Let V be the three dimensional Euclidean vector space and L be the set of all linear maps from V to V. The set of real numbers is denoted by $\mathbb{R}$. Let E be the three dimensional Euclidean point space. It can be identified with V.

Derivatives of fields By *fields* we mean scalar, vector and tensor valued functions defined on position ($\mathbf{X}$) and time (t). In the following we are mainly concerned with the derivatives with respect to the position and therefore dependence of fields on time is suppressed.
A scalar-valued field $\phi(\mathbf{X})$ is differentiable at $\mathbf{X}_0 \in \mathcal{U}(\mathbf{X}_0)$, where $\mathcal{U}(\mathbf{X}_0) \subset E$ is an open neighborhood of $\mathbf{X}_0$, if there exists a unique $\mathbf{c} \in V$ such that

$$\phi(\mathbf{X}) = \phi(\mathbf{X}_0) + \mathbf{c}(\mathbf{X}_0) \cdot (\mathbf{X} - \mathbf{X}_0) + o(|\mathbf{X} - \mathbf{X}_0|), \tag{0.1}$$

where $\frac{o(\epsilon)}{\epsilon} \to 0$ as $\epsilon \to 0$. We call $\mathbf{c}(\mathbf{X}_0) = \nabla\phi|_{\mathbf{X}_0}$ (or $\nabla\phi(\mathbf{X}_0)$) the *gradient* of ϕ at $\mathbf{X}_0$. Consider a curve $\mathbf{X}(u)$ in E parameterized by $u \in \mathbb{R}$. Let $\psi(u) = \phi(\mathbf{X}(u))$ and $\mathbf{X}_1 = \mathbf{X}(u_1)$, $\mathbf{X}_0 = \mathbf{X}(u_0)$ for $\{u_1, u_0\} \in \mathbb{R}$. Then from (0.1),

$$\psi(u_1) - \psi(u_0) = \nabla\phi(\mathbf{X}_0) \cdot (\mathbf{X}_1 - \mathbf{X}_0) + o(|\mathbf{X}_1 - \mathbf{X}_0|). \tag{0.2}$$

Moreover $\mathbf{X}_1 - \mathbf{X}_0 = \mathbf{X}'(u_0)(u_1 - u_0) + o(|u_1 - u_0|)$, where $\mathbf{X}'(u_0)$ is the derivative of $\mathbf{X}$ with respect to u at $u = u_0$. Therefore, $|\mathbf{X}_1 - \mathbf{X}_0| = O(|u_1 - u_0|)$ and consequently we can rewrite (0.2)

$$\frac{\psi(u_1) - \psi(u_0)}{u_1 - u_0} = \nabla\phi(\mathbf{X}_0) \cdot \mathbf{X}'(u_0) + \frac{o(|u_1 - u_0|)}{u_1 - u_0}. \tag{0.3}$$

For $u_1 \to u_0$ we obtain *the chain rule*, $\psi'(u_0) = \nabla\phi(\mathbf{X}(u_0)) \cdot \mathbf{X}'(u_0)$, which can also be expressed as $\frac{d\phi}{du} = \nabla\phi(\mathbf{X}) \cdot \frac{d\mathbf{X}}{du}$ or

$$d\phi(\mathbf{X}) = \nabla\phi(\mathbf{X}) \cdot d\mathbf{X}. \tag{0.4}$$

A vector-valued field $\mathbf{v}(\mathbf{X})$ is differentiable at $\mathbf{X}_0 \in \mathcal{U}(\mathbf{X}_0)$ if there exists a unique tensor $\mathbf{l} \in L$ such that

$$\mathbf{v}(\mathbf{X}) = \mathbf{v}(\mathbf{X}_0) + \mathbf{l}(\mathbf{X}_0)(\mathbf{X} - \mathbf{X}_0) + \mathbf{r}, \tag{0.5}$$

where $|\mathbf{r}| = o(|\mathbf{X} - \mathbf{X}_0|)$. We call $\mathbf{l}(\mathbf{X}_0) = \nabla\mathbf{v}|_{\mathbf{X}_0}$ (or $\nabla\mathbf{v}(\mathbf{X}_0)$) the *gradient* of $\mathbf{v}$ at $\mathbf{X}_0$. The chain rule in this case can be obtained following the procedure preceding equation (0.4):

$$d\mathbf{v}(\mathbf{X}) = (\nabla\mathbf{v})d\mathbf{X}. \tag{0.6}$$

The *divergence* of a vector field is a scalar defined by

$$\operatorname{Div}\mathbf{v} = \operatorname{tr}(\nabla\mathbf{v}). \tag{0.7}$$

The *curl* of a vector field is a vector defined by

$$(\operatorname{Curl}\mathbf{v}) \cdot \mathbf{c} = \operatorname{Div}(\mathbf{v} \times \mathbf{c}) \tag{0.8}$$

for any fixed $\mathbf{c} \in V$.

Differentiability of a tensor-valued function is defined in a similar manner. In particular, for a tensor field $\mathbf{A}(\mathbf{X})$, we write

$$d\mathbf{A}(\mathbf{X}) = (\nabla \mathbf{A})d\mathbf{X}. \tag{0.9}$$

The *divergence* of $\mathbf{A}$ is the vector defined by

$$(\mathrm{Div}\,\mathbf{A}) \cdot \mathbf{c} = \mathrm{Div}(\mathbf{A}^T\mathbf{c}) \tag{0.10}$$

for any fixed $\mathbf{c} \in V$. The *curl* of $\mathbf{A}$ is the tensor defined by

$$(\mathrm{Curl}\,\mathbf{A})\mathbf{c} = \mathrm{Curl}(\mathbf{A}^T\mathbf{c}) \tag{0.11}$$

for any fixed $\mathbf{c} \in V$.

Derivatives of functions on tensor spaces A function $f(\mathbf{A}) : L \to \mathbb{R}$ is said to be differentiable at $\mathbf{A} \in L$ if there exists a linear mapping $\partial_{\mathbf{A}} f$ (which maps tensors in L to scalars) such that for all $\mathbf{B} \in L$ ($\mathbf{B}$ should belong to some open neighborhood of $\mathbf{A}$)

$$f(\mathbf{A} + \mathbf{B}) - f(\mathbf{A}) = \partial_{\mathbf{A}} f(\mathbf{A})[\mathbf{B}] + o(\mathbf{B}). \tag{0.12}$$

This definition is equivalent to

$$\partial_{\mathbf{A}} f(\mathbf{A})[\mathbf{B}] = \frac{d}{ds} f(\mathbf{A} + s\mathbf{B})|_{s=0}. \tag{0.13}$$

A similar definition holds for vector and tensor valued functions (and also for vector arguments). The notation $\alpha[\beta]$ is clear from the context at hand. For example, if $\alpha = \mathbf{A}$ (tensor) and $\beta = \mathbf{B}$ (tensor) then $\mathbf{A}[\mathbf{B}] = \mathbf{A} \cdot \mathbf{B}$. If instead $\beta = \mathbf{b}$ (vector) then $\mathbf{A}[\mathbf{b}] = \mathbf{A}\mathbf{b}$. If $\alpha = \mathbf{a}$ (vector) then $\mathbf{a}[\mathbf{b}] = \mathbf{a} \cdot \mathbf{b}$. See (Gurtin (Introduction to Continuum Mechanics), Chapter 2) for examples.

Kinematic constraints These are a set of local constraints on $\mathbf{F}$ (deformation gradient). Examples include incompressibility, rigidity, etc. The tensor $\mathbf{F}$ can be seen as an element of a nine dimensional space. Therefore the number of constraints cannot exceed nine. However, due to material frame indifference, the dependence of any function (which in the present case defines a constraint) on $\mathbf{F}$ is through $\mathbf{C} = \mathbf{F}^T\mathbf{F}$, thereby reducing the dimension of the space to six. The number of independent kinematic constraints can thus be no more than six. Let these be denoted by

$$\phi^{(i)}(\mathbf{F}) = 0, \quad i = 1, \dots n \leq 6. \tag{0.14}$$

We denote by $\mathcal{M}$, the constrained manifold, which is a subset of L and constitutes all the deformation gradients which satisfy the given constraints. Consider a curve on this manifold, parameterized by u. Let

$$f^{(i)}(u) = \phi^{(i)}(\mathbf{F}(u)). \tag{0.15}$$

Since $f^{(i)}(u) = 0$ identically over the manifold $\mathcal{M}$, we can write (the derivative $\dot{f}$ is with respect to u)

$$0 = \dot{f}^{(i)}(u) = \phi^{(i)}_{\mathbf{F}}(\mathbf{F}) \cdot \dot{\mathbf{F}}(u). \tag{0.16}$$

But $\dot{\mathbf{F}}(u) \in T_{\mathcal{M}}(\mathbf{F})$ i.e. the tangent space associated with $\mathcal{M}$ at $\mathbf{F}$. Therefore $\phi^{(i)}_{\mathbf{F}}$ are orthogonal to $T_{\mathcal{M}}(\mathbf{F})$. If the set of constraints are independent to each other, then the set $\{\phi^{(i)}_{\mathbf{F}}\}$ constitutes a basis for the vector space orthogonal to $T_{\mathcal{M}}(\mathbf{F})$.

We have, for a perfectly elastic material, $\mathbf{P} \cdot \dot{\mathbf{F}} = \dot{W}(\mathbf{F}) = W_{\mathbf{F}} \cdot \dot{\mathbf{F}}$, which implies that

$$(\mathbf{P} - W_{\mathbf{F}}) \cdot \dot{\mathbf{F}} = 0 \tag{0.17}$$

for all $\dot{\mathbf{F}} \in T_{\mathcal{M}}(\mathbf{F})$. Therefore $(\mathbf{P} - W_{\mathbf{F}})$ belongs to the orthogonal vector space to $T_{\mathcal{M}}(\mathbf{F})$, and can be expressed as

$$(\mathbf{P} - W_{\mathbf{F}}) = \sum_{i=1}^{n} \lambda_i \phi^{(i)}_{\mathbf{F}} \tag{0.18}$$

for some scalars λ_i (Lagrange multipliers).

Remark: The function W (contrary to $\phi^{(i)}$) are defined only for $\mathbf{F} \in \mathcal{M}$. The derivative $W_{\mathbf{F}}$ is calculated by first assuming that there exists an extension $\hat{W}$ of W such that $\hat{W} = W$ for all $\mathbf{F} \in \mathcal{M}$. The extension can be differentiated and evaluated on $\mathcal{M}$. It can be shown that any extension can be used without any loss of generality.

Consider incompressibility as an example. Here we have only one constraint of the form $\phi(\mathbf{F}) = \det \mathbf{F} - 1 = 0$. Then $\phi_{\mathbf{F}} = \mathbf{F}^* = \mathbf{F}^{-T}$ on $\mathcal{M}$ (since $\det \mathbf{F} = 1$ on $\mathcal{M}$). Then $\mathbf{P} = W_{\mathbf{F}} + \lambda \mathbf{F}^{-T}$, which on using $\mathbf{P} = \mathbf{T}\mathbf{F}^*$ gives us an expression of the form $\mathbf{T} = W_{\mathbf{F}}\mathbf{F}^T - p\mathbf{1}$ where $p(\mathbf{X}, t)$ is the constraint pressure.