{a} \sin\left[\frac{2\pi}{\lambda}(\mathbf{x} \cdot \mathbf{n} - ct)\right],$$

where $\mathbf{a}$ is the displacement amplitude vector, c is the wave speed, and λ is the wavelength.

Definition 4.5.1. If the body is *long* in one dimension such that $\mathbf{u}(\mathbf{x}, t) = u_1(x, t)$ on a region $x > 0$, such that the only nonzero stress component is $\sigma_{11} = E\varepsilon_{11}$, and the nonzero strain components are ε_{11} and $\varepsilon_{22} = \varepsilon_{33} = -\nu\varepsilon_{11}$, then the solution is (approximately) called a **bar wave** and the governing equation simplifies to

$$\frac{\partial^2 u_1}{\partial x_1^2} = \frac{1}{c_B^2}\frac{\partial^2 u_1}{\partial t_1}, \qquad c_B = \sqrt{\frac{E}{\rho}}.$$

Definition 4.5.2. If the body is infinitely wide in two dimensions (a *half-space*) subject to a time-varying but spatially uniform pressure loading, the displacement field is again $\mathbf{u}(\mathbf{x}, t) = u_1(x_1, t)$, but there are three nonzero stress components, namely σ_{11}, σ_{22}, and σ_{33}. Here $\sigma_{22} = \sigma_{33} = \nu/(1-\nu)\sigma_{11}$. The governing equation for such a **longitudinal wave** is

$$\frac{\partial^2 u_1}{\partial x_1^2} = \frac{1}{c_L^2}\frac{\partial^2 u_1}{\partial t_1}, \qquad c_L = \sqrt{\frac{K + \frac{4}{3}G}{\rho}}.$$

Remark 4.5.3: In seismology, longitudinal earthquake waves are called "P waves". The crust of the earth is essentially an infinitely-wide half-space and thus supports P waves and shear waves (introduced below; also called "S waves"). Because P waves travel faster than S waves, the signal from a P wave is often the first notification of an earthquake. To this end, in seismology the quantity $K + \frac{4}{3}G$ is sometimes called the "P-wave modulus" and denoted M. (We will not use this terminology here.)

[7] Another common case of non-zero acceleration occurs for *bodies spinning at a constant rate*, where the centripetal acceleration must be accounted for in the equilibrium equation

Definition 4.5.4. If the body is infinitely wide in two dimensions (a *half-space*) subject to a time-varying but spatially uniform shear loading (say, parallel to the $\mathbf{e}_3$ direction), the displacement field is $\mathbf{u}(\mathbf{x}, t) = u_3(x_1, t)$, and the only nonzero stress component is σ_{31}. The governing equation for such a **shear wave** is

$$\frac{\partial^2 u_3}{\partial x_1^2} = \frac{1}{c_S^2}\frac{\partial^2 u_3}{\partial t_1}, \qquad c_S = \sqrt{\frac{G}{\rho}}.$$

In each of the above cases, there are two regions in the elastic body, separated by $x^* = c^* t$, where c^* is the appropriate wave speed and x^* is the direction of wave propagation. For $x^* < c^* t$ the displacement is proportional to the *area under the pressure-time or shear-time curve* integrated from $t = 0$ to $t = t - x^*/c^*$. For $x^* > c^* t$ the displacement is zero, meaning the material there has not yet "felt" the elastic wave. As a concrete example, in the case of a bar wave the critical distance x^* at a time t is given by $x_1 = c_B t$, and the displacement at a point $x < c_B t$, at a time t, is given by the integral of the pressure curve $p(\tau)$ on the interval $\tau \in [0, t - x/c_B]$.

Exercise 4.5.5. AKG 2.32 through 2.39.

5 One-Dimensional Linear Viscoelasticity

Recall that an elastic material, as modeled in the previous chapter, (i) experiences a response independent of the deformation rate; (ii) returns instantaneously to the undeformed configuration when applied tractions are removed; and (iii) never dissipates energy. In this section we examine the response of materials that are not purely elastic, but rather possess some of the tendencies of viscous materials. For example, the mechanical response of many polymeric materials can be described using the models developed in this chapter.

In particular, the constitutive relations we will examine in this chapter are called **viscoelastic**. In order to develop the conceptual aspects of the theory without burdensome math, in this chapter we restrict attention to a *one-dimensional* development of the constitutive relation. We will also restrict ourselves to *linear* viscoelasticity, which is sufficient to describe the small-strain behavior of these materials.

In one dimension, the linear elastic constitutive relation reduces to

$$\sigma(t) = E\varepsilon(t),$$

where we have re-introduced a time parameter t. By definition, a viscoelastic material exhibits certain characteristics of such elastic materials. However, a viscoelastic material also exhibits certain characteristics of *linearly viscous* (also called *Newtonian*) materials, for which the stress response depends linearly on the *rate* of strain:

$$\sigma(t) = \eta\dot{\varepsilon}(t),$$

where we call the parameter $\eta > 0$ the **viscosity**.

Phenomenologically, viscoelastic materials:

- experience a stress response that depends on the rate of applied strain;
- fully return to their undeformed configurations some time after the deformation is applied, then removed ("unloaded");
- dissipate a nonzero amount of energy;
- experience **stress relaxation**, where under a constant applied strain, the stress in the material decreases over time;[1]
- experinece **creep**, where under a constant applied stress, the strain in the material increases over time;[2]

Remark 5.0.1: The physical origin of dissipation in viscoelastic materials is a result of internal "drag" during the deformation. For example, an interstitial fluid (e.g., in a foam) or a nonzero amount of chain drag (e.g., in rubbers) contributes to an overall "drag".

Importantly,

[1] An example of stress relaxation occurs in guitar strings, which tend to de-tune (go flat) over time.
[2] An example of creep occurs in *Silly Putty*, which tends to change shape under its own weight.

The *linear* models discussed in this chapter are applicable for *small strains only*; namely, within the range $0\% \leq \varepsilon \leq 5\%$.

5.1 The Stress-Relaxation and Creep Experiments

The canonical experiments which will be used to illustrate the response of viscoelastic material models are that of *stress-relaxation* and *creep*.

Definition 5.1.1. The **Heaviside function** $h(t)$ takes one parameter, the time t, and is defined such that

$$h(t) = \begin{cases} 0, & t \leq 0; \\ 1, & t > 0 \end{cases}.$$

It is like a binary "off-on" switch. The time derivative of the Heaviside function dh/dt is equivalent to the *Dirac delta function*:

Definition 5.1.2. The **Dirac delta function** $\delta(t)$ takes one parameter, the time t, and is defined such that

$$\frac{dh(t)}{dt} = \delta(t) = \begin{cases} 0, & t \neq 0; \\ \infty, & t = 0 \end{cases}, \qquad \int_{-\infty}^{\infty} \delta(t)\, dt = 1$$

Exercise 5.1.3. Sketch a graph of the function $y = h(t-2)$. **Describe** the effect of the parameter τ in the generalized Heaviside function $y = h(t-\tau)$.

5.1.1 Stress-Relaxation

In the **stress-relaxation experiment**, we consider the effect of applying to a one-dimensional body the strain function

$$\varepsilon(t) = \varepsilon_0 h(t).$$

In words, the body experiences no strain for all $t \leq 0$, and the body experiences a constant strain ε_0 for all $t > 0$. Now, we would expect a purely elastic material to have the stress response

$$\sigma(t) = E\varepsilon_0 h(t) \qquad \text{(elastic)}$$

and a purely viscous[3] material to have the stress response

$$\sigma(t) = \eta\varepsilon_0 \delta(t) \qquad \text{(viscous)}.$$

A viscoelastic material behaves somewhere between these two models. In particular, the stress spikes at $t = 0^+$ to a nonzero value, then falls exponentially to a long-term value as $t \to \infty$, where it is constant in the limit. To characterize this behavior, we normalize the time-dependent stress response $\sigma(t)$ we normalize by the input strain:

Definition 5.1.4. In the stress-relaxation experiment, we define the **relaxation modulus** $E_r(t)$ as the normalized stress response,

$$E_r(t) \equiv \frac{\sigma(t)}{\varepsilon_0}.$$

[3] Here and for the rest of this chapter, "viscous" means "linearly viscous", i.e. obeying the linear constitutive response of a Newtonian fluid.

We are often interested in the short- and long-term responses. To this end, we define the **glassy relaxation modulus** $E_{rg} \equiv \lim_{t\to 0^+} E_r(t)$ and the **equilibrium relaxation modulus** $E_{re} \equiv \lim_{t\to\infty} E_r(t)$.

Remark 5.1.5: Notice that once the strain is applied, the elastic model has a constant relaxation modulus equal to the elastic (Young's) modulus, $E_r(t) = Eh(t)$. The glassy and equilibrium moduli are both equal to E in this case. The viscous model has a relaxation modulus $E_r(t) = \eta\delta(t)$, which spikes to infinity at the instant the stress is applied, then immediately goes to (and stays at) zero. Hence, the glassy modulus is infinite and the equilibrium modulus is zero.

5.1.2 Creep

In the **creep experiment**, we consider the effect of applying to a one-dimensional body the stress function

$$\sigma(t) = \sigma_0 h(t).$$

Analagously to the stress-relaxation case, the body remains stress-free for all $t \leq 0$, and experiences a constant one-dimensional stress σ_0 for $t > 0$. Unsurprisingly, we would expect a purely elastic material to have the strain response

$$\varepsilon(t) = \frac{\sigma_0}{E} h(t).$$

The viscous material would have the strain response

$$\varepsilon(t) = \int \frac{\sigma(t)}{\eta}\, dt = \frac{\sigma_0}{\eta} t,$$

which linearly increases with time. In reality, as with the stress-relaxation case, a viscoelastic material behaves in between these two models. At $t = 0^+$, the strain spikes to a nonzero value, and as time progresses, the strain increases nonlinearly (this *creeping* behavior lends its name to the experiment) and eventually asymptotically approaches a long-term constant strain as $t \to \infty$. We thus normalize the time-dependent strain response ε_0 by the input stress:

Definition 5.1.6. In the creep experiment, we define the **creep compliance** $J_c(t)$ as the normalized strain response,

$$J_c(t) \equiv \frac{\varepsilon(t)}{\sigma_0}.$$

The term *compliance* is used because the creep compliance has units of inverse stress, e.g. MPa^{-1}. The **glassy creep compliance** $J_{cg} \equiv \lim_{t\to 0^+} J_c(t)$ characterizes the short-term behavior and the **equilibrium creep compliance** $J_{cg} \equiv \lim_{t\to\infty} J_c(t)$ characterizes the long-term behavior.

Remark 5.1.7: It is worth stating explicitly that $E_r(t)$ and $J_c(t)$ are both *material properties*. These material properties are the *time-dependent* versions of E and $J \equiv 1/E$ for a linear-elastic material. (So far we have not discussed the actual form of these functions, which describe how the viscoelastic response evolves with time; we will do so soon.)

Definition 5.1.8. A **linear viscoelastic material** is one for which $E_r(t)$ does not depend on ε_0, and $J_c(t)$ does not depend on σ_0.

Caution! In general,

$$E_r(t) \neq \frac{1}{J_c(t)},$$

although they are related (just not by this simple relationship). Rather,

$$\int_{0^-}^{t} J_c(t-\tau)\frac{dE_r(\tau)}{d\tau}\,d\tau = h(t)$$

However, it *is* true that $E_{rg} = 1/J_{cg}$ and $E_{re} = 1/J_{re}$.

Example 5.1.9

We are trying to hold together two rigid parts using a polymer bolt (and a rigid nut). The polymeric material has relaxation modulus $E_r(t) = 5e^{-t^{1/3}}$ GPa (time t is in hours). The bolt is tightened rapidly to a tension of 1 kN at $t = 0$. It has a cross-sectional area $A = \pi \times 0.004$ mm^2.

- Find the strain in the bolt as a function of time for $t > 0$.
- Find the bolt tension after 24 hours.

Solution:

- This is a relaxation test (because the plates and nut are fixed). Thus

$$\varepsilon_0 \equiv \varepsilon(0^+) = \frac{\sigma(0^+)}{E_r(0^+)} = \frac{1\text{ kN} \times A}{5\text{ GPa}} = 0.004,$$

  which can be taken to be fixed for all t.

- The stress after 24 hours is

$$\sigma(24\text{ h}) = E_r(24\text{ h})\varepsilon_0 = 1.1\text{ MPa},$$

  so the *force* is $\sigma(24\text{ h})A = 56$ N.

Remark: Actually, this material would have $E_{re} = 0 \implies J_{ce} \to \infty$, which is characteristic of a *viscoelastic fluid*.

5.2 Boltzmann Superposition Principle

Knowledge of $E_r(t)$ (or $J_c(t)$) is sufficient to characterize a material's response to a single *step* strain (or *step* stress) as in the stress-relaxation (or creep) experiment. However, these functions alone cannot be used to describe the response of a viscoelastic material to an *arbitrary* stress or strain input.

Importantly, for a viscoelastic material, the strain (or stress) at a given time depends on the stress (or strain) applied at *all* previous times, not just at the current time. Fortunately, because the theory relies on *linearity*, we can use superposition to fully describe the material response to an arbitrary input. Consider the case of an arbitrary

strain input, $\varepsilon(t)$. The idea is to break this down into finite increments $\Delta\varepsilon_i$ and consider the corresponding incremental response as fully defined by $E_r(t)$. This works because each one of the $\Delta\varepsilon_i$ can be considered to "turn on" at a corresponding t_i and stay on; hence, it can be modeled using a Heaviside step function. Moreover, because the model is linear, the total response is simply the sum of the incremental responses. That is, if

$$\varepsilon(t) = \sum_{i=1}^{N} \Delta\varepsilon_i h(t - t_i),$$

for each value of i the corresponding stress output is

$$\sigma_i(t) = \Delta\varepsilon_i E_r(t - t_i),$$

so the *total* stress output is

$$\sigma(t) = \sum_{i=1}^{N} \sigma_i(t) = \sum_{i=1}^{N} \Delta\varepsilon_i E_r(t - t_i).$$

For a viscoelastic material subject to a time-varying strain input $\varepsilon(t)$, the stress at time $t^* > 0$ depends not only on $\varepsilon(t^*)$, but on $\varepsilon(t')$ *for all* $t' < t^*$.

(The same result holds for a stress input and the corresponding strain output.) In the limit of finer and finer discretizations, $N \to \infty$ and $t_{i+1} - t_i \to 0$ and we can rewrite the sum as an integral. (Explicitly we will let the discrete $\Delta\varepsilon$ become the infinitesimal $\dot{\varepsilon}dt$.)

Key Equation 5.2.1 (Boltzmann Superposition Principle)

For an arbitrary strain input $\varepsilon(t)$ applied at $t = 0$, the response of a viscoelastic material having relaxation modulus $E_r(t)$ is given by

$$\sigma(t) = \int_{0^-}^{t} E_r(t - u) \frac{d\varepsilon(u)}{du}\, du.$$

For an arbitrary stress input $\sigma(t)$ applied at $t = 0$, the response of a viscoelastic material having creep compliance $J_c(t)$ is given by

$$\varepsilon(t) = \int_{0^-}^{t} J_c(t - u) \frac{d\sigma(u)}{du}\, du.$$

Exercise 5.2.2. AKG 5.6

5.3 Standard Linear Solid

Thus far we have treated $E_r(t)$ and $J_c(t)$ as givens. In reality, they are experimentally measured, and there exist several standard models used to describe viscoelastic behavior. Hence, experimental data is usually curve-fit to determine the model parameters, after which the model behavior may be used to extrapolate the specimen behavior. The most common one-dimensional standard model is called the **standard linear solid** (or SLS). The model has two discrete types of one-dimensional elements: a linear spring (i.e., an element which obeys the linear elastic constitutive model $\sigma = E\varepsilon$) and a viscous dashpot (i.e., an element which obeys the linear viscous constitutive model $\sigma = \eta\dot{\varepsilon}$). In general, models with these springs and dashpots are called **analog models** of viscoelasticity.

Exercise 5.3.1. Show that when placed *in parallel* two spring elements having stiffnesses E_1 and E_2 respectively can be replaced by a single equivalent spring having stiffness $E_1 + E_2$. (By *equivalence* we mean that the models have the same strain output for a given stress input, and vice versa.) Repeat the exercise for two springs in *series*, and **show that** the equivalent stiffness is

$$(1/E_1 + 1/E_2)^{-1}.$$

Show that the *opposite* rules hold for dashpots; i.e., that two dashpots of viscosities η_1 and η_2 have equivalent viscosity $\eta_1 + \eta_2$ in *series* and equivalent viscosity $(1/\eta_1 + 1/\eta_2)^{-1}$ in parallel.

Exercise 5.3.2. Graph the stress-relaxation and creep responses of (*i*) a linear spring, (*ii*) a viscous dashpot, (*iii*) a linear spring in series with a viscous dashpot, and (*iv*) a linear spring in parallel in series with a viscous dashpot. **Write down** the governing equation(s) for each system. (Hint: in the models involving a spring and a dashpot, you may need to solve an ODE to describe the response. It is easiest to identify the "viscous part of the strain", ε^V, and use it as an *internal variable* in order to solve the system.)

Key Equation 5.3.3 (Standard linear solid)

The SLS model consists of one spring having stiffness E_1 in parallel with a spring-dashpot combination in series. The spring-dashpot combination consists of a spring of stiffness E_2 and a dashpot of viscosity η.

The total stress experienced by the model consists of the sum of the stress through the branch with only E_1, and the stress through the spring-dashpot (E_2, η) branch. Observe that the stress experienced by the spring E_2 is *equal* to the stress experienced by the dashpot η, whereas the sum of the strains in E_2 and η is together equivalent to the strain in E_1.

The constitutive relation in the SLS model is described by

$$\begin{cases} \sigma &= E_1\varepsilon + E_2(\varepsilon - \varepsilon_v) \\ \dot{\varepsilon}_v &= \frac{E_2}{\eta}(\varepsilon - \varepsilon_v), \end{cases}$$

where ε_v is an "internal variable" that evolves with each increment of time. Equivalently, we can eliminate ε_v, and the constitutive relation becomes

$$\sigma + \tau_R\dot{\sigma} = E_1\varepsilon + (\eta + E_1\tau_R)\dot{\varepsilon},$$

where $\tau_R \equiv \eta/E_2$ is a parameter having dimensions of time, called the *characteristic relaxation time* of the model.

- The response of the SLS model to the stress-relaxation experiment with input strain $\varepsilon(t) = \varepsilon_0 h(t)$ is

$$E_r(t) = E_{re} + (E_{rg} - E_{re}) \exp\left(-\frac{t}{\tau_R}\right),$$

where $E_{rg} = E_1 + E_2$ and $E_{re} = E_1$.

- The response of the SLS model to the creep experiment with input stress $\sigma(t) = \sigma_0 h(t)$ is

$$J_c(t) = J_{ce} + (J_{cg} - J_{ce}) \exp\left(-\frac{t}{\tau_C}\right),$$

where $J_{cg} = 1/(E_1 + E_2)$, $J_{ce} = 1/E_1$, and $\tau_C = \tau_R(E_1 + E_2)/E_1$.

Observe that $E_r(t) \neq 1/J_c(t)$ and $\tau_R \neq \tau_C$ in general, but that $E_{re} = 1/J_{ce}$ and $E_{rg} = 1/J_{cg}$ as promised. Also observe that the stress relaxation and creep compliance functions are independent of the magnitude of the inputs, as promised. Moreover, the physical interpretation of τ_R and τ_C as characteristic time constants is clear. For example, in a time period of τ_R, the stress in the relaxation experiment has decreased by a factor of $1/e$. Hence, the *smaller* the value of τ_R, the *faster* the equilibrium (long-term) condition is reached.

Exercise 5.3.4. Starting with the constitutive relation for the SLS model in "evolution form", namely:

$$\begin{cases} \sigma &= E_1\varepsilon + E_2(\varepsilon - \varepsilon_v) \\ \dot{\varepsilon}_v &= \frac{E_2}{\eta}(\varepsilon - \varepsilon_v), \end{cases}$$

show that the stress and strain and their time derivatives are related by

$$\sigma + \tau_R\dot{\sigma} = E_1\varepsilon + (\eta + E_1\tau_R)\dot{\varepsilon},$$

where $\tau_R \equiv \eta/E_2$. *Hint:* You will need to take a time derivative of the first equation, then eliminate the variables $\dot{\varepsilon}_v$ and ε_v.

Then, **show that** for a step strain input $\varepsilon(t) = \varepsilon_0 h(t)$, we recover the form for $E_r(t)$ shown above. *Hint:* what are $\varepsilon(t)$ and $\dot{\varepsilon}(t)$ for all $t > 0$?

In general, we can see that the SLS model describes a time-varying, exponential-decay type response to a step input. This is commensurate with experimental observations on viscoelastic materials. The SLS model can be further specialized to more basic combinations, although it must be stated that these more basic combinations are not sufficient to accurately model the exponential-decay response to an arbitrary given loading. Nevertheless, we record these models for completeness.

Definition 5.3.5. The **Maxwell model** consists of a single spring and dashpot in series and can be obtained in the limit $E_1 \to 0$ in the SLS model. *The Maxwell model does not exhibit exponential-decay type creep in response to a step stress input.* (Rather, the dashpot exhibits linear creep, which is not arrested by the spring in series with it. This is closer to the response of a viscous *fluid*.)

Definition 5.3.6. The **Kelvin-Voigt model** consists of a single spring and dashpot in parallel and can be obtained in the limit $E_2 \to \infty$ in the SLS model. *The Kelvin-Voigt model does not exhibit exponential-decay type stress relaxation in response to a step strain input.* (Rather, with the strain held constant the dashpot is given no chance to respond, because it only responds to time-varying strains with a nonzero strain rate. Hence the behavior for any $t > 0$ is governed entirely by the spring.)

Exercise 5.3.7. AKG 5.1 and 5.2.

Remark 5.3.8: From these shortcomings it is possible to conclude that in order to predict creep accurately, a spring-dashpot model should include a single spring in a parallel branch (this fixes the problem with the Maxwell model). Moreover, in order to predict stress relaxation accurately, a spring-dashpot model should *not* have a dashpot alone in a parallel branch (this fixes the problem with the Kelvin-Voigt model).

In general, from these arguments we may conclude that the SLS model is the *simplest* analog model that accurately captures both creep and relaxation behavior.

From Remark 5.3.8 it is also possible to extrapolate the *most general* version of an analog viscoelastic model. This model should consist of (i) one spring alone in a parallel branch[4], to accurately model creep; and (ii) an arbitrary number of parallel branches each having a series spring-dashpot combination, to accurately model stress relaxation. This general model is called a **generalized Maxwell model**. Namely, it consists of a single spring of stiffness E_0 in parallel with N series spring-dashpot pairs having stiffnesses E_i and η_i respectively, for $1 \leq i \leq N$. Hence each branch but the first has a corresponding time constant $(\tau_R)_i$, which characterizes how fast that branch responds to a step input. It can then be shown that the stress-relaxation function can be written in terms of a sum,

$$E_r(t) = E_0 + \sum_{i=1}^{N} E_i \exp\left(-\frac{t}{(\tau_R)_i}\right),$$

called a **Prony series solution**. A similar form for $J_c(t)$ can be written. It is this Prony series form to which experimental data is often fit; then, characterization of the specimen reduces to finding the values for E_0, E_i, η_i.

Exercise 5.3.9. Write the constitutive relation (in evolution form) for a generalized Maxwell model having N series spring-dashpot branches and an extra spring in series. *Hint:* Each of the N branches looks like a Maxwell element, so we can express the total stress in terms of N *viscous strain* internal variables. **Show that** a stress-relaxation experiment yields the form of $E_r(t)$ above.

Remark 5.3.10: The "family" of SLS models is only one example of an assumed form for the relaxation function $E_r(t)$. Because this is a game of fitting experimental data, several other models exist to characterize the relaxation profile. For example, a *power-law relaxation function* of the form

$$E_r(t) = E_{re} + \frac{E_{rg} - E_{re}}{(1 + t/\tau_0)^n}$$

with parameters E_{re}, E_{rg}, τ_0, and $n > 0$. Another such example is the *stretched exponential function*

$$E_r(t) = E_{re} + (E_{rg} - E_{re}) \exp\left[-\left(\frac{t}{\tau_0}\right)^{\beta}\right],$$

[4] It can be shown (see the Exercise in this section) that an arbitrary number of springs alone in parallel can be combined to one equivalent spring alone in parallel; hence one spring alone in parallel is sufficient here.

for $0 < \beta \leq 1$.

Exercise 5.3.11. AKG 5.12. Note the different definitions of E_{rg}, E_{re}, etc., when the model is changed! AKG 5.13(a)-(b).

Exercise 5.3.12. Often in stress-relaxation experiments it is impossible to apply a proper Heaviside strain inputs; real testing machines must apply the prescribed strain input *over* a nonzero amount of time. To model this properly, consider the input strain profile

$$\varepsilon(t) = kt - k(t - t_1)h(t - t_1),$$

where $k > 0$ is a constant strain rate (having units of inverse time), and t_1 represents the (known) time over which the strain input is applied. Assume that t_1 and k are constrained such that the total strain is 0.01 after the ramp period.

Suppose the material being tested can be modeled as an SLS material.

1. **Graph** the input strain-time profiles for $k = 1, 2, 5, 10$ sec^{-1}.
2. Use the Boltzmann superposition principle to **write an expression** for the output stress-time profile $\sigma(t)$ in terms of k and the SLS model parameters E_1, E_2, and η.
3. For $E_1 = 0.2$ MPa, $E_2 = 0.8$ MPa, $\eta = 0.8$ MPa sec, **graph** the output stress-time profiles for $k = 1, 2, 5, 10$ sec^{-1}, and the corresponding stress-strain profiles.
4. Now suppose that an N-branch Prony series is required instead of the SLS model. **Write an expression** for the output stress-time profile in this case.

5.3.1 First-order analogy to RC circuits*

In this optional subsection, we develop a first-order electrical/mechanical analogy for the analog viscoelastic model[5]. In particular, we develop an analogy between a linear viscoelastic mechanical system and a basic type of electric circuit. Because the governing differential equations for both systems take the same form, the solutions (and therefore the component behavior) can be directly compared.

Consider an electrical system that consists of a DC voltage (or current) source, a resistor (resistance R), and a capacitor (capacitance C). With respect to the voltage/current source, the resistor and capacitor can be arranged in series, parallel, or a combination thereof. The current-voltage relationship across the resistor obeys

$$I(t) = \frac{1}{R}V(t)$$

and across the capacitor,

$$I(t) = C\frac{dV(t)}{dt},$$

where $I(t)$ represents the current and $V(t)$ represents the voltage, both functions of time. Comparing these relations with the constitutive relations for a spring and dashpot respectively, we can make the following analogies:

$$\boxed{\sigma \iff I; \qquad \varepsilon \iff V; \qquad E \iff 1/R; \qquad C \iff \eta}$$

[5]Note that this first-order analogy does not accurately model the energy (specifically, dissipation) as it is transmitted through circuit components. For a more complete analog, we would introduce a mass to the mechanical system and an inductor to the electrical system, and then compare *second*-order differential equations.

where a *resistor* behaves like a *linear spring* (the resistance acts like the *compliance*, or inverse stiffness), and a *capacitor* behaves like a *dashpot*. Hence the stress-relaxation experiment is analagous to applying a constant voltage across some RC circuit, and the creep experiment is analagous to applying a constant current across some RC circuit. Moreover, the relaxation time $\tau_R = \eta/E$ in a single branch is analagous to the characteristic relaxation time $\tau_{RC} = RC$ which governs capacitive (dis)charging.

Example 5.3.13

Consider a *series* RC circuit, which is initially connected to a DC voltage source. All elements are initially discharged; at time $t = 0$, the voltage source is turned on and from then on provides a constant voltage. That is, the applied voltage profile as a function of time is

$$V(t) = V_0 h(t) \tag{5.1}$$

(recall $h(t)$ is the Heaviside function). For this circuit:

1. **Identify** the corresponding spring-dashpot model and the corresponding mechanical experiment (i.e. *stress-relaxation* or *creep*).
2. **Describe** or sketch a plot of the current through the circuit as a function of time.
3. **What is** the time constant τ for the circuit?

Solution: The series RC circuit corresponds to the *Maxwell model*, having a spring and dashpot in series. The applied quantity is a voltage input, which by the previous part corresponds to a strain input $\varepsilon(t) = \varepsilon_0 h(t)$, which is the *stress-relaxation experiment*. The current response of the circuit as a function of time is similar in nature to the response of the Maxwell model in the stress-relaxation experiment. Initially, the uncharged capacitor admits all the current allowed by the resistor, $I_0 = V_0/R$, but the current then decreases (exponentially) as the capacitor charges. After a long time, the capacitor is fully charged, so no current flows through the circuit.

To determine τ for this series RC circuit, we could set up and solve the system of differential equations, recalling that current is the same through the resistor and capacitor, and that the total voltage drop across both components is equal to the voltage input:

$$\begin{aligned} I_R(t) = I_C(t) \implies \frac{1}{R} V_R(t) &= C \frac{dV_C(t)}{dt} \\ V_R(t) + V_C(t) &= V_0 \end{aligned}$$

After solving the system, we could then inspect the exponential term in your expression for $I_R(t)$ (or $I_C(t)$), which would have the form $e^{-t/\tau}$. But an easier way to work out the time constant is by analogy with the stress-relaxation time constant.

For the stress relaxation experiment the timescale is $\tau_R = \eta/E$ where η and E are the constants corresponding to the series dashpot and spring, respectively. Making the substitution $\eta \to C$ and $E \to 1/R$ we arrive at $\tau = RC$ for the series RC circuit. This is indeed what we would arrive at after solving the system above.

5.4 Correspondence Principle

For small-deformation *structural applications* involving reduced-dimension problems, specifically *beam-bending* and *shaft-torsion* problems, there exists a connection between the classical elastic solutions and the solutions for the same geometry, but with a viscoelastic material model. This rule is called the **correspondence principle**.

Key Equation 5.4.1 (Correspondence Principle)

For a viscoelastic beam loaded at time $t = 0$ with a *stepped* input (i.e., a stepped load $P_0 h(t)$ or a stepped displacement $\delta_0 h(t)$), the viscoelastic beam solution is given by the elastic solution but with either

- E replaced by $E_r(t)$, in problems with a prescribed stepped displacement; or
- $J \equiv 1/E$ replaced by $J_c(t)$, in problems with a prescribed stepped load.

A similar principle holds for viscoelastic shafts with stepped *torsional* inputs, whereby the shear modulus G is replaced by the shear-stress relaxation function $G_r(t)$, and the shear compliance $1/G$ is replaced by the shear-strain creep function $L_c(t)$.

Example 5.4.2 (Viscoelastic cantilevered beam)

Consider a slender cantilevered beam of length L having creep compliance function $J_c(t)$ and moment of inertia I. A prescribed stepped force $F(t) = F_0 h(t)$ is applied at the distal end, producing a time-dependent displacement $w_L \equiv w(x = L, t)$ there. Find an expression for $w_L(t)$.

Solution: The elastic solution is given by

$$w(x) = F_0 \frac{x^2(3L - x)}{6EI} = F_0 \frac{x^2(3L - x)}{6I} J$$

on $0 \le x \le L$ for some elastic modulus E. Then the viscoelastic solution is

$$w(x, t) = F_0 \frac{x^2(3L - x)}{6I} J_c(t),$$

so at the tip

$$w_L(t) = \frac{F_0 L^3}{3I} J_c(t).$$

The correspondence principle can only be used if:

- **the Poisson ratio ν of the material is constant in time, or**
- **regardless of ν, the elastic solution is of the form**

$$\mathbf{u}_0(\mathbf{x}) = E\hat{f}(\mathbf{x}) \qquad \text{or} \qquad \boldsymbol{\sigma}_0(\mathbf{x}) = J\hat{g}(\mathbf{x}),$$

where $\hat{f}$ and $\hat{g}$ are *not* a function of E, ν, or any other elastic constant.

More generally, it can be shown that for *any* boundary value problem concerning a linear viscoelastic material model, the elastic solution can be used to write down the viscoelastic

solution. The procedure is to write down the elastic solution, then substitute

$$\bar{E}_r^*(s) \equiv s\bar{E}_r(s) \equiv s\int_{0^-}^{\infty} e^{-st}E_r(t)\,dt$$

for E, then take the *inverse* Laplace transform to obtain the viscoelastic solution.

The Laplace transform is defined as an invertible function L such that

$$L : f(t) \to \bar{f}(s) \equiv \int_{0^-}^{\infty} e^{-st}f(t)\,dt.$$

Remark 5.4.3: Taking a Laplace transform of $E_r(t)$ and $J_c(t)$ shows that they are related in the following manner:

$$h(t) = \int_{0^-}^{t} J_c(t-\tau)\frac{dE_r(\tau)}{d\tau}\,d\tau = \int_{0^-}^{t} E_r(t-\tau)\frac{dJ_c(\tau)}{d\tau}\,d\tau.$$

Exercise 5.4.4. AKG 5.4, 5.7, 5.8.

5.5 Oscillatory Inputs

In the special case where the stress or strain input is oscillatory in time, i.e., it is of the form

$$\varepsilon(t) = \varepsilon_0\cos(\omega t)$$

for some strain amplitude ε and oscillation frequency ω, it can be shown that

the (stress) output is an oscillatory function with the *same* frequency, but with a nonzero phase lag[a] $\delta > 0$,

[a] Note that δ is defined as a *positive* number in this case such that *the strain always lags behind the stress*, by convention.

such that

$$\sigma(t) = \sigma_0\cos(\omega t + \delta),$$

where σ_0 characterizes the amplitude of the stress response. We call the ratio δ/ω, which has units of time, the **lag time**. In order to relate the response profile to the input profile, we want to re-write $\sigma(t)$ in terms of ε_0 and ω alone. It can be shown that this is always possible, and that the response profile will have the form

$$\sigma(t) = \varepsilon_0\left(E'\cos(\omega t) - E''\sin(\omega t)\right)$$

for

$$E' \equiv \frac{\sigma_0}{\varepsilon_0}\cos(\delta), \qquad E'' \equiv \frac{\sigma_0}{\varepsilon_0}\sin(\delta).$$

The quantity E' is called the **storage modulus** because it has units of stiffness and, being the coefficient of the $\cos(\omega t)$ term, quantifies "how much in phase" the stress response is with the strain input. The quantity E'' is called the **loss modulus**; it quantities how much *out* of phase the stress response is. The *ratio* of these quantities,

$$\tan(\delta) = \frac{E''}{E'},$$

is called the **loss tangent** and represents a measure of *energy loss* as a result of dissipation.

It can be shown that the same applies in the case of an oscillatory stress input, i.e. the input

$$\sigma(t) = \sigma_0 \cos(\omega t)$$

produces a strain output[6]

$$\varepsilon(t) = \varepsilon_0 \cos(\omega t - \delta) = \sigma_0 \left(J' \cos(\omega t) + J'' \sin(\omega t)\right)$$

where the **storage compliance** J' and **loss compliance** J'' are defined to be

$$J' \equiv \frac{\varepsilon_0}{\sigma_0} \cos(\delta), \qquad J'' \equiv \frac{\varepsilon_0}{\sigma_0} \sin(\delta).$$

The loss tangent can be written in the form

$$\tan(\delta) = \frac{J''}{J'}.$$

The material parameters E', E'', J', and J'' are functions of the "excitation frequency" ω.

Remark 5.5.1: For a purely elastic material, $E'' = J'' = 0$, and the input is entirely in phase with the output, $\delta = 0$. For a purely linearly viscous liquid, $E' = J' = 0$, and the input and output are perfectly *out* of phase, $\delta = \pi/2$.

In one cycle of oscillation, which takes a time equivalent to $T = 2\pi/\omega$, the work W done by the input stress or strain profile can be determined by integrating the stress power $\sigma(t)\dot{\varepsilon}(t)$ over the period $t \leq 0 \leq T$. In the oscillatory case, the total work in one full cycle corresponds exactly to the *dissipated energy*, because the material has returned to its original state. Carrying the computation out we can see that

$$W = \pi \sigma_0 \varepsilon_0 \sin(\delta) \qquad \text{(one cycle)},$$

so that the amount of dissipation depends on the value of δ. In terms of the loss modulus or the loss compliance,

$$W = \pi \varepsilon_0^2 E'' = \pi \sigma_0^2 J'' \qquad \text{(one cycle)}.$$

To determine the *ratio* between stored and dissipated energy, we can instead integrate from $0 \leq t \leq T/4$, so that the specimen is in a different configuration from the reference state. In this case we find that

$$W = \underbrace{\frac{1}{2}\sigma_0 \varepsilon_0 \cos(\delta)}_{\text{stored energy}} + \underbrace{\frac{\pi}{4}\sigma_0 \varepsilon_0 \sin(\delta)}_{\text{dissipated energy}}.$$

The ratio of the dissipated energy to the stored energy over this quarter cycle is called the **damping capacity**, and is related to the loss tangent:

$$\text{damping capacity} = \frac{\pi}{2}\tan(\delta) = \frac{\pi}{2}\frac{E''}{E'} = \frac{\pi}{2}\frac{J''}{J'}.$$

[6] Observe the minus sign in the expression $\cos(\omega t - \delta)$ such that again, the *strain always lags the stress*.

The damping capacity $\tan(\delta)$ is a function of the excitation frequency ω.

Exercise 5.5.2. AKG 5.11.

5.6 Complex Number Representation

Being periodic, the stress (strain) inputs and corresponding strain (stress) outputs may be represented in terms of exponentials with complex numbers, which can simplify the form of caclulations. Note that no *new* information is given in this section; the preceding formulations are simply rewritten.

Definition 5.6.1. The **complex modulus** E^* is constructed using the storage modulus E' as the real part and the loss modulus E'' as the imaginary part, i.e.

$$E^* \equiv E' + iE''.$$

Similarly the **complex compliance** J^* can be written as

$$J^* \equiv J' + iJ''.$$

The two are related by $E^*J^* = 1$.

The complex modulus and complex compliance are both functions of the excitation frequency:

$$E^* = E^*(\omega), \qquad J^* = J^*(\omega).$$

Now, we rewrite the oscillatory stress input and corresponding strain output as

$$\sigma(t) = \sigma_0 e^{i\omega t} \implies \varepsilon(t) = J^*\sigma(t).$$

It is clear that the "physical" interpretation of J^* is the strain output normalized to the stress input, just as we had defined before:

$$J^*(\omega) = \frac{\varepsilon(t)}{\sigma_0 e^{i\omega t}}.$$

Similarly for an oscillatory strain input and corresponding stress output

$$\varepsilon(t) = \varepsilon_0 e^{i\omega t} \implies \sigma(t) = E^*\varepsilon(t).$$

Example 5.6.2

Using the definitions above:

- For a linear spring, $\sigma(t) = E\varepsilon(t) \implies \sigma_0 e^{i\omega t} = E\varepsilon_0 e^{i\omega t}$, so it follows that $E^* = E$ for a linear spring. That is, for a spring, $E' = E$ and $E'' = 0$.
- For a Newtonian dashpot, $\sigma(t) = \eta\dot{\varepsilon}(t) \implies \sigma_0 e^{i\omega t} = \eta i\omega\varepsilon_0 e^{i\omega t}$, so it follows that $E^* = i\omega\eta$ for a linear dashpot. That is, for a dashpot, $E' = 0$ and $E'' = \eta\omega = E\tau_R\omega$.

Remark 5.6.3: A complex number written in the form $z = Ae^{i\theta}$ has corresponding *amplitude* A and *phase angle* θ. The same complex number can be written as $z = x + iy$, so that there is an additive decomposition into *real* and *imaginary* parts. The parameters A, θ, x, and y here are related by

$$x = A\cos(\theta); y = A\sin(\theta); A = x^2 + y^2; \tan(\theta) = \frac{y}{x}.$$

Specifically, E^* can be expressed two ways:

$$E^*(\omega) = |E^*(\omega)|\exp(i\delta(\omega)) = E'(\omega) + iE''(\omega);$$

from which it follows that

$$E' = |E'|\cos(\delta); \qquad E'' = |E''|\sin(\delta); \qquad \tan(\delta) = \frac{E''}{E'}$$

as before.

Definition 5.6.4. The **Deborah number** De is the dimensionless excitation frequency normalized to the characteristic relaxation time,

$$\mathrm{De} = \omega\tau_R.$$

Because E', E'', and $\tan\delta$ are functions of the excitation frequency, it is common to plot their trends as a function of the nondimensional De instead.

- For $\mathrm{De} \to 0$ (the excitation frequency is very, very small compared to the inverse of the characteristic relaxation time), the polymer behaves like *rubber* and the storage modulus E' is low. Namely, $E' = E_{re}$, a value independent of frequency. In this case $\tan\delta \approx 0$.
- For $\mathrm{De} \to \infty$ (the excitation frequency is very, very large compared to the inverse of the characteristic relaxation time), the polymer behaves like *glass* ("is glassy") and the storage modulus E' is high. Namely, $E' = E_{rg}$, a value independent of frequency. In this case $\tan\delta \approx 0$ as well.
- For $\mathrm{De} \approx 10^0$ (the excitation and relaxation timescales coincide), the material behaves viscoelastically, with E' increasing with frequency. Here, E'' and $\tan\delta$ are near their highest values.

Remark 5.6.5: The frequency-dependent response of a Prony series is a natural extension of the previous observations. Specifically, if the Prony series has N branches, the graph of $E'(\omega)$ will have N plateaus, and the graph of $(\tan(\delta))(\omega)$ will have N peaks, each corresponding with the characteristic stiffness E_i and the inverse of the characteristic time constant τ_i of an individual branch.

Exercise 5.6.6. AKG 5.9, 5.13(c), 5.16.

The following exercises are more general and cover the extension of our one-dimensional framework to three dimensions.

Exercise 5.6.7. AKG 5.22, 5.23, 5.24, 5.25.

5.7 Temperature Dependence of Viscoelastic Effects

In general, the behavior of polymeric materials is highly temperature dependent. Thus, all material properties δ, E', E'', J', J'' are not only functions of frequency ω, but also of temperature. It is important to recall that solid polymers are made up of long polymer chains, which are held together by van der Waals bonds and covalent crosslinks. In the case of semi-crystalline polymers, there is also a crystalline structure (order) in a fraction of the material.

Definition 5.7.1. The **glass transition temperature** T_g is the temperature at which the van der Waals bonds melt, and only entanglements are left between chains; hence, the entangled polymer chains begin to slide relative to their neighbors, resulting in an abrupt loss of structural properties. In actuality the glass transition occurs over a range of temperatures, typically within a $20°C$ window.

For amorphous polymers (e.g. polycarbonate), below the glass transition temperature the material behavior is highly elastic (i.e., $E' > E''$). Near the glass transition temperature, van der Waals bonds melt and entanglements slightly loosen, causing the sliding of polymer chains and frictional dissipation (which causes a spike in E''). Above the glass transition temperature, the polymer has effectively changed state: both E' and E'' decrease. But E' has decreased *more* (because the van der Waals bonds remain melted at such high temperatures), so $\tan\delta$ remains elevated.

For a viscoelastic polymer, the graph of the storage modulus $E'(T)$ as a function of temperature T will be monotone decreasing. Moreover, the graph of the loss modulus $E''(T)$ will exhibit a peak at the glass transition temperature T_g. At this temperature the storage modulus E' is in its most rapid rate of descent.

For semi-crystalline polymers, there is an additional amount of temperature required to melt the *crystals* and remove all elastic contributions. Therefore after the glass transition temperature where the van der Waals bonds have melted but the crystal structure still exists, there is a *plateau* in the plot of E'. This plateau ends at the *melting temperature* of the polymer.

Definition 5.7.2. The **melting temperature** T_m *for semi-crystalline polymers*[7] is the temperature at which the crystal structure begins to experience sufficient thermal vibrations to cease being an ordered solid. Above the melting temperature, viscoelastic solids act more like viscous fluids. The graph of $E'(T)$ for a typical semi-crystalline polymer will have the same characteristics as a typical amorphous polymer, with an *additional* drop near the melting temperature.

5.7.1 Time-Temperature Superposition for Amorphous Polymers

Remarkably, for amorphous polymers, the graphs of the relaxation modulus as a function of time, $E_r(t)$, as measured at different temperatures, coincide in such a way that they can be *shifted* along the *time* axis using a **shift factor**, which is a material property.

[7] Typically a polymer is considered semi-crystalline if the degree of crystallinity is at a minimum 30-35%.

To do so, one measurement temperature is chosen as a *reference temperature* T_{ref}, and for this reference temperature the shift factor at all other measurement temperatures is determined. There exist empirical correlations for the shift factor as a function of temperatures which hold for almost all amorphous polymers. Once the shift factors are determined, the corresponding relaxation curves are shifted by that amount to create a "master curve". The key point is that time-dependent effects like the reduction in $E_r(t)$ with time t happens *faster* as a result of *higher* temperatures, but "predictably" so.

Exercise 5.7.3. AKG 5.26, 5.27.

6 Limits to Elastic Response and Plasticity

6.1 Limits to Elastic Response

As discussed in the previous sections, the elastic (or viscoelastic) theory is useful in describing the behavior of most materials *up to a "small" displacement limit*[1]. If (1) pre-existing cracks are not present, and (2) the load is not cyclic, the elastic limit is typically followed by one of the two following types of responses:

- brittle failure, which is marked by catastrophic and fast fracture; or
- ductile yielding, which is marked by an increase in strain with a corresponding change in stress which is nonlinear, often accompanied by permanent deformation of the specimen before failure.

In the case of brittle failure, the mathematical failure criterion is simple: simply take

$$\sigma_1 = \sigma_{1,\text{cr}} \qquad \text{(onset of failure)}$$

where σ_1 is the maximum principal stress, $\sigma_1 = \max(\sigma_i^P)$, and $\sigma_{1,\text{cr}}$ is a material property, obtained from a laboratory experiment on a standardized sample of the material. If $\sigma_1 < \sigma_{1,\text{cr}}$, we assume the material remains elastic. (Note that in reality, for most materials which fail in a brittle manner, the *compressive* failure stress is much higher than the tensile failure stress. This is due to the predominant failure mode being crack propagation.)

For ductile yielding, the situation is a bit more nuanced. To define a criterion for the onset of non-linearity, we still would like to define a scalar function

$$f : \text{Lin} \to \mathbb{R}, \boldsymbol{\sigma} \mapsto \bar{\sigma},$$

which "captures the essence" of the entire stress state $\boldsymbol{\sigma}$ in a single number $\bar{\sigma}$ which we can then compare to a threshold value Y that represents a material property. This *yield criterion* should therefore require

$$\bar{\sigma} = Y \qquad \text{(onset of yield)}$$

with $\bar{\sigma} < Y$ corresponding with elastic behavior. It would be especially simple if we could connect the material property Y to a physical experiment which can be done easily in the laboratory.

It turns out that the simplest — and most important — laboratory experiment that can characterize the properties of both brittle and ductile materials is the *uniaxial tension test*. We have already mentioned the resulting stress state before; recall that in an *elastic* state of uniaxial tension the stress tensor is given as $\boldsymbol{\sigma}_{11} = \sigma \mathbf{e}_1 \otimes \mathbf{e}_1$ and the resulting strain tensor is $\boldsymbol{\varepsilon} = \varepsilon_{11}\mathbf{e}_1 \otimes \mathbf{e}_1 + \varepsilon_{22}\mathbf{e}_2 \otimes \mathbf{e}_2 + \varepsilon_{33}\mathbf{e}_3 \otimes \mathbf{e}_3$, whereby we were able to define the elastic modulus $E \equiv \sigma_{11}/\varepsilon_{11}$.

[1] except in the case of finite elasticity, which is applicable to some rubber materials for large deformations

6.1.1 Phenomenological response of polycrystalline metals in uniaxial tension

Let us first consider the inelastic behavior of polycrystalline metals, which represent a significant proportion of practical engineering materials. Consider a metallic specimen to which we apply a uniaxial tensile displacement δ; correspondingly, and measure the load P in that direction.

Definition 6.1.1. We normalize the displacement to **engineering strain** $e \equiv \delta/L_0$ and normalize the load to **engineering stress** $s \equiv P/A_0$, where L_0 and A_0 are the length and area, respectively, of the gauge section of the specimen *in the reference* (undeformed) *configuration*, which is taken to be before the test begins.

Qualitatively, a typical plot of engineering stress versus engineering strain looks like the following:

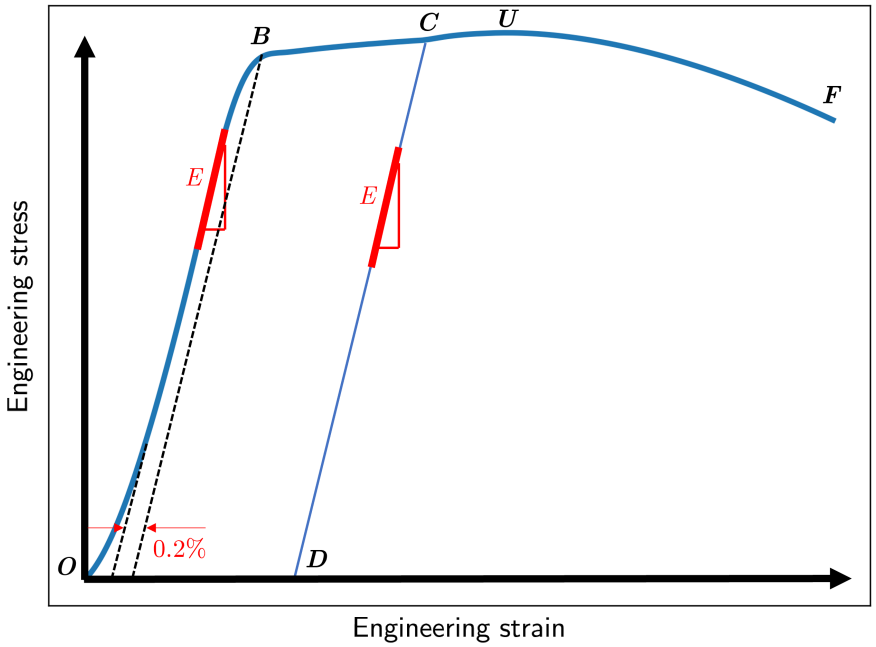


Figure 6.1: Experimental stress-strain data obtained from a sample of aluminum 6061-T6. Labeled are the yield point B, the point of ultimate tensile strength U, and the point of failure F, alongside a hypothetical unloading-reloading curve CD.

Experimentally, we make the following observations alongside the recorded data:

- The initial non-linear region (sometimes called the "toe" region) represents the take-up of slack by the specimen, grips, and testing machine, and is not considered to be representative of any material property.
- The elastic regime is taken to span from the end of this toe region until point B. Behavior within this limit is entirely reversible upon unloading and reloading.
- From $B \to C \to U$ the specimen volume is conserved. This region is characterized by **strain hardening**, whereby the specimen is able to withstand a greater load P despite a reduction in the cross-sectional area.

- From $U \to F$, the deformation is localized at a **neck**. The strain hardening capacity is unable to keep up with the rapid rate of reduction in cross-section at the neck, and thus the load P decreases.
- The onset of yielding (point B) is often hard to determine experimentally, so the **0.2% offset method** is often employed, whereby the yield point is taken to be the point where the experimental data curve intersects a line drawn with slope equal to the computed elastic modulus E and with a horizontal intercept at $e = 0.002 = 0.2\%$.
- If the material were unloaded at point C, it would follow line CD, which has slope equal to the elastic modulus. Subsequent *re-loading* would also follow line DC, and the material would appear to have a higher yield strength after this unload-reload cycle.
- The area under the engineering stress-engineering strain curve is a measure of *energy absorption* per unit reference volume $A_0 L_0$, and is called **toughness**.

Remark 6.1.2: Notice that in the elastic-plastic universe, a given state of stress may correspond to more than one state of strain, and a given state of strain may correspond to more than one state of stress. This is in explicit contrast to the theory of linear elasticity, for example, where there is a one-to-one correspondence between stress and strain. This motivates the need for a more complicated constitutive model for plasticity, and explains why we cannot simply write an simple stress-strain relation to model the plastic regime.

Definition 6.1.3. We define **true stress** $\sigma \equiv P/A$, where A is the *deformed* cross-sectional area at the instant that P was measured, and define **true strain** $\varepsilon = \ln(1+e) \equiv \ln\left(\frac{L_0+\delta}{L_0}\right)$, where $L_0 + \delta$ represents the *deformed* gauge length at the instant that ε was measured.

Key Equation 6.1.4

Until the onset of necking, we can use the fact that volume is conserved to write a relationship between engineering strain and engineering stress. Then, in this regime, we have

$$\sigma = s(1+e)$$
$$\varepsilon = \ln(1+e)$$

Up until the yield point, σ and ε coincide with s and e, respectively, Between the yield point and the ultimate tensile stress, which marks the onset of necking, we always have $s \leq \sigma$. After the onset of necking, we generally do not refer to σ, because the deformation is no longer uniform across the gauge section of the specimen.

Remark 6.1.5: To explain the physical phenomenon of *necking*, consider the gauge section of a tensile dogbone specimen. Suppose that during a displacement-controlled tensile test, one small "slice" of the cross-section happens to deform a small bit more than its neighbors, perhaps due to an internal flaw or other imperfection. Then,

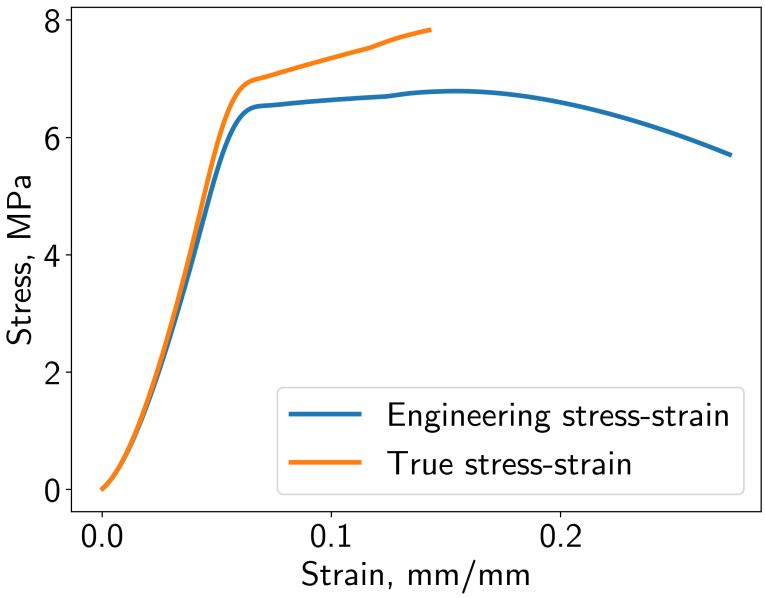


Figure 6.2: Engineering versus true stress-strain curves for the aluminum sample. The true stress and true strain are computed using (6.1.4) only up to the onset of necking.

this slice must support a higher true stress than any neighboring slice, because it carries the same load over a smaller cross-sectional area. Now, if this slice has not yet reached its ultimate tensile strength, the extra deformation will lead to *strengthening*, because the slice has reached a higher point on the hardening curve relative to its neighbors, in the regime where the additional amount of hardening is more than enough to compensate for the additional amount of true stress. Hence in a subsequent increment of elongation the *other* slices will take up the deformation, bringing the entire specimen back into a homogeneous state. However, if the ultimate tensile strength has already been attained by the slice, any subsequent localized deformation *cannot* be sustained by further hardening. The strain-hardening capacity is outmatched by the increase in true stress, and the resulting deformation will be localized to this unlucky slice. This is precisely the necking instability. Therefore, the critical condition for the onset of necking is attained precisely when the tensile strength is reached, which is by definition the maximum of the engineering stress-strain curve:

$$dP = 0 \implies \frac{ds}{de} = 0.$$

This is called the "Considere condition" for necking. Equivalently on the true stress-true strain curve we have

$$\frac{d\sigma}{d\varepsilon} = \sigma.$$

Example 6.1.6

Suppose a metal's true stress-true strain curve can be modeled as $\sigma = A\varepsilon^n$, for constants A and n. What is the ultimate tensile strength?

Solution: The ultimate tensile strength is defined to be the engineering stress at the onset of necking. The condition for the onset of necking is given by

$$\frac{d\sigma}{d\varepsilon} = \sigma \implies nA\varepsilon^{n-1} = A\varepsilon^n \implies \varepsilon = n,$$

and it remains to find the engineering stress at this value of true strain. Using $\sigma = s(1+e)$ and $\varepsilon = \ln(1+e)$, we have $\sigma = s\exp(\varepsilon) \implies s = \sigma/\exp(\varepsilon) = An^n/\exp(n)$.

Consider a particular nonzero state marked by a point $(\sigma_0, \varepsilon_0)$ on the stress-strain curve. If a specimen is currently at this state and then *unloaded*, as discussed previously the stress-strain relation would follow a line with slope equal to the elastic modulus until the stress falls to zero. If the specimen had passed the yield point at all during initial loading, the *strain* will be nonzero when the stress falls to zero. We call this residual strain the **plastic strain** associated with the strain ε_0, and we write ε_0^P to denote this value. The quantity $\varepsilon_0 - \varepsilon_0^P$ is called the **elastic strain** and is written ε_0^E. This is the amount of strain "recovered" as the specimen is unloaded. Trivially, then, for all strain states ε_0, we have

$$\varepsilon_0 = \varepsilon_0^E + \varepsilon_0^P,$$

where $\varepsilon_0^E = \sigma_0/E$.

Remark 6.1.7: For *small* strains, we can define an "engineering strain increment" based on the incremental displacement normalized by the original gauge length, which produces a nice linear response:

$$de = \frac{dL}{L_0} \implies e = \int_{L=L_0}^{L=L} \frac{dL}{L_0} = \frac{L - L_0}{L_0}.$$

For *large strains*, we can define a similar "true strain increment" based on the incremental displacement normalized by the deformed length, which is not a constant. Hence, the ratio dL/L produces a nonlinear response when integrated:

$$d\varepsilon = \frac{dL}{L} \implies \varepsilon = \int_{L=L_0}^{L=L} \frac{dL}{L} = \ln\left(\frac{L}{L_0}\right).$$

The two are related: in fact, the engineering strain is simply the first term in the Taylor expansion of the true strain:

$$\ln\left(\frac{L}{L_0}\right) = \ln\left(1 + \frac{\delta}{L_0}\right) = \frac{\delta}{L_0} - \underbrace{\frac{1}{2}\left(\frac{\delta}{L_0}\right)^2 + \ldots}_{\text{small when } \delta \ll L_0}$$

In what follows we will specalize the theory of nonlinear deformation to the case of polycrystalline metals, which undergo *plasticity*. In doing so we will favor the true stress and true strain measures.

Plasticity theory is formulated with respect to true stress and true strain. This is because:

- **Tension and compression look identical in the true stress-true strain framework.**

- **The true strain decomposes additively by definition.**
- **True strain accommodates "large" deformations, whereas the infinitesimal strain tensor is only an approximation.**

Remark 6.1.8: For problems worked *in a principal basis*, if the original reference dimensions in the basis directions are x_0, y_0, and z_0, and the deformed dimensions are x, y, z respectively, the logarithmic (Hencky) true strain tensor can be represented as

$$[\ln \mathbf{U}] = \begin{bmatrix} \ln\left(\frac{x}{x_0}\right) & 0 & 0 \\ 0 & \ln\left(\frac{y}{y_0}\right) & 0 \\ 0 & 0 & \ln\left(\frac{z}{z_0}\right) \end{bmatrix}.$$

6.1.2 Yield Criteria

Now that we have described the phenomenological behavior of a ductile metallic specimen to a uniaxial tension test, let us return to the question of defining a yield function, $f(\boldsymbol{\sigma})$, which we can use as a criterion to determine the onset of inelastic behavior. We require that

1. the yield function must depend only on the *deviatoric* stress components, because experimental observations suggest that plasticity is volume-conserving (incompressible), a claim we will physically justify later;
2. the yield function must be invariant of the basis of $\boldsymbol{\sigma}$, if the material is isotropic;
3. the yield function must output $|\sigma_{11}|$ when the state of stress $\boldsymbol{\sigma}$ corresponds to uniaxial tension, so that we can take the yield criterion to correspond with the critical value of σ_{11} at the onset of yield in a uniaxial tension test.

These three requirements are necessary, but not sufficient, to entirely determine a form of the function f. Choosing different forms for f in terms of various stress components leads to the definition of various yield criteria; we record the most popular ones here. We denote the critical value of the appropriate stress measure by Y. For a sample which has never undergone plasticity, Y corresponds to the *yield strength* σ_y obtained in uniaxial tension.

Definition 6.1.9. The **von Mises yield criterion** defines

$$f(\boldsymbol{\sigma}) = f(\boldsymbol{\sigma}') = \sqrt{\frac{3}{2}(\sigma_1'^2 + \sigma_2'^2 + \sigma_3'^2)} \equiv \bar{\sigma}$$

to be the **von Mises stress** (also called the **equivalent tensile stress**) and requires that

$$\bar{\sigma} = Y \qquad \text{(onset of yield)}.$$

Remark 6.1.10: The scalar $\bar{\sigma}$ is a scalar multiple of the *second invariant* of the *deviatoric* stress tensor.

Definition 6.1.11. The **Tresca yield criterion** says that the onset of yield corresponds to a limiting value of the maximum shear stress, which is defined to be *half* of the difference between the maximum and minimum principal stresses:

$$\tau_{\max} \equiv \frac{1}{2}(\sigma_1^P - \sigma_3^P).$$

Specifically, the Tresca criterion stipulates that yield occurs when the maximum shear stress in the specimen reaches the *maximum shear stress in a tensile specimen at yield*, for which $\tau_y = Y/2$. (Here τ_y is a material property called the Tresca **shear yield strength**.) Hence in the uniaxial tensile experiment, the yield criterion is, explicitly,

$$(\sigma_1^P - \sigma_3^P) = Y \qquad \text{(onset of yield)}.$$

Remark 6.1.12: The Tresca yield criterion is more conservative than the Mises yield criterion. In uniaxial or equibiaxial stress, the two criteria are identical. However, for any other stress state, yielding occurs at lower stress values according to the Tresca criterion, and for a given stress state, the Tresca criterion predicts larger plastic deformation than the Mises criterion. For example, in the case of pure shear, the Mises yield strength is reached at a stress which is a factor of $2/\sqrt{3} \approx 1.15$ larger than the stress at which the Tresca yield strength is reached.

However, calculating the principal stresses to compute the Tresca stress is involved, whereas there is a direct formula to compute the Mises stress for any given stress state. Thus, the remainder of this section uses the Mises yield criterion to develop plasticity theory.

Stress invariants

We remarked earlier that $\bar{\sigma}$ is related to an invariant of the deviatoric stress tensor. More broadly, several invariants of the stress tensor or of the deviatoric stress tensor are widely used in theory; we record them here. In this section, let σ_{ij} represent the components of stress in any basis, and let σ_i represent the *principal* components of stress. Moreover, we define the principal components of deviatoric stress to be

$$\sigma_i' \equiv \sigma_i - \frac{1}{3}(\sigma_1 + \sigma_2 + \sigma_3).$$

Definition 6.1.13. The **mean normal pressure** or **equivalent pressure stress** is defined to be

$$\begin{aligned}\bar{p} &\equiv -\frac{1}{3}\mathrm{tr}\boldsymbol{\sigma} = -\frac{1}{3}\sigma_{kk} \\ &= -\frac{1}{3}(\sigma_{11} + \sigma_{22} + \sigma_{33}) \\ &= -\frac{1}{3}(\sigma_1 + \sigma_2 + \sigma_3)\end{aligned}$$

In a state of pure hydrostatic pressure ($\sigma_{11} = \sigma_{22} = \sigma_{33}$, all other $\sigma_{ij} = 0$), the mean normal pressure is exactly equal to the *negative* of the nonzero stress components. In general, $\bar{p} > 0$ when the hydrostatic stress is compressive.

Definition 6.1.14. The **equivalent shear stress** is defined to be

$$\begin{aligned}\bar{\tau} &\equiv \sqrt{\frac{1}{2}\mathrm{tr}(\boldsymbol{\sigma'^2})} = \sqrt{\frac{1}{2}\sigma'_{ij}\sigma'_{ij}} \\ &= \sqrt{\frac{1}{6}\left[(\sigma_{11}-\sigma_{22})^2+(\sigma_{22}-\sigma_{33})^2+(\sigma_{33}-\sigma_{11})^2\right]+(\sigma_{12}^2+\sigma_{23}^2+\sigma_{31}^2)} \\ &= \sqrt{\frac{1}{6}\left[(\sigma_1-\sigma_2)^2+(\sigma_2-\sigma_3)^2+(\sigma_3-\sigma_1)^2\right]} \\ &= \sqrt{\frac{1}{2}(\sigma_1'^2+\sigma_2'^2+\sigma_3'^2)}\end{aligned}$$

In a state of pure shear ($\sigma_{12} \neq 0$, all other $\sigma_{ij} = 0$), the equivalent shear stress $\bar{\tau} = |\sigma_{12}|$. Note that this invariant is *not* equal to the limiting value τ_Y in the Tresca yield criterion, which is a material property of a given specimen.

Definition 6.1.15. The von Mises or **equivalent tensile stress** is defined to be

$$\begin{aligned}\bar{\sigma} &\equiv \sqrt{\frac{3}{2}\mathrm{tr}(\boldsymbol{\sigma'^2})} = \sqrt{\frac{3}{2}\sigma'_{ij}\sigma'_{ij}} \\ &= \sqrt{\frac{1}{2}\left[(\sigma_{11}-\sigma_{22})^2+(\sigma_{22}-\sigma_{33})^2+(\sigma_{33}-\sigma_{11})^2\right]+3(\sigma_{12}^2+\sigma_{23}^2+\sigma_{31}^2)} \\ &= \sqrt{\frac{1}{2}\left[(\sigma_1-\sigma_2)^2+(\sigma_2-\sigma_3)^2+(\sigma_3-\sigma_1)^2\right]} \\ &= \sqrt{\frac{3}{2}(\sigma_1'^2+\sigma_2'^2+\sigma_3'^2)} \\ &= \sqrt{3}\bar{\tau}\end{aligned}$$

In a state of pure tension ($\sigma_{11} \neq 0$, all other $\sigma_{ij} = 0$), the equivalent tensile stress $\bar{\sigma} = |\sigma_{11}|$.

Definition 6.1.16. The **third stress invariant** is defined to be

$$\begin{aligned}\bar{r} &\equiv \left[\frac{9}{2}\mathrm{tr}(\boldsymbol{\sigma'^3})\right]^{1/3} = \left[\frac{9}{2}\sigma'_{ik}\sigma'_{kj}\sigma'_{ji}\right]^{1/3} \\ &= \left[\frac{9}{2}(\sigma_1'^3+\sigma_2'^3+\sigma_3'^3)\right]^{1/3}\end{aligned}$$

Exercise 6.1.17. AKG 2.9, 2.11 through 2.16.

6.2 Strain Hardening

After yield, if the material is not unloaded, it will become *harder to further plastically deform* the more it plastically deforms. This phenomenon is called **hardening** (or sometimes *strain hardening*). Specifically, experiments suggest that the stress required to *continue* plastic deformation increases with increasing deformation. We call this evolving "resistance to plastic flow", as measured by the stress required for an additional infinitesimal increment of plastic strain, the **flow strength**, and to reflect this, we recast Y not as a single number, but as a *flow strength function* that evolves with the plastic deformation. In the most general case, the flow strength is a function of the strain *history*, because it must capture the idea that a material hardens as deformation continues.

Notice that because we must record the strain *history* in order to fully understand how hard it is to continue plastic deformation, we cannot simply assume Y is a function of the current plastic strain state, i.e., we cannot simply write $Y = Y(\varepsilon^P)$. To this end we must define some parameter that captures the total strain history in a manner that "accumulates" all increments of plastic strain that have, in some way, contributed to prior hardening.

This parameter is called the **equivalent tensile plastic strain** and is written $\bar{\varepsilon}^P$, so that we have

$$Y \equiv Y(\bar{\varepsilon}^P) > 0, \qquad Y(\bar{\varepsilon}^P = 0) = \sigma_Y.$$

Note that we have not yet given a definition of $\bar{\varepsilon}^P$. This must be done carefully and with nuance, so we postpone the formal definition until we have introduced a few more aspects of plasticity. Nevertheless, we observe that the second relation here formalizes the idea that the flow strength Y is always *at least* the yield strength σ_Y during plastic strain; in other words, when there has been *no* plastic strain history, the amount of stress it takes to begin plastic deformation is precisely the yield stress. Finally, we note that the way the hardening character of a particular sample evolves depending on the current (and past) loading, or in other words the specific model for $Y(\bar{\varepsilon}^P)$, depends on the material composition and the ambient temperature. In this section we will first note some general results and then later mention a few particular constitutive relations.

6.2.1 Types of strain hardening

There are two models for strain hardening: *isotropic hardening* and *kinematic hardening*. The isotropic hardening model is an *idealization* of what occurs for most metals, whereas the kinematic hardening model accurately captures certain experimentally-observed effects. In reality, metals exhibit a behavior that has elements of both models. However, for particular cases it may be acceptable to use the simpler isotropic hardening model.

The key difference between the two models is best illustrated in the case of cyclic loading. Consider a loading case where a specimen is brought in tension past the yield stress σ_Y, further loaded into the plastic regime, then unloaded before necking. Suppose the value of the flow strength Y just before unloading is recorded as σ_f. The specimen is fully unloaded to zero stress and then loaded in *compression* until it begins to plastically deform again. Suppose now that the value of the flow strength at the onset of plasticity in the compressive stage is recorded as σ_r.

Definition 6.2.1. **Isotropic hardening** is characterized as the situation where $|\sigma_f| = |\sigma_r|$, i.e., where the yield strength upon reversal of load is exactly equal in magnitude to the maximum flow strength attained before initial unloading.

Definition 6.2.2. **Kinematic hardening** is chracterized as the situation where $|\sigma_f| > |\sigma_r|$, i.e., yielding upon reversal happens at an *earlier* magnitude of the flow strength compared to the maximum flow strength attained before initial unloading. In other words, the apparent compressive yield stress is reduced after tensile loading. Experimentally, this phenomenon is called the *Bauschinger effect*.

Importantly, in general, we observe that *both* accumulated tensile plastic strain *and* accumulated compressive plastic strain contribute to successive hardening. That is, *regardless of sign*, with every increment of plastic strain, the strength increases. Hence, using the current value of the plastic strain alone to characterize the hardening behavior (and the stress in the material) is insufficient — we need a parameter that captures the accumulation of plasticity. This is a key departure from the constitutive modeling we have seen before, in elasticity, where the stress was independent of the loading history.

6.2.2 Temperature dependence

Let T_m denote the melting temperature of a sample in absolute units, e.g. Kelvin. Then,

if $T_{\text{test}} > 0.35T_m$, plastic flow is *temperature*- and *rate*-dependent.

Namely, materials exhibit relaxation and creep behavior in the plastic regime. We call this rate-dependence **viscoplasticity**. In general, the flow resistance increases with strain-rate, and the flow resistance decreases with increasing temperature.

Remark 6.2.3: The analog model for viscoplasticity is a spring in series with a dashpot-slider combination. The slider's static friction represents the yield strength of the material: the dashpot-slider combination does not flow until this stress is attained. Once the slider does admit strain, however, it must move with the dashpot, which captures the relaxation or creep aspect of viscoplastic flow.

6.3 Physical basis for metal plasticity

Metals are polycrystalline, made of crystalline *grains* which meet at *grain boundaries*. Each grain consists of an orderly atomic lattice, with atoms spaced in a regular pattern. However, all real metals contain defects called *dislocations* in their lattices. For example, a *line dislocation* occurs when one row of the lattice has one fewer atom than its neighboring rows, causing a local distortion of the lattice. It turns out that these dislocations cause the overall strength of the metal to decrease from the theoretical value calculated by assuming a perfect network of bonds. In fact,

plastic deformation is due to the motion of dislocations, which we call "slip". That is, plasticity occurs because the grains themselves plastically deform.

When we say that a dislocation has moved (by one atomic distance), we are actually describing the rupture of a bond and a formation of another bond in a neighboring atomic site that "transfers" the location of the defect. In this way, it is possible to cause plastic strain in a crystal without an existing dislocation by *creating* a dislocation that moves through the crystal as strain is applied. Moreover, because dislocation motion does not affect the average interatomic spacing in a metal lattice,

plastic deformation is an entirely incompressible (volume-preserving) process.

Slip generally occurs along atomic *planes* where the atomic density is highest. On those planes, the slip direction is generally the atomic direction in which, again, the density of atoms is highest. In particular, Schmid (1935) proposed that

- slip occurs when the magnitude of the shear stress resolved along a slip direction on a slip place reaches a *critical value*, enough to move a dislocation.

Mathematically, if $\mathbf{s}^\alpha$ is a slip direction and $\mathbf{m}^\alpha$ is the normal vector to a slip plane, then the slip criterion can be represented as

$$\tau^\alpha \equiv \mathbf{s}^\alpha \cdot \boldsymbol{\sigma} \mathbf{m}^\alpha = \tau^\alpha_{\text{cr}} \qquad \text{(onset of slip)},$$

where τ_{cr} is the **critical resolved shear stress**, a material property. In general, the critical resolved shear stress is a function of interatomic bond strength, spacing, etc., but it is worth repeating that

due to the presence of dislocations, the experimentally observed critical resolved shear stress τ_{cr} is much less than the ideal value[a].

[a] The ideal shear strength value is the estimated shear strength required to move an entire plane of atoms over by one atomic distance, overcoming all the interatomic bonds simultaneously, as in a theoretically perfect lattice.

Therefore, it is the "dislocation strength" that sets the value of τ_{cr}, and hence there are techniques to *engineer* the strength of metals based on how dislocations can (or cannot) move. Every metal lattice is endowed with an **intrinsic lattice resistance** τ_l, which corresponds with the energy to break any bonds at all. Covalently bonded materials like diamond, carbides, oxides, and nitrides have very high intrinsic lattice resistances, whereas metals have low intrinsic lattice resistances. Hence, strengthening of metals focuses on inhibiting dislocation motion. To this end, the following strengthening mechanisms for metals are commonly seen:

- **solid solution strengthening**: adding alloying elements either substitutionally (replacing a native atom in the lattice) or interstitially (fitting an atom between native lattice atoms). Alloy elements "roughen" the slip plane (as described by Ashby) and cause *lattice mismatch strains* ε_s which alter the native spacing c of atoms on slip planes. In both (substitutional and interstitial) cases, the presence of the alloying atom inhibits the passage of a dislocation relative to the intrinsic strength of the native atoms. The contribution of solid solution strengthening to the critical resolved shear stress scales as $\tau_{ss} \propto \varepsilon_s^{3/2} c^{1/2}$.

- **obstacle strengthening**: adding closely-spaced, small, hard particles which make it more difficult to physically move dislocations around. (For example, impurities can be dissolved at high temperature, and the resulting alloy is allowed to cool, which results in precipitation. Alternatively, oxides can be mixed into metal powder before it is sintered.) The strengthening contribution scales as $\tau_o \propto Gb/r$, where G is the shear modulus of the particle, b is the magnitude of the Burgers vector, which represents the characteristic length scale of a dislocation, and r is the characteristic length of the obstacle particle.

- **strain hardening**: intentionally creating more dislocations by plastically deforming the material causes the dislocations to be entangled and obstruct one another, which hinders the movement of *subsequent* dislocations. (This is the reason why materials which are unloaded post-yielding, then reloaded, appear to have higher yield strengths upon reloading.) If ρ is the dislocation density, the contribution of strain hardening to the critical resolved shear stress is $\tau_{sh} \propto \alpha G b \sqrt{\rho}$, where $0 < \alpha < 1$ is a constant.

Remark 6.3.1: A typical value of the dislocation density (measured as the total dislocation line length $\sum b_i$ per unit specimen volume) in a *well-annealed* high-purity single crystal is about $\rho = 10^8$ m/m^3. After plastic deformation, the density can grow to $\rho = 10^{15}$ m/m^3.

A good estimate for the critical resolved shear stress in a single crystal is then

$$\tau_{\text{cr}} = \tau_l + \tau_{ss} + \tau_o + \tau_{sh}.$$

From this, it has been experimentally found that for a *polycrystalline* metal, the critical resolved shear stress is approximately 1.5 times that of a single crystal. The reason is because polycrystals contain grains oriented in all directions —the favorably oriented grains (i.e., grains with slip planes closest to the direction of maximal shear) will yield first.

In a uniaxial tension test of a polycrystal, the plane of maximum shear stress is oriented at 45° to the tensile axis and the magnitude of the shear stress on this plane is half the magnitude of the applied tensile load. Therefore the *tensile stress at yielding* has twice the magnitude of the shear stress on the most favorable (45°-oriented) slip plane, and in turn this shear stress is (about) 1.5 times the critical resolved shear stress in a single crystal. Putting this all together, the tensile yield strength σ_y of a polycrystal can be related to τ_{cr} for a single crystal by

$$\sigma_y \approx 3\tau_{cr}.$$

Finally, *grain boundaries* contribute to strengthening by acting as obstacles against dislocation movement. In simple terms, dislocations can be thought of as "piling up" at grain boundaries, at which point additional energy must be expended in order to transmit slip across grains. The **Hall-Petch effect** quantifies the size effect of grains. If the characteristic grain size (i.e., the diameter of the average polycrystal grain) is D, then the yield strength of the specimen is approximately given by

$$\sigma_y \approx \sigma_0 + \frac{k}{\sqrt{D}},$$

where σ_0 and k are material constants. (Here, σ_0 is approximately the yield strength of a *large-grained* polycrystal.)

Remark 6.3.2: One method of experimentally determining the strength of a sample without destroying it is through *hardness testing*. Hardness represents the material's resistance to local plastic deformation. In general, the hardness H is related to the yield strength by

$$H = 3\sigma_y.$$

Exercise 6.3.3. AKG 3.14.

6.4 Constitutive theories for one-dimensional plasticity

In this section we will record the standard theories for plasticity in one spatial dimension. In accord with our previous discussions on the evolution of hardening and the lack of a one-to-one correspondence between stress and strain for a general elastic-plastic material, we cannot directly prescribe a "stress-strain relationship" as we have been doing for elastic and viscoelastic materials. Instead, our goal will be to describe how an *increment* of plastic strain, as described mathematically by the notion of a plastic strain *rate* $\dot{\bar{\varepsilon}}^P$, evolves based on the current and former states of the system. Our material model will thus be an *assumed form* for the hardening function $Y = Y(\bar{\varepsilon}^P)$, together with the rule

that $\bar{\sigma} = Y$ during plastic deformation, accompanied by an evolution equation for $\dot{\bar{\varepsilon}}^P$ in terms of stress components.

Because we are working first in one dimension, arguments will be made in terms of scalar quantities, with reference to the direction of testing. Later, we will generalize these results to three dimensions, working with tensors instead. Recall that we will work only in terms of true strain and true stress.

Remark 6.4.1: Several common models for the form of $Y = Y(\bar{\varepsilon}^P)$ include:

- the power-law hardening function, which has three parameters Y_0, K, n:

$$Y(\bar{\varepsilon}^P) = Y_0 + K(\bar{\varepsilon}^P)^n,$$

  which reduces to a *linear* model when $n = 1$, and

- the Voce equation, which has three parameters Y_0, Y_s, H_0:

$$Y(\bar{\varepsilon}^P) = Y_s - (Y_s - Y_0)\exp\left(-\frac{H_0}{Y_s}\bar{\varepsilon}^P\right)$$

In both cases Y_0 represents an initial value of the yield function, which coincides with the yield strength of the material. The most appropriate model is typically chosen using the aid of curve-fitting routines; recall that it is possible to extract experimental hardening data simply from a *uniaxial tension* test by plotting the true stress $\sigma(= Y)$ as a function of the plastic strain $\varepsilon^P(= \bar{\varepsilon}^P)$ after the onset of yield.

Exercise 6.4.2. AKG A.12 and A.19. Here, the *perfectly-plastic* assumption is equivalent to setting $Y(\bar{\varepsilon}^P) = \text{const.} = \sigma_y$.

6.4.1 Kinematics of strain

We assume that the total strain ε can be decomposed into an elastic part and a plastic part, as can the strain rate:

$$\varepsilon = \varepsilon^E + \varepsilon^P = \frac{\sigma}{E} + \varepsilon^P$$
$$\dot{\varepsilon} = \dot{\varepsilon}^E + \dot{\varepsilon}^P$$

Thus, the rate-of-working per unit volume $\dot{\psi} = \sigma\dot{\varepsilon}$ can be likewise broken down:

$$\dot{\psi} = \sigma\dot{\varepsilon} = \sigma\dot{\varepsilon}^E + \sigma\dot{\varepsilon}^P.$$

The term $\sigma\dot{\varepsilon}^E$ represents the rate at which work must be applied to a unit volume in order to cause elastic deformation (i.e., stretching of interatomic bonds) at an elastic strain rate ε^E. Hence, this stress power is *recoverable*. However, the plastic stress power is *dissipative*, $\mathcal{D} \equiv \sigma\dot{\varepsilon}^P \geq 0$, because in plasticity bonds are physically broken as dislocations move.

6.4.2 Flow strength

Recall that under plastic flow, the flow strength is defined in terms of the equivalent tensile plastic strain, $Y \equiv Y(\bar{\varepsilon}^P)$. Moreover, recall that in order for plastic flow to occur,

the material must be meeting the yield criterion, $\bar{\sigma} = Y$. In the one-dimensional case, there is only one stress component, so $\bar{\sigma} = |\sigma|$, so a criterion for plastic flow to occur is

$$\bar{\sigma} = |\sigma| = Y \quad \text{and} \quad \sigma\dot{\varepsilon} > 0 \qquad \text{(plastic flow)}.$$

To properly define the equivalent tensile plastic strain $\bar{\varepsilon}^P$, we make the observation that the flow strength depends on the integral of the time history of the plastic strain rate, since for an arbitrary loading path *both* negative and positive increments of plastic strain contribute to the present flow strength (see the discussion above). To that end we must first define the equivalent tensile plastic strain *rate*.

Definition 6.4.3. The **equivalent tensile plastic strain rate** is defined to be the absolute value of the plastic strain rate at a given time,

$$\dot{\bar{\varepsilon}}^P \equiv |\dot{\varepsilon}^P| \geq 0.$$

Definition 6.4.4. The **equivalent tensile plastic strain** is defined to be the time integral of the equivalent tensile plastic strain rate,

$$\bar{\varepsilon}^P \equiv \int_0^t \dot{\bar{\varepsilon}}^P(\tau)\, d\tau = \int_0^t |\dot{\varepsilon}^P(\tau)|\, d\tau.$$

For the special case of one-dimensional loading, $\bar{\varepsilon}^P = |\varepsilon^P|$.

Remark 6.4.5: In rate-independent plasticity, time has no constitutive significance. Therefore we may speak of "increments of plastic strain" $d\varepsilon^P$ and hence define the equivalent tensile plastic strain to be

$$\bar{\varepsilon}^P \equiv \int_{\mathcal{C}} |d\varepsilon^P|,$$

where the path $\mathcal{C}$ to integrate over is a curve within the true stress-true strain space. Clearly, $\bar{\varepsilon}^P$ is a nondecreasing function.

Example 6.4.6

Consider a one-meter long rigid-plastic bar. Extend it in tension past the tensile yield strength, then compress it past the corresponding compressive yield strength, then pull in tension again until it returns to one meter. In this final state $\varepsilon = 0 \implies \varepsilon^P = 0$, *but* $\bar{\varepsilon}^P > 0$. Namely, if the bar is *then* re-loaded in tension, the yield point will be higher than the original yield strength: the bar has been hardened.

Remark 6.4.7: We can build up an *analog model* for plasticity, in accordance with the kinematic decomposition of strain, $\varepsilon = \varepsilon^E + \varepsilon^P$. In particular we can combine an elastic spring (with characteristic modulus E) and a *plastic slider*, which is an element with a threshold strength σ_Y along with the property that when the load is released, any deformation is permanent. (The plastic slider is like a box with static friction and dynamic friction that evolves with the strain.) Hence if the load is removed, only the spring relaxes, but the slider remains in a state of permanent deformation.

Definition 6.4.8. The **hardening modulus** or *strain-hardening rate*[2] is denoted H and is a function of the equivalent plastic strain, $H = H(\bar{\varepsilon}^P)$. It is defined to be the instantaneous slope of the graph of flow strength versus equivalent plastic strain,

$$H(\bar{\varepsilon}^P) \equiv \frac{dY(\bar{\varepsilon}^P)}{d\bar{\varepsilon}^P} \geq 0,$$

with equality for *non-hardening* models such as the "perfectly plastic model", which has $H(\bar{\varepsilon}^P) = 0$ for all $\bar{\varepsilon}^P$, and inequality otherwise, which corresponds to models of *strain hardening*.

Finally, we observe that in plasticity,

the stress and the rate of plastic strain are *co-directional*, i.e.,

$$\sigma > 0 \iff \dot{\varepsilon}^P > 0; \quad \sigma < 0 \iff \dot{\varepsilon}^P < 0,$$

or in words, a tensile-directed stress must tend to cause elongation, and a compressive-directed stress must tend to cause contraction. As such

$$\frac{\dot{\varepsilon}^P}{|\dot{\varepsilon}^P|} = \frac{\sigma}{|\sigma|} \implies \dot{\varepsilon}^P = |\dot{\varepsilon}^P|\frac{\sigma}{|\sigma|} = \dot{\bar{\varepsilon}}^P \frac{\sigma}{|\sigma|}.$$

6.4.3 Rate-independent, isotropic hardening

Assume we are given a form of $Y(\bar{\varepsilon}^P)$ such that for every attainable value of the equivalent plastic strain, we can identify the value of stress associated with yielding there. We then have two governing principles:

1. Plastic strain cannot increase when $|\sigma| < Y(\bar{\varepsilon}^P)$. This is said to be the *elastic state*. Hence
$$|\sigma| < Y(\bar{\varepsilon}^P) \implies \dot{\varepsilon}^P = 0.$$
2. Plastic strain may increase when $|\sigma| = Y(\bar{\varepsilon}^P)$. This is said to be the *elastic-plastic state*. Hence
$$|\sigma| = Y(\bar{\varepsilon}^P) \implies \dot{\varepsilon}^P \neq 0 \quad \text{possible.}$$

Note the careful wording in the second principle. It is not *guaranteed*, only *possible*, that we have a nonzero plastic strain rate when we are "on the yield surface". In fact, when $|\sigma| = Y(\bar{\varepsilon}^P)$, we may *either* have:

- Elastic unloading: $|\sigma| = Y(\bar{\varepsilon}^P)$ and $\sigma\dot{\varepsilon} < 0$, such that $\dot{\varepsilon} \neq 0$ but $\dot{\bar{\varepsilon}}^P = 0$, or
- Plastic loading: $|\sigma| = Y(\bar{\varepsilon}^P)$ and $\sigma\dot{\varepsilon} > 0$, such that $\dot{\varepsilon} \neq 0$ and $\dot{\bar{\varepsilon}}^P > 0$.

The plastic loading case is the "interesting" case here, because it represents an *evolution* of the plastic strain. Namely, in order to *maintain* a state where $\dot{\bar{\varepsilon}}^P > 0$, we must have that $|\sigma| = Y(\bar{\varepsilon}^P)$ *and* that $\frac{d}{dt}|\sigma| = \frac{d}{dt}Y(\bar{\varepsilon}^P)$. Combined with the consistency condition, we arrive at the following **flow rule**, or evolution equation, for the plastic strain rate $\dot{\varepsilon}^P$:

[2] Observe that this is a "rate" of hardening *per increment in equivalent plastic strain*, *not* a rate with respect to time.

Key Equation 6.4.9 (Flow rule, one-dimensional, rate-independent, isotropic hardening)

In rate-independent, isotropic hardening, the plastic strain rate $\dot{\varepsilon}^P$ evolves according to the equivalent plastic strain rate $|\dot{\varepsilon}^P|$, the equivalent stress $|\sigma|$ relative to the hardening function $Y(\bar{\varepsilon}^P)$, and the direction of the stress power $\sigma\dot{\varepsilon}$ as:

$$\dot{\varepsilon}^P = \begin{cases} 0, & |\sigma| < Y(\bar{\varepsilon}^P) \quad \text{(elastic state)} \\ 0, & |\sigma| = Y(\bar{\varepsilon}^P) \quad \text{but} \quad \sigma\dot{\varepsilon} < 0 \quad \text{(elastic unloading)} \\ \frac{E}{E+H(\bar{\varepsilon}^P)}\dot{\varepsilon}, & |\sigma| = Y(\bar{\varepsilon}^P) \quad \text{and} \quad \sigma\dot{\varepsilon} > 0 \quad \text{(plastic loading)} \end{cases}$$

Remark 6.4.10: It may be helpful to think of time derivatives instead as strain increments, i.e.:

- *Elastic state*: if $|\sigma| < Y(\bar{\varepsilon}^P)$, and an increment of strain $d\varepsilon$ is applied, then $d\varepsilon^P = 0$ (and $d\bar{\varepsilon}^P = 0$). Note that $d\sigma = Ed\varepsilon$.
- *Elastic unloading*: if $|\sigma| = Y(\bar{\varepsilon}^P)$, and $\sigma d\varepsilon < 0 \iff \text{sign}(\sigma) \neq \text{sign}(d\varepsilon)$, and an increment of strain $d\varepsilon$ is applied, then $d\varepsilon^P = 0$ (and $d\bar{\varepsilon}^P = 0$). Again, $d\sigma = Ed\varepsilon$.
- *Plastic loading*: if $|\sigma| = Y(\bar{\varepsilon}^P)$, and $\sigma d\varepsilon > 0 \iff \text{sign}(\sigma) = \text{sign}(d\varepsilon)$, and an increment of strain $d\varepsilon$ is applied, then $d\varepsilon^P \neq 0$ (and $d\bar{\varepsilon}^P > 0$). In this case, particularly, $d\varepsilon^P = d\varepsilon - d\sigma/E$, with $|\sigma + d\sigma| = Y(\bar{\varepsilon}^P + d\bar{\varepsilon}^P)$ and $d\bar{\varepsilon}^P = |d\varepsilon^P|$.

Remark 6.4.11: During the elastic state or during elastic unloading, the slope of the (true) stress-(true) strain curve is equal to the elastic modulus E. However, during plastic loading, the slope is given by a *tangent modulus*,

$$E_{\text{tan}} = \frac{EH(\bar{\varepsilon}^P)}{E + H(\bar{\varepsilon}^P)}.$$

Exercise 6.4.12. AKG 3.13, 3.11.

6.4.4 Rate-dependent, isotropic hardening

Some materials (like hot metals) exhibit a dependence of the flow strength not only on the plastic strain, but on the plastic strain *rate*. In general, to achieve a faster plastic strain rate, a higher stress must be applied. Accordingly, we generalize the concept of the hardening function Y to include a rate-dependent term. Specifically, we assume that the *viscoplastic* flow strength $\mathcal{S}(\bar{\varepsilon}^P, \dot{\bar{\varepsilon}}^P)$ can be written in the form

$$\mathcal{S}(\bar{\varepsilon}^P, \dot{\bar{\varepsilon}}^P) = Y(\bar{\varepsilon}^P)\left(\frac{\dot{\bar{\varepsilon}}^P}{\dot{\varepsilon}_0}\right)^m,$$

with the usual rate-independent hardening function $Y(\bar{\varepsilon}^P)$ being multiplied by a term which depends on the strain rate $\dot{\bar{\varepsilon}}^P$ nondimensionalized by a *reference strain rate* $\dot{\varepsilon}_0 > 0$, raised to a *rate sensitivity parameter* $0 < m \leq 1$. Note that $\dot{\varepsilon}_0$ and m are usually taken to be material constants. When that as $m \to 0$, we recover the rate-independent formulation.

When $m = 1$, the behavior is linearly dependent on the strain rate, as in a Newtonian fluid.

When the material is in a state of plastic loading, we now have the condition that $|\sigma| = \mathcal{S}(\bar{\varepsilon}^P, \dot{\bar{\varepsilon}}^P)$, which when combined with the consistency condition leads to the following flow rule:

$$\dot{\varepsilon}^P = \underbrace{\dot{\varepsilon}_0 \left(\frac{|\sigma|}{Y(\bar{\varepsilon}^P)} \right)^{1/m}}_{\dot{\bar{\varepsilon}}^P} \frac{\sigma}{|\sigma|}$$

Recall that at higher temperatures, the plastic behavior of metals becomes heavily dependent on the strain rate. Above $T > 0.5T_m$, the general form given previously is not specific enough. In particular, we need to build in an explicit temperature dependence. This happens by replacing the previously *constant* material parameter $\dot{\varepsilon}_0$ with a *temperature-dependent rate sensitivity parameter*,

$$\dot{\varepsilon}_0 = \dot{\varepsilon}_0(T) \equiv A \exp\left(-\frac{Q}{RT} \right),$$

where we have used the familiar Arrhenius relationship as a model. Here Q is an *activation energy* corresponding to the energy required for lattice self-diffusion, and R is the gas constant. Note that the parameter A has units of inverse time, or frequency.

The following flow rule summarizes both cases:

Key Equation 6.4.13 (Flow rule, one-dimensional, rate-dependent, isotropic hardening)

In rate-dependent, isotropic hardening, the plastic strain rate $\dot{\varepsilon}^P$ evolves according to the equivalent plastic strain rate $|\dot{\varepsilon}^P|$, the stress σ, and the strain rate sensitivity parameter m according to

$$\dot{\varepsilon}^P = \dot{\varepsilon}_0 \left(\frac{|\sigma|}{Y(\bar{\varepsilon}^P)} \right)^{1/m} \frac{\sigma}{|\sigma|},$$

where

$$\dot{\varepsilon}_0 = \begin{cases} \text{const.}, & T \leq 0.35T_m \\ A \exp(-Q/RT), & T \geq 0.5T_m \end{cases}$$

Note that in the high-temperature case, the flow rule can be rewritten to solve for $Y(\bar{\varepsilon}^P)$ in order to show that it, too, also evolves with temperature. For $0.35T_m < T < 0.5T_m$, the activation energy is a complex function and the simple Arrhenius relationship cannot be used.

In the high-temperature ($T > 0.5T_m$) viscoplastic model, there is no more formal yield stress, i.e., $Y(0) = 0$. The immediate consequence is that *for all nonzero stresses, plastic strain occurs*:

$$|\sigma| > 0 \implies |\dot{\varepsilon}^P| > 0$$

Exercise 6.4.14. AKG 3.18. Show that if $n = 1$, we recover the Maxwell model (hint: let $B = \eta$). Moreover show that if $n \gg 1$ we recover a rate-independent model with S like an effective yield stress.

Yield threshold

For some materials ($0.35T_m \leq T \leq 0.5T_m$), it is necessary to introduce a **yield threshold** $Y_{\text{th}}(\bar{\varepsilon}^P) > 0$ for each value of $\bar{\varepsilon}^P$, such that yield may *only* occur when

$$|\sigma| > Y_{\text{th}}(\bar{\varepsilon}^P);$$

note the strict inequality. This formulation is more general than the standard isotropic hardening model, which results in plastic flow for any stress level.

With this yield threshold the flow strength becomes

$$\mathcal{S}(\bar{\varepsilon}^P, \dot{\bar{\varepsilon}}^P) = Y_{\text{th}}(\bar{\varepsilon}^P) + Y(\bar{\varepsilon}^P) \left(\frac{\dot{\bar{\varepsilon}}^P}{\dot{\varepsilon}_0} \right)^m,$$

so that

$$|\sigma| - Y_{\text{th}}(\bar{\varepsilon}^P) = Y(\bar{\varepsilon}^P) \left(\frac{\dot{\bar{\varepsilon}}^P}{\dot{\varepsilon}_0} \right)^m,$$

which yields the following form of the flow rule:

$$\dot{\varepsilon}^P = \dot{\varepsilon}_0 \left\langle \frac{|\sigma| - Y_{\text{th}}(\bar{\varepsilon}^P)}{Y(\bar{\varepsilon}^P)} \right\rangle^{1/m} \frac{\sigma}{|\sigma|},$$

where the Macauley angle-bracket notation means

$$\langle u \rangle \equiv \begin{cases} 0, & u \leq 0 \\ u, & u > 0. \end{cases}$$

Hence, when $|\sigma| \leq Y_{\text{th}}(\bar{\varepsilon}^P)$, the response is *purely elastic*, and plastic flow *only occurs* when $|\sigma| > Y_{\text{th}}(\bar{\varepsilon}^P)$. In the special case when $Y_{\text{th}}(\bar{\varepsilon}^P) = 0$, we recover the usual rate-dependent, isotropic hardening yield threshold. Recall that in that formulation, plastic flow occurs for *all* non-zero values of $|\sigma|$.

Remark 6.4.15: Another special case is worth mentioning. If we take the threshold strength to be a constant, say $Y_{\text{th}} = \sigma_y = \text{const.}$, and assume a perfectly plastic response, $Y = \text{const.}$, we are left with

$$|\sigma| = \sigma_y + k|\dot{\varepsilon}^P|^m,$$

where k acts like a *viscosity*. The units of k are *exactly* those of a viscosity when the sensitivity parameter $m = 1$, in which case we have the **Bingham viscoplastic model**,

$$|\sigma| = \sigma_y + \mu_p|\dot{\varepsilon}^P|,$$

where we have rewritten k as the plastic viscosity μ_p. This model is appropriate for mayonnaise and toothpaste, among other materials.

6.4.5 Rate-independent, kinematic and isotropic hardening

Recall that for most materials, the isotropic hardening model is only an idealization. Rather, if the direction of stress is reversed post-yielding, the yield strength in the reverse direction is typically experimentally seen to be lower than the flow strength upon initial de-loading (the *Bauschinger effect*). The physical explanation for this effect is two-fold. For the sake of concreteness, let us assume that the initial direction of stress was tensile, and thus the reverse direction of loading is compressive.

1. Bringing the material past its initial yield point in tension resulted in cold-working (strain-hardening) of the material, which generated a great deal of dislocations, most of which "got stuck" at grain boundaries or at obstacles like precipitates. This generates **back stress** in the material, because the dislocations are "repelled" from each other where they are piled up. Then, during the reversal of loading, the back stress makes it *easier* to move these dislocations in the reverse direction.

2. Upon reversal of loading, some of the dislocations generated as a result of the compressive loading are "equal and opposite" to those generated as a result of the tensile loading. Therefore, they can "cancel each other out", and thus the net effect of the original (tensile) strain-hardening is lessened, resulting in a lower apparent yield strength in compression.

Therefore, in order to describe the flow rule for a kinematically hardening material, it is necessary to incorporate the effect of the back stress in the *free-energy function* which is then used to derive the dissipation inequality. Namely, we will leave the elastic free-energy contribution unchanged, but prescribe a model for the plastic free-energy function which records the effect of the back stress. Specifically, we write

$$\begin{aligned}\psi &= \psi^E + \psi^P \\ &= \frac{1}{2}E(\varepsilon^E)^2 + \frac{1}{2}CA^2,\end{aligned}$$

where $C > 0$ is the **back stress modulus** and A is a dimensionless strain-like internal variable that is work conjugate to the back stress. We will assume that A follows an evolution equation

$$\dot{A} = \dot{\varepsilon}^P - \gamma A\dot{\bar{\varepsilon}}^P, \qquad A(0) = 0,$$

where $\gamma > 0$ characterizes the *dynamic recovery* of the model. The important consquence of this formula is the development of the back stress term and the **effective stress** term

$$\sigma_{\text{back}} \equiv CA \implies \sigma_{\text{eff}} = \sigma - \sigma_{\text{back}}.$$

Essentially, we will use σ_{eff} everywhere we previously used σ, as this effective stress now accounts for the "offset" of the back stress[3]. Hence the co-directionality assumption becomes

$$n^P \equiv \frac{\dot{\varepsilon}^P}{|\dot{\varepsilon}^P|} = \frac{\sigma_{\text{eff}}}{|\sigma_{\text{eff}}|} = \frac{\sigma - \sigma_{\text{back}}}{|\sigma - \sigma_{\text{back}}|}.$$

Moreover we will recast $dY/d\bar{\varepsilon}^P$ to be the *isotropic hardening modulus* H_{iso}, and introduce an **overall hardening modulus** which includes the dynamic recovery constribution,

$$H \equiv H_{\text{iso}} + C(1 - \gamma A n^P).$$

Then the steps of the computation of the flow rule matches those for the case of isotropic hardening, and we have the following:

Key Equation 6.4.16 (Flow rule, one-dimensional, rate-independent, kinematic hardening)

In rate-independent, kinematic hardening, the plastic strain rate $\dot{\varepsilon}^P$ evolves according to the equivalent plastic strain rate $|\dot{\varepsilon}^P|$, the effective stress σ_{eff} relative to the

[3]Then, the rate of change of free energy can be expressed as $\dot{\psi} = \sigma\dot{\varepsilon}^E + CA\dot{A} = \sigma\dot{\varepsilon}^E + \sigma_{\text{back}}\dot{\varepsilon}^P - C\gamma A^2\dot{\bar{\varepsilon}}^P$, for which the dissipation inequality becomes $\mathcal{D} = (\sigma_{\text{eff}}n^P + C\gamma A^2)\dot{\bar{\varepsilon}}^P \geq 0$.

hardening function $Y(\bar{\varepsilon}^P)$, and the direction of the effective stress power $\sigma_{\text{eff}}\dot{\varepsilon}$ as:

$$\dot{\varepsilon}^P = \begin{cases} 0, & \sigma_{\text{eff}} < Y(\bar{\varepsilon}^P) \quad \text{(elastic state)} \\ 0, & \sigma_{\text{eff}} = Y(\bar{\varepsilon}^P) \quad \text{but} \quad \sigma_{\text{eff}}\dot{\varepsilon} < 0 \quad \text{(elastic unloading)} \\ \frac{E}{E+H}\dot{\varepsilon}, & \sigma_{\text{eff}} = Y(\bar{\varepsilon}^P) \quad \text{and} \quad \sigma_{\text{eff}}\dot{\varepsilon} > 0 \quad \text{(plastic loading)} \end{cases}$$

We can still use the usual models (e.g., power-law, Voce) for the form of the hardening function Y.

6.4.6 Rate-dependent, kinematic and isotropic hardening

We can apply the same generalizations as in the purely-isotropic case to a model which includes temperature dependence and an (optional) yield threshold to attain the flow rule for rate-dependent viscoplasticity with *kinematic hardening*. The major difference is the use of σ_{eff} in place of σ:

$$\dot{\varepsilon}^P = \dot{\varepsilon}_0 \left\langle \frac{\sigma_{\text{eff}} - Y_{\text{th}}(\bar{\varepsilon}^P)}{Y(\bar{\varepsilon}^P)} \right\rangle^{1/m} \frac{\sigma_{\text{eff}}}{|\sigma_{\text{eff}}|},$$

where for $T \geq 0.5T_m$,

$$\dot{\varepsilon_0} = \dot{\varepsilon_0}(T) = A \exp\left(-\frac{Q}{RT}\right).$$

6.5 Constitutive theories for three-dimensional plasticity

Now we can extend the concepts developed for one-dimensional plasticity to a full three-dimensional theory. The reasoning is identical to the one-dimensional case, but in general we will be working in terms of tensors, not just scalars. However, we will make use of the **von Mises** (or equivalent tensile) stress $\bar{\sigma}$,

$$\bar{\sigma} \equiv \sqrt{\frac{3}{2}}|\boldsymbol{\sigma}'| = \sqrt{\frac{1}{2}\left[(\sigma_{11}-\sigma_{22})^2 + (\sigma_{22}-\sigma_{33})^2 + (\sigma_{33}-\sigma_{11})^2\right] + 3(\sigma_{12}^2 + \sigma_{23}^2 + \sigma_{31}^2)},$$

and analagously we will define the **equivalent tensile plastic strain rate** $\dot{\bar{\varepsilon}}^P$ as

$$\dot{\bar{\varepsilon}}^P \equiv \sqrt{\frac{2}{3}}|\dot{\boldsymbol{\varepsilon}}^P| = \sqrt{\frac{2}{9}\left[(\dot{\varepsilon}_{11}-\dot{\varepsilon}_{22})^2 + (\dot{\varepsilon}_{22}-\dot{\varepsilon}_{33})^2 + (\dot{\varepsilon}_{33}-\dot{\varepsilon}_{11})^2\right] + \frac{4}{3}(\dot{\varepsilon}_{12}^2 + \dot{\varepsilon}_{23}^2 + \dot{\varepsilon}_{31}^2)}.$$

Here, the tensor $\dot{\boldsymbol{\varepsilon}}^P$ is the three-dimensional plastic strain rate tensor, which comes from the additive decomposition of the strain tensor into elastic and plastic parts,

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^E + \boldsymbol{\varepsilon}^P \implies \dot{\boldsymbol{\varepsilon}} = \dot{\boldsymbol{\varepsilon}}^E + \dot{\boldsymbol{\varepsilon}}^P.$$

Analagous to the factor $\sqrt{3/2}$ in the definition of the equivalent tensile stress, the factor $\sqrt{2/3}$ in the definition of the equivalent tensile strain rate connects this function of a three-dimensional strain state back to a uniaxial test. Moreover, the factor of $\sqrt{2/3}$ ensures that the rate of plastic working can be written entirely in terms of the equivalent tensile stress and the equivalent tensile strain rate, $\boldsymbol{\sigma} \cdot \dot{\boldsymbol{\varepsilon}}^P = \bar{\sigma}\dot{\bar{\varepsilon}}^P$.

Finally,, the **equivalent tensile plastic strain** is the integral (accumulation) of the equivalent plastic strain rate,

$$\bar{\varepsilon}^P \equiv \int_0^t \dot{\bar{\varepsilon}}^P(\tau)\, d\tau.$$

Again, the integration variable τ need not correspond to time; rather, it represents an integration over the *plastic loading history* of the specimen.

Example 6.5.1

Consider two rigid-plastic cubes with initial side lengths 1. One cube is stretched in *simple tension* such that the new side lengths are 4, 1/2, and 1/2. The other cube is stretched in *plane strain tension* so that the new side lengths are 4, 1, and 1/4. Which cube has a higher equivalent plastic strain?

Solution. For a rigid-plastic material with a constant (non-changing) flow direction,

$$\Delta\bar{\varepsilon}^P = \Delta\bar{\varepsilon} = \sqrt{\frac{2}{3}}|\Delta\boldsymbol{\varepsilon}|.$$

In the first case $\boldsymbol{\varepsilon} = \mathbf{0}$ in the reference configuration, and the final strain is

$$[\boldsymbol{\varepsilon}] = \begin{bmatrix} \ln(4/1) & 0 & 0 \\ 0 & \ln((1/2)/1) & 0 \\ 0 & 0 & \ln((1/2)/1) \end{bmatrix}$$

so that

$$\bar{\varepsilon}^P = \sqrt{\frac{2}{3}}|\boldsymbol{\varepsilon}| = \sqrt{\frac{2}{3}}\{(\ln 4)^2 + (\ln(1/2))^2 + (\ln(1/2))^2\}^{1/2} = \ln(4)$$

.

For the second case

$$[\boldsymbol{\varepsilon}] = \begin{bmatrix} \ln(4/1) & 0 & 0 \\ 0 & \ln(1/1) & 0 \\ 0 & 0 & \ln((1/4)/1) \end{bmatrix}$$

and the same computation yields $\bar{\varepsilon}^P = \sqrt{\frac{4}{3}}\ln(4) > \ln(4)$.

Therefore the plane strain tension sample has attained a higher equivalent plastic strain, and has therefore hardened more.

Remark 6.5.2: It will often happen that both sides of an equation in a plasticity problem will be expressible as *rates*, i.e., as time derivatives of quantities. If there is no explicit rate dependence, then both sides can be freely integrated with respect to time.

6.5.1 Rate-independent, isotropic hardening

For small deformations (strains up to about 5%) wherein the small elastic strain tensor $\boldsymbol{\varepsilon}^E$ can be used, the theory is very closely related to the one-dimensional case. In the case of the plastic strain tensor, we typically use the Hencky strain $\boldsymbol{\varepsilon} = \ln \mathbf{U}$. Then, the total strain follows the decomposition

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^E + \boldsymbol{\varepsilon}^P, \qquad \text{tr } \boldsymbol{\varepsilon}^P = 0,$$

from which it follows that

$$\dot{\boldsymbol{\varepsilon}} = \dot{\boldsymbol{\varepsilon}}^E + \dot{\boldsymbol{\varepsilon}}^P, \qquad \text{tr } \dot{\boldsymbol{\varepsilon}}^P = 0.$$

The rate-of-working per unit volume is then

$$\dot{\psi} = \boldsymbol{\sigma} \cdot \dot{\varepsilon} = \boldsymbol{\sigma} \cdot \dot{\varepsilon}^E + \boldsymbol{\sigma} \cdot \dot{\varepsilon}^P,$$

but the elastic stress power $\boldsymbol{\sigma} \cdot \dot{\varepsilon}^E$ is recoverable, so the dissipation is entirely due to the plastic stress power. Moreover, because $\dot{\boldsymbol{\varepsilon}}^P$ is *deviatoric*, the tensor contraction $\boldsymbol{\sigma} \cdot \dot{\varepsilon}^P$ is equivalent to $\boldsymbol{\sigma}' \cdot \dot{\varepsilon}^P$, so the dissipation inequality reads $\mathcal{D} = \boldsymbol{\sigma}' \cdot \dot{\varepsilon}^P \geq 0$. The stress $\boldsymbol{\sigma}$ depends only on the *elastic* part of the strain, with the usual small-deformation elastic constitutive relation

$$\boldsymbol{\sigma} = 2G(\boldsymbol{\varepsilon}^E)' + \kappa(\text{tr } \boldsymbol{\varepsilon}^E)\mathbf{I}.$$

On the other hand, the *plastic* part of the strain evolves according to the flow rule, which in three dimensions is given by the codirectionality *between the stress deviator and the plastic strain rate*:

$$\mathbf{N}^P = \frac{\dot{\boldsymbol{\varepsilon}}^P}{|\dot{\boldsymbol{\varepsilon}}^P|} = \frac{\boldsymbol{\sigma}'}{|\boldsymbol{\sigma}'|} \implies \dot{\boldsymbol{\varepsilon}}^P = \frac{3}{2}\dot{\bar{\varepsilon}}^P \frac{\boldsymbol{\sigma}'}{\bar{\sigma}}.$$

Observe that the result of the three-dimensional co-directionality assumption is a relation between the stress and the plastic strain *rate*; this is a hallmark of plasticity theory. Putting this together and rearranging, we see that

$$\begin{aligned}\dot{\boldsymbol{\varepsilon}} &= \dot{\boldsymbol{\varepsilon}}^E + \dot{\boldsymbol{\varepsilon}}^P \\ &= \left[\frac{1+\nu}{E}\dot{\boldsymbol{\sigma}} - \frac{\nu}{E}(\text{tr } \dot{\boldsymbol{\sigma}})\mathbf{I}\right] + \left[\frac{3}{2}\dot{\bar{\varepsilon}}^P \frac{\boldsymbol{\sigma}'}{\bar{\sigma}}\right].\end{aligned}$$

We can integrate both sides in time and work in terms of *plastic strain increments*, which is more useful *if the stress state is known a priori*. Namely,

$$d\varepsilon_{ij} = \left[\frac{1+\nu}{E}d\sigma_{ij} - \frac{\nu}{E}\sigma_{kk}\delta_{ij}\right] + \left[\frac{3}{2}d\bar{\varepsilon}^P \frac{\sigma'_{ij}}{\bar{\sigma}}\right],$$

where the first bracketed term is the increment in elastic strain $d\varepsilon_{ij}^E$, and the second bracketed term is the increment in plastic strain $d\varepsilon_{ij}^P$.

The codirectionality flow rule is sufficient to describe the behavior of a specimen when it is possible to go "directly" between an initial and a final state, for example by computing an integral over the plastic strain increment equation. For other situations when a distinct *evolution equation* is required, the flow rule has the evolutionary form

$$\begin{aligned}\dot{\boldsymbol{\varepsilon}}^P &= \frac{3}{2}\dot{\bar{\varepsilon}}^P \frac{\boldsymbol{\sigma}'}{\bar{\sigma}}, \\ \dot{\bar{\varepsilon}}^P &= \chi\left(\frac{3G}{3G + H(\bar{\varepsilon}^P)}\right)\left(\frac{\boldsymbol{\sigma}' \cdot \dot{\boldsymbol{\varepsilon}}}{\bar{\sigma}}\right),\end{aligned}$$

where $H(\bar{\varepsilon}^P) = dY(\bar{\varepsilon}^P)/d\bar{\varepsilon}^P$ is the strain-hardening rate as before, and the switching parameter χ is given by

$$\chi = \begin{cases} 0, & \bar{\sigma} < Y(\bar{\varepsilon}^P) \qquad \text{(elastic state)}; \\ 0, & \bar{\sigma} = Y(\bar{\varepsilon}^P) \text{ but } \boldsymbol{\sigma}' \cdot \dot{\boldsymbol{\varepsilon}} \leq 0 \qquad \text{(elastic unloading or neutral loading)}; \\ 1, & \bar{\sigma} = Y(\bar{\varepsilon}^P) \text{ and } \boldsymbol{\sigma}' \cdot \dot{\boldsymbol{\varepsilon}} > 0 \qquad \text{(plastic loading)}. \end{cases}$$

Also as before, a particular model for $Y(\bar{\varepsilon}^P)$ (for example, a power law relationship) may be specified. The benefit of this type of formulation is that even though the strain state is a tensor, the hardening function is still entirely specified by one scalar, the equivalent plastic strain $\bar{\varepsilon}^P$.

Beyond small deformations

In some engineering metals which can undergo large plastic deformations, it is common and convenient to neglect the elastic part of the strain. In this **rigid-plastic** approximation, it is assumed that $\varepsilon_{ij}^E \ll \varepsilon_{ij}^P$ so that *all the strain is plastic strain.* Equivalently, the rigid-plastic model is the limit where $E \to \infty$. Either way, the evolution for strain simplifies to

$$d\varepsilon_{ij} = \frac{3}{2} d\bar{\varepsilon}^P \frac{\sigma'_{ij}}{\bar{\sigma}} \qquad \text{(rigid-plastic)}.$$

If the strain rate tensor is known, the deviatoric stress may be expressed by the inverse of this equation,

$$\sigma'_{ij} = \left(\frac{2Y(\bar{\varepsilon})}{3\dot{\bar{\varepsilon}}^P} \right) \dot{\varepsilon}_{ij},$$

because $\bar{\sigma} = Y(\bar{\varepsilon})$ during plastic flow. Note also that we have re-introduced the strain rate tensor and accordingly have replaced $\bar{\varepsilon}^P$ with $\dot{\bar{\varepsilon}}^P$.

Remark 6.5.3: In this case (and in plasticity in general), only the *deviatoric* stress state can be computed directly from a given strain state. The hydrostatic part of the stress can only be represented by an undetermined scalar field,

$$\sigma_{ij} = \left(\frac{2Y(\bar{\varepsilon})}{3\dot{\bar{\varepsilon}}^P} \right) \dot{\varepsilon}_{ij} - P\delta_{ij},$$

where P must be determined by traction boundary conditions.

Exercise 6.5.4. AKG 3.2 through 3.10, 3.12.

6.5.2 Rate-dependent, isotropic hardening

The basic principles of the three-dimensional rate-dependent model are exactly the same as in the three-dimensional rate-independent model. In particular, we have the same decomposition of the strain tensor and the strain rate tensor, and we still require a flow rule based on codirectionality,

$$\dot{\boldsymbol{\varepsilon}}^P = \frac{3}{2} \dot{\bar{\varepsilon}}^P \frac{\boldsymbol{\sigma}'}{\bar{\sigma}}.$$

The difference comes in when we specify a form for the flow strength. Just like in the one-dimensional case, instead of $Y(\bar{\varepsilon}^P)$, we will use a viscoplastic flow strength model $\mathcal{S}(\bar{\varepsilon}^{\mathcal{P}}, \dot{\bar{\varepsilon}}^{\mathcal{P}})$ that depends on the equivalent plastic strain rate in an *explicit* power-law form:

$$\mathcal{S}(\bar{\varepsilon}^{\mathcal{P}}, \dot{\bar{\varepsilon}}^{\mathcal{P}}) = Y(\bar{\varepsilon}^P) \times \left(\frac{\dot{\bar{\varepsilon}}^P}{\dot{\varepsilon}_0} \right)^m,$$

where again $0 < m \leq 1$ is the *rate sensitivity*, and $\dot{\varepsilon}_0$ is a *reference strain rate.*

For low absolute temperatures, the reference strain rate $\dot{\varepsilon}_0$ is roughly constant, but when $T \geq 0.5T_m$, then the reference strain rate must explicitly depend on temperature,

$$\dot{\varepsilon}_0 = \dot{\varepsilon}_0(T) = A \exp\left(-\frac{Q}{RT} \right) \qquad (T > 0.5T_m).$$

In either case the flow strength model can be arranged to provide the evolution equation for $\dot{\bar{\varepsilon}}^P$:

$$\dot{\bar{\varepsilon}}^P = \dot{\varepsilon}_0 \left(\frac{\bar{\sigma}}{Y(\bar{\varepsilon}^P)} \right)^{1/m}.$$

This, together with the statement of codirectionality,

$$\dot{\varepsilon}_{ij}^P = \frac{3}{2}\frac{\dot{\bar{\varepsilon}}^P}{\bar{\sigma}}\sigma'_{ij},$$

fully specifies the constitutive model. Importantly, as in the one-dimensional case,

the rate-dependent (viscoplastic) model has the result that the plastic strain rate is non-zero whenever the stress is nonzero; equivalently, there is *no purely elastic range* at all!

If, however, a yield threshold $Y_{th}(\bar{\varepsilon}^P)$ is modeled, it may be incorporated in the same manner as the one-dimensional formulation, namely that the evolution equation for $\dot{\bar{\varepsilon}}^P$ becomes

$$\dot{\bar{\varepsilon}}^P = \dot{\varepsilon_0}\left\langle \frac{\bar{\sigma} - Y_{th}(\bar{\varepsilon}^P)}{Y(\bar{\varepsilon}^P)} \right\rangle^{1/m}.$$

Remark 6.5.5: The flow rule, together with any other constitutive equations of any plasticity model, require initial conditions. The standard set of initial conditions set all initial strain to zero at time zero, and set the value of the hardening function to the yield stress at time zero,

$$\boldsymbol{\varepsilon}(\mathbf{x}, 0) = \boldsymbol{\varepsilon}^P(\mathbf{x}, 0) = \mathbf{0}; \quad Y(\mathbf{x}, 0) = \sigma_y.$$

Exercise 6.5.6. AKG 3.19 through 3.24.

6.5.3 Rate-independent, kinematic and isotropic hardening

To account for kinematic hardening, we again must modify the free-energy function in the same way we did for the one-dimensional case: by adding an explicit term that accounts of "plastic free energy", so that $\psi = \psi^E + \psi^P$. We may assume the usual elastic theory holds, so the elastic free energy is quadratic in the elastic strain. For the plastic free energy we define a dimensionless strain-tensor-like internal variable $\mathbf{A}$, which is symmetric and deviatoric (like $\boldsymbol{\varepsilon}^P$), as well as a **back stress** $\boldsymbol{\sigma}_{\text{back}}$, such that

$$\psi^P \equiv \frac{1}{2}C|\mathbf{A}|^2, \qquad \boldsymbol{\sigma}_{\text{back}} \equiv C\mathbf{A}$$

for some **back stress modulus** $C > 0$. Then, the **effective stress** which accounts for the driving force that creates the Bauschinger effect, as well as its direction $\mathbf{N}^P$, are defined to be

$$\boldsymbol{\sigma}_{\text{eff}} \equiv \boldsymbol{\sigma}' - \boldsymbol{\sigma}_{\text{back}}, \qquad \mathbf{N}^P \equiv \frac{\boldsymbol{\sigma}_{\text{eff}}}{|\boldsymbol{\sigma}_{\text{eff}}|}$$

with the **equivalent tensile effective stress**

$$\bar{\sigma}_{\text{eff}} \equiv \sqrt{\frac{3}{2}}|\boldsymbol{\sigma}_{\text{eff}}|.$$

Then the codirectionality flow rule reads

$$\dot{\boldsymbol{\varepsilon}}^P = \frac{3}{2}\dot{\bar{\varepsilon}}^P\frac{\boldsymbol{\sigma}_{\text{eff}}}{\bar{\sigma}_{\text{eff}}}.$$

If plastic loading is happening (equivalently, if both $\bar{\sigma}_{\text{eff}} = Y(\bar{\varepsilon}^P)$ *and* $\mathbf{N}^P \cdot \dot{\boldsymbol{\varepsilon}} > 0$, note the strict inequality), then the strain state evolves as

$$\begin{aligned}\dot{\boldsymbol{\varepsilon}}^P &= \left(\frac{3G}{3G+H}\right)\left(\mathbf{N}^P \cdot \dot{\boldsymbol{\varepsilon}}\right)\mathbf{N}^P, \\ \dot{\mathbf{A}} &= \dot{\boldsymbol{\varepsilon}}^P - \gamma \mathbf{A}\dot{\bar{\varepsilon}}^P\end{aligned}$$

where

$$H \equiv \underbrace{\frac{dY(\bar{\varepsilon}^P)}{d\bar{\varepsilon}^P}}_{\equiv H_{\text{iso}}} + C\left(\frac{3}{2} - \sqrt{\frac{3}{2}}\gamma \mathbf{A}\cdot\mathbf{N}^P\right)$$

is the total hardening modulus. Note that the plastic strain rate does not change, $\dot{\boldsymbol{\varepsilon}}^P = 0$, if *either* $\bar{\sigma}_{\text{eff}} < Y(\bar{\varepsilon}^P)$ (corresponding to the elastic state), *or* $\bar{\sigma}_{\text{eff}} = Y(\bar{\varepsilon}^P)$ but $\mathbf{N}^P \cdot \dot{\boldsymbol{\varepsilon}} \leq 0$ (corresponding to elastic unloading in the plastic state or neutral loading in the plastic state, respectively).

6.5.4 Rate-dependent, kinematic and isotropic hardening

Hopefully the pattern is clear by now. In the case of rate-dependent plasticity with kinematic hardening, we combine the features of the three-dimensional rate-dependent theory — namely the lack of yield condition, which is replaced by an explicit equation for $\dot{\bar{\varepsilon}}^P$ (for example in power law form) — and the features of kinematic hardening, which means that everything is now dependent on the effective stress $\boldsymbol{\sigma}_{\text{eff}}$ and the equivalent tensile effective stress $\bar{\sigma}_{\text{eff}}$. Explicitly, the flow rule (assuming a power-law rate dependence) is given by

$$\begin{aligned}\dot{\boldsymbol{\varepsilon}}^P &= \frac{3}{2}\dot{\bar{\varepsilon}}^P \frac{\boldsymbol{\sigma}_{\text{eff}}}{\bar{\sigma}_{\text{eff}}}, \\ \dot{\mathbf{A}} &= \dot{\boldsymbol{\varepsilon}}^P - \gamma \mathbf{A}\dot{\bar{\varepsilon}}^P, \\ \dot{\bar{\varepsilon}}^P &= \dot{\varepsilon}_0 \left\langle \frac{\bar{\sigma} - Y_{th}(\bar{\varepsilon}^P)}{Y(\bar{\varepsilon}^P)} \right\rangle^{1/m},\end{aligned}$$

where $Y_{th}(\bar{\varepsilon}^P)$ is the optional threshold yield strength. For high temperatures, it is appropriate to take the reference strain rate $\dot{\varepsilon}_0$ as the usual Arrhenius form.

6.6 Large-deformation plasticity

The previous sections fundamentally assume an *additive* decomposition of the strain

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^E + \boldsymbol{\varepsilon}^P$$

and hence an additive decomposition of the strain rate

$$\dot{\boldsymbol{\varepsilon}} = \dot{\boldsymbol{\varepsilon}}^E + \dot{\boldsymbol{\varepsilon}}^P.$$

Even if the logarithmic strain measure (i.e., $\boldsymbol{\varepsilon} = \ln \mathbf{U}$) is used, the concept of an additive decomposition of strain cannot account for the most general case where both elastic and plastic parts of the deformation can be made arbitrarily large. For example, a broad class of polymers can undergo such large elastic *and* large plastic deformations.

The "finite" characterization of such a general, large-deformation model for elastic-plastic deformation assumes a multiplicative **Kröner decomposition** of the total deformation gradient $\mathbf{F}$,

$$\mathbf{F} = \mathbf{F}^E \mathbf{F}^P,$$

where the **elastic part of the deformation gradient** (or the *elastic distortion*) $\mathbf{F}^E$ physically represents local stretching and rotation of the microstructure, and the **plastic part of the deformation gradient** (or the *plastic distortion*) $\mathbf{F}^p$ physically represents the local inelastic deformation, for example due to the flow of defects in a metal or due to the change in free volume in a polymer.

Remark 6.6.1: Although we define $\mathbf{F}$ as the gradient of a motion function, the two tensors $\mathbf{F}^E$ and $\mathbf{F}^P$ are *not* necessarily the gradients of anything; in other words, there does not necessarily exist an "elastic motion" $\boldsymbol{\chi}^E$ for which $\mathbf{F}^E = \nabla \boldsymbol{\chi}^E$.

Rather, we can only think of $\mathbf{F}^E$ and $\mathbf{F}^P$ as linear transformations that operate sequentially on a vector in the reference configuration $d\mathbf{X}$. Occasionally the vector $d\mathbf{l} \equiv \mathbf{F}^P d\mathbf{X}$ is said to reside in an **intermediate state** (or, in the context of polymers, in a *relaxed state*). In this framework $\mathbf{F}^P$ is seen as a linear transformation from the reference configuration to this intermediate configuration, and $\mathbf{F}^E$ is seen to transform vectors from the intermediate configuration to the deformed configuration.

Correspondingly, we can define

$$J^E = \det \mathbf{F}^E > 0$$

and

$$J^P = \det \mathbf{F}^P > 0;$$

if the plasticity is modeled as incompressible (common for metals, but not necessarily the case, as for polymers) it follows that $J^P = 1$ and therefore that $J = J^E$.

Moreover, recalling that the spatial velocity gradient $\mathbf{L} \equiv \nabla \mathbf{v}, \mathbf{L} : d\mathbf{x} \mapsto d\mathbf{v}$ is related to $\mathbf{F}$ by

$$\mathbf{L} = \dot{\mathbf{F}} \mathbf{F}^{-1},$$

we can use the Kröner decomposition to arrive at the relation

$$\mathbf{L} = \dot{\mathbf{F}} \mathbf{F}^{-1} = \dot{\mathbf{F}}^E \mathbf{F}^{E-1} + \mathbf{F}^E (\dot{\mathbf{F}}^P \mathbf{F}^{P-1}) \mathbf{F}^{E-1}$$

(note that here, the notation $\mathbf{F}^{E-1}$ refers to the inverse of the tensor $\mathbf{F}^E$). If, by analogy with $\mathbf{L}$, we define an **elastic rate of distortion** $\mathbf{L}^E \equiv \dot{\mathbf{F}}^E \mathbf{F}^{E-1}$ and a **plastic rate of distortion** $\mathbf{L}^P \equiv \dot{\mathbf{F}}^P \mathbf{F}^{P-1}$, we have

$$\mathbf{L} = \mathbf{L}^E + \mathbf{F}^E \mathbf{L}^P \mathbf{F}^{E-1}.$$

Recalling also that $\mathbf{L}$ can be decomposed into its symmetric part $\mathbf{D}$, which represents a *rate of stretching*, and its skew part $\mathbf{W}$, which represents a *rate of spinning*, we can also analogously define tensors $\mathbf{D}^E$, $\mathbf{W}^E$, $\mathbf{D}^P$, and $\mathbf{W}^P$ such that e.g.,

$$\mathbf{L}^P = \mathbf{D}^P + \mathbf{W}^P, \qquad \mathbf{D}^P = \operatorname{skw} \mathbf{L}^P.$$

The rate of stretching tensor $\mathbf{D}$ is the large-deformation analog to the strain-rate tensor $\dot{\boldsymbol{\varepsilon}}$. By analogy, then, it can be see that $\mathbf{D}^P$ plays the role of $\dot{\boldsymbol{\varepsilon}}$ in the large-deformation theory. For example, we will see that the evolution of the equivalent plastic strain can be written in terms of $\mathbf{D}^P$ (and stress-like quantities similar to our old friend $\boldsymbol{\sigma}'$).

From the definition of $\mathbf{L}^P$, it can be seen that $\mathbf{F}^P$ evolves according to

$$\dot{\mathbf{F}}^P = \mathbf{L}^P \mathbf{F}^P.$$

For *isotropic* materials it is commonly assumed that

$$\mathbf{W}^P = 0 \qquad \text{(isotropic materials)},$$

which specializes the preceding evolution equation to

$$\dot{\mathbf{F}}^P = \mathbf{D}^P \mathbf{F}^P.$$

In other words, a flow rule on $\mathbf{D}^P$ is necessary and sufficient to prescribe how $\mathbf{F}^P$ behaves.

Remark 6.6.2: In this model, prescribing the evolution of $\mathbf{F}^P$ is sufficient to prescribe *everything* in the model assuming (as is typical) that the total deformation $\mathbf{F}$ is known. Moreover, we can then determine the body's stresses which are traditionally taken to depend only on the elastic part of the deformation, and $\mathbf{F}^E = \mathbf{F}\mathbf{F}^{P-1}$.

To complete the plastic part of the theory, we need to specify exactly *how* $\mathbf{D}^P$ (or, more generally, $\mathbf{L}^P$) relates to the other flow variables. This requires the identification of a driving stress for plasticity in the general large-deformation case. In the small-deformation case we assumed that the deviatoric part of the Mises stress was the driving stress for plasticity, and this formed the crux of our co-directionality assumption.

Since we are working with large deformations, recall that the Cauchy stress is denoted by $\mathbf{T}$. We define an **elastic second Piola stress**

$$\mathbf{S}^E \equiv J^E \mathbf{F}^{E-1} \mathbf{T} \mathbf{F}^{E-T}$$

where $\mathbf{F}^{E-T}$ denotes the inverse transpose of the elastic part of the deformation gradient. Then, defining an **elastic right Cauchy-Green tensor**

$$\mathbf{C}^E \equiv \mathbf{F}^{E^T} \mathbf{F}^E$$

we can define the **Mandel stress**,

$$\mathbf{M}^E \equiv \mathbf{C}^E \mathbf{S}^E.$$

We denote the deviatoric part of the Mandel stress by $\mathbf{M}_0^E$, i.e.,

$$\mathbf{M}_0^E = \mathbf{M}^E - \frac{1}{3}\mathrm{tr}\,(\mathbf{M}^E)\mathbf{I}.$$

Key Equation 6.6.3 (Co-directionality assumption for large-deformation plasticity)

In a large-deformation plasticity theory, the rate of plastic stretching $\mathbf{D}^P$ is assumed to be co-directional with the deviatoric part of the Mandel stress, $\mathbf{M}_0^E$, with the plastic flow direction $\mathbf{N}^P$ thus given by

$$\mathbf{N}^P = \frac{\mathbf{D}^P}{|\mathbf{D}^P|} = \frac{\mathbf{M}_0^E}{|\mathbf{M}_0^E|}.$$

Derivation 6.6.4 (Where does the Mandel stress come from?)

An explanation for the sudden appearance of these new stress measures, as well as the justification for the new co-directionality assumption, requires a derivation of the equations of equilibrium for this case of large-deformation elastic-plastic deformation.

We approach this by the *virtual power* technique. Consider a part $\mathcal{P}$ of a body undergoing deformation. There is a traction field $\mathbf{t}$ and a body-force field $\mathbf{b}$ such that the total *external virtual power* is given by

$$\mathcal{W}_{\text{ext}} = \int_{\partial\mathcal{P}} \mathbf{t}(\mathbf{n}) \cdot \tilde{\mathbf{v}} \, dA + \int_{\mathcal{P}} \mathbf{b} \cdot \tilde{\mathbf{v}} \, dV$$

for a particular *virtual velocity* $\tilde{\mathbf{v}}$. Mathematically, the virtual velocity represents an arbitrary vector with units of velocity. Physically, it represents any admissible velocity of the part of the body, and we can make particular choices for $\tilde{\mathbf{v}}$ to specalize the virtual work equation.

Meanwhile, to compute the *internal virtual power*, we make the assumption that the internal virtual power has two sources: one due to a *macroscopic stress* $\mathbf{S}^{\text{macr}}$ conjugate to the elastic distortion rate $\mathbf{L}^E$, and another due to a *microscopic stress* $\mathbf{S}^{\text{micr}}$ conjugate to the plastic distortion rate $\mathbf{L}^P$. To this end, writing

$$\operatorname{grad} \tilde{\mathbf{v}} \equiv \tilde{\mathbf{L}}^E + \mathbf{F}^E \tilde{\mathbf{L}}^P \mathbf{F}^{E-1}$$

for our virtual velocity $\tilde{\mathbf{v}}$, we compute the internal virtual power as

$$\mathcal{W}_{\text{int}} = \int_{\mathcal{P}} \left(\mathbf{S}^{\text{macr}} \cdot \tilde{\mathbf{L}}^E + \frac{1}{J^E} \mathbf{S}^{\text{micr}} \cdot \tilde{\mathbf{L}}^P \right) dV$$

The factor $1/J^E$ is due to the fact that $\tilde{\mathbf{L}}^P$ is a tensor in the intermediate space, and hence we need to map the "plastic stress power term" to the deformed space.

The statement of the principle of virtual power has two parts:

1. For any choice of $\tilde{\mathbf{v}}$,
$$\mathcal{W}_{\text{int}} = \mathcal{W}_{\text{ext}}.$$
2. For any choice of $\tilde{\mathbf{v}}$ subject to the constraint that $\tilde{\mathbf{v}}(\mathbf{x}) = \mathbf{a} + \boldsymbol{\Omega}\mathbf{x}$ for all $\mathbf{x}$ where $\mathbf{a}$ and $\boldsymbol{\Omega}$ are spatially constant, together with $\mathbf{L}^e = \boldsymbol{\Omega}$ and $\mathbf{L}^p = \mathbf{0}$,
$$\mathcal{W}_{\text{int}} = 0.$$

Then, let us first make a choice of $\tilde{\mathbf{v}}$ such that $\tilde{\mathbf{L}}^P = \mathbf{0}$. For this choice, the first part of the principle of virtual power yields the result

$$\begin{cases} \mathbf{t}(\mathbf{n}) = \mathbf{S}^{\text{macr}} \mathbf{n}, \\ \operatorname{div} \mathbf{S}^{\text{macr}} + \mathbf{b} = 0. \end{cases}$$

The second part yields the result

$$\mathbf{S}^{\text{macr}} = (\mathbf{S}^{\text{macr}})^T.$$

It follows that this macroscopic stress, or in other words the stress tensor which is work-conjugate with the elastic disortion rate, is none other than the Cauchy stress!

Let us separately choose $\tilde{\mathbf{v}} = \mathbf{0}$ with $\tilde{\mathbf{L}}^P$ arbitrary. Using $\mathbf{T}$ now in place of $\mathbf{S}^{\text{macr}}$ the second part of the principle of virtual power yields the result

$$\mathbf{S}^{\text{micr}} = J^E \mathbf{F}^{ET} \mathbf{T} \mathbf{F}^{E-T}.$$

This definition motivates the identification of $\mathbf{S}^{\text{micr}}$ as the Mandel stress. In other terms, the Mandel stress *is the stress measure which is power conjugate to the plastic distortion rate.*

With these results, the *actual* internal expenditure of power can be written as

$$\mathcal{W}_{\text{int}} = \int_{\mathcal{P}} \left(\mathbf{T} \cdot \mathbf{L}^E + \frac{1}{J^E} \mathbf{M}^E \cdot \mathbf{L}^P \right) dV,$$

or in terms of the elastic second Piola stress $\mathbf{S}^E$ introduced previously,

$$\mathcal{W}_{\text{int}} = \int_{\mathcal{P}} \left(\frac{1}{2J^E} \mathbf{S}^E \cdot \dot{\mathbf{C}}^E + \frac{1}{J^E} \mathbf{M}^E \cdot \mathbf{L}^P \right) dV.$$

Having defined these quantities, the rest of the statement of the flow rule is much the same. In particular, we can define equivalent scalar quantities for the plastic strain and the driving stress as before (which we will still refer to as the **equivalent stress** and the **equivalent plastic strain**):

$$\bar{\sigma} \equiv \sqrt{\frac{3}{2}} |\mathbf{M}_0^E|, \qquad \dot{\bar{\varepsilon}}^P \equiv \sqrt{\frac{2}{3}} |\mathbf{D}^p|.$$

Remark 6.6.5: Although we have defined the equivalent plastic strain rate in the same way as was done in the small-deformation plasticity theory, here the quantity $\sqrt{2/3}|\mathbf{D}^p|$ cannot be directly interpreted as the time rate of change of some strain measure. More generally, $\mathbf{D}^P$ is a "rate-like descriptor" quantity, but unlike $\dot{\boldsymbol{\varepsilon}}$, it is not the time derivative of a meaningful quantity.

It then follows that the basic flow rule for isotropic materials becomes

$$\mathbf{D}^P = \frac{3}{2} \frac{\dot{\bar{\varepsilon}}^P}{\bar{\sigma}} \mathbf{M}_0^E.$$

Since we have re-introduced the same scalar quantities as the small-deformation model,

the same one-dimensional flow-strength relations can be used to describe the evolution of the equivalent plastic strain in terms of an evolving hardening rule and other internal variables. Moreover, in a rate-independent theory, the same loading-unloading and consistency conditions apply, and can be written in terms of the equivalent stress and equivalent plastic strain measures given here.

Example 6.6.6

As a simple but explicit example, a complete statement of the flow rule (i.e., the evolution of plastic deformation) for a rate-dependent model is given by

$$\dot{\mathbf{F}}^P = \mathbf{D}^P \mathbf{F}^P,$$

with

$$\mathbf{D}^P = \frac{3}{2}\dot{\bar{\varepsilon}}^P \frac{\mathbf{M}_0^E}{\bar{\sigma}}$$

where

$$\dot{\bar{\varepsilon}}^P = \dot{\varepsilon}_0 \left(\frac{\bar{\sigma}}{Y(\bar{\varepsilon}^P)} \right)^{1/m},$$

for some material parameters $\dot{\varepsilon}_0$ and m, and a prescribed hardening function $Y(\bar{\varepsilon}^P)$, for example

$$Y(\bar{\varepsilon}^P) = K(\bar{\varepsilon}^P)^n$$

for some K, n.

The large-deformation theory can be closed by specifying how the stresses are related to the deformation. As mentioned previously, it is generally taken to be the case that the stresses are related to only the elastic part of the deformation, described by $\mathbf{F}^E$. In view of this, we can prescribe a free-energy function that depends only on quantities derived from $\mathbf{F}^E$. Let the polar decomposition of $\mathbf{F}^E$ be denoted by $\mathbf{R}^E\mathbf{U}^E$.

For *small elastic strains* it is appropriate to simply assume a logarithmic strain measure similar to that used in the small-deformation theory, i.e.,

$$\mathbf{E}^E \equiv \ln \mathbf{U}^E,$$

and to assume a strain-energy function which is quadratic in this strain measure,

$$\psi = G|\mathbf{E}_0^E|^2 + \frac{1}{2}\kappa(\mathrm{tr}\mathbf{E}^E)^2,$$

for a shear modulus G and a bulk modulus κ. Then,

$$\mathbf{M}^E = \frac{\partial \psi}{\partial \mathbf{E}^E} = \mathbb{C}\mathbf{E}^E$$

with

$$\mathbb{C} \equiv 2G\left(\mathbb{I}^{\mathrm{sym}} - \frac{1}{3}\mathbf{I}\otimes\mathbf{I} \right) + \kappa \mathbf{I}\otimes\mathbf{I}$$

and finally

$$\mathbf{T} = \frac{1}{J^E}\mathbf{R}^E\mathbf{M}^E\mathbf{R}^{E^T}.$$

For *larger elastic strains*, as in the case of polymers, this theory is nicely compatible with the usual hyperelastic material models, now expressed in terms of $\mathbf{F}^E$. In this framework, if the free-energy function is specified in terms of the invariants of $\mathbf{C}^E$, it follows that

$$\mathbf{S}^E = 2\frac{\partial \psi}{\partial \mathbf{C}^E}, \quad \mathbf{T} = \mathbf{C}^E\mathbf{S}^E.$$

Example 6.6.7

For a simple example, we consider take the compressible neo-Hookean free-energy function, expressed as the sum of a deviatoric part and a volumetric part, given by

$$\psi = \frac{1}{2}G(\mathrm{tr}\,\bar{\mathbf{C}}^E - 3) + \frac{1}{2}\kappa(J-1)^2$$

where $\bar{\mathbf{C}}^E = J^{-2/3}\mathbf{C}^E$ represents the isochoric part of $\mathbf{C}^E$. The corresponding

Cauchy stress can be expressed in terms of the deviator of the isochoric part of $\mathbf{B}^E$ as

$$\mathbf{T} = \frac{G}{J^e}\bar{\mathbf{B}}_0^E + \kappa(J-1)\mathbf{I}.$$

7 Fracture

A common failure mode of solid objects is **fracture**, which is the *parting of the solid into two or more pieces*. Thus far, we have considered the mechanics of homogeneous solid materials, but in reality all solid bodies contain *cracks* (or other small defects) at some length scale, which grow during service and eventually cause fracture. The goal of this chapter is to develop a *fracture criterion* and discuss the stress field around a crack.

Remark 7.0.1 (Characterizing fracture)**:** Fracture can be characterized over two length scales: *globally* at the length scale of the part or specimen, and *locally* at the length scale of the initial crack. At each length scale, the fracture can be classified as *ductile*, meaning it is accompanied by plastic flow and deformation before failure, or *brittle*, meaning the failure is sudden, catastrophic, and accompanied by little to no deformation. Specifically:

- **globally brittle fracture** is characterized by little or no macroscopic inelastic deformation, with a plot of load as a function of displacement behaving linearly until fracture
- **globally ductile fracture** is characterized by considerable macroscopic inelastic deformation, with a nonlinear regime of the load-displacement curve developing before fracture
- **locally brittle fracture** is due to the cleavage of grains in the microstructure, the decohesion along grain boundaries, or both, which leads to negligble amounts of plastic flow within grains
- **locally ductile fracture** is due to the nucleation and growth of voids at microstructural inclusions or precipitates in the material, accompanied by plastic flow and tearing of the microstructure near the crack tip

It is worth emphasizing that fracture can be, e.g., locally brittle and globally ductile, or any combination of the above. In general, ductile fracture requires more energy than brittle fracture. The extra energy expended in ductile fracture is dissipated as a result of plasticity. (Ductile failure is also "safer" in that it is not catastrophic, and a larger amount of plastic deformation is easier to observe.) In this section we will consider *globally brittle fracture* with a *small* local plastic region *at most*. This assumption allows us to work from the theory of linear elasticity, which in turn allows the following development to be a *universal* picture for crack propagation.

Remark 7.0.2: In metals having BCC or HCP crystal structures (like most steels), but *not* in metals having FCC crystal structure (like aluminum), the fracture behavior is temperature-dependent. At high temperatures, thermal agitation assists the motion of dislocations, and fracture is ductile. But below a critical **ductile-to-brittle transition temperature**, the thermal motion of dislocations is reduced, the intrinsic lattice resistance increases, and the mechanism of fracture turns brittle.

For certain steels, the ductile-brittle transition temperature can be as high as 0°C, which is of importance in common engineering situations.

Recall that in isotropic linear elasticity, if a specimen is pulled in *tension* with far-field applied stress σ_∞, the stress concentration factor K_t around a feature with characteristic length a and radius ρ scales as $K_t \propto \sqrt{a/\rho}$, such that $\sigma_{\max} \propto K_t\sigma_\infty \propto \sigma_\infty\sqrt{a/\rho}$. A crack is really just a special case of this stress concentration situation with $\rho \to a_0$, for a_0 the interatomic spacing between lattice rows. We therefore expect $\sigma_{\max}$ to scale as $\sigma_\infty\sqrt{a}$. To this end we define a *stress intensity factor* which depends on the far-field applied stress and the crack geometry exactly in this way. We can then consider a failure criterion on this stress intensity factor to be an equivalent criterion on $\sigma_{\max}$, which is exactly the condition for brittle failure.

Definition 7.0.3. For a crack of characteristic length a subject to a tensile far-field stress σ_∞, the **mode I stress intensity factor**[1] K_{I} is defined to be[2]

$$K_{\mathrm{I}} \equiv Q\sigma_\infty\sqrt{\pi a},$$

where Q is a nondimensional geometeric "configuration correction factor" that accounts for the particular crack geometry. The development of K_{I} in terms of a geometry-dependent Q provides the elegant result that the theory based on K_{I} is applicable for *all* mode I cracks!

Remark 7.0.4: The configuration correction factor Q is a nondimensional quantity that is, in general, a function of nondimensional geometric parameters, such as a/w, where w is a characteristic specimen dimension in the direction of the crack. Values of Q are tabulated for a variety of "standard" or common crack configurations (see, for example, Appendix H of AKG); the order of magnitude of Q is typically unity. In the special case where $a \ll w$, i.e. when the crack is located sufficiently far away from the edges of the specimen that the specimen can be considered *infinite* relative to the crack, we have $Q = 1$.

Example 7.0.5

Consider a two-dimensional plate with width w and length $L > 3w$. If there is a crack of length $2a$ in the center of the plate aligned with the width direction, and a far-field stress σ_∞ applied parallel to the length direction, the configuration correction factor is given by

$$Q = \hat{Q}(a/w) = \sqrt{\sec\left(\pi a/w\right)}.$$

Observe that if $w \to \infty$, we recover $Q = 1$, the infinite-plate configuration correction factor.

The qualifier "mode I" corresponds to a tensile far-field applied stress perpendicular to the direction of the crack. In general, the far-field stress may also be applied in an

[1] Here, the subscript I represents a Roman numeral, and we say "K one".

[2] Although their names are similar, it is important to distinguish between this *stress intensity factor* (with units of stress times square-root-length, e.g. MPa $\mathrm{m}^{1/2}$), and the *stress concentration factor* (which is a nondimensional stress ratio).

in-plane sliding "mode II", or an *anti-plane tearing* "mode III" (or some combination of the above) when viewed from the crack. However, in most engineering applications, it is phenomenologically seen that globally brittle fracture occurs in mode I loading whereby the crack propagates in a direction *perpendicular to the local direction of maximum principal stress*. As such, the mode I fracture criterion dominates:

Key Equation 7.0.6 (Mode I linear elastic fracture criterion)

In mode I loading, the onset of crack propagation occurs if and only if

$$K_{\mathrm{I}} \equiv Q\sigma_\infty\sqrt{\pi a} = K_{\mathrm{Ic}},$$

where a is a length scale associated with the crack[a], σ_∞ is the far-field applied tensile stress, and K_{Ic} is a *material property* called the "critical mode I stress intensity factor" or more simply the **fracture toughness**.

[a] Usually, a is exactly the length of the crack, but not always! The values for Q (e.g. in a table) will also specify the corresponding way to define a and σ_∞.

The fracture toughness K_{Ic} measures the material's resistance to crack propagation. Just like values of σ_y, values for K_{Ic} are obtained from standardized mechanical testing. For metals, see the standard ASTM E399.

Remark 7.0.7: The measured fracture toughness for most materials is *orders of magnitude* smaller than the theoretical *ideal cleavage strength* required to separate neighboring planes of atoms, as in a perfect crystal. The culprit, as always, is the unavoidable presence of micro-cracks, imperfections, and other flaws present in every real specimen.

7.1 Elastic stress fields around cracks

It is instructive to describe how the elastic stress field changes from the far-field value near a crack, which effectively acts like a perturbation; moreover, a *sharp* crack produces a theoretical singularity in the elastic solution, a situation which we will rectify in the next section.

Consider a crack with length $2a$ in an two-dimensional infinite body oriented in the direction $\mathbf{e}_1$, and suppose the body is subject to mode I loading with a far-field applied stress σ_∞ applied in the $\mathbf{e}_2$ direction. Take the origin in polar coordinates to be at the center of the crack. If the material is isotropic and linear elastic, then for $\theta = 0$ the stress along the crack axis is given by

$$\sigma_{22} = \sigma_{22}(x = a + r, \theta = 0) = \frac{\sigma_\infty\sqrt{a}(1 + r/a)}{\sqrt{2r}\sqrt{1 + r/2a}}.$$

This analytical solution was found by Charles E. Inglis in 1913; it was a foundational result in the subject of *linear-elastic fracture mechanics*. Near the crack, when r/a is small, the expression is approximately

$$\sigma_{22}(x = a + r, \theta = 0) \approx \frac{\sigma_\infty\sqrt{a}}{\sqrt{2r}} = \frac{K_{\mathrm{I}}}{\sqrt{2\pi r}}.$$

This argument can actually be *generalized* to develop a so-called "asymptotic crack tip stress field" for this mode I loading, as long as $r \ll a$. If this assumption is true, then

$$\begin{pmatrix}\sigma_{11}\\ \sigma_{22}\\ \sigma_{12}\end{pmatrix} = \frac{K_{\mathrm{I}}}{\sqrt{2\pi r}}\cos\left(\frac{\theta}{2}\right)\begin{pmatrix}1-\sin\left(\frac{\theta}{2}\right)\sin\left(\frac{3\theta}{2}\right)\\ 1+\sin\left(\frac{\theta}{2}\right)\sin\left(\frac{3\theta}{2}\right)\\ \sin\left(\frac{\theta}{2}\right)\cos\left(\frac{3\theta}{2}\right)\end{pmatrix} + [\text{h.o.t.}],$$

with $\sigma_{33} = \sigma_{13} = \sigma_{23} = 0$ for plane stress, and $\sigma_{33} = \nu(\sigma_{11} + \sigma_{22}) \neq 0$, $\sigma_{13} = \sigma_{23} = 0$ for plane strain. Similarly, it can be shown that the displacement field (u_1, u_2) has components

$$u_i = \frac{K_{\mathrm{I}}}{2G}\sqrt{\frac{r}{2\pi}}f_i(\theta,\nu) + u_0$$

for $i = 1, 2$, where f_i is a scalar-valued function of the angle and Poisson ratio. Similar expressions for σ_{ij} and u_i can be written for mode II and mode III loadings, as well. Importantly,

since the development of K_{I} and the local elastic stress fields comes from a *linear* theory, superposition applies:

- **to stresses,**

$$\sigma_{ij} = \frac{1}{\sqrt{2\pi r}}[K_{\mathrm{I}}f_{\mathrm{I}}(\theta) + K_{\mathrm{II}}f_{\mathrm{II}}(\theta) + K_{\mathrm{III}}f_{\mathrm{III}}(\theta)];$$

- **to displacement fields,**

$$u_i = \frac{1}{4G}\sqrt{\frac{r}{2\pi}}[K_{\mathrm{I}}g_{\mathrm{I}}(\theta,\nu) + K_{\mathrm{II}}g_{\mathrm{II}}(\theta,\nu) + K_{\mathrm{III}}g_{\mathrm{III}}(\theta,\nu)] + u_0;$$

- **and to stress intensity factors,**

$$K_{\mathrm{I}} = K_{\mathrm{I}}^{(1)} + K_{\mathrm{I}}^{(2)} + \ldots + K_{\mathrm{I}}^{(m)},$$

as long as the far-field loading can be decomposed as

$$\sigma_\infty = \sigma_\infty^{(1)} + \sigma_\infty^{(2)} + \ldots + \sigma_\infty^{(m)}.$$

7.2 The plastic zone

So far we have developed the requirement that in order for the linear-elastic fracture criterion $K_{\mathrm{I}} = K_{\mathrm{Ic}}$ to hold on a region of size r near the crack tip, we must have $r \ll a$. Clearly a is the characteristic length of the crack, but what determines r? We know that in a globally brittle situation, any amount of plasticity cannot occur close to the specimen boundaries, and hence must be localized to a region near the crack *tip*. In this case we can call the maximum value of r such that the asymptotic formulation still applies the *radius of the K-field*, r_{K}.

Remark 7.2.1: Mathematically, the statement that *the asymptotic formulation still applies at the radius of the K-field* says that the K-field solution must give the correct tractions at r_{K}. In particular, suppose we let

$$\tilde{\boldsymbol{\sigma}}(r_{\mathrm{K}},\theta) \equiv \frac{K_{\mathrm{I}}}{\sqrt{2\pi r_{\mathrm{K}}}} f(\theta)$$

be the stress field in the region where the asymptotic formulation is valid (i.e., the K-field). Then, we require that

$$\mathbf{t}(r_{\mathrm{K}},\theta) = \tilde{\boldsymbol{\sigma}}(r_{\mathrm{K}},\theta)\,\mathbf{n},$$

where $\mathbf{t}(r_{\mathrm{K}},\theta)$ represents the actual traction vector at the point (r_{K},θ).

Moreover, *within* r_{K}, we must limit the amount of local plasticity. We know that *some* local plasticity exists; if not, the K-field solution would "naively" suggest that $\sigma_{22} \to \infty$ as $r \to a$ (a *stress singularity*). In reality, σ_{22} will attain a maximum value, namely the *yield strength* σ_y, *everywhere* within some (hopefully small) distance from the crack tip. We call this the *size of the plastic zone*, r_{Ip}, and it can be found by requiring[3] $\sigma_{22}(r,\theta=0) = \sigma_y$:

$$\sigma_{22} \equiv \frac{K_{\mathrm{I}}}{\sqrt{2\pi r_{\mathrm{Ip}}}} = \sigma_y \implies r_{\mathrm{Ip}} = \frac{1}{2\pi}\left(\frac{K_{\mathrm{I}}}{\sigma_y}\right)^2.$$

Hence r_{Ip} represents a characteristic length-scale of local plasticity. Only when $r_{\mathrm{Ip}} \ll r_K \ll a$ do all of the assumptions in the linear-elastic theory hold, so this so-called **small-scale yielding criterion** is a major requirement for the applicability of the linear elastic theory. It can be similarly argued that *all other* specimen dimensions must be much, much larger than r_{Ip} as well, so as not to induce edge effects that lead to global ductility.

Key Equation 7.2.2 (Small-scale yielding)

The theory of linear-elastic fracture mechanics, namely the fracture criterion given by $K_{\mathrm{I}} = K_{\mathrm{Ic}}$, together with the asymptotic formulations of the stress and displacement fields, holds only if

$$a, (W-a), h \geq 15 \times \left[\frac{1}{2\pi}\left(\frac{K_{\mathrm{I}}}{\sigma_y}\right)^2\right],$$

where a is the crack dimension, W is the specimen dimension in the direction of the crack, and h is the transverse dimension between the crack and the edge of the specimen[a]. Note that the crack geometry and far-field loading strength determine K_{I}, so this condition is really a check on the applied load levels.

[a] The factor of 15 is chosen as a "rule of thumb". It can be shown that when $r/a \approx 0.1$, or equivalently $r_K = a/10$, the asymptotic solution produces about a 7% error. Hence, if $r_{\mathrm{Ip}} = a/15$, we have $r_{\mathrm{Ip}} \ll r_K \ll a$, as desired.

In words, the elastic solution only holds for distances $r > r_{\mathrm{Ip}}$. But when r_{Ip} is small *compared to any other length scales in the problem*, the error in assuming the elastic solution holds everywhere is negligible.

[3] For completeness we would also need to look at the non-zero σ_{33} component (in the case of plane strain), as well as the other non-zero components of stress near the crack tip, which contribute to $\bar{\sigma}$. These values slightly change the size of r_P, but our formulation based solely on σ_{22} remains a good approximation. This is wrapped up in the "insurance" factor of 15 that will be introduced shortly.

7.3 Fracture toughness testing

In standardized testing (e.g. under ASTM E399) to obtain values of K_{Ic}, the following general procedure is followed:

1. A *notched* and pre-cracked specimen is fabricated and tested until the onset of crack propagation is detected.
2. At the corresponding value of σ_∞ which causes the onset of crack propagation, the apparent value of K_c is computed. This value of K_c is then used to compute the characteristic length scale of plasticity in the specimen,
$$r_c \equiv \frac{1}{2\pi}\left(\frac{K_c}{\sigma_y}\right)^2.$$
3. If the characteristic length scale of plasticity r_c is small enough (i.e., at least 15 times smaller) compared to the crack length a and other in-plane specimen dimensions $(w-a), h$, the test is valid. Otherwise the specimen must be resized.
4. The out-of-plane thickness B changes the apparent value of K_c obtained in the test. In general, *as the thickness B increases from zero, K_c first increases, then decreases to an asymptotic value* that corresponds with the *plane-strain* fracture toughness. This asymptotic value of K_c is taken to be K_{Ic}.

The reason for the thickness dependence is that when B is small (i.e., for thin samples), the size of the plastic zone in the out-of-plane direction becomes commensurate with the specimen thickness. Hence large-scale "dimpling" occurs, which causes macroscopic plastic deformation near the area of interest, which is undesirable. This case corresponds with *plane stress*. However, for large B, the plastic deformation is localized (as required by LEFM), and the conditions approximate *plane strain*. Hence K_{Ic} is also called the *plane strain fracture toughness*. In general it is required that in order for the measured value of K_c to be taken as K_{Ic}, the thickness B must exceed 15 times the size of the asymptotic plastic zone r_{Ic},
$$B \geq 15 \times \left[\frac{1}{2\pi}\left(\frac{K_{\mathrm{Ic}}}{\sigma_y}\right)^2\right] \equiv 15 \times r_{\mathrm{Ic}}.$$

Example 7.3.1

A standard sample of aluminum 6061-T651 has $E = 72$ GPa and $\sigma_y = 275$ MPa. A valid plane strain fracture toughness test reveals that the value of K_{Ic} is 34 MPa $\sqrt{\mathrm{m}}$, which corresponds to an asymptotic plastic zone of dimension $r_{\mathrm{Ic}} = 2.4$ mm, so that a valid specimen must be at least 38.4 mm thick.

Exercise 7.3.2. AKG 4.1 through 4.15.

7.4 Energy-based formulation

As an alternative to the onset of crack propagation criterion $K_{\mathrm{I}} = K_{\mathrm{Ic}}$, Griffith (1921) proposed a criterion

$$\mathcal{G} = \mathcal{G}_c \qquad \text{(onset of crack propagation)},$$

where $\mathcal{G}$ is the **energy release rate**, a function of the free energy and the stress and displacement fields, and $\mathcal{G}_c$ is a material property called the **critical energy release rate** or simply the *toughness*[4].

For an ideally brittle material, $\mathcal{G}_c$ can be thought of as exactly the energy required to create two new atomic surfaces, i.e., $\mathcal{G}_c = 2\gamma_s$, where γ_s is the surface energy of the solid. For other classes of materials, the formulation is much more complex, but the general idea is that $\mathcal{G}_c$ is made up of (1) the energy per unit area of crack face which goes into creating new surfaces, and (2) the energy dissipated to plastic deformation and other localized fracture processes near the crack tip.

Meanwhile, the energy release rate $\mathcal{G}$ is the work done by external tractions *minus* the change in free energy of the body, which is exactly the difference between the work done on the body and the work stored as elastic energy in the body. This total is always positive, and hence $\mathcal{G} > 0$. In particular *per unit specimen depth*,

$$\mathcal{G} \equiv \int_{\partial B} (\boldsymbol{\sigma}\mathbf{n}) \cdot \frac{\partial \mathbf{u}}{\partial a} ds - \frac{d}{da} \int_B \psi dA,$$

where da represents an increment of the crack length a. Here B is the (two-dimensional) body and ∂B is its boundary. Thus, $\mathcal{G}$ represents the energy *per unit area of crack* that is "available" for crack extension. Only when this amount of energy per unit area is sufficiently high does the crack advance.

When the external tractions are specified in terms of *generalized forces* P_i with corresponding *generalized displacements* Δ_i, we can write[5]

$$\mathcal{G} = \sum_{i=1}^{N} P_i \frac{d\Delta_i}{da} - \frac{d}{da} \int_B \psi dA.$$

In the most general case, $\mathcal{G}$ is computed using a ***J*-integral**, which is the integral of a similar energy-difference quantity but taken over an *arbitrary* path surrounding the crack tip. It can be shown that the value of the integral is *path-independent* whenever body forces (gravitational or inertial) are negligible. This makes the integral much easier to compute compared to the definition of $\mathcal{G}$, because the path of J can be cleverly chosen to simplify terms.

Suppose the body is planar (i.e., two-dimensional), and let Γ be a closed contour surrounding the crack-tip. Then, per unit specimen depth,

$$\mathcal{G} = J(\Gamma) = \int_\Gamma \left(\psi \mathbf{n} - (\nabla \mathbf{u})^T (\boldsymbol{\sigma}\mathbf{n}) \right) \, ds,$$

where $\mathbf{n}$ represents the outward normal vector to the line element $ds \in \Gamma$.

Finally, it can be shown that the energy-based formulation and the stress-intensity formulation actually coincide. This connection was made by Rice (1968), who showed that for an isotropic, linear elastic material under small-scale yielding

$$\mathcal{G} = \frac{1}{\bar{E}}(K_{\mathrm{I}}^2 + K_{\mathrm{II}}^2) + \frac{1+\nu}{E} K_{\mathrm{III}}^2,$$

[4] Be careful! This use of the word *toughness* is in a completely different context compared with the area-under-the-curve toughness. Rather, the critical energy release rate is a measure of *fracture* toughness for a material

[5] Here, P_i could represent a concentrated point force with Δ_i a corresponding displacement, *or* P_i could represent a concentrated moment (e.g., from two very closely spaced equal and opposite forces), with Δ_i now the corresponding rotation.

where

$$\bar{E} = \begin{cases} E, & \text{plane stress;} \\ \frac{E}{1-\nu^2} & \text{plane strain.} \end{cases}$$

Hence if the loading is purely in mode I, then $\mathcal{G} = K_{\mathrm{I}}^2/\bar{E}$. We can define $K_{\mathrm{Ic}} \equiv \sqrt{\bar{E}\mathcal{G}_c}$ such that

$$K_{\mathrm{I}} = K_{\mathrm{Ic}} \iff \mathcal{G} = \mathcal{G}_c$$

and the onset-of-fracture criterion is consistent.

8 Fatigue

Fatigue is the failure of components under the action of *repeated* fluctuating stresses or strains. The key point is that if loads are cyclic or otherwise repeated, failure can occur *even when the magnitude of loading is smaller than the yield strength as measured in monotonic tests*! The underlying principle is that even if the macroscopic behavior remains elastic, and even if the stresses are well below a material's yield strength, *localized and permanent* microstructral change occurs in the vicinity of defects, which *progressively* grow until a critical crack size is reached. In this way, fatigue can be seen as a *process* that occurs over a number of cycles. In summary,

even if the structure deforms elastically, multiple repetitions of *very small* inelastic strains leads to cumulative damage ending in failure.

The classical diagnostic tool against fatigue failure is termed *non-destructive evaluation* (NDE), whereby small cracks that are candidates for fatigue initiation are identified and fixed before failure can occur. Typically, NDE is done several times over the service lifetime of a part. To be conservative, the usual assumption is that if *no* crack is found as a result of NDE, cracks having length equal to the resolution of the NDE tool are assumed to exist. Then, the failure lifetime is dependent on the number of cycles it takes this small crack to propagate to a critical crack size whereupon $K_I = K_{Ic}$. However, this **defect-tolerant approach** is only one perspective.

In contrast, the **defect-free approach** assumes *no* cracks or defects. In this approach, the failure lifetime is taken to be the number of cycles it takes to *develop* a small crack, and the assumptions are mostly based on empirical data gathered at different applied stresses or strains. The number of cycles between such a crack initiation and subsequent failure is taken to be much smaller than the number of cycles to crack initiation. (The defect-free approach is less conservative and is generally used for components which are not safety-critical.) This chapter examines both approaches separately.

Remark 8.0.1: The concepts relating to fatigue were first developed in connection to the railroad industry, during rapid railroad expansion in the mid-19th century. However, as any MBTA commuter will recognize, fatigue is a problem that still plagues rail systems today.

8.1 Defect-free approach

In the defect-free approach the game is to predict the *number of cycles to failure*, typically denoted N_f, that corresponds to the period from the beginning of service to the *initiation of a crack* as a result of fatigue. In this approach it is assumed that subsequent failure is imminent after crack initiates[1]. In general, N_f can be found either by looking at the stress in a region of interest (for example, near notches or other stress concentrators), or by looking at the strain. Particularly, since fatigue is a cyclic process, we will look at stress *amplitudes* and strain *amplitudes* in our modeling.

[1] More precisely, the defect-free approach assumes that the number of cycles to *propagate* a crack once it has initiated is much less than the number of cycles needed to initiate the crack.

8.1.1 Stress perspective (S-N curves)

Consider a specimen which undergoes a cyclic applied stress in the range $\sigma_{\min} \leq \sigma \leq \sigma_{\max}$.

Definition 8.1.1. The **stress range** $\Delta\sigma$ and **stress amplitude** σ_a are defined to be

$$\Delta\sigma \equiv \sigma_{\max} - \sigma_{\min}, \qquad \sigma_a \equiv \frac{1}{2}\Delta\sigma.$$

Definition 8.1.2. The **mean stress** σ_m is the average of the minimum and maximum stresses,

$$\sigma_m \equiv \frac{1}{2}(\sigma_{\max} + \sigma_{\min}).$$

If $\sigma_a < \sigma_{\text{UTS}}$, the ultimate tensile strength, then experimental data for σ_a versus N_f *for a given value of* σ_e can be plotted in a so-called **S-N curve**. Note that by convention, it is the "S" (i.e., σ_a) that is plotted on the vertical axis, and the "N" (i.e., N_f) that is plotted on the horizontal axis. This presentation allows N_f to be read off if σ_a is known, or the maximum allowable σ_a to be read off if N_f is a design parameter.

In experimental data of *ferrous alloys* (steels), an interesting phenomenon occurs. There exists an asymptotic value of σ_a below which $N_f \to \infty$; that is, for sufficiently low values of stress the specimen appears virtually immune to fatigue failure! The threshold value of stress is called the **endurance limit** and is denoted σ_e.

Remark 8.1.3: For ferrous alloys, empirical evidence has suggested the following correlation between the ultimate tensile strength σ_{UTS} in monotonic loading and the endurance limit σ_e:

$$\sigma_e = \min\left(\frac{\sigma_{\text{UTS}}}{2},\ 700\ \text{MPa}\right)$$

Remark 8.1.4: For engineering purposes, materials which do not have an endurance limit (for example, aluminum alloys) are typically assigned a "pseudo-endurance limit", which is taken to be the stress amplitude corresponding to $N_f = 10^7$ cycles.

8.1.2 Strain perspective (strain-life approach)

Consider a specimen which undergoes a cyclic applied strain in the range $\varepsilon_{\min} \leq \varepsilon \leq \varepsilon_{\max}$.

Definition 8.1.5. The **strain range** $\Delta\varepsilon$ and **strain amplitude** ε_a are defined to be

$$\Delta\varepsilon \equiv \varepsilon_{\max} - \varepsilon_{\min}, \qquad \varepsilon_a = \frac{1}{2}\Delta\varepsilon.$$

Moreover the usual additive decomposition of strain into elastic and plastic parts yields the following:

Definition 8.1.6. The **elastic strain range** and **plastic strain range** are defined as the elastic and plastic parts of the strain range, respectively:

$$\Delta\varepsilon = \Delta\varepsilon^E + \Delta\varepsilon^P.$$

The **elastic strain amplitude** and **plastic strain amplitude** are defined as the elastic and plastic parts of the strain amplitude, respectively:

$$\varepsilon_a = \varepsilon_a^E + \varepsilon_a^P.$$

Note that if the material has elastic modulus E, then $\Delta\varepsilon^E = \Delta\sigma/E$, and $\varepsilon_a^E = \sigma_a/E$.

Basquin (1910), Coffin (1954), and Manson (1953) observed the following facts about strain-driven defect-free fatigue:

- **When the stress amplitude is largely below the macroscopic yield strength of the material, the fatigue lifetime is high, $N_f \gtrsim 10^4$ cycles. In this "high-cycle" regime, the lifetime N_f primarily depends on the elastic strain amplitude, and the plastic strain amplitude is negligible[a]. Failure is dominated by the strength of the material.**
- **When the stress amplitude exceeds the yield strength, the fatigue lifetime is low, $N_f \lesssim 10^4$ cycles. In this "low-cycle" regime, the lifetime N_f primarily depends on the plastic strain amplitude, which is larger than the elastic strain amplitude. Failure is dominated by the ductility of the material.**

[a] Despite the plastic strain amplitude being small in comparison to the elastic strain amplitude, it is important to re-emphasize that *microscale* plasticity is still present, and this is the source of crack growth.

Experimentally, it has been found that power-law relationships work well for describing the fatigue lifetimes, in both regimes. Let $2N_f$ denote the number of *reversals* to failure (in one *cycle* of loading there are two *reversals* of the strain direction). Then, in the high-cycle regime,

$$\varepsilon_a^E \propto \sigma_a = \sigma_f' \cdot (2N_f)^b,$$

and in the low-cycle regime,

$$\varepsilon_a^P = \varepsilon_f' \cdot (2N_f)^c,$$

where $\{\sigma_f', b\}$, and $\{\varepsilon_f', c\}$ are material property coefficients called the **fatigue strength** and **fatigue ductility** parameters, respectively.

Finally, we note that a *tensile* mean stress, $\sigma_m > 0$, causes a reduction in the fatigue life, because of the tendency for tensile stresses to propagate cracks. This effect is particularly relevant for the high-cycle regime, when the overall plastic strain is small. Hence, the effective fatigue strength coefficient σ_f' is smaller. The usual assumption is that σ_f' can be replaced simply by $\sigma_f' - \sigma_m$.

For a complete (approximate) description of the strain-life behavior, we can combine these formulations into a "master curve" using the additive decomposition $\varepsilon_a = \varepsilon_a^E + \varepsilon_a^P$:

Key Equation 8.1.7 (Strain-life formula, defect-free fatigue approach)

For $\sigma_m \leq 0$, the total strain amplitude ε_a is related to the number of reversals to failure $2N_f$ by

$$\varepsilon_a = \frac{\sigma_f'}{E}(2N_f)^b + \varepsilon_f'(2N_f)^c, \qquad (\sigma_m \leq 0),$$

by the elastic modulus E and the fatigue strength and fatigue ductility parameters, all material constants.

For $\sigma_m > 0$, the formula accounts for the tensile mean stress effects:

$$\varepsilon_a = \frac{\sigma'_f - \sigma_m}{E}(2N_f)^b + \varepsilon'_f(2N_f)^c, \qquad (\sigma_m > 0).$$

8.1.3 Miner's Rule

The strain-life formula is appropriate when the value of ε_a is constant throughout the entire service lifetime. But this, in general, is *not* the case for most parts. Rather, a part is subjected to a series of, say, N "types" of cyclic loading, each at strain amplitude $(\varepsilon_a)_i$ for n_i cycles, $i = 1, 2, 3, ..., N$. **Miner's rule** says that we can treat this complicated history as a *linear* one in which we only care about the *cumulative fatigue damage*. Specifically, say that strain amplitude $(\varepsilon_a)_i$ is associated with a fatigue life of $(N_f)_i$ cycles, by the usual strain-life formula. Then, the failure criterion becomes

$$\sum_{i=1}^{N} \frac{n_i}{(N_f)_i} = 1.$$

In words, each of the N "blocks" of loading contributes a nondimensional fraction of total damage $n_i/(N_f)_i$, which is normalized by the computed fatigue life for that loading. When the cumulative damage equals one, the part is said to have reached its fatigue life.

8.2 Defect-tolerant approach

In the defect-tolerant approach, we are concerned with *the number of cycles to failure*, again N_f, that corresponds to the number of cycles to *propagate* an (existing or assumed-to-exist) initial crack of length a_i to a final length a_c, which corresponds with the condition $K_I = K_{Ic}$. The length a_i is either measured using NDE, or assumed to be the largest undetectable crack with the present NDE measurement technique, if no crack is explicitly found. The length a_c is given in terms of the material's fracture toughness and the maximum far-field stress,

$$a_c = \frac{1}{\pi}\left(\frac{K_{Ic}}{Q\sigma_{\max}}\right)^2.$$

To determine the number of cycles it takes to propagate a crack from a_i to a_c, it is necessary to model the *rate* at which the crack grows, with respect to the number of cycles. In particular, let da/dN represent the incremental extension Δa of a crack of length a in a single cycle. In terms of fracture properties, the applied cyclic stress with range $\Delta\sigma$ corresponds to an applied *cyclic stress intensity* with range ΔK_I:

$$\Delta K_I = Q\Delta\sigma\sqrt{\pi a} = Q(\sigma_{\max} - \sigma_{\min})\sqrt{\pi a}.$$

Remark 8.2.1: In the defect-tolerant approach the stress range $\Delta\sigma$ is only the *tensile* stress range, because compressive stresses are assumed *not* to contribute to crack growth. Hence, if $\sigma_{\min} < 0$, we take $\Delta K_I = Q(\sigma_{\max} - 0)\sqrt{\pi a}$. For example, if the stress on a component is cycled between -3 MPa and $+3$ MPa, use $\Delta\sigma = 3$ MPa.

Experimental data of da/dN versus ΔK_I reveals the following trend:

1. At low values of ΔK_I (corresponding to low values of $\Delta\sigma$), there is negligible crack growth. Typically $da/dN \leq 10^{-9}$ m/cycle, so the crack is only advancing by nanometers every second.

2. Past a *threshold value* of ΔK_I, which is denoted $\Delta K_{I,th}$, the plot of da/dN versus ΔK_I is roughly log-linear, suggesting a power law relationship,

$$\frac{da}{dN} = C(\Delta K_I)^m.$$

   The fit parameters C and m are experimentally determined and represent material parameters. In this regime, typically 10^{-9} m/cycle $\leq da/dN \leq 10^{-6}$ m/cycle.

3. As $K_{I,max}$ approaches K_{Ic}, the crack growth *rate* increases without bound until $K_{I,max} = K_{Ic}$ and failure occurs. The upper limit of ΔK_I that corresponds with a deviation from log-linearity is usually taken to be the design upper limit for crack growth.

Therefore the following simple form for the crack growth rate is assumed.

Key Equation 8.2.2 (Paris' Law)

Let $\Delta K_I = Q\Delta\sigma\sqrt{\pi a}$ be the cyclic stress intensity range. Then the crack growth rate da/dN is given by

$$\frac{da}{dN} = \begin{cases} 0, & \Delta K_I < \Delta K_{I,th}; \\ C(\Delta K_I)^m, & \Delta K_I \geq \Delta K_{I,th}. \end{cases}$$

Remark 8.2.3: The dimensions of C are complicated and depend on the value of m. In particular, C has units of [(meters per cycle) per $(\text{MPa}\sqrt{\text{m}})^m$].

Remark 8.2.4: In general, the configuration correction factor Q is a function of a, and thus Q changes as the crack grows. This makes it difficult to write an analytical formula for $a = a(N)$ without making assumptions. One common simplifying assumption is that Q is roughly *constant* during the crack growth process, which is appropriate for *small* cracks.

Computing the number of cycles to failure is thus simplified to the problem of finding $N_i = 0 \to N_f$ corresponding to the growth of the crack from $a_i \to a_c$ according to Paris' law. To that end, we rearrange the differential equation as

$$dN = \frac{1}{C}\frac{1}{(\Delta K_I)^m}\,da = \frac{1}{C}\frac{1}{(Q(a)\Delta\sigma\sqrt{\pi a})^m}\,da,$$

from which it follows that

$$N_f = \int_0^{N_f} dN = \int_{a_i}^{a_c} \frac{1}{C}\frac{1}{(Q(a)\Delta\sigma\sqrt{\pi a})^m}\,da.$$

If Q and $\Delta\sigma$ are constant over the entire loading period, then the integrand simplifies and an analytical solution can be written out.

Key Equation 8.2.5 (Integrated form of Paris' Law)

If $\Delta K_I \geq \Delta K_{I,th}$, for constant Q and $\Delta\sigma$, the number of cycles to failure N_f, corresponding to the number of cycles it takes to grow a crack from size a_i to size $a_c = (1/\pi)(K_{Ic}/Q\sigma_{\max})^2$, is given by

$$N_f = \frac{1}{C}\frac{1}{(Q\Delta\sigma\sqrt{\pi})^m}\int_{a_i}^{a_c} a^{-m/2}\, da.$$

For $m = 2$, this yields

$$N_f = \frac{1}{C}\frac{1}{(Q\Delta\sigma\sqrt{\pi})^m}\left[\ln\left(\frac{a_c}{a_i}\right)\right],$$

and for $m \neq 2$ this yields

$$N_f = \frac{2}{(m-2)C(Q\Delta\sigma\sqrt{\pi})^m}\left[a_i^{(2-m)/2} - a_c^{(2-m)/2}\right].$$

Clearly N_f is affected by the value of K_{Ic} (through a_c), the initial crack size a_i, and the stress range $\Delta\sigma$. Typically, K_{Ic} (a material property) and $\Delta\sigma$ (a property of the application) are hard to control, so the best knob to tune in practical engineering practices is a_i, either by an improved manufacturing process and/or NDE techniques with better resolution.

Exercise 8.2.6. AKG 4.16 through 4.28.