Chapter 6

Differential Calculus

In this chapter, it is assumed that all linear spaces and flat spaces under consideration are finite-dimensional.

61 Differentiation of Processes

Let $\mathcal{E}$ be a flat space with translation space $\mathcal{V}$. A mapping $p : I \to \mathcal{E}$ from some interval $I \in \operatorname{Sub}\mathbb{R}$ to $\mathcal{E}$ will be called a **process**. It is useful to think of the value $p(t) \in \mathcal{E}$ as describing the *state* of some physical system at time t. In special cases, the mapping p describes the motion of a particle and $p(t)$ is the *place* of the particle at time t. The concept of differentiability for real-valued functions (see Sect.08) extends without difficulty to processes as follows:

Definition 1: *The process $p : I \to \mathcal{E}$ is said to be* **differentiable at** $t \in I$ *if the limit*

$$\partial_t p := \lim_{s \to 0} \frac{1}{s}(p(t+s) - p(t)) \tag{61.1}$$

exists. Its value $\partial_t p \in \mathcal{V}$ is then called the **derivative of** p **at** t. *We say that p is* **differentiable** *if it is differentiable at all $t \in I$. In that case, the mapping $\partial p : I \to \mathcal{V}$ defined by $(\partial p)(t) := \partial_t p$ for all $t \in I$ is called the* **derivative** *of p.*

Given $n \in \mathbb{N}^{\times}$, we say that p is n **times differentiable** *if $\partial^n p : I \to \mathcal{V}$ can be defined by the recursion*

$$\partial^1 p := \partial p, \quad \partial^{k+1} p := \partial(\partial^k p) \quad \text{for all} \quad k \in (n-1)^{]}. \tag{61.2}$$

We say that p is **of class C^n** *if it is n times differentiable and $\partial^n p$ is continuous. We say that p is* **of class C$^\infty$** *if it is of class C^n for all $n \in \mathbb{N}^{\times}$.*

As for a real-valued function, it is easily seen that a process p is continuous at $t \in \operatorname{Dom} p$ if it is differentiable at t. Hence p is continuous if it is differentiable, but it may also be continuous without being differentiable.

In analogy to (08.34) and (08.35), we also use the notation

$$p^{(k)} := \partial^k p \quad \text{for all} \quad k \in (n-1)^] \tag{61.3}$$

when p is an n-times differentiable process, and we use

$$p^{\bullet} := p^{(1)} = \partial p, \quad p^{\bullet\bullet} := p^{(2)} = \partial^2 p, \quad p^{\bullet\bullet\bullet} := p^{(3)} = \partial^3 p, \tag{61.4}$$

if meaningful.

We use the term "process" also for a mapping $p : I \to \mathcal{D}$ from some interval I into a *subset* $\mathcal{D}$ of the flat space $\mathcal{E}$. In that case, we use poetic license and ascribe to p any of the properties defined above if $p|^{\mathcal{E}}$ has that property. Also, we write ∂p instead of $\partial(p|^{\mathcal{E}})$ if p is differentiable, etc. (If $\mathcal{D}$ is included in some flat $\mathcal{F}$, then one can take the direction space of $\mathcal{F}$ rather than all of $\mathcal{V}$ as the codomain of ∂p. This ambiguity will usually not cause any difficulty.)

The following facts are immediate consequences of Def.1, and Prop.5 of Sect.56 and Prop.6 of Sect.57.

Proposition 1: *The process $p : I \to \mathcal{E}$ is differentiable at $t \in I$ if and only if, for each $\boldsymbol{\lambda}$ in some basis of $\mathcal{V}^*$, the function $\boldsymbol{\lambda}(p - p(t)) : I \to \mathbb{R}$ is differentiable at t.*

The process p is differentiable if and only if, for every flat function $a \in$ $\operatorname{Flf} \mathcal{E}$*, the function $a \circ p$ is differentiable.*

Proposition 2: *Let $\mathcal{E}, \mathcal{E}'$ be flat spaces and $\alpha : \mathcal{E} \to \mathcal{E}'$ a flat mapping. If $p : I \to \mathcal{E}$ is a process that is differentiable at $t \in I$, then $\alpha \circ p : I \to \mathcal{E}'$ is also differentiable at $t \in I$ and*

$$\partial_t(\alpha \circ p) = (\nabla\alpha)(\partial_t p). \tag{61.5}$$

If p is differentiable then (61.5) *holds for all $t \in I$ and we get*

$$\partial(\alpha \circ p) = (\nabla\alpha)\partial p. \tag{61.6}$$

Let $p : I \to \mathcal{E}$ and $q : I \to \mathcal{E}'$ be processes having the same domain I. Then $(p, q) : I \to \mathcal{E} \times \mathcal{E}'$, defined by value-wise pair formation, (see (04.13)) is another process. It is easily seen that p and q are both differentiable at $t \in I$ if and only if (p, q) is differentiable at t. If this is the case we have

$$\partial_t(p, q) = (\partial_t q, \partial_t q). \tag{61.7}$$

Both p and q are differentiable if and only if (p,q) is, and, in that case,

$$\partial(p,q) = (\partial p, \partial q). \tag{61.8}$$

Let p and q be processes having the same domain I and the same codomain $\mathcal{E}$. Since the point-difference $(x,y) \mapsto x - y$ is a flat mapping from $\mathcal{E} \times \mathcal{E}$ into $\mathcal{V}$ whose gradient is the vector-difference $(\mathbf{u},\mathbf{v}) \mapsto \mathbf{u} - \mathbf{v}$ from $\mathcal{V} \times \mathcal{V}$ into $\mathcal{V}$ we can apply Prop.2 to obtain

Proposition 3: *If $p : I \to \mathcal{E}$ and $q : I \to \mathcal{E}$ are both differentiable at $t \in I$, so is the value-wise difference $p - q : I \to \mathcal{V}$, and*

$$\partial_t(p - q) = (\partial_t p) - (\partial_t q). \tag{61.9}$$

If p and q are both differentiable, then (61.9) *holds for all $t \in I$ and we get*

$$\partial(p - q) = \partial p - \partial q. \tag{61.10}$$

The following result generalizes the Difference-Quotient Theorem stated in Sect.08.

Difference-Quotient Theorem: *Let $p : I \to \mathcal{E}$ be a process and let $t_1, t_2 \in I$ with $t_1 < t_2$. If $p|_{[t_1,t_2]}$ is continuous and if p is differentiable at each $t \in]t_1, t_2[$ then*

$$\frac{p(t_2) - p(t_1)}{t_2 - t_1} \in \operatorname{Clo}\operatorname{Cxh}\{\partial_t p \mid t \in]t_1, t_2[\}. \tag{61.11}$$

Proof: Let $a \in \operatorname{Flf}\mathcal{E}$ be given. Then $(a \circ p)\ |_{[t_1,t_2]}$ is continuous and, by Prop.1, $a \circ p$ is differentiable at each $t \in]t_1, t_2[$. By the elementary Difference-Quotient Theorem (see Sect.08) we have

$$\frac{(a \circ p)(t_2) - (a \circ p)(t_2)}{t_2 - t_1} \in \{\partial_t(a \circ p) \mid t \in]t_1, t_2[\}.$$

Using (61.5) and (33.4), we obtain

$$\nabla a \left(\frac{p(t_2) - p(t_1)}{t_2 - t_1} \right) \in (\nabla a)_>(\mathcal{S}), \tag{61.12}$$

where

$$\mathcal{S} := \{\partial_t p \mid t \in]t_1, t_2[\}.$$

Since (61.12) holds for all $a \in \operatorname{Flf}\mathcal{E}$ we can conclude that $b(\frac{p(t_2)-p(t_1)}{t_2-t_1}) \geq 0$ holds for all those $b \in \operatorname{Flf}\mathcal{V}$ that satisfy $b_>(\mathcal{S}) \subset \mathbb{P}$. Using the Half-Space Intersection Theorem of Sect.54, we obtain the desired result (61.11). ∎

Notes 61

(1) See Note (8) to Sect.08 concerning notations such as $\partial_t p$, $\partial p, \partial^n p, p^{\cdot}, p^{(n)}$, etc.

62 Small and Confined Mappings

Let $\mathcal{V}$ and $\mathcal{V}'$ be linear spaces of strictly positive dimension. Consider a mapping $\mathbf{n}$ from a neighborhood of zero in $\mathcal{V}$ to a neighborhood of zero in $\mathcal{V}'$. If $\mathbf{n}(\mathbf{0}) = \mathbf{0}$ and if $\mathbf{n}$ is continuous at $\mathbf{0}$, then we can say, intuitively, that $\mathbf{n}(\mathbf{v})$ approaches $\mathbf{0}$ in $\mathcal{V}'$ as $\mathbf{v}$ approaches $\mathbf{0}$ in $\mathcal{V}$. We wish to make precise the idea that $\mathbf{n}$ is *small* near $\mathbf{0} \in \mathcal{V}$ in the sense that $\mathbf{n}(\mathbf{v})$ approaches $\mathbf{0} \in \mathcal{V}'$ *faster* than $\mathbf{v}$ approaches $\mathbf{0} \in \mathcal{V}$.

Definition 1: *We say that a mapping* $\mathbf{n}$ *from a neighborhood of* $\mathbf{0}$ *in* $\mathcal{V}$ *to a neighborhood of* $\mathbf{0}$ *in* $\mathcal{V}'$ *is* **small near 0** *if* $\mathbf{n}(\mathbf{0}) = \mathbf{0}$ *and, for all norms* ν *and* ν' *on* $\mathcal{V}$ *and* $\mathcal{V}'$*, respectively, we have*

$$\lim_{\mathbf{u}\to\mathbf{0}} \frac{\nu'(\mathbf{n}(\mathbf{u}))}{\nu(\mathbf{u})} = 0. \tag{62.1}$$

The set of all such small mappings will be denoted by $\text{Small}(\mathcal{V}, \mathcal{V}')$.

Proposition 1: *Let* $\mathbf{n}$ *be a mapping from a neighborhood of* $\mathbf{0}$ *in* $\mathcal{V}$ *to a neighborhood of* $\mathbf{0}$ *in* $\mathcal{V}'$*. Then the following conditions are equivalent:*

(i) $\mathbf{n} \in \text{Small}(\mathcal{V}, \mathcal{V}')$.

(ii) $\mathbf{n}(\mathbf{0}) = \mathbf{0}$ *and the limit-relation* (62.1) *holds for some norm* ν *on* $\mathcal{V}$ *and some norm* ν' *on* $\mathcal{V}'$.

(iii) *For every bounded subset* $\mathcal{S}$ *of* $\mathcal{V}$ *and every* $\mathcal{N}' \in \text{Nhd}_0(\mathcal{V}')$ *there is a* $\delta \in \mathbb{P}^\times$ *such that*

$$\mathbf{n}(s\mathbf{v}) \in s\mathcal{N}' \quad \text{for all} \quad s \in]-\delta, \delta[\tag{62.2}$$

and all $\mathbf{v} \in \mathcal{S}$ *such that* $s\mathbf{v} \in \text{Dom}\,\mathbf{n}$.

Proof: (i) $\Rightarrow$ (ii): This implication is trivial.

(ii) $\Rightarrow$ (iii): Assume that (ii) is valid. Let $\mathcal{N}' \in \text{Nhd}_{\mathbf{0}}(\mathcal{V}')$ and a bounded subset $\mathcal{S}$ of $\mathcal{V}$ be given. By Cor.1 to the Cell-Inclusion Theorem of Sect.52, we can choose $b \in \mathbb{P}^\times$ such that

$$\nu(\mathbf{v}) \le b \quad \text{for all} \quad \mathbf{v} \in \mathcal{S}. \tag{62.3}$$

By Prop.3 of Sect.53 we can choose $\varepsilon \in \mathbb{P}^\times$ such that

$$\varepsilon b \text{Ce}(\nu') \subset \mathcal{N}'. \tag{62.4}$$

Applying Prop.4 of Sect.57 to the assumption (ii) we obtain $\delta \in \mathbb{P}^\times$ such that, for all $\mathbf{u} \in \text{Dom}\,\mathbf{n}$,

$$\nu'(\mathbf{n}(\mathbf{u})) < \varepsilon\nu(\mathbf{u}) \quad \text{if} \quad 0 < \nu(\mathbf{u}) < \delta b. \tag{62.5}$$

Now let $\mathbf{v} \in \mathcal{S}$ be given. Then $\nu(s\mathbf{v}) = |s|\nu(\mathbf{v}) \leq |s|b < \delta b$ for all $s \in]-\delta, \delta[$ such that $s\mathbf{v} \in \mathrm{Dom}\,\mathbf{n}$. Therefore, by (62.5), we have

$$\nu'(\mathbf{n}(s\mathbf{v})) < \varepsilon\nu(s\mathbf{v}) = \varepsilon|s|\nu(\mathbf{v}) \leq |s|\varepsilon b$$

if $s\mathbf{v} \neq \mathbf{0}$, and hence

$$\mathbf{n}(s\mathbf{v}) \in s\varepsilon b\mathrm{Ce}(\nu')$$

for all $s \in]-\delta, \delta[$ such that $s\mathbf{v} \in \mathrm{Dom}\,\mathbf{n}$. The desired conclusion (62.2) now follows from (62.4).

(iii) ⇒ (i): Assume that (iii) is valid. Let a norm ν on $\mathcal{V}$, a norm ν' on $\mathcal{V}'$, and $\varepsilon \in \mathbb{P}^\times$ be given. We apply (iii) to the choices $\mathcal{S} := \mathrm{Ce}(\nu)$, $\mathcal{N}' := \varepsilon\mathrm{Ce}(\nu')$ and determine $\delta \in \mathbb{P}^\times$ such that (62.2) holds. If we put $s := 0$ in (62.2) we obtain $\mathbf{n}(\mathbf{0}) = \mathbf{0}$. Now let $\mathbf{u} \in \mathrm{Dom}\,\mathbf{n}$ be given such that $0 < \nu(\mathbf{u}) < \delta$. If we apply (62.2) with the choices $s := \nu(\mathbf{u})$ and $\mathbf{v} := \frac{1}{s}\mathbf{u}$, we see that $\mathbf{n}(\mathbf{u}) \in \nu(\mathbf{u})\varepsilon\mathrm{Ce}(\nu')$, which yields

$$\frac{\nu'(\mathbf{n}(\mathbf{u}))}{\nu(\mathbf{u})} < \varepsilon.$$

The assertion follows by applying Prop.4 of Sect.57. ∎

The condition (iii) of Prop.1 states that

$$\lim_{s \to 0} \frac{1}{s}\mathbf{n}(s\mathbf{v}) = 0 \tag{62.6}$$

for all $\mathbf{v} \in \mathcal{V}$ and, roughly, that the limit is approached uniformly as $\mathbf{v}$ varies in an arbitrary bounded set.

We also wish to make precise the intuitive idea that a mapping $\mathbf{h}$ from a neighborhood of $\mathbf{0}$ in $\mathcal{V}$ to a neighborhood of $\mathbf{0}$ in $\mathcal{V}'$ is *confined* near zero in the sense that $\mathbf{h}(\mathbf{v})$ approaches $\mathbf{0} \in \mathcal{V}'$ *not more slowly* than $\mathbf{v}$ approaches $\mathbf{0} \in \mathcal{V}$.

Definition 2: *A mapping* $\mathbf{h}$ *from a neighborhood of* $\mathbf{0}$ *in* $\mathcal{V}$ *to a neighborhood of* $\mathbf{0}$ *in* $\mathcal{V}'$ *is said to be* **confined near 0** *if for every norm* ν *on* $\mathcal{V}$ *and every norm* ν' *on* $\mathcal{V}'$ *there is* $\mathcal{N} \in \mathrm{Nhd}_{\mathbf{0}}(\mathcal{V})$ *and* $\kappa \in \mathbb{P}^\times$ *such that*

$$\nu'(\mathbf{h}(\mathbf{u})) \leq \kappa\nu(\mathbf{u}) \quad \text{for all} \quad \mathbf{u} \in \mathcal{N} \cap \mathrm{Dom}\,\mathbf{h}. \tag{62.7}$$

The set of all such confined mappings will be denoted by $\mathrm{Conf}(\mathcal{V}, \mathcal{V}')$.

Proposition 2: *Let* $\mathbf{h}$ *be a mapping from a neighborhood of* $\mathbf{0}$ *in* $\mathcal{V}$ *to a neighborhood of* $\mathbf{0}$ *in* $\mathcal{V}'$. *Then the following are equivalent:*

(i) $\mathbf{h} \in \mathrm{Conf}(\mathcal{V}, \mathcal{V}')$.

(ii) *There exists a norm* ν *on* $\mathcal{V}$*, a norm* ν' *on* $\mathcal{V}'$*, a neighborhood* $\mathcal{N}$ *of* $\mathbf{0}$ *in* $\mathcal{V}$*, and* $\kappa \in \mathbb{P}^\times$ *such that* (62.7) *holds.*

(iii) *For every bounded subset* $\mathcal{S}$ *of* $\mathcal{V}$ *there is* $\delta \in \mathbb{P}^\times$ *and a bounded subset* $\mathcal{S}'$ *of* $\mathcal{V}'$ *such that*

$$\mathbf{h}(s\mathbf{v}) \in s\mathcal{S}' \quad \text{for all} \quad s \in \,]-\delta, \delta[\tag{62.8}$$

and all $\mathbf{v} \in \mathcal{S}$ *such that* $s\mathbf{v} \in \operatorname{Dom} \mathbf{h}$.

Proof: (i) ⇒ (ii): This is trivial.

(ii) ⇒ (iii): Assume that (ii) holds. Let a bounded subset $\mathcal{S}$ of $\mathcal{V}$ be given. By Cor.1 to the Cell-Inclusion Theorem of Sect.52, we can choose $b \in \mathbb{P}^\times$ such that

$$\nu(\mathbf{v}) \leq b \quad \text{for all} \quad \mathbf{v} \in \mathcal{S}.$$

By Prop.3 of Sect.53, we can determine $\delta \in \mathbb{P}^\times$ such that $\delta b \operatorname{Ce}(\nu) \subset \mathcal{N}$. Hence, by (62.7), we have for all $\mathbf{u} \in \operatorname{Dom} \mathbf{h}$

$$\nu'(\mathbf{h}(\mathbf{u})) \leq \kappa\nu(\mathbf{u}) \quad \text{if} \quad \nu(\mathbf{u}) < \delta b. \tag{62.9}$$

Now let $\mathbf{v} \in \mathcal{S}$ be given. Then $\nu(s\mathbf{v}) = |s|\nu(\mathbf{v}) \leq |s|b < \delta b$ for all $s \in \,]-\delta, \delta[$ such that $s\mathbf{v} \in \operatorname{Dom} \mathbf{h}$. Therefore, by (62.9) we have

$$\nu'(\mathbf{h}(s\mathbf{v}) \leq \kappa\nu(s\mathbf{v}) = \kappa|s|\nu(\mathbf{v}) < |s|\kappa b$$

and hence

$$\mathbf{h}(s\mathbf{v}) \in s\kappa b \operatorname{Ce}(\nu')$$

for all $s \in \,]-\delta, \delta[$ such that $s\mathbf{v} \in \operatorname{Dom} \mathbf{h}$. If we put $\mathcal{S}' := \kappa b \operatorname{Ce}(\nu')$, we obtain the desired conclusion (62.8).

(iii) ⇒ (i): Assume that (iii) is valid. Let a norm ν on $\mathcal{V}$ and a norm ν' on $\mathcal{V}'$ be given. We apply (iii) to the choice $\mathcal{S} := \operatorname{Ce}(\nu)$ and determine $\mathcal{S}'$ and $\delta \in \mathbb{P}^\times$ such that (62.8) holds. Since $\mathcal{S}'$ is bounded, we can apply the Cell-Inclusion Theorem of Sect.52 and determine $\kappa \in \mathbb{P}^\times$ such that $\mathcal{S}' \subset \kappa \operatorname{Ce}(\nu')$. We put $\mathcal{N} := \delta \operatorname{Ce}(\nu) \cap \operatorname{Dom} \mathbf{h}$, which belongs to $\operatorname{Nhd}_{\mathbf{0}}(\mathcal{V})$. If we put $s := 0$ in (62.8) we obtain $\mathbf{h}(\mathbf{0}) = \mathbf{0}$, which shows that (62.7) holds for $\mathbf{u} := \mathbf{0}$. Now let $\mathbf{u} \in \mathcal{N}^\times$ be given, so that $0 < \nu(\mathbf{u}) < \delta$. If we apply (62.8) with the choices $s := \nu(\mathbf{u})$ and $\mathbf{v} := \frac{1}{s}\mathbf{u}$, we see that

$$\mathbf{h}(\mathbf{u}) \in \nu(\mathbf{u})\mathcal{S}' \subset \nu(\mathbf{u})\kappa \operatorname{Ce}(\nu'),$$

which yields the assertion (62.7). ∎

The following results are immediate consequences of the definition and of the properties on linear mappings discussed in Sect.52:

(I) Value-wise sums and value-wise scalar multiples of mappings that are small [confined] near zero are again small [confined] near zero.

(II) Every mapping that is small near zero is also confined near zero, i.e.

$$\mathrm{Small}(\mathcal{V},\mathcal{V}') \subset \mathrm{Conf}(\mathcal{V},\mathcal{V}').$$

(III) If $\mathbf{h} \in \mathrm{Conf}(\mathcal{V},\mathcal{V}')$, then $\mathbf{h}(\mathbf{0}) = \mathbf{0}$ and $\mathbf{h}$ is continuous at $\mathbf{0}$.

(IV) Every linear mapping is confined near zero, i.e.

$$\mathrm{Lin}(\mathcal{V},\mathcal{V}') \subset \mathrm{Conf}(\mathcal{V},\mathcal{V}').$$

(V) The only linear mapping that is small near zero is the zero-mapping, i.e.

$$\mathrm{Lin}(\mathcal{V},\mathcal{V}') \cap \mathrm{Small}(\mathcal{V},\mathcal{V}') = \{\mathbf{0}\}.$$

Proposition 3: *Let $\mathcal{V},\mathcal{V}',\mathcal{V}''$ be linear spaces and let $\mathbf{h} \in \mathrm{Conf}(\mathcal{V},\mathcal{V}')$ and $\mathbf{k} \in \mathrm{Conf}(\mathcal{V}',\mathcal{V}'')$ be such that $\mathrm{Dom}\,\mathbf{k} = \mathrm{Cod}\,\mathbf{h}$. Then $\mathbf{k}\circ\mathbf{h} \in \mathrm{Conf}(\mathcal{V},\mathcal{V}'')$. Moreover, if one of $\mathbf{k}$ or $\mathbf{h}$ is small near zero so is $\mathbf{k}\circ\mathbf{h}$.*

Proof: Let norms ν,ν',ν'' on $\mathcal{V},\mathcal{V}',\mathcal{V}''$, respectively, be given. Since $\mathbf{h}$ and $\mathbf{k}$ are confined we can find $\kappa,\kappa' \in \mathbb{P}^\times$ and $\mathcal{N} \in \mathrm{Nhd}_{\mathbf{0}}(\mathcal{V})$, $\mathcal{N}' \in \mathrm{Nhd}_{\mathbf{0}}(\mathcal{V}')$ such that

$$\nu''((\mathbf{k}\circ\mathbf{h})(\mathbf{u}) \leq \kappa'\nu'(\mathbf{h}(\mathbf{u})) \leq \kappa'\kappa\nu(\mathbf{u}) \tag{62.10}$$

for all $\mathbf{u} \in \mathcal{N}\cap\mathrm{Dom}\,\mathbf{h}$ such that $\mathbf{h}(\mathbf{u}) \in \mathcal{N}'\cap\mathrm{Dom}\,\mathbf{k}$, i.e. for all $\mathbf{u} \in \mathcal{N}\cap\mathbf{h}^<(\mathcal{N}'\cap\mathrm{Dom}\,\mathbf{k})$. Since $\mathbf{h}$ is continuous at $\mathbf{0} \in \mathcal{V}$, we have $\mathbf{h}^<(\mathcal{N}'\cap\mathrm{Dom}\,\mathbf{k}) \in \mathrm{Nhd}_{\mathbf{0}}(\mathcal{V})$ and hence $\mathcal{N}\cap\mathbf{h}^<(\mathcal{N}'\cap\mathrm{Dom}\,\mathbf{k}) \in \mathrm{Nhd}_{\mathbf{0}}(\mathcal{V})$. Thus, (62.7) remains satisfied when we replace $\mathbf{h},\kappa$ and $\mathcal{N}$ by $\mathbf{k}\circ\mathbf{h},\kappa'\kappa$, and $\mathcal{N}\cap\mathbf{h}^<(\mathcal{N}'\cap\mathrm{Dom}\,\mathbf{k})$, respectively, which shows that $\mathbf{k}\circ\mathbf{h} \in \mathrm{Conf}(\mathcal{V},\mathcal{V}'')$.

Assume now, that one of $\mathbf{k}$ and $\mathbf{h}$, say $\mathbf{h}$, is small. Let $\varepsilon \in \mathbb{P}^\times$ be given. Then we can choose $\mathcal{N} \in \mathrm{Nhd}_{\mathbf{0}}(\mathcal{V})$ such that $\nu'(\mathbf{h}(\mathbf{u})) \leq \kappa\nu(\mathbf{u})$ holds for all $\mathbf{u} \in \mathcal{N}\cap\mathrm{Dom}\,\mathbf{h}$ with $\kappa := \frac{\varepsilon}{\kappa'}$. Therefore, (62.10) gives $\nu''((\mathbf{k}\circ\mathbf{h})(\mathbf{u})) \leq \varepsilon\nu(\mathbf{u})$ for all $\mathbf{u} \in \mathcal{N}\cap\mathbf{h}^<(\mathcal{N}'\cap\mathrm{Dom}\,\mathbf{k})$. Since $\varepsilon \in \mathbb{P}^\times$ was arbitrary this proves that

$$\lim_{\mathbf{u}\to\mathbf{0}} \frac{\nu''((\mathbf{k}\circ\mathbf{h})(\mathbf{u}))}{\nu(\mathbf{u})} = 0,$$

i.e. that $\mathbf{k}\circ\mathbf{h}$ is small near zero. ■

Now let $\mathcal{E}$ and $\mathcal{E}'$ be flat spaces with translation spaces $\mathcal{V}$ and $\mathcal{V}'$, respectively.

Definition 3: *Let $x \in \mathcal{E}$ be given. We say that a mapping $\boldsymbol{\sigma}$ from a neighborhood of $x \in \mathcal{E}$ to a neighborhood of $\mathbf{0} \in \mathcal{V}'$ is* **small near** *x if the mapping $\mathbf{v} \mapsto \boldsymbol{\sigma}(x + \mathbf{v})$ from $(\mathrm{Dom}\, \boldsymbol{\sigma}) - x$ to $\mathrm{Cod}\, \boldsymbol{\sigma}$ is small near $\mathbf{0}$. The set of all such small mappings will be denoted by $\mathrm{Small}_x(\mathcal{E}, \mathcal{V}')$.*

We say that a mapping φ from a neighborhood of $x \in \mathcal{E}$ to a neighborhood of $\varphi(x) \in \mathcal{E}'$ is **confined near** *x if the mapping $\mathbf{v} \mapsto (\varphi(x+\mathbf{v}) - \varphi(x))$ from $(\mathrm{Dom}\, \varphi) - x$ to $(\mathrm{Cod}\, \varphi) - \varphi(x)$ is confined near zero.*

The following characterization is immediate.

Proposition 4: *The mapping φ is confined near $x \in \mathcal{E}$ if and only if for every norm ν on $\mathcal{V}$ and every norm ν' on $\mathcal{V}'$ there is $\mathcal{N} \in \mathrm{Nhd}_x(\mathcal{E})$ and $\kappa \in \mathbb{P}^\times$ such that*

$$\nu'(\varphi(y) - \varphi(x)) \leq \kappa \nu(y - x) \quad \text{for all} \quad y \in \mathcal{N}. \tag{62.11}$$

We now state a few facts that are direct consequences of the definitions, the results (I)–(V) stated above, and Prop.3:

(VI) Value-wise sums and differences of mappings that are small [confined] near x are again small [confined] near x. Here, "sum" can mean either the sum of two vectors or sum of a point and a vector, while "difference" can mean either the difference of two vectors or the difference of two points.

(VII) Every $\boldsymbol{\sigma} \in \mathrm{Small}_x(\mathcal{E}, \mathcal{V}')$ is confined near x.

(VIII) If a mapping is confined near x it is continuous at x.

(IX) A flat mapping $\alpha : \mathcal{E} \to \mathcal{E}'$ is confined near every $x \in \mathcal{E}$.

(X) The only flat mapping $\boldsymbol{\beta} : \mathcal{E} \to \mathcal{V}'$ that is small near some $x \in \mathcal{E}$ is the constant $\mathbf{0}_{\mathcal{E} \to \mathcal{V}'}$.

(XI) If φ is confined near $x \in \mathcal{E}$ and if ψ is a mapping with $\mathrm{Dom}\, \psi = \mathrm{Cod}\, \varphi$ that is confined near $\varphi(x)$ then $\psi \circ \varphi$ is confined near x.

(XII) If $\boldsymbol{\sigma} \in \mathrm{Small}_x(\mathcal{E}, \mathcal{V}')$ and $\mathbf{h} \in \mathrm{Conf}(\mathcal{V}', \mathcal{V}'')$ with $\mathrm{Cod}\, \boldsymbol{\sigma} = \mathrm{Dom}\, \mathbf{h}$, then $\mathbf{h} \circ \boldsymbol{\sigma} \in \mathrm{Small}_x(\mathcal{E}, \mathcal{V}'')$.

(XIII) If φ is confined near $x \in \mathcal{E}$ and if $\boldsymbol{\sigma}$ is a mapping with $\mathrm{Dom}\, \boldsymbol{\sigma} = \mathrm{Cod}\, \varphi$ that is small near $\varphi(x)$ then $\boldsymbol{\sigma} \circ \varphi \in \mathrm{Small}_x(\mathcal{E}, \mathcal{V}'')$, where $\mathcal{V}''$ is the linear space for which $\mathrm{Cod}\, \boldsymbol{\sigma} \in \mathrm{Nhd}_{\mathbf{0}}(\mathcal{V}'')$.

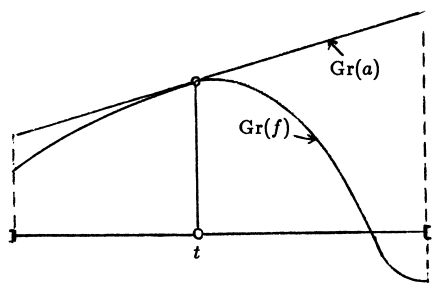
(XIV) An adjustment of a mapping that is small [confined] near x is again small [confined] near x, provided only that the concept small [confined] near x remains meaningful after the adjustment.

Notes 62

(1) In the conventional treatments, the norms ν and ν' in Defs.1 and 2 are assumed to be prescribed and fixed. The notation $\mathbf{n} = o(\nu)$, and the phrase "$\mathbf{n}$ is small oh of ν", are often used to express the assertion that $\mathbf{n} \in \text{Small}(\mathcal{V}, \mathcal{V}')$. The notation $\mathbf{h} = O(\nu)$, and the phrase "$\mathbf{h}$ is big oh of ν", are often used to express the assertion that $\mathbf{h} \in \text{Conf}(\mathcal{V}, \mathcal{V}')$. I am introducing the terms "small" and "confined" here for the first time because I believe that the conventional terminology is intolerably awkward and involves a misuse of the = sign.

63 Gradients, Chain Rule

Let I be an open interval in $\mathbb{R}$. One learns in elementary calculus that if a function $f : I \to \mathbb{R}$ is differentiable at a point $t \in I$, then the graph of f has a tangent at $(t, f(t))$. This tangent is the graph of a flat function $a \in \text{Flf}(\mathbb{R})$. Using poetic license, we refer to this function itself as the *tangent to f at $t \in I$*. In this sense, the tangent a is given by $a(r) := f(t) + (\partial_t f)(r - t)$ for all $r \in \mathbb{R}$.



If we put $\sigma := f - a|_I$, then $\sigma(r) = f(r) - f(t) - (\partial_t f)(r - t)$ for all $r \in I$. We have $\lim_{s\to 0} \frac{\sigma(t+s)}{s} = 0$, from which it follows that $\sigma \in \text{Small}_t(\mathbb{R}, \mathbb{R})$. One can use the existence of a tangent to *define* differentiability at t. Such a definition generalizes directly to mappings involving flat spaces.

Let $\mathcal{E}, \mathcal{E}'$ be flat spaces with translation spaces $\mathcal{V}, \mathcal{V}'$, respectively. We consider a mapping $\varphi : \mathcal{D} \to \mathcal{D}'$ from an open subset $\mathcal{D}$ of $\mathcal{E}$ into an open subset $\mathcal{D}'$ of $\mathcal{E}'$.

Proposition 1: *Given $x \in \mathcal{D}$, there can be at most one flat mapping $\alpha : \mathcal{E} \to \mathcal{E}'$ such that the value-wise difference $\varphi - \alpha|_{\mathcal{D}} : \mathcal{D} \to \mathcal{V}'$ is small near x.*

Proof: If the flat mappings α_1, α_2 both have this property, then the value-wise difference $(\alpha_2 - \alpha_1)|_{\mathcal{D}} = (\varphi - \alpha_1|_{\mathcal{D}}) - (\varphi - \alpha_2|_{\mathcal{D}})$ is small near

$x \in \mathcal{E}$. Since $\alpha_2 - \alpha_1$ is flat, it follows from (X) and (XIV) of Sect.62 that $\alpha_2 - \alpha_1$ is the zero mapping and hence that $\alpha_1 = \alpha_2$. ■

Definition 1: *The mapping $\varphi : \mathcal{D} \to \mathcal{D}'$ is said to be* **differentiable** *at $x \in \mathcal{D}$ if there is a flat mapping $\alpha : \mathcal{E} \to \mathcal{E}'$ such that*

$$\varphi - \alpha|_{\mathcal{D}} \in \mathrm{Small}_x(\mathcal{E}, \mathcal{V}'). \tag{63.1}$$

This (unique) flat mapping α is then called the **tangent to** φ **at** x. *The* **gradient of** φ *at x is defined to be the gradient of α and is denoted by*

$$\nabla_x \varphi := \nabla \alpha. \tag{63.2}$$

We say that φ is **differentiable** *if it is differentiable at all $x \in \mathcal{D}$. If this is the case, the mapping*

$$\nabla \varphi : \mathcal{D} \to \mathrm{Lin}(\mathcal{V}, \mathcal{V}') \tag{63.3}$$

defined by

$$(\nabla \varphi)(x) := \nabla_x \varphi \quad \text{for all} \quad x \in \mathcal{D} \tag{63.4}$$

is called the **gradient** *of φ. We say that φ is* **of class $\mathbf{C}^1$** *if it is differentiable and if its gradient $\nabla\varphi$ is continuous. We say that φ is* **twice differentiable** *if it is differentiable and if its gradient $\nabla\varphi$ is also differentiable. The gradient of $\nabla\varphi$ is then called the* **second gradient** *of φ and is denoted by*

$$\nabla^{(2)} \varphi := \nabla(\nabla \varphi) : \mathcal{D} \to \mathrm{Lin}(\mathcal{V}, \mathrm{Lin}(\mathcal{V}, \mathcal{V}')) \cong \mathrm{Lin}_2(\mathcal{V}^2, \mathcal{V}'). \tag{63.5}$$

We say that φ is **of class $\mathbf{C}^2$** *if it is twice differentiable and if $\nabla^{(2)}\varphi$ is continuous.*

If the subsets $\mathcal{D}$ and $\mathcal{D}'$ are arbitrary, not necessarily open, and if $x \in \mathrm{Int}\,\mathcal{D}$, we say that $\varphi : \mathcal{D} \to \mathcal{D}'$ is *differentiable at x* if $\varphi|^{\mathcal{E}}_{\mathrm{Int}\,\mathcal{D}}$ is differentiable at x and we write $\nabla_x \varphi$ for $\nabla_x(\varphi|^{\mathcal{E}}_{\mathrm{Int}\,\mathcal{D}})$.

The differentiability properties of a mapping φ remain unchanged if the codomain of φ is changed to any open subset of $\mathcal{E}'$ that includes $\mathrm{Rng}\,\varphi$. The gradient of φ remains unaltered. If $\mathrm{Rng}\,\varphi$ is included in some flat $\mathcal{F}'$ in $\mathcal{E}'$, one may change the codomain to a subset that is open in $\mathcal{F}'$. in that case, the gradient of φ at a point $x \in \mathcal{D}$ must be replaced by the adjustment $\nabla_x \varphi|^{\mathcal{U}'}$ of $\nabla_x \varphi$, where $\mathcal{U}'$ is the direction space of $\mathcal{F}'$.

The differentiability and the gradient of a mapping at a point depend only on the values of the mapping near that point. To be more precise, let φ_1 and φ_2 be two mappings whose domains are neighborhoods of a given $x \in \mathcal{E}$ and whose codomains are open subsets of $\mathcal{E}'$. Assume that φ_1 and φ_2 agree on some neighborhood of x, i.e. that $\varphi_1|_{\mathcal{N}}^{\mathcal{E}'} = \varphi_2|_{\mathcal{N}}^{\mathcal{E}'}$ for some $\mathcal{N} \in \mathrm{Nhd}_x(\mathcal{E})$. Then φ_1 is differentiable at x if and only if φ_2 is differentiable at x. If this is the case, we have $\nabla_x\varphi_1 = \nabla_x\varphi_2$.

Every flat mapping $\alpha : \mathcal{E} \to \mathcal{E}'$ is differentiable. The tangent of α at every point $x \in \mathcal{E}$ is α itself. The gradient of α as defined in this section is the constant $(\nabla\alpha)_{\mathcal{E}\to\mathrm{Lin}(\mathcal{V},\mathcal{V}')}$ whose value is the gradient $\nabla\alpha$ of α in the sense of Sect.33. The gradient of a linear mapping is the constant whose value is this linear mapping itself.

In the case when $\mathcal{E} := \mathcal{V} := \mathbb{R}$ and when $\mathcal{D} := I$ is an open interval, differentiability at $t \in I$ of a process $p : I \to \mathcal{D}'$ in the sense of Def.1 above reduces to differentiability of p at t in the sense of Def.1 of Sect.1. The gradient $\nabla_t p \in \mathrm{Lin}(\mathbb{R}, \mathcal{V}')$ becomes associated with the derivative $\partial_t p \in \mathcal{V}'$ by the natural isomorphism from $\mathcal{V}'$ to $\mathrm{Lin}(\mathbb{R}, \mathcal{V}')$, so that $\nabla_t p = (\partial_t p)\otimes$ and $\partial_t p = (\nabla_t p)1$ (see Sect.25). If I is an interval that is not open and if t is an endpoint of it, then the derivative of $p : I \to \mathcal{D}'$ at t, if it exists, cannot be associated with a gradient.

If φ is differentiable at x then it is confined and hence continuous at x. This follows from the fact that its tangent α, being flat, is confined near x and that the difference $\varphi - \alpha|_{\mathcal{D}}$, being small near x, is confined near x. The converse is not true. For example, it is easily seen that the absolute-value function $(t \mapsto |t|) : \mathbb{R} \to \mathbb{R}$ is confined near $0 \in \mathbb{R}$ but not differentiable at 0.

The following criterion is immediate from the definition:

Characterization of Gradients: *The mapping $\varphi : \mathcal{D} \to \mathcal{D}'$ is differentiable at $x \in \mathcal{D}$ if and only if there is an $\mathbf{L} \in \mathrm{Lin}(\mathcal{V}, \mathcal{V}')$ such that $\mathbf{n} : (\mathcal{D} - x) \to \mathcal{V}$, defined by*

$$\mathbf{n}(\mathbf{v}) := (\varphi(x + \mathbf{v}) - \varphi(x)) - \mathbf{L}\mathbf{v} \quad \text{for all} \quad \mathbf{v} \in \mathcal{D} - x, \tag{63.6}$$

is small near $\mathbf{0}$ in $\mathcal{V}$. If this is the case, then $\nabla_x\varphi = \mathbf{L}$.

Let $\mathcal{D}, \mathcal{D}_1, \mathcal{D}_2$ be open subsets of flat spaces $\mathcal{E}, \mathcal{E}_1, \mathcal{E}_2$ with translation spaces $\mathcal{V}, \mathcal{V}_1, \mathcal{V}_2$, respectively. The following result follows immediately from the definitions if we use the term-wise evaluations (04.13) and (14.12).

Proposition 2: *The mapping $(\varphi_1, \varphi_2) : \mathcal{D} \to \mathcal{D}_1 \times \mathcal{D}_2$ is differentiable at $x \in \mathcal{D}$ if and only if both φ_1 and φ_2 are differentiable at x. If this is the case, then*

$$\nabla_x(\varphi_1, \varphi_2) = (\nabla_x\varphi_1, \nabla_x\varphi_2) \in \mathrm{Lin}(\mathcal{V}, \mathcal{V}_1) \times \mathrm{Lin}(\mathcal{V}, \mathcal{V}_2) \cong \mathrm{Lin}(\mathcal{V}, \mathcal{V}_1 \times \mathcal{V}_2) \tag{63.7}$$

General Chain Rule: *Let $\mathcal{D}, \mathcal{D}', \mathcal{D}''$ be open subsets of flat spaces $\mathcal{E}, \mathcal{E}', \mathcal{E}''$ with translation spaces $\mathcal{V}, \mathcal{V}', \mathcal{V}''$, respectively. If $\varphi : \mathcal{D} \to \mathcal{D}'$ is* *differentiable at $x \in \mathcal{D}$ and if $\psi : \mathcal{D}' \to \mathcal{D}''$ is differentiable at $\varphi(x)$, then the composite $\psi \circ \varphi : \mathcal{D} \to \mathcal{D}''$ is differentiable at x. The tangent to the composite $\psi \circ \varphi$ at x is the composite of the tangent to φ at x and the tangent to ψ at $\varphi(x)$ and we have*

$$\nabla_x(\psi \circ \varphi) = (\nabla_{\varphi(x)}\psi)(\nabla_x \varphi). \tag{63.8}$$

If φ and ψ are both differentiable, so is $\psi \circ \varphi$, and we have

$$\nabla(\psi \circ \varphi) = (\nabla\psi \circ \varphi)(\nabla\varphi), \tag{63.9}$$

where the product on the right is understood as value-wise composition.

Proof: Let α be the tangent to φ at x and β the tangent to ψ at $\varphi(x)$. Then

$$\boldsymbol{\sigma} := \varphi - \alpha|_{\mathcal{D}} \in \mathrm{Small}_x(\mathcal{E}, \mathcal{V}'),$$

$$\boldsymbol{\tau} := \psi - \beta|_{\mathcal{D}'} \in \mathrm{Small}_{\varphi(x)}(\mathcal{E}', \mathcal{V}'').$$

We have

$$\begin{aligned} \psi \circ \varphi &= (\beta + \boldsymbol{\tau}) \circ (\alpha + \boldsymbol{\sigma}) \\ &= \beta \circ (\alpha + \boldsymbol{\sigma}) + \boldsymbol{\tau} \circ (\alpha + \boldsymbol{\sigma}) \\ &= \beta \circ \alpha + (\nabla\beta) \circ \boldsymbol{\sigma} + \boldsymbol{\tau} \circ (\alpha + \boldsymbol{\sigma}) \end{aligned}$$

where domain restriction symbols have been omitted to avoid clutter. It follows from (VI), (IX), (XII), and (XIII) of Sect.62 that $(\nabla\beta) \circ \boldsymbol{\sigma} + \boldsymbol{\tau} \circ (\alpha + \boldsymbol{\sigma}) \in \mathrm{Small}_x(\mathcal{E}, \mathcal{V}'')$, which means that

$$\psi \circ \varphi - \beta \circ \alpha \in \mathrm{Small}_x(\mathcal{E}, \mathcal{V}'').$$

If follows that $\psi \circ \varphi$ is differentiable at x with tangent $\beta \circ \alpha$. The assertion (63.8) follows from the Chain Rule for Flat Mappings of Sect.33. ■

Let $\varphi : \mathcal{D} \to \mathcal{E}'$ and $\psi : \mathcal{D} \to \mathcal{E}'$ both be differentiable at $x \in \mathcal{D}$. Then the value-wise difference $\varphi - \psi : \mathcal{D} \to \mathcal{V}'$ is differentiable at x and

$$\nabla_x(\varphi - \psi) = \nabla_x\varphi - \nabla_x\psi. \tag{63.10}$$

This follows from Prop.2, the fact that the point-difference $(x', y') \mapsto (x' - y')$ is a flat mapping from $\mathcal{E}' \times \mathcal{E}'$ into $\mathcal{V}'$, and the General Chain Rule.

When the General Chain Rule is applied to the composite of a vector-valued mapping with a linear mapping it yields

Proposition 3: *If* $\mathbf{h} : \mathcal{D} \to \mathcal{V}'$ *is differentiable at* $x \in \mathcal{D}$ *and if* $\mathbf{L} \in \mathrm{Lin}(\mathcal{V}', \mathcal{W})$*, where* $\mathcal{W}$ *is some linear space, then* $\mathbf{Lh} : \mathcal{D} \to \mathcal{W}$ *is differentiable at* $x \in \mathcal{D}$ *and*

$$\nabla_x(\mathbf{Lh}) = \mathbf{L}(\nabla_x \mathbf{h}). \tag{63.11}$$

Using the fact that vector-addition, transpositions of linear mappings, and the trace operations are all linear operations, we obtain the following special cases of Prop.3:

(I) Let $\mathbf{h} : \mathcal{D} \to \mathcal{V}'$ and $\mathbf{k} : \mathcal{D} \to \mathcal{V}'$ both be differentiable at $x \in \mathcal{D}$. Then the value-wise sum $\mathbf{h} + \mathbf{k} : \mathcal{D} \to \mathcal{V}'$ is differentiable at x and

$$\nabla_x(\mathbf{h} + \mathbf{k}) = \nabla_x \mathbf{h} + \nabla_x \mathbf{k}. \tag{63.12}$$

(II) Let $\mathcal{W}$ and $\mathcal{Z}$ be linear spaces and let $\mathbf{F} : \mathcal{D} \to \mathrm{Lin}(\mathcal{W}, \mathcal{Z})$ be differentiable at x. If $\mathbf{F}^\top : \mathcal{D} \to \mathrm{Lin}(\mathcal{Z}^*, \mathcal{W}^*)$ is defined by value-wise transposition, then $\mathbf{F}^\top$ is differentiable at $x \in \mathcal{D}$ and

$$\nabla_x(\mathbf{F}^\top)\mathbf{v} = ((\nabla_x \mathbf{F})\mathbf{v})^\top \quad \text{for all} \quad \mathbf{v} \in \mathcal{V}. \tag{63.13}$$

In particular, if I is an open interval and if $\mathbf{F} : I \to \mathrm{Lin}(\mathcal{W}, \mathcal{Z})$ is differentiable, so is $\mathbf{F}^\top : I \to \mathrm{Lin}(\mathcal{Z}^*, \mathcal{W}^*)$, and

$$(\mathbf{F}^\top)^\bullet = (\mathbf{F}^\bullet)^\top. \tag{63.14}$$

(III) Let $\mathcal{W}$ be a linear space and let $\mathbf{F} : \mathcal{D} \to \mathrm{Lin}(\mathcal{W})$ be differentiable at x. If $\mathrm{tr}\mathbf{F} : \mathcal{D} \to \mathbb{R}$ is the value-wise trace of $\mathbf{F}$, then $\mathrm{tr}\mathbf{F}$ is differentiable at x and

$$(\nabla_x(\mathrm{tr}\mathbf{F}))\mathbf{v} = \mathrm{tr}((\nabla_x \mathbf{F})\mathbf{v}) \quad \text{for all} \quad \mathbf{v} \in \mathcal{V}. \tag{63.15}$$

In particular, if I is an open interval and if $\mathbf{F} : I \to \mathrm{Lin}(\mathcal{W})$ is differentiable, so is $\mathrm{tr}\mathbf{F} : I \to \mathbb{R}$, and

$$(\mathrm{tr}\mathbf{F})^\bullet = \mathrm{tr}(\mathbf{F}^\bullet). \tag{63.16}$$

We note three special cases of the General Chain Rule: Let I be an open interval and let $\mathcal{D}$ and $\mathcal{D}'$ be open subsets of $\mathcal{E}$ and $\mathcal{E}'$, respectively. If $p : I \to \mathcal{D}$ and $\varphi : \mathcal{D} \to \mathcal{D}'$ are differentiable, so is $\varphi \circ p : I \to \mathcal{D}'$, and

$$(\varphi \circ p)^\bullet = ((\nabla\varphi) \circ p)p^\bullet. \tag{63.17}$$

If $\varphi : \mathcal{D} \to \mathcal{D}'$ and $f : \mathcal{D}' \to \mathbb{R}$ are differentiable, so is $f \circ \varphi : \mathcal{D} \to \mathbb{R}$, and

$$\nabla(f \circ \varphi) = (\nabla\varphi)^{\top}((\nabla f) \circ \varphi) \tag{63.18}$$

(see (21.3)). If $f : \mathcal{D} \to I$ and $p : I \to \mathcal{D}'$ are differentiable, so is $p \circ f$, and

$$\nabla(p \circ f) = (p^{\bullet} \circ f) \otimes \nabla f. \tag{63.19}$$

Notes 63

(1) Other common terms for the concept of "gradient" of Def.1 are "differential", "Fréchet differential", "derivative", and "Fréchet derivative". Some authors make an artificial distinction between "gradient" and "differential". We cannot use "derivative" because, for processes, "gradient" and "derivative" are distinct though related concepts.

(2) The conventional definitions of gradient depend, at first view, on the prescription of a norm. Many texts never even mention the fact that the gradient is a norm-invariant concept. In some contexts, as when one deals with genuine Euclidean spaces, this norm-invariance is perhaps not very important. However, when one deals with mathematical models for space-time in the theory of relativity, the norm-invariance is crucial because it shows that the concepts of differential calculus have a "Lorentz-invariant" meaning.

(3) I am introducing the notation $\nabla_x\varphi$ for the gradient of φ at x because the more conventional notation $\nabla\varphi(x)$ suggests, incorrectly, that $\nabla\varphi(x)$ is necessarily the value at x of a gradient-mapping $\nabla\varphi$. In fact, one cannot define the gradient-mapping $\nabla\varphi$ without first having a notation for the gradient at a point (see 63.4).

(4) Other notations for $\nabla_x\varphi$ in the literature are $\mathrm{d}\varphi(x)$, $\mathrm{D}\varphi(x)$, and $\varphi'(x)$.

(5) I conjecture that the "Chain" of "Chain Rule" comes from an old terminology that used "chaining" (in the sense of "concatenation") for "composition". The term "Composition Rule" would need less explanation, but I retained "Chain Rule" because it is more traditional and almost as good.

64 Constricted Mappings

In this section, $\mathcal{D}$ and $\mathcal{D}'$ denote arbitrary subsets of flat spaces $\mathcal{E}$ and $\mathcal{E}'$ with translation spaces $\mathcal{V}$ and $\mathcal{V}'$, respectively.

Definition 1: *We say that the mapping $\varphi : \mathcal{D} \to \mathcal{D}'$ is* **constricted** *if for every norm ν on $\mathcal{V}$ and every norm ν' on $\mathcal{V}'$ there is $\kappa \in \mathbb{P}^{\times}$ such that*

$$\nu'(\varphi(y) - \varphi(x)) \leq \kappa\nu(y - x) \tag{64.1}$$

holds for all $x, y \in \mathcal{D}$. *The infimum of the set of all* $\kappa \in \mathbb{P}^\times$ *for which* (64.1) *holds for all* $x, y \in \mathcal{D}$ *is called the* **striction** *of* φ *relative to* ν *and* ν'*; it is denoted by* $\mathrm{str}(\varphi; \nu, \nu')$.

It is clear that

$$\mathrm{str}(\varphi; \nu, \nu') = \sup\left\{ \frac{\nu'(\varphi(y) - \varphi(x))}{\nu(y-x)} \;\middle|\; x, y \in \mathcal{D}, x \neq y \right\}. \tag{64.2}$$

Proposition 1: *For every mapping* $\varphi : \mathcal{D} \to \mathcal{D}'$*, the following are equivalent:*

(i) φ *is constricted.*

(ii) *There exist norms* ν *and* ν' *on* $\mathcal{V}$ *and* $\mathcal{V}'$*, respectively, and* $\kappa \in \mathbb{P}^\times$ *such that* (64.1) *holds for all* $x, y \in \mathcal{D}$.

(iii) *For every bounded subset* $\mathcal{C}$ *of* $\mathcal{V}$ *and every* $\mathcal{N}' \in \mathrm{Nhd}_0(\mathcal{V}')$ *there is* $\rho \in \mathbb{P}^\times$ *such that*

$$x - y \in s\mathcal{C} \quad \Longrightarrow \quad \varphi(x) - \varphi(y) \in s\rho\mathcal{N}' \tag{64.3}$$

for all $x, y \in \mathcal{D}$ *and* $s \in \mathbb{P}^\times$.

Proof: (i) ⇒ (ii): This implication is trivial.

(ii) ⇒ (iii): Assume that (ii) holds, and let a bounded subset $\mathcal{C}$ of $\mathcal{V}$ and $\mathcal{N}' \in \mathrm{Nhd}_{\mathbf{0}}(\mathcal{V}')$ be given. By Cor.1 of the Cell-Inclusion Theorem of Sect.52, we can choose $b \in \mathbb{P}^\times$ such that

$$\nu(\mathbf{u}) \leq b \quad \text{for all} \quad \mathbf{u} \in \mathcal{C}. \tag{64.4}$$

By Prop.3 of Sect.53, we can choose $\sigma \in \mathbb{P}^\times$ such that $\sigma\overline{\mathrm{Ce}}(\nu') \subset \mathcal{N}'$. Now let $x, y \in \mathcal{D}$ and $s \in \mathbb{P}^\times$ be given and assume that $x - y \in s\mathcal{C}$. Then $\frac{1}{s}(x-y) \in \mathcal{C}$ and hence, by (64.4), we have $\nu(\frac{1}{s}(x-y)) \leq b$, which gives $\nu(x-y) \leq sb$. Using (64.1) we obtain $\nu'(\varphi(y) - \varphi(x)) \leq s\kappa b$. If we put $\rho := \frac{\kappa b}{\sigma}$, this means that

$$\varphi(y) - \varphi(x) \in s\rho\sigma\overline{\mathrm{Ce}}(\nu') \subset s\rho\mathcal{N}',$$

i.e. that (64.3) holds.

(iii) ⇒ (i): Assume that (iii) holds and let a norm ν on $\mathcal{V}$ and a norm ν' on $\mathcal{V}'$ be given. We apply (iii) with the choices $\mathcal{C} := \mathrm{Bdy}\,\mathrm{Ce}(\nu)$ and $\mathcal{N}' := \mathrm{Ce}(\nu')$. Let $x, y \in \mathcal{D}$ with $x \neq y$ be given. If we put $s := \nu(x-y)$

then $x-y \in s\text{Bdy}\,\text{Ce}(\nu)$ and hence, by (64.3), $\varphi(x)-\varphi(y) \in s\rho\text{Ce}(\nu')$, which implies

$$\nu'(\varphi(x)-\varphi(y)) < s\rho = \rho\nu(x-y).$$

Thus, if we put $\kappa := \rho$, then (64.1) holds for all $x, y \in \mathcal{D}$. ∎

If $\mathcal{D}$ and $\mathcal{D}'$ are open sets and if $\varphi : \mathcal{D} \to \mathcal{D}'$ is constricted, it is confined near every $x \in \mathcal{D}$, as is evident from Prop.4 of Sect.62. The converse is not true. For example, one can show that the function $f :]-1,1[\to \mathbb{R}$ defined by

$$f(t) := \left\{ \begin{array}{ccc} t\sin(\frac{1}{t}) & \text{if} & t \in]0,1[\\ 0 & \text{if} & t \in]-1,0] \end{array} \right\} \tag{64.5}$$

is not constricted, but is confined near every $t \in]-1,1[$.

Every flat mapping $\alpha : \mathcal{E} \to \mathcal{E}'$ is constricted and

$$\text{str}(\alpha; \nu, \nu') = ||\nabla\alpha||_{\nu,\nu'},$$

where $||\ ||_{\nu,\nu'}$ is the operator norm on $\text{Lin}(\mathcal{V},\mathcal{V}')$ corresponding to ν and ν' (see Sect.52).

Constrictedness and strictions remain unaffected by a change of codomain. If the domain of a constricted mapping is restricted, then it remains constricted and the striction of the restriction is less than or equal to the striction of the original mapping. (Pardon the puns.)

Proposition 2: *If $\varphi : \mathcal{D} \to \mathcal{D}'$ is constricted then it is uniformly continuous.*

Proof: We use condition (iii) of Prop.1. Let $\mathcal{N}' \in \text{Nhd}_0(\mathcal{V}')$ be given. We choose a bounded neighborhood $\mathcal{C}$ of $\mathbf{0}$ in $\mathcal{V}$ and determine $\rho \in \mathbb{P}^\times$ according to (iii). Putting $\mathcal{N} := \frac{1}{\rho}\mathcal{C} \in \text{Nhd}_{\mathbf{0}}(\mathcal{V})$ and $s := \frac{1}{\rho}$, we see that (64.3) gives

$$x-y \in \mathcal{N} \quad \Longrightarrow \quad \varphi(x)-\varphi(y) \in \mathcal{N}' \quad \text{for all} \quad x,y \in \mathcal{D}. ∎$$

Pitfall: The converse of Prop.2 is not valid. A counterexample is the square-root function $\sqrt{\ } : \mathbb{P} \to \mathbb{P}$, which is uniformly continuous but not constricted. Another counterexample is the function defined by (64.5). ∎

The following result is the most useful criterion for showing that a given mapping is constricted.

Striction Estimate for Differentiable Mappings: *Assume that $\mathcal{D}$ is an open convex subset of $\mathcal{E}$, that $\varphi : \mathcal{D} \to \mathcal{D}'$ is differentiable and that the gradient $\nabla\varphi : \mathcal{D} \to \text{Lin}(\mathcal{V},\mathcal{V}')$ has a bounded range. Then φ is constricted and*

$$\text{str}(\varphi; \nu, \nu') \le \sup\{||\nabla_z\varphi||_{\nu,\nu'} \mid z \in \mathcal{D}\} \tag{64.6}$$

for all norms ν, ν' on $\mathcal{V}, \mathcal{V}'$, respectively.

Proof: Let $x, y \in \mathcal{D}$. Since $\mathcal{D}$ is convex, we have $tx + (1-t)y \in \mathcal{D}$ for all $t \in [0,1]$ and hence we can define a process

$$p : [0,1] \to \mathcal{D}' \quad \text{by} \quad p(t) := \varphi(tx + (1-t)y).$$

By the Chain Rule, p is differentiable at t when $t \in]0,1[$ and we have

$$\partial_t p = (\nabla_z \varphi)(x - y) \quad \text{with} \quad z := tx + (1-t)y \in \mathcal{D}.$$

Applying the Difference-Quotient Theorem of Sect.61 to p, we obtain

$$\varphi(x) - \varphi(y) = p(1) - p(0) \in \operatorname{Clo}\operatorname{Cxh}\{(\nabla_z \varphi)(x-y) \mid z \in \mathcal{D}\}. \tag{64.7}$$

If ν, ν' are norms on $\mathcal{V}, \mathcal{V}'$ then, by (52.7),

$$\nu'((\nabla_z \varphi)(x-y)) \le ||\nabla_z \varphi||_{\nu,\nu'} \nu(x-y) \tag{64.8}$$

for all $z \in \mathcal{D}$. To say that $\nabla\varphi$ has a bounded range is equivalent, by Cor.1 to the Cell-Inclusion Theorem of Sect.52, to

$$\kappa := \sup\{||\nabla_z \varphi||_{\nu,\nu'} \mid z \in \mathcal{D}\} < \infty. \tag{64.9}$$

It follows from (64.8) that $\nu'((\nabla_z \varphi)(x-y) \le \kappa\nu(x-y)$ for all $z \in \mathcal{D}$, which can be expressed in the form

$$(\nabla_z \varphi)(x-y) \in \kappa\nu(x-y)\overline{\operatorname{Ce}}(\nu') \quad \text{for all} \quad z \in \mathcal{D}.$$

Since the set on the right is closed and convex we get

$$\operatorname{Clo}\operatorname{Cxh}\{(\nabla_z \varphi)(y-x) \mid z \in \mathcal{D}\} \subset \kappa\nu(x-y)\overline{\operatorname{Ce}}(\nu')$$

and hence, by (64.7), $\varphi(x) - \varphi(y) \in \kappa\nu(x-y)\overline{\operatorname{Ce}}(\nu')$. This may be expressed in the form

$$\nu'(\varphi(x) - \varphi(y)) \le \kappa\nu(x-y).$$

Since $x, y \in \mathcal{D}$ were arbitrary it follows that φ is constricted. The definition (64.9) shows that (64.6) holds. ■

Remark: It is not hard to prove that the inequality in (64.6) is actually an equality. ■

Proposition 3: *If $\mathcal{D}$ is a non-empty open convex set and if $\varphi : \mathcal{D} \to \mathcal{D}'$ is differentiable with gradient zero, then φ is constant.*

Proof: Choose norms ν, ν' on $\mathcal{V}, \mathcal{V}'$, respectively. The assumption $\nabla\varphi = \mathbf{0}$ gives $||\nabla_z \varphi||_{\nu,\nu'} = 0$ for all $z \in \mathcal{D}$. Hence, by (64.6), we have $\operatorname{str}(\varphi, \nu, \nu') =$

0. Using (64.2), we conclude that $\nu'(\varphi(y)-\varphi(x)) = 0$ and hence $\varphi(x) = \varphi(y)$ for all $x, y \in \mathcal{D}$. ∎

Remark: In Prop.3, the condition that $\mathcal{D}$ be convex can be replaced by the weaker one that $\mathcal{D}$ be "connected". This means, intuitively, that every two points in $\mathcal{D}$ can be connected by a continuous curve entirely within $\mathcal{D}$. ∎

Proposition 4: *Assume that $\mathcal{D}$ and $\mathcal{D}'$ are open subsets and that $\varphi : \mathcal{D} \to \mathcal{D}'$ is of class C^1. Let $\mathfrak{k}$ be a compact subset of $\mathcal{D}$. For every norm ν on $\mathcal{V}$, every norm ν' on $\mathcal{V}'$, and every $\varepsilon \in \mathbb{P}^\times$ there exists $\delta \in \mathbb{P}^\times$ such that $\mathfrak{k} + \delta\overline{\mathrm{Ce}}(\nu) \subset \mathcal{D}$ and such that, for every $x \in \mathfrak{k}$, the function $\mathbf{n}_x : \delta\mathrm{Ce}(\nu) \to \mathcal{V}'$ defined by*

$$\mathbf{n}_x(\mathbf{v}) := \varphi(x + \mathbf{v}) - \varphi(x) - (\nabla_x\varphi)\mathbf{v} \tag{64.10}$$

is constricted with

$$\mathrm{str}(\mathbf{n}_x; \nu, \nu') \leq \varepsilon. \tag{64.11}$$

Proof: Let a norm ν on $\mathcal{V}$ be given. By Prop.6 of Sect.58, we can obtain $\delta_1 \in \mathbb{P}^\times$ such that $\mathfrak{k} + \delta_1\overline{\mathrm{Ce}}(\nu) \subset \mathcal{D}$. For each $x \in \mathfrak{k}$, we define $\mathbf{m}_x : \delta_1\mathrm{Ce}(\nu) \to \mathcal{V}'$ by

$$\mathbf{m}_x(\mathbf{v}) := \varphi(x + \mathbf{v}) - \varphi(x) - (\nabla_x\varphi)\mathbf{v}. \tag{64.12}$$

Differentiation gives

$$\nabla_{\mathbf{v}}\mathbf{m}_x = \nabla\varphi(x + \mathbf{v}) - \nabla\varphi(x) \tag{64.13}$$

for all $x \in \mathfrak{k}$ and all $\mathbf{v} \in \delta_1\mathrm{Ce}(\nu)$. Since $\mathfrak{k} + \delta_1\overline{\mathrm{Ce}}(\nu)$ is compact by Prop.6 of Sect.58 and since $\nabla\varphi$ is continuous, it follows by the Uniform Continuity Theorem of Sect.58 that $\nabla\varphi|_{\mathfrak{k}+\delta_1\overline{\mathrm{Ce}}(\nu)}$ is uniformly continuous. Now let a norm ν' on $\mathcal{V}'$ and $\varepsilon \in \mathbb{P}^\times$ be given. By Prop.4 of Sect.56, we can then determine $\delta_2 \in \mathbb{P}^\times$ such that

$$\nu(y - x) < \delta_2 \implies ||\nabla\varphi(y) - \nabla\varphi(x)||_{\nu,\nu'} < \varepsilon$$

for all $x, y \in \mathfrak{k} + \delta_1\overline{\mathrm{Ce}}(\nu)$. In view of (64.13), it follows that

$$\mathbf{v} \in \delta_2\mathrm{Ce}(\nu) \implies ||\nabla_{\mathbf{v}}\mathbf{m}_x||_{\nu,\nu'} < \varepsilon$$

for all $x \in \mathfrak{k}$ and all $\mathbf{v} \in \delta_1\mathrm{Ce}(\nu)$. If we put $\delta := \min\{\delta_1, \delta_2\}$ and if we define $\mathbf{n}_x := \mathbf{m}_x|_{\delta\mathrm{Ce}(\nu)}$ for every $x \in \mathfrak{k}$, we see that $\{||\nabla_{\mathbf{v}}\mathbf{n}_x||_{\nu,\nu'} \mid \mathbf{v} \in \delta\mathrm{Ce}(\nu)\}$ is bounded by ε for all $x \in \mathfrak{k}$. By (64.12) and the Striction Estimate for Differentiable Mappings, the desired result follows. ∎

Definition 2: *Let $\varphi : \mathcal{D} \to \mathcal{D}$ be a constricted mapping from a set $\mathcal{D}$ into itself. Then*

$$\mathrm{str}(\varphi) := \inf\{\mathrm{str}(\varphi;\nu,\nu) \mid \nu \text{ a norm on } \mathcal{V}\} \tag{64.14}$$

is called the **absolute striction** *of φ. If $\mathrm{str}(\varphi) < 1$ we say that φ is a* **contraction**.

Contraction Fixed Point Theorem: *Every contraction has at most one fixed point and, if its domain is closed and not empty, it has exactly one fixed point.*

Proof: Let $\varphi : \mathcal{D} \to \mathcal{D}$ be a contraction. We choose a norm ν on $\mathcal{V}$ such that $\kappa := \mathrm{str}(\varphi;\nu,\nu) < 1$ and hence, by (64.1),

$$\nu(\varphi(x) - \varphi(y)) \le \kappa\nu(x - y) \quad \text{for all} \quad x, y \in \mathcal{D}. \tag{64.15}$$

If x and y are fixed points of φ, so that $\varphi(x) = x$, $\varphi(y) = y$, then (64.15) gives $\nu(x-y)(1-\kappa) \le 0$. Since $1-\kappa > 0$ this is possible only when $\nu(x - y) = 0$ and hence $x = y$. Therefore, φ can have at most one fixed point.

We now assume $\mathcal{D} \ne \emptyset$, choose $s_0 \in \mathcal{D}$ arbitrarily, and define

$$s_n := \varphi^{\circ n}(s_0) \quad \text{for all} \quad n \in \mathbb{N}^\times$$

(see Sect.03). It follows from (64.15) that

$$\nu(s_{m+1} - s_m) = \nu(\varphi(s_m) - \varphi(s_{m-1})) \le \kappa\nu(s_m - s_{m-1})$$

for all $m \in \mathbb{N}^\times$. Using induction, one concludes that

$$\nu(s_{m+1} - s_m) \le \kappa^m \nu(s_1 - s_0) \quad \text{for all} \quad m \in \mathbb{N}.$$

Now, if $n \in \mathbb{N}$ and $r \in \mathbb{N}$, then

$$s_{n+r} - s_n = \sum_{k\in r^[} (s_{n+k+1} - s_{n+k})$$

and hence

$$\nu(s_{n+r} - s_n) \le \sum_{k\in r^[} \kappa^{n+k}\nu(s_1 - s_0) \le \frac{\kappa^n}{1-\kappa}\nu(s_1 - s_0).$$

Since $\kappa < 1$, we have $\lim_{n\to\infty} \kappa^n = 0$, and it follows that for every $\varepsilon \in \mathbb{P}^\times$ there is an $m \in \mathbb{N}$ such that $\nu(s_{n+r} - s_n) < \varepsilon$ whenever $n \in m+\mathbb{N}$, $r \in \mathbb{N}$. By the Basic Convergence Criterion of Sect.55 it follows that $s := (s_n \mid n \in \mathbb{N})$ converges. Since $\varphi(s_n) = s_{n+1}$ for all $n \in \mathbb{N}$, we have $\varphi \circ s = (s_{n+1} \mid n \in \mathbb{N})$, which converges to the same limit as s. We put

$$x := \lim s = \lim (\varphi \circ s).$$

Now assume that $\mathcal{D}$ is closed. By Prop.6 of Sect.55 it follows that $x \in \mathcal{D}$. Since φ is continuous, we can apply Prop.2 of Sect.56 to conclude that $\lim (\varphi \circ s) = \varphi(\lim s)$, i.e. that $\varphi(x) = x$. ■

Notes 64

(1) The traditional terms for "constricted" and "striction" are "Lipschitzian" and "Lipschitz number (or constant)", respectively. I am introducing the terms "constricted" and "striction" here because they are much more descriptive.

(2) It turns out that the absolute striction of a lineon coincides with what is often called its "spectral radius".

(3) The Contraction Fixed Point Theorem is often called the "Contraction Mapping Theorem" or the "Banach Fixed Point Theorem".

65 Partial Gradients, Directional Derivatives

Let $\mathcal{E}_1, \mathcal{E}_2$ be flat spaces with translation spaces $\mathcal{V}_1, \mathcal{V}_2$. As we have seen in Sect.33, the set-product $\mathcal{E} := \mathcal{E}_1 \times \mathcal{E}_2$ is then a flat space with translation space $\mathcal{V} := \mathcal{V}_1 \times \mathcal{V}_2$. We consider a mapping $\varphi : \mathcal{D} \to \mathcal{D}'$ from an open subset $\mathcal{D}$ of $\mathcal{E}$ into an open subset $\mathcal{D}'$ of a flat space $\mathcal{E}'$ with translation space $\mathcal{V}'$. Given any $x_2 \in \mathcal{E}_2$, we define $(\cdot, x_2) : \mathcal{E}_1 \to \mathcal{E}$ according to (04.21), put

$$\mathcal{D}_{(\cdot,x_2)} := (\cdot, x_2)^{<}(\mathcal{D}) = \{z \in \mathcal{E}_1 \mid (z, x_2) \in \mathcal{D}\}, \qquad (65.1)$$

which is an open subset of $\mathcal{E}_1$ because $(\cdot, x_1)$ is flat and hence continuous, and define $\varphi(\cdot, x_2) : \mathcal{D}_{(\cdot,x_2)} \to \mathcal{D}'$ according to (04.22). If $\varphi(\cdot, x_2)$ is differentiable at x_1 for all $(x_1, x_2) \in \mathcal{D}$ we define the **partial 1-gradient** $\nabla_{(1)}\varphi : \mathcal{D} \to \text{Lin}(\mathcal{V}_1, \mathcal{V}')$ of φ by

$$\nabla_{(1)}\varphi(x_1, x_2) := \nabla_{x_1}\varphi(\cdot, x_2) \quad \text{for all} \quad (x_1, x_2) \in \mathcal{D}. \qquad (65.2)$$

In the special case when $\mathcal{E}_1 := \mathbb{R}$, we have $\mathcal{V}_1 = \mathbb{R}$, and the **partial 1-derivative** $\varphi_{,1} : \mathcal{D} \to \mathcal{V}'$, defined by

$$\varphi_{,1}(t, x_2) := (\partial\varphi(\cdot, x_2))(t) \quad \text{for all} \quad (t, x_2) \in \mathcal{D}, \qquad (65.3)$$

is related to $\nabla_{(1)}\varphi$ by $\nabla_{(1)}\varphi = \varphi_{,1}\otimes$ and $\varphi_{,1} = (\nabla_{(1)}\varphi)1$ (value-wise).

Similar definitions are employed to define $\mathcal{D}_{(x_1,\cdot)}$, the mapping $\varphi(x_1,\cdot) : \mathcal{D}_{(x_1,\cdot)} \to \mathcal{D}'$, the **partial 2-gradient** $\nabla_{(2)}\varphi : \mathcal{D} \to \mathrm{Lin}(\mathcal{V}_2, \mathcal{V}')$ and the **partial 2-derivative** $\varphi_{,2} : \mathcal{D} \to \mathcal{V}'$.

The following result follows immediately from the definitions.

Proposition 1: *If $\varphi : \mathcal{D} \to \mathcal{D}'$ is differentiable at $x := (x_1, x_2) \in \mathcal{D}$, then the gradients $\nabla_{x_1}\varphi(\cdot, x_2)$ and $\nabla_{x_2}\varphi(x_1,\cdot)$ exist and*

$$(\nabla_x\varphi)\mathbf{v} = (\nabla_{x_1}\varphi(\cdot, x_2))\mathbf{v}_1 + (\nabla_{x_2}\varphi(x_1,\cdot))\mathbf{v}_2 \tag{65.4}$$

for all $\mathbf{v} := (\mathbf{v}_1, \mathbf{v}_2) \in \mathcal{V}$.

If φ is differentiable, then the partial gradients $\nabla_{(1)}\varphi$ and $\nabla_{(2)}\varphi$ exist and we have

$$\nabla\varphi = \nabla_{(1)}\varphi \oplus \nabla_{(2)}\varphi \tag{65.5}$$

where the operation $\oplus$ on the right is understood as value-wise application of (14.13).

Pitfall: The converse of Prop.1 is not true: A mapping can have partial gradients without being differentiable. For example, the mapping $\varphi : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$, defined by

$$\varphi(s,t) := \left\{ \begin{array}{ccc} \frac{st}{s^2+t^2} & \text{if} & (s,t) \neq (0,0) \\ 0 & \text{if} & (s,t) = (0,0) \end{array} \right\},$$

has partial derivatives at $(0,0)$ since both $\varphi(\cdot,0)$ and $\varphi(0,\cdot)$ are equal to the constant 0. Since $\varphi(t,t) = \frac{1}{2}$ for $t \in \mathbb{R}^\times$ but $\varphi(0,0) = 0$, it is clear that φ is not even continuous at $(0,0)$, let alone differentiable. ■

The second assertion of Prop.1 shows that if φ is of class C^1, then $\nabla_{(1)}\varphi$ and $\nabla_{(2)}\varphi$ exist and are continuous. The converse of this statement is true, but the proof is highly nontrivial:

Proposition 2: *Let $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}'$ be flat spaces. Let $\mathcal{D}$ be an open subset of $\mathcal{E}_1 \times \mathcal{E}_2$ and $\mathcal{D}'$ an open subset of $\mathcal{E}'$. A mapping $\varphi : \mathcal{D} \to \mathcal{D}'$ is of class C^1 if (and only if) the partial gradients $\nabla_{(1)}\varphi$ and $\nabla_{(2)}\varphi$ exist and are continuous.*

Proof: Let $(x_1, x_2) \in \mathcal{D}$. Since $\mathcal{D} - (x_1, x_2)$ is a neighborhood of $(\mathbf{0},\mathbf{0})$ in $\mathcal{V}_1 \times \mathcal{V}_2$, we may choose $\mathcal{M}_1 \in \mathrm{Nhd}_{\mathbf{0}}(\mathcal{V}_1)$ and $\mathcal{M}_2 \in \mathrm{Nhd}_{\mathbf{0}}(\mathcal{V}_2)$ such that

$$\mathcal{M} := \mathcal{M}_1 \times \mathcal{M}_2 \subset (\mathcal{D} - (x_1, x_2)).$$

We define $\mathbf{m} : \mathcal{M} \to \mathcal{V}'$ by

$$\begin{aligned} \mathbf{m}(\mathbf{v}_1, \mathbf{v}_2) \quad := \quad & \varphi((x_1, x_2) + (\mathbf{v}_1, \mathbf{v}_2)) - \varphi(x_1, x_2) \\ & - (\nabla_{(1)}\varphi(x_1, x_2)\mathbf{v}_1 + \nabla_{(2)}\varphi(x_1, x_2)\mathbf{v}_2). \end{aligned}$$

It suffices to show that $\mathbf{m} \in \mathrm{Small}(\mathcal{V}_1 \times \mathcal{V}_2, \mathcal{V}')$, for if this is the case, then the Characterization of Gradients of Sect.63 tells us that φ is differentiable at (x_1, x_2) and that its gradient at (x_1, x_2) is given by (65.4). In addition, since $(x_1, x_2) \in \mathcal{D}$ is arbitrary, we can then conclude that (65.5) holds and, since both $\nabla_{(1)}\varphi$ and $\nabla_{(2)}\varphi$ are continuous, that $\nabla\varphi$ is continuous.

We note that $\mathbf{m} := \mathbf{h}+\mathbf{k}$, when $\mathbf{h} : \mathcal{M} \to \mathcal{V}'$ and $\mathbf{k} : \mathcal{M} \to \mathcal{V}'$ are defined by

$$\left.\begin{aligned} \mathbf{h}(\mathbf{v}_1, \mathbf{v}_2) &:= \varphi(x_1, x_2 + \mathbf{v}_2) - \varphi(x_1, x_2) - \nabla_{(2)}\varphi(x_1, x_2)\mathbf{v}_2, \\ \mathbf{k}(\mathbf{v}_1, \mathbf{v}_2) &:= \varphi(x_1 + \mathbf{v}_1, x_2 + \mathbf{v}_2) - \varphi(x_1, x_2 + \mathbf{v}_2) - \nabla_{(1)}\varphi(x_1, x_2)\mathbf{v}_1. \end{aligned}\right\} \tag{65.6}$$

The differentiability of $\varphi(x_1, \cdot)$ at x_2 insures that $\mathbf{h}$ belongs to $\mathrm{Small}(\mathcal{V}_1 \times \mathcal{V}_2, \mathcal{V}')$ and it only remains to be shown that $\mathbf{k}$ belongs to $\mathrm{Small}(\mathcal{V}_1 \times \mathcal{V}_2, \mathcal{V}')$.

We choose norms ν_1, ν_2, and ν' on $\mathcal{V}_1, \mathcal{V}_2$ and $\mathcal{V}'$, respectively. Let $\varepsilon \in \mathbb{P}^\times$ be given. Since $\nabla_{(1)}\varphi$ is continuous at (x_1, x_2), there are open convex neighborhoods $\mathcal{N}_1$ and $\mathcal{N}_2$ of $\mathbf{0}$ in $\mathcal{V}_1$ and $\mathcal{V}_2$, respectively, such that $\mathcal{N}_1 \subset \mathcal{M}_1$, $\mathcal{N}_2 \subset \mathcal{M}_2$, and

$$||\nabla_{(1)}\varphi((x_1, x_2) + (\mathbf{v}_1, \mathbf{v}_2)) - \nabla_{(1)}\varphi(x_1, x_2)\,||_{\nu_1,\nu'} < \varepsilon$$

for all $(\mathbf{v}_1, \mathbf{v}_2) \in \mathcal{N}_1 \times \mathcal{N}_2$. Noting (65.6), we see that this is equivalent to

$$||\nabla_{\mathbf{v}_1}\mathbf{k}(\cdot, \mathbf{v}_2)\,||_{\nu_1,\nu'} < \varepsilon \quad \text{for all} \quad \mathbf{v}_1 \in \mathcal{N}_1,\ \ \mathbf{v}_2 \in \mathcal{N}_2.$$

Applying the Striction Estimate of Sect.64 to $\mathbf{k}(\cdot, \mathbf{v}_2)|_{\mathcal{N}_1}$, we infer that $\mathrm{str}(\mathbf{k}(\cdot, \mathbf{v}_2)|_{\mathcal{N}_1}; \nu_1, \nu') \leq \varepsilon$. It is clear from (65.6) that $\mathbf{k}(\mathbf{0}, \mathbf{v}_2) = \mathbf{0}$ for all $\mathbf{v}_2 \in \mathcal{N}_2$. Hence we can conclude, in view of (64.1), that

$$\nu'(\mathbf{k}(\mathbf{v}_1, \mathbf{v}_2)) \leq \varepsilon\nu_1(\mathbf{v}_1) \leq \varepsilon \max\{\nu_1(\mathbf{v}_1), \nu_2(\mathbf{v}_2)\} \text{ for all } \mathbf{v}_1 \in \mathcal{N}_1,\ \ \mathbf{v}_2 \in \mathcal{N}_2.$$

Since $\varepsilon \in \mathbb{P}^\times$ was arbitrary, it follows that

$$\lim_{(\mathbf{v}_1,\mathbf{v}_2)\to(\mathbf{0},\mathbf{0})} \frac{\nu'(\mathbf{k}(\mathbf{v}_1, \mathbf{v}_2))}{\max\{\nu_1(\mathbf{v}_1), \nu_2(\mathbf{v}_2)\}} = 0,$$

which, in view of (62.1) and (51.23), shows that $\mathbf{k} \in \mathrm{Small}(\mathcal{V}_1 \times \mathcal{V}_2, \mathcal{V}')$ as required. ∎

Remark: In examining the proof above one observes that one needs merely the existence of the partial gradients and the continuity of one of them in order to conclude that the mapping is differentiable. ∎

We now generalize Prop.2 to the case when $\mathcal{E} := \bigtimes(\mathcal{E}_i \mid i \in I)$ is the set-product of a finite family $(\mathcal{E}_i \mid i \in I)$ of flat spaces. This product $\mathcal{E}$ is a flat space whose translation space is the set-product $\mathcal{V} := \times(\mathcal{V}_i \mid i \in I)$ of the translation spaces $\mathcal{V}_i$ of the $\mathcal{E}_i$, $i \in I$. Given any $x \in \mathcal{E}$ and $j \in I$, we define the mapping $(x.j) : \mathcal{E}_j \to \mathcal{E}$ according to (04.24).

We consider a mapping $\varphi : \mathcal{D} \to \mathcal{D}'$ from an open subset $\mathcal{D}$ of $\mathcal{E}$ into an open subset $\mathcal{D}'$ of $\mathcal{E}'$. Given any $x \in \mathcal{D}$ and $j \in I$, we put

$$\mathcal{D}_{(x.j)} := (x.j)^{<}(\mathcal{D}) \subset \mathcal{E}_j \tag{65.7}$$

and define $\varphi(x.j) : \mathcal{D}_{(x.j)} \to \mathcal{D}'$ according to (04.25). If $\varphi(x.j)$ is differentiable at x_j for all $x \in \mathcal{D}$, we define the **partial j-gradient** $\nabla_{(j)}\varphi : \mathcal{D} \to \mathrm{Lin}(\mathcal{V}_j, \mathcal{V}')$ of φ by $\nabla_{(j)}\varphi(x) := \nabla_{x_j}\varphi(x.j)$ for all $x \in \mathcal{D}$.

In the special case when $\mathcal{E}_j := \mathbb{R}$ for some $j \in I$, the **partial j-derivative** $\varphi_{,j} : \mathcal{D} \to \mathcal{V}'$, defined by

$$\varphi_{,j}(x) := (\partial\varphi(x.j))(x_j) \quad \text{for all} \quad x \in \mathcal{D}, \tag{65.8}$$

is related to $\nabla_{(j)}\varphi$ by $\nabla_{(j)}\varphi = \varphi_{,j}\otimes$ and $\varphi_{,j} = (\nabla_{(j)}\varphi)1$.

In the case when $I := \{1, 2\}$, these notations and concepts reduce to the ones explained in the beginning (see (04.21)).

The following result generalizes Prop.1.

Proposition 3: *If $\varphi : \mathcal{D} \to \mathcal{D}'$ is differentiable at $x = (x_i \mid i \in I) \in \mathcal{D}$, then the gradients $\nabla_{x_j}\varphi(x.j)$ exist for all $j \in I$ and*

$$(\nabla_x\varphi)\mathbf{v} = \sum_{j\in I}(\nabla_{x_j}\varphi(x.j))\mathbf{v}_j \tag{65.9}$$

for all $\mathbf{v} = (\mathbf{v}_i \mid i \in I) \in \mathcal{V}$.

If φ is differentiable, then the partial gradients $\nabla_{(j)}\varphi$ exist for all $j \in I$ and we have

$$\nabla\varphi = \bigoplus_{j\in I}(\nabla_{(j)}\varphi), \tag{65.10}$$

where the operation $\bigoplus$ on the right is understood as value-wise application of (14.18).

In the case when $\mathcal{E} := \mathbb{R}^I$, we can put $\mathcal{E}_i := \mathbb{R}$ for all $i \in I$ and (65.10) reduces to

$$\nabla\varphi = \mathrm{lnc}_{(\varphi_{,i} \mid i\in I)}, \tag{65.11}$$

where the value at $x \in \mathcal{D}$ of the right side is understood to be the linear-combination mapping of the family $(\varphi_{,i}(x) \mid i \in I)$ in $\mathcal{V}'$. If moreover, $\mathcal{E}'$ is also a space of the form $\mathcal{E}' := \mathbb{R}^K$ with some finite index set K, then

$\mathrm{lnc}_{(\varphi_{,i} \mid i \in I)}$ can be identified with the function which assigns to each $x \in \mathcal{D}$ the matrix

$$(\varphi_{k,i}(x) \mid (k,i) \in K \times I) \in \mathbb{R}^{K \times I} \cong \mathrm{Lin}(\mathbb{R}^I, \mathbb{R}^K).$$

Hence $\nabla\varphi$ can be identified with the matrix

$$\nabla\varphi = (\varphi_{k,i} \mid (k,i) \in K \times I). \tag{65.12}$$

of the partial derivatives $\varphi_{k,i} : \mathcal{D} \to \mathbb{R}$ of the component functions. As we have seen, the mere existence of these partial derivatives is not sufficient to insure the existence of $\nabla\varphi$. Only if $\nabla\varphi$ is known a priori to exist is it possible to use the identification (65.12).

Using the natural isomorphism between

$$\bigtimes (\mathcal{E}_i \mid i \in I) \quad \text{and} \quad \mathcal{E}_j \times (\bigtimes (\mathcal{E}_i \mid i \in I \setminus \{j\}))$$

and Prop.2, one easily proves, by induction, the following generalization.

Partial Gradient Theorem: *Let I be a finite index set and let $\mathcal{E}_i$, $i \in I$ and $\mathcal{E}'$ be flat spaces. Let $\mathcal{D}$ be an open subset of $\mathcal{E} := \bigtimes (\mathcal{E}_i \mid i \in I)$ and $\mathcal{D}'$ an open subset of $\mathcal{E}'$. A mapping $\varphi : \mathcal{D} \to \mathcal{D}'$ is of class C^1 if and only if the partial gradients $\nabla_{(i)}\varphi$ exist and are continuous for all $i \in I$.*

The following result is an immediate corollary:

Proposition 4: *Let I and K be finite index sets and let $\mathcal{D}$ and $\mathcal{D}'$ be open subsets of $\mathbb{R}^I$ and $\mathbb{R}^K$, respectively. A mapping $\varphi : \mathcal{D} \to \mathcal{D}'$ is of class C^1 if and only if the partial derivatives $\varphi_{k,i} : \mathcal{D} \to \mathbb{R}$ exist and are continuous for all $k \in K$ and all $i \in I$.*

Let $\mathcal{E}, \mathcal{E}'$ be flat spaces with translation spaces $\mathcal{V}$ and $\mathcal{V}'$, respectively. Let $\mathcal{D}, \mathcal{D}'$ be open subsets of $\mathcal{E}$ and $\mathcal{E}'$, respectively, and consider a mapping $\varphi : \mathcal{D} \to \mathcal{D}'$. Given any $x \in \mathcal{D}$ and $\mathbf{v} \in \mathcal{V}^{\times}$, the range of the mapping $(s \mapsto (x + s\mathbf{v})) : \mathbb{R} \to \mathcal{E}$ is a line through x whose direction space is $\mathbb{R}\mathbf{v}$. Let $S_{x,\mathbf{v}} := \{s \in \mathbb{R} \mid x + s\mathbf{v} \in \mathcal{D}\}$ be the pre-image of $\mathcal{D}$ under this mapping and $x + 1_{S_{x,\mathbf{v}}}\mathbf{v} : S_{x,\mathbf{v}} \to \mathcal{D}$ a corresponding adjustment of the mapping. Since $\mathcal{D}$ is open, $S_{x,\mathbf{v}}$ is an open neighborhood of zero in $\mathbb{R}$ and $\varphi \circ (x + 1_{S_{x,\mathbf{v}}}\mathbf{v}) : S_{x,\mathbf{v}} \to \mathcal{D}'$ is a process. The derivative at $0 \in \mathbb{R}$ of this process, if it exists, is called the **directional derivative of** φ **at** x and is denoted by

$$(\mathrm{dd}_{\mathbf{v}}\varphi)(x) := \partial_0(\varphi \circ (x + 1_{S_{x,\mathbf{v}}}\mathbf{v})) = \lim_{s \to 0} \frac{\varphi(x + s\mathbf{v}) - \varphi(x)}{s}. \tag{65.13}$$

If this directional derivative exists for all $x \in \mathcal{D}$, it defines a function

$$\mathrm{dd}_{\mathbf{v}}\varphi : \mathcal{D} \to \mathcal{V}'$$

which is called the **directional v-derivative** of φ. The following result is immediate from the General Chain Rule:

Proposition 5: *If $\varphi : \mathcal{D} \to \mathcal{D}'$ is differentiable at $x \in \mathcal{D}$ then the directional* $\mathbf{v}$*-derivative of φ at x exists for all $\mathbf{v} \in \mathcal{V}$ and is given by*

$$(\mathrm{dd}_{\mathbf{v}}\varphi)(x) = (\nabla_x\varphi)\mathbf{v}. \tag{65.14}$$

Pitfall: The converse of this Proposition is false. In fact, a mapping can have directional derivatives in all directions at all points in its domain without being differentiable. For example, the mapping $\varphi : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\varphi(s,t) := \left\{ \begin{array}{ccc} \frac{s^2 t}{s^4+t^2} & \text{if} & (s,t) \neq (0,0) \\ 0 & \text{if} & (s,t) = (0,0) \end{array} \right\}$$

has the directional derivatives

$$(\mathrm{dd}_{(\alpha,\beta)}\varphi)(0,0) = \left\{ \begin{array}{ccc} \frac{\alpha^2}{\beta} & \text{if} & \beta \neq 0 \\ 0 & \text{if} & \beta = 0 \end{array} \right\}$$

at $(0,0)$. Using Prop.4 one easily shows that $\varphi|_{\mathbb{R}^2\setminus\{(0,0)\}}$ is of class C^1 and hence, by Prop.5, φ has directional derivatives in all directions at all $(s,t) \in \mathbb{R}^2$. Nevertheless, since $\varphi(s,s^2) = \frac{1}{2}$ for all $s \in \mathbb{R}^\times$, φ is not even continuous at $(0,0)$, let alone differentiable. ■

Proposition 6: *Let $\mathfrak{b}$ be a set basis of $\mathcal{V}$. Then $\varphi : \mathcal{D} \to \mathcal{D}'$ is of class C^1 if and only if the directional* $\mathbf{b}$*-derivatives $\mathrm{dd}_{\mathbf{b}}\varphi : \mathcal{D} \to \mathcal{V}'$ exist and are continuous for all $\mathbf{b} \in \mathfrak{b}$.*

Proof: We choose $q \in \mathcal{D}$ and define $\alpha : \mathbb{R}^{\mathfrak{b}} \to \mathcal{E}$ by $\alpha(\lambda) := q + \sum_{\mathbf{b}\in\mathfrak{b}} \lambda_{\mathbf{b}}\mathbf{b}$ for all $\lambda \in \mathbb{R}^{\mathfrak{b}}$. Since $\mathfrak{b}$ is a basis, α is a flat isomorphism. If we define $\overline{\mathcal{D}} := \alpha^{<}(\mathcal{D}) \subset \mathbb{R}^{\mathfrak{b}}$ and $\overline{\varphi} := \varphi \circ \alpha|_{\overline{\mathcal{D}}}^{\mathcal{D}} : \overline{\mathcal{D}} \to \mathcal{D}'$, we see that the directional $\mathbf{b}$-derivatives of φ correspond to the partial derivatives of $\overline{\varphi}$. The assertion follows from the Partial Gradient Theorem, applied to the case when $I := \mathfrak{b}$ and $\mathcal{E}_i := \mathbb{R}$ for all $i \in I$. ■

Combining Prop.5 and Prop.6, we obtain

Proposition 7: *Assume that $\mathcal{S} \in \mathrm{Sub}\,\mathcal{V}$ spans $\mathcal{V}$. The mapping $\varphi : \mathcal{D} \to \mathcal{D}'$ is of class C^1 if and only if the directional derivatives $\mathrm{dd}_{\mathbf{v}}\varphi : \mathcal{D} \to \mathcal{V}'$ exist and are continuous for all $\mathbf{v} \in \mathcal{S}$.*

Notes 65

(1) The notation $\nabla_{(i)}$ for the partial i-gradient is introduced here for the first time. I am not aware of any other notation in the literature.

(2) The notation $\varphi_{,i}$ for the partial i-derivative of φ is the only one that occurs frequently in the literature and is not objectionable. Frequently seen notations such as $\partial\varphi/\partial x_i$ or φ_{x_i} for $\varphi_{,i}$ are poison to me because they contain dangling dummies (see Part D of the Introduction).

(3) Some people use the notation $\nabla_{\mathbf{v}}$ for the directional derivative $\mathrm{dd}_{\mathbf{v}}$. I am introducing $\mathrm{dd}_{\mathbf{v}}$ because (by Def.1 of Sect.63) $\nabla_{\mathbf{v}}$ means something else, namely the gradient at $\mathbf{v}$.

66 The General Product Rule

The following result shows that bilinear mappings (see Sect.24) are of class C^1 and hence continuous. (They are not uniformly continuous except when zero.)

Proposition 1: *Let $\mathcal{V}_1, \mathcal{V}_2$ and $\mathcal{W}$ be linear spaces. Every bilinear mapping $\mathbf{B} : \mathcal{V}_1 \times \mathcal{V}_2 \to \mathcal{W}$ is of class C^1 and its gradient $\nabla\mathbf{B} : \mathcal{V}_1 \times \mathcal{V}_2 \to \mathrm{Lin}(\mathcal{V}_1 \times \mathcal{V}_2, \mathcal{W})$ is the linear mapping given by*

$$\nabla\mathbf{B}(\mathbf{v}_1, \mathbf{v}_2) = (\mathbf{B}^{\sim}\mathbf{v}_2) \oplus (\mathbf{B}\mathbf{v}_1) \tag{66.1}$$

for all $\mathbf{v}_1 \in \mathcal{V}_1, \mathbf{v}_2 \in \mathcal{V}_2$.

Proof: Since $\mathbf{B}(\mathbf{v}_1, \cdot) = \mathbf{B}\mathbf{v}_1 : \mathcal{V}_2 \to \mathcal{W}$ (see (24.2)) is linear for each $\mathbf{v}_1 \in \mathcal{V}$, the partial 2-gradient of $\mathbf{B}$ exists and is given by $\nabla_{(2)}\mathbf{B}(\mathbf{v}_1, \mathbf{v}_2) = \mathbf{B}\mathbf{v}_1$ for all $(\mathbf{v}_1, \mathbf{v}_2) \in \mathcal{V}_1 \times \mathcal{V}_2$. It is evident that $\nabla_{(2)}\mathbf{B} : \mathcal{V}_1 \times \mathcal{V}_2 \to \mathrm{Lin}(\mathcal{V}_2, \mathcal{W})$ is linear and hence continuous. A similar argument shows that $\nabla_{(1)}\mathbf{B}$ is given by $\nabla_{(1)}\mathbf{B}(\mathbf{v}_1, \mathbf{v}_2) = \mathbf{B}^{\sim}\mathbf{v}_2$ for all $(\mathbf{v}_1, \mathbf{v}_2) \in \mathcal{V}_1 \times \mathcal{V}_2$ and hence is also linear and continuous. By the Partial Gradient Theorem of Sect.65, it follows that $\mathbf{B}$ is of class C^1. The formula (66.1) is a consequence of (65.5). ∎

Let $\mathcal{D}$ be an open set in a flat space $\mathcal{E}$ with translation space $\mathcal{V}$. If $\mathbf{h}_1 : \mathcal{D} \to \mathcal{V}_1$, $\mathbf{h}_2 : \mathcal{D} \to \mathcal{V}_2$ and $\mathbf{B} \in \mathrm{Lin}_2(\mathcal{V}_1 \times \mathcal{V}_2, \mathcal{W})$ we write

$$\mathbf{B}(\mathbf{h}_1, \mathbf{h}_2) := \mathbf{B} \circ (\mathbf{h}_1, \mathbf{h}_2) : \mathcal{D} \to \mathcal{W}$$

so that

$$\mathbf{B}(\mathbf{h}_1, \mathbf{h}_2)(x) = \mathbf{B}(\mathbf{h}_1(x), \mathbf{h}_2(x)) \quad \text{for all} \quad x \in \mathcal{D}.$$

The following theorem, which follows directly from Prop.1 and the General Chain Rule of Sect.63, is called "Product Rule" because the term "product" is often used for the bilinear forms to which the theorem is applied.

General Product Rule: *Let $\mathcal{V}_1, \mathcal{V}_2$ and $\mathcal{W}$ be linear spaces and let $\mathbf{B} \in \mathrm{Lin}_2(\mathcal{V}_1 \times \mathcal{V}_2, \mathcal{W})$ be given. Let $\mathcal{D}$ be an open subset of a flat space and let mappings $\mathbf{h}_1 : \mathcal{D} \to \mathcal{V}_1$ and $\mathbf{h}_2 : \mathcal{D} \to \mathcal{V}_2$ be given. If $\mathbf{h}_1$ and $\mathbf{h}_2$ are both differentiable at $x \in \mathcal{D}$ so is $\mathbf{B}(\mathbf{h}_1, \mathbf{h}_2)$, and we have*

$$\nabla_x \mathbf{B}(\mathbf{h}_1, \mathbf{h}_2) = (\mathbf{B}\mathbf{h}_1(x))\nabla_x \mathbf{h}_2 + (\mathbf{B}^{\sim}\mathbf{h}_2(x))\nabla_x \mathbf{h}_1. \tag{66.2}$$

If $\mathbf{h}_1$ and $\mathbf{h}_2$ are both differentiable [of class C^1], so is $\mathbf{B}(\mathbf{h}_1, \mathbf{h}_2)$, and we have

$$\nabla \mathbf{B}(\mathbf{h}_1, \mathbf{h}_2) = (\mathbf{B}\mathbf{h}_1)\nabla \mathbf{h}_2 + (\mathbf{B}^{\sim}\mathbf{h}_2)\nabla \mathbf{h}_1, \tag{66.3}$$

where the products on the right are understood as value-wise compositions.

We now apply the General Product Rule to special bilinear mappings, namely to the ordinary product in $\mathbb{R}$ and to the four examples given in Sect.24. In the following list of results $\mathcal{D}$ denotes an open subset of a flat space having $\mathcal{V}$ as its translation space; $\mathcal{W}, \mathcal{W}'$ and $\mathcal{W}''$ denote linear spaces.

(I) If $f : \mathcal{D} \to \mathbb{R}$ and $g : \mathcal{D} \to \mathbb{R}$ are differentiable [of class C^1], so is the value-wise product $fg : \mathcal{D} \to \mathbb{R}$ and we have

$$\nabla(fg) = f\nabla g + g\nabla f. \tag{66.4}$$

(II) If $f : \mathcal{D} \to \mathbb{R}$ and $\mathbf{h} : \mathcal{D} \to \mathcal{W}$ are differentiable [of class C^1], so is the value-wise scalar multiple $f\mathbf{h} : \mathcal{D} \to \mathcal{W}$, and we have

$$\nabla(f\mathbf{h}) = \mathbf{h} \otimes \nabla f + f\nabla \mathbf{h}. \tag{66.5}$$

(III) If $\mathbf{h} : \mathcal{D} \to \mathcal{W}$ and $\boldsymbol{\eta} : \mathcal{D} \to \mathcal{W}^*$ are differentiable [of class C^1], so is the function $\boldsymbol{\eta}\mathbf{h} : \mathcal{D} \to \mathbb{R}$ defined by $(\boldsymbol{\eta}\mathbf{h})(x) := \boldsymbol{\eta}(x)\mathbf{h}(x)$ for all $x \in \mathcal{D}$, and we have

$$\nabla(\boldsymbol{\eta}\mathbf{h}) = (\nabla \mathbf{h})^{\top}\boldsymbol{\eta} + (\nabla \boldsymbol{\eta})^{\top}\mathbf{h}. \tag{66.6}$$

(IV) If $\mathbf{F} : \mathcal{D} \to \mathrm{Lin}(\mathcal{W}, \mathcal{W}')$ and $\mathbf{h} : \mathcal{D} \to \mathcal{W}$ are differentiable at $x \in \mathcal{D}$, so is $\mathbf{F}\mathbf{h}$ (defined by $(\mathbf{F}\mathbf{h})(y) := \mathbf{F}(y)\mathbf{h}(y)$ for all $y \in \mathcal{D}$) and we have

$$\nabla_x(\mathbf{F}\mathbf{h})\mathbf{v} = ((\nabla_x \mathbf{F})\mathbf{v})\mathbf{h}(x) + (\mathbf{F}(x)\nabla_x \mathbf{h})\mathbf{v} \tag{66.7}$$

for all $\mathbf{v} \in \mathcal{V}$. If $\mathbf{F}$ and $\mathbf{h}$ are differentiable [of class C^1], so is $\mathbf{F}\mathbf{h}$ and (66.7) holds for all $x \in \mathcal{D}, \mathbf{v} \in \mathcal{V}$.

(V) If $\mathbf{F} : \mathcal{D} \to \mathrm{Lin}(\mathcal{W}, \mathcal{W}')$ and $\mathbf{G} : \mathcal{D} \to \mathrm{Lin}(\mathcal{W}', \mathcal{W}'')$ are differentiable at $x \in \mathcal{D}$, so is $\mathbf{GF}$, defined by value-wise composition, and we have

$$\nabla_x(\mathbf{GF})\mathbf{v} = ((\nabla_x\mathbf{G})\mathbf{v})\mathbf{F}(x) + \mathbf{G}(x)((\nabla_x\mathbf{F})\mathbf{v}) \tag{66.8}$$

for all $\mathbf{v} \in \mathcal{V}$. If $\mathbf{F}$ and $\mathbf{G}$ are differentiable [of class C^1], so is $\mathbf{GF}$, and (66.8) holds for all $x \in \mathcal{D}, \mathbf{v} \in \mathcal{V}$.

If $\mathcal{W}$ is an inner-product space, then $\mathcal{W}^*$ can be identified with $\mathcal{W}$, (see Sect.41) and (III) reduces to the following result.

(VI) If $\mathbf{h}, \mathbf{k} : \mathcal{D} \to \mathcal{W}$ are differentiable [of class C^1], so is their value-wise inner product $\mathbf{h} \cdot \mathbf{k}$ and

$$\nabla(\mathbf{h} \cdot \mathbf{k}) = (\nabla\mathbf{h})^\top\mathbf{k} + (\nabla\mathbf{k})^\top\mathbf{h}. \tag{66.9}$$

In the case when $\mathcal{D}$ reduces to an interval in $\mathbb{R}$, (66.4) becomes the Product Rule $(08.32)_2$ of elementary calculus. The following formulas apply to differentiable processes $f, \mathbf{h}, \eta, \mathbf{F}$, and $\mathbf{G}$ with values in $\mathbb{R}, \mathcal{W}, \mathcal{W}^*, \mathrm{Lin}(\mathcal{W}, \mathcal{W}')$, and $\mathrm{Lin}(\mathcal{W}', \mathcal{W}'')$, respectively:

$$(f\mathbf{h})^\bullet = f^\bullet\mathbf{h} + f\mathbf{h}^\bullet, \tag{66.10}$$

$$(\boldsymbol{\eta}\mathbf{h})^\bullet = \boldsymbol{\eta}^\bullet\mathbf{h} + \boldsymbol{\eta}\mathbf{h}^\bullet, \tag{66.11}$$

$$(\mathbf{Fh})^\bullet = \mathbf{F}^\bullet\mathbf{h} + \mathbf{F}\mathbf{h}^\bullet, \tag{66.12}$$

$$(\mathbf{GF})^\bullet = \mathbf{G}^\bullet\mathbf{F} + \mathbf{G}\mathbf{F}^\bullet. \tag{66.13}$$

If $\mathbf{h}$ and $\mathbf{k}$ are differentiable processes with values in an inner product space, then

$$(\mathbf{h} \cdot \mathbf{k})^\bullet = \mathbf{h}^\bullet \cdot \mathbf{k} + \mathbf{h} \cdot \mathbf{k}^\bullet. \tag{66.14}$$

Proposition 2: *Let $\mathcal{W}$ be an inner-product space and let $\mathbf{R} : \mathcal{D} \to \mathrm{Lin}\mathcal{W}$ be given. If $\mathbf{R}$ is differentiable at $x \in \mathcal{D}$ and if $\mathrm{Rng}\,\mathbf{R} \subset \mathrm{Orth}\mathcal{W}$ then*

$$\mathbf{R}(x)^\top((\nabla_x\mathbf{R})\mathbf{v}) \in \mathrm{Skew}\mathcal{W} \quad \text{for all} \quad \mathbf{v} \in \mathcal{V}. \tag{66.15}$$

Conversely, if $\mathbf{R}$ is differentiable, if $\mathcal{D}$ is convex, if (66.15) holds for all $x \in \mathcal{D}$, and if $\mathbf{R}(q) \in \mathrm{Orth}\mathcal{W}$ for some $q \in \mathcal{D}$ then $\mathrm{Rng}\,\mathbf{R} \subset \mathrm{Orth}\mathcal{W}$.

Proof: Assume that $\mathbf{R}$ is differentiable at $x \in \mathcal{D}$. Using (66.8) with the choice $\mathbf{F} := \mathbf{R}$ and $\mathbf{G} := \mathbf{R}^\top$ we find that

$$\nabla_x(\mathbf{R}^\top\mathbf{R})\mathbf{v} = ((\nabla_x\mathbf{R}^\top)\mathbf{v})\mathbf{R}(x) + \mathbf{R}^\top(x)((\nabla_x\mathbf{R})\mathbf{v})$$

holds for all $\mathbf{v} \in \mathcal{V}$. By (63.13) we obtain

$$\nabla_x(\mathbf{R}^\top\mathbf{R})\mathbf{v} = \mathbf{W}^\top + \mathbf{W} \quad \text{with} \quad \mathbf{W} := \mathbf{R}^\top(x)((\nabla_x\mathbf{R})\mathbf{v}) \tag{66.16}$$

for all $\mathbf{v} \in \mathcal{V}$. Now, if $\operatorname{Rng}\mathbf{R} \subset \operatorname{Orth}\mathcal{V}$ then $\mathbf{R}^\top\mathbf{R} = \mathbf{1}_\mathcal{W}$ is constant (see Prop.2 of Sect.43). Hence the left side of (66.16) is zero and $\mathbf{W}$ must be skew for all $\mathbf{v} \in \mathcal{V}$, i.e. (66.15) holds.

Conversely, if (66.15) holds for all $x \in \mathcal{D}$, then, by (66.16) $\nabla_x(\mathbf{R}^\top\mathbf{R}) = \mathbf{0}$ for all $x \in \mathcal{D}$. Since $\mathcal{D}$ is convex, we can apply Prop.3 of Sect.64 to conclude that $\mathbf{R}^\top\mathbf{R}$ must be constant. Hence, if $\mathbf{R}(q) \in \operatorname{Orth}\mathcal{W}$, i.e. if $(\mathbf{R}^\top\mathbf{R})(q) = \mathbf{1}_\mathcal{W}$, then $\mathbf{R}^\top\mathbf{R}$ is the constant $\mathbf{1}_\mathcal{W}$, i.e. $\operatorname{Rng}\mathbf{R} \subset \operatorname{Orth}\mathcal{W}$. ■

Remark: In the second part of Prop.2, the condition that $\mathcal{D}$ be convex can be replaced by the one that $\mathcal{D}$ be "connected" as explained in the Remark after Prop.3 of Sect.64. ■

Corollary: *Let I be an open interval and let $\mathbf{R} : I \to \operatorname{Lin}\mathcal{W}$ be a differentiable process. Then $\operatorname{Rng}\mathbf{R} \subset \operatorname{Orth}\mathcal{W}$ if and only if $\mathbf{R}(t) \in \operatorname{Orth}\mathcal{W}$ for some $t \in I$ and $\operatorname{Rng}(\mathbf{R}^\top\mathbf{R}^\bullet) \subset \operatorname{Skew}\mathcal{W}$.*

The following result shows that quadratic forms (see Sect.27) are of class C^1 and hence continuous.

Proposition 3: *Let $\mathcal{V}$ be a linear space. Every quadratic form $\mathbf{Q} : \mathcal{V} \to \mathbb{R}$ is of class C^1 and its gradient $\nabla\mathbf{Q} : \mathcal{V} \to \mathcal{V}^*$ is the linear mapping*

$$\nabla\mathbf{Q} = 2\,\breve{\mathbf{Q}}\,, \tag{66.17}$$

where $\breve{\mathbf{Q}}$ is identified with the symmetric bilinear form $\breve{\mathbf{Q}} \in \operatorname{Sym}_2(\mathcal{V}^2, \mathbb{R}) \cong \operatorname{Sym}(\mathcal{V}, \mathcal{V}^*)$ associated with $\mathbf{Q}$ (see (27.14)).

Proof: By (27.14), we have

$$\mathbf{Q}(\mathbf{u}) = \breve{\mathbf{Q}}\,(\mathbf{u}, \mathbf{u}) = (\breve{\mathbf{Q}}\,(\mathbf{1}_\mathcal{V}, \mathbf{1}_\mathcal{V}))(\mathbf{u})$$

for all $\mathbf{u} \in \mathcal{V}$. Hence, if we apply the General Product Rule with the choices $\mathbf{B} := \breve{\mathbf{Q}}$ and $\mathbf{h}_1 := \mathbf{h}_2 := \mathbf{1}_\mathcal{V}$, we obtain $\nabla\mathbf{Q} = \breve{\mathbf{Q}} + \breve{\mathbf{Q}}^\sim$. Since $\breve{\mathbf{Q}}$ is symmetric, this reduces to (66.17). ■

Let $\mathcal{V}$ be a linear space. The **lineonic** n**th power** $\operatorname{pow}_n : \operatorname{Lin}\mathcal{V} \to \operatorname{Lin}\mathcal{V}$ on the algebra $\operatorname{Lin}\mathcal{V}$ of lineons (see Sect.18) is defined by

$$\operatorname{pow}_n(\mathbf{L}) := \mathbf{L}^n \quad \text{for all} \quad n \in \mathbb{N}. \tag{66.18}$$

The following result is a generalization of the familiar differentiation rule $(\iota^n)^\bullet = n\iota^{n-1}$.

Proposition 4: *For every* $n \in \mathbb{N}$ *the lineonic nth power* pow_n *on* $\text{Lin}\mathcal{V}$ *defined by* (66.18) *is of class* C^1 *and, for each* $\mathbf{L} \in \text{Lin}\mathcal{V}$*, its gradient* $\nabla_{\mathbf{L}}\text{pow}_n \in \text{Lin}(\text{Lin}\mathcal{V})$ *at* $\mathbf{L} \in \text{Lin}\mathcal{V}$ *is given by*

$$(\nabla_{\mathbf{L}}\text{pow}_n)\mathbf{M} = \sum_{k \in n]} \mathbf{L}^{k-1}\mathbf{M}\mathbf{L}^{n-k} \quad \text{for all} \quad \mathbf{M} \in \text{Lin}\mathcal{V}. \tag{66.19}$$

Proof: For $n = 0$ the assertion is trivial. For $n = 1$, we have $\text{pow}_1 = \mathbf{1}_{\text{Lin}\mathcal{V}}$ and hence $\nabla_{\mathbf{L}}\text{pow}_1 = \mathbf{1}_{\text{Lin}\mathcal{V}}$ for all $\mathbf{L} \in \text{Lin}\mathcal{V}$, which is consistent with (66.19). Also, since ∇pow_1 is constant, pow_1 is of class C^1. Assume, then, that the assertion is valid for a given $n \in \mathbb{R}$. Since $\text{pow}_{n+1} = \text{pow}_1\text{pow}_n$ holds in terms of value-wise composition, we can apply the result (V) above to the case when $\mathbf{F} := \text{pow}_n$, $\mathbf{G} := \text{pow}_1$ in order to conclude that pow_{n+1} is of class C^1 and that

$$(\nabla_{\mathbf{L}}\text{pow}_{n+1})\mathbf{M} = ((\nabla_{\mathbf{L}}\text{pow}_1)\mathbf{M})\text{pow}_n(\mathbf{L}) + \text{pow}_1(\mathbf{L})((\nabla_{\mathbf{L}}\text{pow}_n)\mathbf{M})$$

for all $\mathbf{M} \in \text{Lin}\mathcal{V}$. Using (66.18) and (66.19) we get

$$(\nabla_{\mathbf{L}}\text{pow}_{n+1})\mathbf{M} = \mathbf{M}\mathbf{L}^n + \mathbf{L}(\sum_{k \in n]} \mathbf{L}^{k-1}\mathbf{M}\mathbf{L}^{n-k}),$$

which shows that (66.19) remains valid when n is replaced by $n + 1$. The desired result follows by induction. ■

If $\mathbf{M}$ commutes with $\mathbf{L}$, then (66.19) reduces to

$$(\nabla_{\mathbf{L}}\text{pow}_n)\mathbf{M} = (n\mathbf{L}^{n-1})\mathbf{M}.$$

Using Prop.4 and the form (63.17) of the Chain Rule, we obtain

Proposition 5: *Let* I *be an interval and let* $\mathbf{F} : I \to \text{Lin}\mathcal{V}$ *be a process. If* $\mathbf{F}$ *is differentiable [of class* C^1*], so is its value-wise* n*th power* $\mathbf{F}^n : I \to \text{Lin}\mathcal{V}$ *and we have*

$$(\mathbf{F}^n)^\bullet = \sum_{k \in n]} \mathbf{F}^{k-1}\mathbf{F}^\bullet\mathbf{F}^{n-k} \quad \text{for all} \quad n \in \mathbb{N}. \tag{66.20}$$

67 Divergence, Laplacian

In this section, $\mathcal{D}$ denotes an open subset of a flat space $\mathcal{E}$ with translation space $\mathcal{V}$, and $\mathcal{W}$ denotes a linear space.

If $\mathbf{h} : \mathcal{D} \to \mathcal{V}$ is differentiable at $x \in \mathcal{D}$, we can form the trace (see Sect.26) of the gradient $\nabla_x \mathbf{h} \in \mathrm{Lin}\mathcal{V}$. The result

$$\mathrm{div}_x \mathbf{h} := \mathrm{tr}(\nabla_x \mathbf{h}) \tag{67.1}$$

is called the **divergence** of $\mathbf{h}$ *at* x. If $\mathbf{h}$ is differentiable we can define the **divergence** $\mathrm{div}\ \mathbf{h} : \mathcal{D} \to \mathbb{R}$ of $\mathbf{h}$ by

$$(\mathrm{div}\ \mathbf{h})(x) := \mathrm{div}_x \mathbf{h} \quad \text{for all} \quad x \in \mathcal{D}. \tag{67.2}$$

Using the product rule (66.5) and (26.3) we obtain

Proposition 1: *Let* $\mathbf{h} : \mathcal{D} \to \mathcal{V}$ *and* $f : \mathcal{D} \to \mathbb{R}$ *be differentiable. Then the divergence of* $f\mathbf{h} : \mathcal{D} \to \mathcal{V}$ *is given by*

$$\mathrm{div}\ (f\mathbf{h}) = (\nabla f)\mathbf{h} + f \mathrm{div}\ \mathbf{h}. \tag{67.3}$$

Consider now a mapping $\mathbf{H} : \mathcal{D} \to \mathrm{Lin}(\mathcal{V}^*, \mathcal{W})$ that is differentiable at $x \in \mathcal{D}$. For every $\boldsymbol{\omega} \in \mathcal{W}^*$ we can form the value-wise composite $\boldsymbol{\omega}\mathbf{H} : \mathcal{D} \to \mathrm{Lin}(\mathcal{V}^*, \mathbb{R}) = \mathcal{V}^{**} \cong \mathcal{V}$ (see Sect.22). Since $\boldsymbol{\omega}\mathbf{H}$ is differentiable at x for every $\boldsymbol{\omega} \in \mathcal{W}^*$ (see Prop.3 of Sect.63) we may consider the mapping

$$(\boldsymbol{\omega} \mapsto \mathrm{tr}(\nabla_x(\boldsymbol{\omega}\mathbf{H}))) : \mathcal{W}^* \to \mathbb{R}.$$

It is clear that this mapping is linear and hence an element of $\mathcal{W}^{**} \cong \mathcal{W}$.

Definition 1: *Let* $\mathbf{H} : \mathcal{D} \to \mathrm{Lin}(\mathcal{V}^*, \mathcal{W})$ *be differentiable at* $x \in \mathcal{D}$. *Then the* **divergence** *of* $\mathbf{H}$ *at* x *is defined to be the (unique) element* $\mathrm{div}_x \mathbf{H}$ *of* $\mathcal{W}$ *which satisfies*

$$\boldsymbol{\omega}(\mathrm{div}_x \mathbf{H}) = \mathrm{tr}(\nabla_x(\boldsymbol{\omega}\mathbf{H})) \quad \text{for all} \quad \boldsymbol{\omega} \in \mathcal{W}^*. \tag{67.4}$$

If $\mathbf{H}$ *is differentiable, then its* **divergence** $\mathrm{div}\ \mathbf{H} : \mathcal{D} \to \mathcal{W}$ *is defined by*

$$(\mathrm{div}\ \mathbf{H})(x) := \mathrm{div}_x \mathbf{H} \quad \text{for all} \quad x \in \mathcal{D}. \tag{67.5}$$

In the case when $\mathcal{W} := \mathbb{R}$, we also have $\mathcal{W}^* \cong \mathbb{R}$ and $\mathrm{Lin}(\mathcal{V}^*, \mathcal{W}) = \mathcal{V}^{**} \cong \mathcal{V}$. Thus, using (67.4) with $\boldsymbol{\omega} := 1 \in \mathbb{R}$, we see that the definition just given is consistent with (67.1) and (67.2).

If we replace $\mathbf{h}$ and $\mathbf{L}$ in Prop.3 of Sect.63 by $\mathbf{H}$ and

$$(\mathbf{K} \mapsto \boldsymbol{\omega}\mathbf{K}) \in \mathrm{Lin}(\mathrm{Lin}(\mathcal{V}^*, \mathcal{W}), \mathcal{V}),$$

respectively, we obtain

$$\nabla_x(\boldsymbol{\omega}\mathbf{H}) = \boldsymbol{\omega}\nabla_x \mathbf{H} \quad \text{for all} \quad \boldsymbol{\omega} \in \mathcal{W}^*, \tag{67.6}$$

where the right side must be interpreted as the composite of $\nabla_x \mathbf{H} \in \mathrm{Lin}_2(\mathcal{V}\times\mathcal{V}^*, \mathcal{W})$ with $\boldsymbol{\omega} \in \mathrm{Lin}(\mathcal{W}, \mathbb{R})$. This composite, being an element of $\mathrm{Lin}_2(\mathcal{V}\times\mathcal{V}^*, \mathbb{R})$, must be reinterpreted as an element of $\mathrm{Lin}\mathcal{V}$ via the identifications $\mathrm{Lin}_2(\mathcal{V}\times\mathcal{V}^*, \mathbb{R}) \cong \mathrm{Lin}(\mathcal{V}, \mathrm{Lin}(\mathcal{V}^*, \mathbb{R})) = \mathrm{Lin}(\mathcal{V}, \mathcal{V}^{**}) \cong \mathrm{Lin}\mathcal{V}$ to make (67.6) meaningful. Using (67.6), (67.4), and (67.1) we see that div $\mathbf{H}$ satisfies

$$\boldsymbol{\omega}(\mathrm{div}_x \mathbf{H}) = \mathrm{div}_x(\boldsymbol{\omega}\mathbf{H}) = \mathrm{tr}(\boldsymbol{\omega}\nabla_x \mathbf{H}) \tag{67.7}$$

for all $\boldsymbol{\omega} \in \mathcal{W}^*$.

The following results generalize Prop.1.

Proposition 2: *Let* $\mathbf{H} : \mathcal{D} \to \mathrm{Lin}(\mathcal{V}^*, \mathcal{W})$ *and* $\boldsymbol{\rho} : \mathcal{D} \to \mathcal{W}^*$ *be differentiable. Then the divergence of the value-wise composite* $\boldsymbol{\rho}\mathbf{H} : \mathcal{D} \to \mathrm{Lin}(\mathcal{V}^*, \mathbb{R}) \cong \mathcal{V}$ *is given by*

$$\mathrm{div}\,(\boldsymbol{\rho}\mathbf{H}) = \boldsymbol{\rho}\mathrm{div}\,\mathbf{H} + \mathrm{tr}(\mathbf{H}^\top \nabla\boldsymbol{\rho}), \tag{67.8}$$

where value-wise evaluation and composition are understood.

Proof: Let $x \in \mathcal{D}$ and $\mathbf{v} \in \mathcal{V}$ be given. Using (66.8) with the choices $\mathbf{G} := \boldsymbol{\rho}$ and $\mathbf{F} := \mathbf{H}$ we obtain

$$\nabla_x(\boldsymbol{\rho}\mathbf{H})\mathbf{v} = ((\nabla_x\boldsymbol{\rho})\mathbf{v})\mathbf{H}(x) + \boldsymbol{\rho}(x)((\nabla_x\mathbf{H})\mathbf{v}). \tag{67.9}$$

Since $(\nabla_x\boldsymbol{\rho})\mathbf{v} \in \mathcal{W}^*$, (21.3) gives

$$((\nabla_x\boldsymbol{\rho})\mathbf{v})\mathbf{H}(x) = (\mathbf{H}(x)^\top\nabla_x\boldsymbol{\rho})\mathbf{v}.$$

On the other hand, we have

$$\boldsymbol{\rho}(x)((\nabla_x\mathbf{H})\mathbf{v}) = (\boldsymbol{\rho}(x)(\nabla_x\mathbf{H}))\mathbf{v}$$

if we interpret $\nabla_x\mathbf{H}$ on the right as an element of $\mathrm{Lin}_2(\mathcal{V}\times\mathcal{V}^*, \mathcal{W})$. Hence, since $\mathbf{v} \in \mathcal{V}$ was arbitrary, (67.9) gives

$$\nabla_x(\boldsymbol{\rho}\mathbf{H}) = \boldsymbol{\rho}(x)\nabla_x\mathbf{H} + \mathbf{H}(x)^\top\nabla_x\boldsymbol{\rho}.$$

Taking the trace, using (67.1), and using (67.7) with $\boldsymbol{\omega} := \boldsymbol{\rho}(x)$, we get

$$\mathrm{div}_x(\boldsymbol{\rho}\mathbf{H}) = \boldsymbol{\rho}(x)\mathrm{div}_x\mathbf{H} + \mathrm{tr}(\mathbf{H}(x)^\top\nabla_x\boldsymbol{\rho}).$$

Since $x \in \mathcal{D}$ was arbitrary, the desired result (67.8) follows. ∎

Proposition 3: *Let* $\mathbf{h} : \mathcal{D} \to \mathcal{V}$ *and* $\mathbf{k} : \mathcal{D} \to \mathcal{W}$ *be differentiable. The divergence of* $\mathbf{k}\otimes\mathbf{h} : \mathcal{D} \to \mathrm{Lin}(\mathcal{V}^*, \mathcal{W})$*, defined by taking the value-wise tensor product* (see Sect.25)*, is then given by*

$$\mathrm{div}\,(\mathbf{k}\otimes\mathbf{h}) = (\nabla\mathbf{k})\mathbf{h} + (\mathrm{div}\,\mathbf{h})\mathbf{k}. \tag{67.10}$$

Proof: By (67.7) we have

$$\boldsymbol{\omega}\,\mathrm{div}_x(\mathbf{k}\otimes\mathbf{h}) = \mathrm{div}_x(\boldsymbol{\omega}(\mathbf{k}\otimes\mathbf{h})) = \mathrm{div}_x((\boldsymbol{\omega}\,\mathbf{k})\mathbf{h})$$

for all $x \in \mathcal{D}$ and all $\boldsymbol{\omega} \in \mathcal{W}^*$. Hence, by Prop.1,

$$\boldsymbol{\omega}\,\mathrm{div}\,(\mathbf{k}\otimes\mathbf{h}) = (\nabla(\boldsymbol{\omega}\mathbf{k}))\mathbf{h} + (\boldsymbol{\omega}\mathbf{k})\mathrm{div}\,\mathbf{h} = \boldsymbol{\omega}((\nabla\mathbf{k})\mathbf{h} + \mathbf{k}\,\mathrm{div}\,\mathbf{h})$$

holds for all $\boldsymbol{\omega} \in \mathcal{W}^*$, and (67.10) follows. ∎

Proposition 4: *Let* $\mathbf{H} : \mathcal{D} \to \mathrm{Lin}(\mathcal{V}^*, \mathcal{W})$ *and* $f : \mathcal{D} \to \mathbb{R}$ *be differentiable. The divergence of* $f\mathbf{H} : \mathcal{D} \to \mathrm{Lin}(\mathcal{V}^*, \mathcal{W})$ *is then given by*

$$\mathrm{div}\,(f\mathbf{H}) = f\mathrm{div}\,\mathbf{H} + \mathbf{H}(\nabla f). \tag{67.11}$$

Proof: Let $\boldsymbol{\omega} \in \mathcal{W}^*$ be given. If we apply Prop.2 to the case when $\boldsymbol{\rho} := f\boldsymbol{\omega}$ we obtain

$$\mathrm{div}\,(\boldsymbol{\omega}(f\mathbf{H})) = \boldsymbol{\omega}(f\mathrm{div}\,\mathbf{H}) + \mathrm{tr}(\mathbf{H}^\top\nabla(f\boldsymbol{\omega})). \tag{67.12}$$

Using $\nabla(f\boldsymbol{\omega}) = \boldsymbol{\omega}\otimes\nabla f$ and using (25.9), (26.3), and (22.3), we obtain $\mathrm{tr}(\mathbf{H}^\top\nabla(f\boldsymbol{\omega})) = \mathrm{tr}(\mathbf{H}^\top(\boldsymbol{\omega}\otimes\nabla f)) = \nabla f(\mathbf{H}^\top\boldsymbol{\omega}) = \boldsymbol{\omega}(\mathbf{H}\nabla f)$. Hence (67.12) and (67.7) yield $\boldsymbol{\omega}\,\mathrm{div}\,(f\mathbf{H}) = \boldsymbol{\omega}(f\mathrm{div}\,\mathbf{H} + \mathbf{H}\nabla f)$. Since $\boldsymbol{\omega} \in \mathcal{W}^*$ was arbitrary, (67.11) follows. ∎

From now on we assume that $\mathcal{E}$ has the structure of a Euclidean space, so that $\mathcal{V}$ becomes an inner-product space and we can use the identification $\mathcal{V}^* \cong \mathcal{V}$ (see Sect.41). Thus, if $\mathbf{k} : \mathcal{D} \to \mathcal{W}$ is twice differentiable, we can consider the divergence of $\nabla\mathbf{k} : \mathcal{D} \to \mathrm{Lin}(\mathcal{V}, \mathcal{W}) \cong \mathrm{Lin}(\mathcal{V}^*, \mathcal{W})$.

Definition 2: *Let* $\mathbf{k} : \mathcal{D} \to \mathcal{W}$ *be twice differentiable. Then the* **Laplacian** *of* $\mathbf{k}$ *is defined to be*

$$\Delta\mathbf{k} := \mathrm{div}\,(\nabla\mathbf{k}). \tag{67.13}$$

If $\Delta\mathbf{k} = 0$*, then* $\mathbf{k}$ *is called a* **harmonic** *function.*

Remark: In the case when $\mathcal{E}$ is not a genuine Euclidean space, but one whose translation space has index 1 (see Sect.47), the term *D'Alembertian* or *Wave-Operator* and the symbol $\Box$ are often used instead of Laplacian and Δ. ∎

The following result is a direct consequence of (66.5) and Props.3 and 4.

Proposition 5: *Let* $\mathbf{k} : \mathcal{D} \to \mathcal{W}$ *and* $f : \mathcal{D} \to \mathbb{R}$ *be twice differentiable. The Laplacian of* $f\mathbf{k} : \mathcal{D} \to \mathcal{W}$ *is then given by*

$$\Delta(f\mathbf{k}) = (\Delta f)\mathbf{k} + f\Delta\mathbf{k} + 2(\nabla\mathbf{k})(\nabla f). \tag{67.14}$$

The following result follows directly from the form (63.19) of the Chain Rule and Prop.3.

Proposition 6: *Let I be an open interval and let $f : \mathcal{D} \to I$ and $\mathbf{g} : I \to \mathcal{W}$ both be twice differentiable. The Laplacian of $\mathbf{g} \circ f : \mathcal{D} \to \mathcal{W}$ is then given by*

$$\Delta(\mathbf{g} \circ f) = (\nabla f)^{\cdot 2}(\mathbf{g}^{\bullet\bullet} \circ f) + (\Delta f)(\mathbf{g}^{\bullet} \circ f). \tag{67.15}$$

We now discuss an important application of Prop.6. We consider the case when $f : \mathcal{D} \to I$ is given by

$$f(x) := (x - q)^{\cdot 2} \quad \text{for all} \quad x \in \mathcal{D}, \tag{67.16}$$

where $q \in \mathcal{E}$ is given. We wish to solve the following problem: How can $\mathcal{D}, I$ and $\mathbf{g} : I \to \mathcal{W}$ be chosen such that $\mathbf{g} \circ f$ is harmonic?

It follows directly from (67.16) and Prop.3 of Sect.66, applied to $\mathbf{Q} := \text{sq}$, that $\nabla_x f = 2(x-q)$ for all $x \in \mathcal{D}$ and that hence $(\nabla f)^{\bullet 2} = 4f$. Also, $\nabla(\nabla f)$ is the constant with value $2\mathbf{1}_{\mathcal{V}}$ and hence, by (26.9),

$$\Delta f = \text{div}\,(\nabla f) = 2\text{tr}\mathbf{1}_{\mathcal{V}} = 2\dim \mathcal{V} = 2\dim \mathcal{E}.$$

Substitution of these results into (67.15) gives

$$\Delta(\mathbf{g} \circ f) = 4f(\mathbf{g}^{\bullet\bullet} \circ f) + 2(\dim \mathcal{E})\mathbf{g}^{\bullet} \circ f. \tag{67.17}$$

Hence, $\mathbf{g} \circ f$ is harmonic if and only if

$$(2\iota\mathbf{g}^{\bullet\bullet} + n\mathbf{g}^{\bullet}) \circ f = 0 \quad \text{with} \quad n := \dim \mathcal{E}, \tag{67.18}$$

where ι is the identity mapping of $\mathbb{R}$, suitably adjusted (see Sect.08). Now, if we choose $I := \text{Rng}\, f$, then (67.18) is satisfied if and only if $\mathbf{g}$ satisfies the ordinary differential equation $2\iota\mathbf{g}^{\bullet\bullet} + n\mathbf{g}^{\bullet} = 0$. It follows that $\mathbf{g}$ must be of the form

$$\mathbf{g} = \left\{ \begin{array}{lll} \mathbf{a}\iota^{-(\frac{n}{2}-1)} + \mathbf{b} & \text{if} & n \neq 2 \\ \mathbf{a}\log + \mathbf{b} & \text{if} & n = 2 \end{array} \right\}, \tag{67.19}$$

where $\mathbf{a}, \mathbf{b} \in \mathcal{W}$, provided that I is an interval.

If $\mathcal{E}$ is a genuine Euclidean space, we may take $\mathcal{D} := \mathcal{E} \setminus \{q\}$ and $I := \text{Rng}\, f = \mathbb{P}^{\times}$. We then obtain the harmonic function

$$\mathbf{h} = \left\{ \begin{array}{lll} \mathbf{a}r^{-(n-2)} + \mathbf{b} & \text{if} & n \neq 2 \\ \mathbf{a}(\log \circ r) + \mathbf{b} & \text{if} & n = 2 \end{array} \right\}, \tag{67.20}$$

where $\mathbf{a}, \mathbf{b} \in \mathcal{W}$ and where $r : \mathcal{E} \setminus \{q\} \to \mathbb{P}^{\times}$ is defined by $r(x) := |x - q|$.

If the Euclidean space $\mathcal{E}$ is not genuine and $\mathcal{V}$ is double-signed we have two possibilities. For all $n \in \mathbb{N}^\times$ we may take $\mathcal{D} := \{x \in \mathcal{E} \mid (x - q)^{\bullet 2} > 0\}$ and $I := \mathbb{P}^\times$. If n is even and $n \neq 2$, we may instead take $\mathcal{D} := \{x \in \mathcal{E} \mid (x-q)^{\bullet 2} < 0\}$ and $I := -\mathbb{P}^\times$.

Notes 67

(1) In some of the literature on "Vector Analysis", the notation $\nabla \bullet \mathbf{h}$ instead of $\operatorname{div} \mathbf{h}$ is used for the divergence of a vector field $\mathbf{h}$, and the notation $\nabla^2 f$ instead of Δf for the Laplacian of a function f. These notations come from a formalistic and erroneous understanding of the meaning of the symbol ∇ and should be avoided.

68 Local Inversion, Implicit Mappings

In this section, $\mathcal{D}$ and $\mathcal{D}'$ denote open subsets of flat spaces $\mathcal{E}$ and $\mathcal{E}'$ with translation spaces $\mathcal{V}$ and $\mathcal{V}'$, respectively.

To say that a mapping $\varphi : \mathcal{D} \to \mathcal{D}'$ is differentiable at $x \in \mathcal{D}$ means, roughly, that φ may be approximated, near x, by a flat mapping α, the tangent of φ at x. One might expect that if the tangent α is invertible, then φ itself is, in some sense, "locally invertible near x". To decide whether this expectation is justified, we must first give a precise meaning to "locally invertible near x".

Definition: *Let a mapping $\varphi : \mathcal{D} \to \mathcal{D}'$ be given. We say that ψ is a* **local inverse** *of φ if $\psi = (\varphi|_{\mathcal{N}}^{\mathcal{N}'})^{\leftarrow}$ for suitable open subsets $\mathcal{N} = \operatorname{Cod}\psi = \operatorname{Rng}\psi$ and $\mathcal{N}' = \operatorname{Dom}\psi$ of $\mathcal{D}$ and $\mathcal{D}'$, respectively. We say that ψ is a* **local inverse** *of φ* **near** *$x \in \mathcal{D}$ if $x \in \operatorname{Rng}\psi$. We say that φ is* **locally invertible near** *$x \in \mathcal{D}$ if it has some local inverse near x.*

To say that ψ is a local inverse of φ means that

$$\psi \circ \varphi|_{\mathcal{N}}^{\mathcal{N}'} = 1_{\mathcal{N}} \quad \text{and} \quad \varphi|_{\mathcal{N}}^{\mathcal{N}'} \circ \psi = 1_{\mathcal{N}'} \tag{68.1}$$

for suitable open subsets $\mathcal{N} = \operatorname{Cod}\psi$ and $\mathcal{N}' = \operatorname{Dom}\psi$ of $\mathcal{D}$ and $\mathcal{D}'$, respectively.

If ψ_1 and ψ_2 are local inverses of φ and $\mathcal{M} := (\operatorname{Rng}\psi_1) \cap (\operatorname{Rng}\psi_2)$, then ψ_1 and ψ_2 must agree on $\varphi_{>}(\mathcal{M}) = \psi_1^{<}(\mathcal{M}) = \psi_2^{<}(\mathcal{M})$, i.e.

$$\psi_1|_{\varphi_>(\mathcal{M})}^{\mathcal{M}} = \psi_2|_{\varphi_>(\mathcal{M})}^{\mathcal{M}} \quad \text{if} \quad \mathcal{M} := (\operatorname{Rng}\psi_1) \cap (\operatorname{Rng}\psi_2). \tag{68.2}$$

Pitfall: A mapping $\varphi : \mathcal{D} \to \mathcal{D}'$ may be locally invertible near *every* point in $\mathcal{D}$ without being invertible, even if it is surjective. In fact, if ψ is a local inverse of φ, $\varphi^{<}(\operatorname{Dom}\psi)$ need not be included in $\operatorname{Rng}\psi$.

For example, the function $f :]-1,1[^{\times} \to]0,1[$ defined by $f(s) := |s|$ for all $s \in]-1,1[^{\times}$ is easily seen to be locally invertible, surjective, and even continuous. The identity function $1_{]0,1[}$ of $]0,1[$ is a local inverse of f, and we have

$$f^{<}(\mathrm{Dom}\, 1_{]0,1[}) = f^{<}(\,]0,1[\,) =]-1,1[^{\times} \not\subset]0,1[= \mathrm{Rng}\, 1_{[0,1]}.$$

The function $h :]0,1[\, \to\,]-1,0[$ defined by $h(s) = -s$ for all $s \in]0,1[$ is another local inverse of f.

If $\dim \mathcal{E} \geq 2$, one can even give counter-examples of mappings $\varphi : \mathcal{D} \to \mathcal{D}'$ of the type described above for which $\mathcal{D}$ is convex (see Sect.74). ■

The following two results are recorded for later use.

Proposition 1: *Assume that $\varphi : \mathcal{D} \to \mathcal{D}'$ is differentiable at $x \in \mathcal{D}$ and locally invertible near x. Then, if some local inverse near x is differentiable at $\varphi(x)$, so are all others and all have the same gradient, namely $(\nabla_x \varphi)^{-1}$.*

Proof: We choose a local inverse ψ_1 of φ near x such that ψ_1 is differentiable at $\varphi(x)$. Applying the Chain Rule to (68.1) gives

$$(\nabla_{\varphi(x)}\psi_1)(\nabla_x\varphi) = \mathbf{1}_{\mathcal{V}} \quad \text{and} \quad (\nabla_x\varphi)(\nabla_{\varphi(x)}\psi_1) = \mathbf{1}_{\mathcal{V}'},$$

which shows that $\nabla_x\varphi$ is invertible and $\nabla_{\varphi(x)}\psi_1 = (\nabla_x\varphi)^{-1}$.

Let now a local inverse ψ_2 of φ near x be given. Let $\mathcal{M}$ be defined as in (68.2). Since $\mathcal{M}$ is open and since ψ_1, being differentiable at x, is continuous at x, it follows that $\varphi_{>}(\mathcal{M}) = \psi_1^{<}(\mathcal{M})$ is a neighborhood of $\varphi(x)$ in $\mathcal{E}'$. By (68.2), ψ_2 agrees with ψ_1 on the neighborhood $\varphi_{>}(\mathcal{M})$ of $\varphi(x)$. Hence ψ_2 must also be differentiable at $\varphi(x)$, and $\nabla_{\varphi(x)}\psi_2 = \nabla_{\varphi(x)}\psi_1 = (\nabla_x\varphi)^{-1}$. ■

Proposition 2: *Assume that $\varphi : \mathcal{D} \to \mathcal{D}'$ is continuous and that ψ is a local inverse of φ near x.*

(i) *If $\mathcal{M}'$ is an open neighborhood of $\varphi(x)$ and $\mathcal{M}' \subset \mathrm{Dom}\,\psi$, then $\psi_{>}(\mathcal{M}') = \varphi^{<}(\mathcal{M}') \cap \mathrm{Rng}\,\psi$ is an open neighborhood of x; hence the adjustment $\psi|_{\mathcal{M}'}^{\psi_{>}(\mathcal{M}')}$ is again a local inverse of φ near x.*

(ii) *Let $\mathcal{G}$ be an open subset of $\mathcal{E}$ with $x \in \mathcal{G}$. If ψ is continuous at $\varphi(x)$ then there is an open neighborhood $\mathcal{M}$ of x with $\mathcal{M} \subset \mathrm{Rng}\,\psi \cap \mathcal{G}$ such that $\psi^{<}(\mathcal{M}) = \varphi_{>}(\mathcal{M})$ is open; hence the adjustment $\psi|_{\psi^{<}(\mathcal{M})}^{\mathcal{M}}$ is again a local inverse of φ near x.*

Proof: Part (i) is an immediate consequence of Prop.3 of Sect.56. To prove (ii), we observe first that $\psi^{<}(\mathrm{Rng}\,\psi \cap \mathcal{G})$ must be a neighborhood of $\varphi(x) = \psi^{\leftarrow}(x)$. We can choose an open subset $\mathcal{M}'$ of $\psi^{<}(\mathrm{Rng}\,\psi \cap \mathcal{G})$ with

$\varphi(x) \in \mathcal{M}'$ (see Sect.53). Application of (i) gives the desired result with $\mathcal{M} := \psi_>(\mathcal{M}')$. ■

Remark: It is in fact true that if $\varphi : \mathcal{D} \to \mathcal{D}'$ is continuous, then every local inverse of φ is also continuous, but the proof is highly non-trivial and goes beyond the scope of this presentation. Thus, in Part (ii) of Prop.2, the requirement that ψ be continuous at $\varphi(x)$ is, in fact, redundant. ■

Pitfall: The expectation mentioned in the beginning is *not* justified. A continuous mapping $\varphi : \mathcal{D} \to \mathcal{D}'$ can have an invertible tangent at $x \in \mathcal{D}$ without being locally invertible near x. An example is the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(t) := \left\{ \begin{array}{lll} 2t^2 \sin\left(\frac{1}{t}\right) + t & \text{if} & t \in \mathbb{R}^\times \\ 0 & \text{if} & t = 0 \end{array} \right\}. \tag{68.3}$$

It is differentiable and hence continuous. Since $f^\bullet(0) = 1$, the tangent to f at 0 is invertible. However, f is not monotone, let alone locally invertible, near 0, because one can find numbers s arbitrarily close to 0 such that $f^\bullet(s) < 0$. ■

Let I be an open interval and let $f : I \to \mathbb{R}$ be a function of class C^1. If the tangent to f at a given $t \in I$ is invertible, i.e. if $f^\bullet(t) \neq 0$, one can easily prove, not only that f is locally invertible near t, but also that it has a local inverse near t that is of class C^1. This result generalizes to mappings $\varphi : \mathcal{D} \to \mathcal{D}'$, but the proof is far from easy.

Local Inversion Theorem: *Let $\mathcal{D}$ and $\mathcal{D}'$ be open subsets of flat spaces, let $\varphi : \mathcal{D} \to \mathcal{D}'$ be of class C^1 and let $x \in \mathcal{D}$ be such that the tangent to φ at x is invertible. Then φ is locally invertible near x and every local inverse of φ near x is differentiable at $\varphi(x)$.*

Moreover, there exists a local inverse ψ of φ near x that is of class C^1 and satisfies

$$\nabla_y \psi = (\nabla_{\psi(y)}\varphi)^{-1} \quad \text{for all} \quad y \in \operatorname{Dom}\psi. \tag{68.4}$$

Before proceeding with the proof, we state two important results that are closely related to the Local Inversion Theorem.

Implicit Mapping Theorem: *Let $\mathcal{E}, \mathcal{E}', \mathcal{E}''$ be flat spaces, let $\mathcal{A}$ be an open subset of $\mathcal{E} \times \mathcal{E}'$ and let $\omega : \mathcal{A} \to \mathcal{E}''$ be a mapping of class C^1. Let $(x_o, y_o) \in \mathcal{A}$, $z_o := \omega(x_o, y_o)$, and assume that $\nabla_{y_o}\omega(x_o, \bullet)$ is invertible. Then there exist an open neighborhood $\mathcal{D}$ of x_o and a mapping $\varphi : \mathcal{D} \to \mathcal{E}'$, differentiable at x_o, such that $\varphi(x_o) = y_o$ and $\mathrm{Gr}(\varphi) \subset \mathcal{A}$, and*

$$\omega(x, \varphi(x)) = z_o \quad \text{for all} \quad x \in \mathcal{D}. \tag{68.5}$$

Moreover, $\mathcal{D}$ and φ can be chosen such that φ is of class C^1 and

$$\nabla\varphi(x) = -(\nabla_{(2)}\omega(x, \varphi(x)))^{-1}\nabla_{(1)}\omega(x, \varphi(x)) \quad \text{for all} \quad x \in \mathcal{D}. \tag{68.6}$$

Remark: The mapping φ is determined *implicitly* by an equation in this sense: for any given $x \in \mathcal{D}$, $\varphi(x)$ is a solution of the equation

$$? \, y \in \mathcal{E}', \quad \omega(x, y) = z_o. \tag{68.7}$$

In fact, one can find a neighborhood $\mathcal{M}$ of (x_o, y_o) in $\mathcal{E} \times \mathcal{E}'$ such that $\varphi(x)$ is the *only* solution of

$$? \, y \in \mathcal{M}_{(x,\bullet)}, \quad \omega(x, y) = z_o, \tag{68.8}$$

where $\mathcal{M}_{(x,\bullet)} := \{y \in \mathcal{E}' \mid (x, y) \in \mathcal{M}\}$. ■

Differentiation Theorem for Inversion Mappings: *Let* $\mathcal{V}$ *and* $\mathcal{V}'$ *be linear spaces of equal dimension. Then the set* $\mathrm{Lis}(\mathcal{V}, \mathcal{V}')$ *of all linear isomorphorisms from* $\mathcal{V}$ *onto* $\mathcal{V}'$ *is a (non-empty) open subset of* $\mathrm{Lin}(\mathcal{V}, \mathcal{V}')$*, the inversion mapping* $\mathrm{inv} : \mathrm{Lis}(\mathcal{V}, \mathcal{V}') \to \mathrm{Lin}(\mathcal{V}', \mathcal{V})$ *defined by* $\mathrm{inv}(\mathbf{L}) := \mathbf{L}^{-1}$ *is of class* C^1*, and its gradient is given by*

$$(\nabla_{\mathbf{L}}\mathrm{inv})\mathbf{M} = -\mathbf{L}^{-1}\mathbf{M}\mathbf{L}^{-1} \quad \text{for all} \quad \mathbf{M} \in \mathrm{Lin}(\mathcal{V}, \mathcal{V}'). \tag{68.9}$$

The proof of these three theorems will be given in stages, which will be designated as lemmas. After a basic preliminary lemma, we will prove a weak version of the first theorem and then derive from it weak versions of the other two. The weak version of the last theorem will then be used, like a bootstrap, to prove the final version of the first theorem; the final versions of the other two theorems will follow.

Lemma 1: *Let* $\mathcal{V}$ *be a linear space, let* ν *be a norm on* $\mathcal{V}$*, and put* $\mathcal{B} := \mathrm{Ce}(\nu)$*. Let* $\mathbf{f} : \mathcal{B} \to \mathcal{V}$ *be a mapping with* $\mathbf{f}(\mathbf{0}) = \mathbf{0}$ *such that* $\mathbf{f} - \mathbf{1}_{\mathcal{B}\subset\mathcal{V}}$ *is constricted with*

$$\kappa := \mathrm{str}(\mathbf{f} - 1_{\mathcal{B}\subset\mathcal{V}}; \nu, \nu) < 1. \tag{68.10}$$

Then the adjustment $\mathbf{f}|^{(1-\kappa)\mathcal{B}}$ (see Sect.03) *of* $\mathbf{f}$ *has an inverse and this inverse is confined near* $\mathbf{0}$.

Proof: We will show that for every $\mathbf{w} \in (1-\kappa)\mathcal{B}$, the equation

$$? \, \mathbf{z} \in \mathcal{B}, \quad \mathbf{f}(\mathbf{z}) = \mathbf{w} \tag{68.11}$$

has a unique solution.

To prove uniqueness, suppose that $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{B}$ are solutions of (68.11) for a given $\mathbf{w} \in \mathcal{V}$. We then have

$$(\mathbf{f} - \mathbf{1}_{\mathcal{B}\subset\mathcal{V}})(\mathbf{z}_1) - (\mathbf{f} - 1_{\mathcal{B}\subset\mathcal{V}})(\mathbf{z}_2) = \mathbf{z}_2 - \mathbf{z}_1$$

and hence, by (68.10), $\nu(\mathbf{z}_2 - \mathbf{z}_1) \leq \kappa\nu(\mathbf{z}_2 - \mathbf{z}_1)$, which amounts to $(1-\kappa)\nu(\mathbf{z}_2 - \mathbf{z}_1) \leq 0$. This is compatible with $\kappa < 1$ only if $\nu(\mathbf{z}_1 - \mathbf{z}_2) = 0$ and hence $\mathbf{z}_1 = \mathbf{z}_2$.

To prove existence, we assume that $\mathbf{w} \in (1-\kappa)\mathcal{B}$ is given and we put $\alpha := \frac{\nu(\mathbf{w})}{1-\kappa}$, so that $\alpha \in [0, 1[$. For every $\mathbf{v} \in \alpha\overline{\mathcal{B}}$, we have $\nu(\mathbf{v}) \leq \alpha$ and hence, by (68.10)

$$\nu((\mathbf{f} - 1_{\mathcal{B}\subset\mathcal{V}})(\mathbf{v})) \leq \kappa\nu(\mathbf{v}) \leq \kappa\alpha,$$

which implies that

$$\begin{aligned} \nu(\mathbf{w} - (\mathbf{f}(\mathbf{v}) - \mathbf{v})) &\leq \nu(\mathbf{w}) + \nu((\mathbf{f} - \mathbf{1}_{\mathcal{B}-\mathcal{V}})(\mathbf{v})) \\ &\leq (1-\kappa)\alpha + \kappa\alpha = \alpha. \end{aligned}$$

Therefore, it makes sense to define $\mathbf{h}_\mathbf{w} : \alpha\overline{\mathcal{B}} \to \alpha\overline{\mathcal{B}}$ by

$$\mathbf{h}_\mathbf{w}(\mathbf{v}) := \mathbf{w} - (\mathbf{f}(\mathbf{v}) - \mathbf{v}) \quad \text{for all} \quad \mathbf{v} \in \alpha\overline{\mathcal{B}}. \tag{68.12}$$

Since $\mathbf{h}_\mathbf{w}$ is the difference of a constant with value $\mathbf{w}$ and a restriction of $\mathbf{f} - 1_{\mathcal{B}-\mathcal{V}}$, it is constricted and has an absolute striction no greater than κ. Therefore, it is a contraction, and since its domain $\alpha\overline{\mathcal{B}}$ is closed, the Contraction Fixed Point Theorem states that $\mathbf{h}_\mathbf{w}$ has a fixed point $\mathbf{z} \in \alpha\overline{\mathcal{B}}$. It is evident from (68.12) that a fixed point of $\mathbf{h}_\mathbf{w}$ is a solution of (68.11).

We proved that (68.11) has a unique solution $\mathbf{z} \in \mathcal{B}$ and that this solution satisfies

$$\nu(\mathbf{z}) \leq \alpha := \frac{\nu(\mathbf{w})}{1-\kappa}. \tag{68.13}$$

If we define $\mathbf{g} : (1-\kappa)\mathcal{B} \to \mathbf{f}^<((1-\kappa)\mathcal{B})$ in such a way that $\mathbf{g}(\mathbf{w})$ is this solution of (68.11), then $\mathbf{g}$ is the inverse we are seeking. It follows from (68.13) that $\nu(\mathbf{g}(\mathbf{w})) \leq \frac{1}{1-\kappa}\nu(\mathbf{w})$ for all $\mathbf{w} \in (1-\kappa)\mathcal{B}$. In view of part (ii) of Prop.2 of Sect.62, this proves that $\mathbf{g}$ is confined. ∎

Lemma 2: *The assertion of the Local Inversion Theorem, except possibly the statement starting with "Moreover", is valid.*

Proof: Let $\alpha : \mathcal{E} \to \mathcal{E}'$ be the (invertible) tangent to φ at x. Since $\mathcal{D}$ is open, we may choose a norming cell $\mathcal{B}$ such that $x + \mathcal{B} \subset \mathcal{D}$. We define $\mathbf{f} : \mathcal{B} \to \mathcal{V}$ by

$$\mathbf{f} := (\alpha^{\leftarrow}|_{\mathcal{D}'} \circ \varphi|_{x+\mathcal{B}} \circ (x + \mathbf{1}_\mathcal{V})|_\mathcal{B}^{x+\mathcal{B}}) - x. \tag{68.14}$$

Since φ is of class C^1, so is $\mathbf{f}$. Clearly, we have

$$\mathbf{f}(\mathbf{0}) = \mathbf{0}, \quad \nabla_\mathbf{0}\mathbf{f} = \mathbf{1}_\mathcal{V}. \tag{68.15}$$

Put $\nu := \mathrm{no}_{\mathcal{B}}$, so that $\mathcal{B} = \mathrm{Ce}(\nu)$. Choose $\varepsilon \in]0,1[$ ($\varepsilon := \frac{1}{2}$ would do). If we apply Prop.4 of Sect.64 to the case when φ there is replaced by $\mathbf{f}$ and $\mathfrak{k}$ by $\{\mathbf{0}\}$, we see that there is $\delta \in]0,1]$ such that

$$\kappa := \mathrm{str}(\mathbf{f}|_{\delta\mathcal{B}} - \mathbf{1}_{\delta\mathcal{B}\subset\mathcal{V}}; \nu, \nu) \leq \varepsilon. \tag{68.16}$$

Now, it is clear from (64.2) that a striction relative to ν and ν' remains unchanged if ν and ν' are multiplied by the same strictly positive factor. Hence, (68.16) remains valid if ν there is replaced by $\frac{1}{\delta}\nu$. Since $\mathrm{Ce}(\frac{1}{\delta}\nu) = \delta\mathrm{Ce}(\nu) = \delta\mathcal{B}$ (see Prop.6 of Sect.51), it follows from (68.16) that Lemma 1 can be applied when ν, $\mathcal{B}$, and $\mathbf{f}$ there are replaced by $\frac{1}{\delta}\nu, \delta\mathcal{B}$ and $\mathbf{f}|_{\delta\mathcal{B}}$, respectively. We infer that if we put

$$\mathcal{M}_o := (1-\kappa)\delta\mathcal{B}, \quad \mathcal{N}_o := \mathbf{f}^{<}(\mathcal{M}_o),$$

then $\mathbf{f}|_{\mathcal{N}_o}^{\mathcal{M}_o}$ has an inverse $\mathbf{g} : \mathcal{M}_o \to \mathcal{N}_o$ that is confined near zero. Note that $\mathcal{M}_o$ and $\mathcal{N}_o$ are both open neighborhoods of zero, the latter because it is the pre-image of an open set under a continuous mapping. We evidently have

$$\mathbf{g}|^{\mathcal{V}} - \mathbf{1}_{\mathcal{M}_o\subset\mathcal{V}} = (\mathbf{1}_{\mathcal{N}_o\subset\mathcal{V}} - \mathbf{f}|_{\mathcal{N}_o}) \circ \mathbf{g}.$$

In view of (68.15), $\mathbf{1}_{\mathcal{N}_o\subset\mathcal{V}} - \mathbf{f}|_{\mathcal{N}_o}$ is small near $\mathbf{0} \in \mathcal{V}$. Since $\mathbf{g}$ is confined near $\mathbf{0}$, it follows by Prop.3 of Sect.62 that $\mathbf{g}|^{\mathcal{V}} - \mathbf{1}_{\mathcal{M}_o\subset\mathcal{V}}$ is small near zero. By the Characterization of Gradients of Sect.63, we conclude that $\mathbf{g}$ is differentiable at $\mathbf{0}$ with $\nabla_{\mathbf{0}}\mathbf{g} = \mathbf{1}_{\mathcal{V}}$.

We now define

$$\mathcal{N} := x + \mathcal{N}_o, \quad \mathcal{N}' := \varphi(x) + (\nabla\alpha)_{>}(\mathcal{M}_o).$$

These are open neighborhoods of x and $\varphi(x)$, respectively. A simple calculation, based on (68.14) and $\mathbf{g} = (\mathbf{f}|_{\mathcal{N}_o}^{\mathcal{M}_o})^{\leftarrow}$, shows that

$$\psi := (x + 1_{\mathcal{V}})|_{\mathcal{N}_o}^{\mathcal{N}} \circ \mathbf{g} \circ (\alpha^{\leftarrow} - x)|_{\mathcal{N}'}^{\mathcal{M}_o} \tag{68.17}$$

is the inverse of $\varphi|_{\mathcal{N}}^{\mathcal{N}'}$. Using the Chain Rule and the fact that $\mathbf{g}$ is differentiable at $\mathbf{0}$, we conclude from (68.17) that ψ is differentiable at $\varphi(x)$. ■

Lemma 3: *The assertion of the Implicit Mapping Theorem, except possibly the statement starting with "Moreover", is valid.*

Proof: We define $\tilde{\omega} : \mathcal{A} \to \mathcal{E} \times \mathcal{E}''$ by

$$\tilde{\omega}(x, y) := (x, \omega(x, y)) \quad \text{for all} \quad (x, y) \in \mathcal{A}. \tag{68.18}$$

Of course, $\tilde{\omega}$ is of class C^1. Using Prop.2 of Sect.63 and (65.4), we see that

$$(\nabla_{(x_o,y_o)}\tilde{\omega})(\mathbf{u},\mathbf{v}) = (\mathbf{u}, (\nabla_{x_o}\omega(\cdot, y_o))\mathbf{u} + (\nabla_{y_o}\omega(x_o,\cdot))\mathbf{v})$$

for all $(\mathbf{u},\mathbf{v}) \in \mathcal{V}\times\mathcal{V}'$. Since $\nabla_{y_o}\omega(x_o,\cdot)$ is invertible, so is $\nabla_{(x_o,y_o)}\tilde{\omega}$. Indeed,we have

$$(\nabla_{(x_o,y_o)}\tilde{\omega})^{-1}(\mathbf{u},\mathbf{w}) = (\mathbf{u}, (\nabla_{y_o}\omega(x_o,\cdot))^{-1}(\mathbf{w} - (\nabla_{x_o}\omega(\cdot,y_o))\mathbf{u}))$$

for all $(\mathbf{u},\mathbf{w}) \in \mathcal{V}\times\mathcal{V}''$. Lemma 2 applies and we conclude that $\tilde{\omega}$ has a local inverse ρ with $(x_o,y_o) \in \mathrm{Rng}\,\rho$ that is differentiable at $\tilde{\omega}(x_o,y_o) = (x_o,z_o)$. In view of Prop.2, (i), we may assume that the domain of ρ is of the form $\mathcal{D}\times\mathcal{D}''$, where $\mathcal{D}$ and $\mathcal{D}''$ are open neighborhoods of x_o and z_o, respectively. Let $\psi : \mathcal{D}\times\mathcal{D}'' \to \mathcal{E}$ and $\psi' : \mathcal{D}\times\mathcal{D}'' \to \mathcal{E}'$ be the value-wise terms of ρ, so that

$$\rho(x,z) = (\psi(x,z), \psi'(x,z)) \quad \text{for all} \quad x \in \mathcal{D},\ z \in \mathcal{D}''.$$

Then, by (68.18),

$$\tilde{\omega}(\rho(x,y)) = (\psi(x,z), \omega(\psi(x,z), \psi'(x,z))) = (x,z)$$

and hence

$$\psi(x,z) = x, \quad \omega(x,\psi'(x,z)) = z \quad \text{for all} \quad x \in \mathcal{D},\ z \in \mathcal{D}''. \tag{68.19}$$

Since $\rho(x_o,z_o) = (\psi(x_o,z_o), \psi'(x_o,z_o)) = (x_o,y_o)$, we have $\psi'(x_o,z_o) = y_o$. Therefore, if we define $\varphi : \mathcal{D} \to \mathcal{E}''$ by $\varphi(x) := \psi'(x,z_o)$, we see that $\varphi(x_o) = y_o$ and (68.5) are satisfied. The differentiability of φ at x_o follows from the differentiablity of ρ at (x_o,z_o).

If we define $\mathcal{M} := \mathrm{Rng}\,\rho$, it is immediate that $\varphi(x)$ is the only solution of (68.8). ∎

Lemma 4: *Let $\mathcal{V}$ and $\mathcal{V}'$ be linear spaces of equal dimension. Then* $\mathrm{Lis}(\mathcal{V},\mathcal{V}')$ *is an open subset of* $\mathrm{Lin}(\mathcal{V},\mathcal{V}')$ *and the inversion mapping* $\mathrm{inv} : \mathrm{Lis}(\mathcal{V},\mathcal{V}') \to \mathrm{Lin}(\mathcal{V}',\mathcal{V})$ *defined by* $\mathrm{inv}(\mathbf{L}) := \mathbf{L}^{-1}$ *is differentiable.*

Proof: We apply Lemma 3 with $\mathcal{E}$ replaced by $\mathrm{Lin}(\mathcal{V},\mathcal{V}')$, $\mathcal{E}'$ by $\mathrm{Lin}(\mathcal{V}',\mathcal{V})$, and $\mathcal{E}''$ by $\mathrm{Lin}\mathcal{V}$. For ω of Lemma 3 we take the mapping $(\mathbf{L},\mathbf{M}) \mapsto \mathbf{ML}$ from $\mathrm{Lin}(\mathcal{V},\mathcal{V}') \times \mathrm{Lin}(\mathcal{V}',\mathcal{V})$ to $\mathrm{Lin}\mathcal{V}$. Being bilinear, this mapping is of class C^1 (see Prop.1 of Sect.66). Its partial 2-gradient at $(\mathbf{L},\mathbf{M})$ does not depend on $\mathbf{M}$. It is the right-multiplication $\mathrm{Ri}_{\mathbf{L}} \in \mathrm{Lin}(\mathrm{Lin}(\mathcal{V}',\mathcal{V}), \mathrm{Lin}\mathcal{V})$ defined by $\mathrm{Ri}_{\mathbf{L}}(\mathbf{K}) := \mathbf{KL}$ for all $\mathbf{K} \in \mathrm{Lin}(\mathcal{V},\mathcal{V}')$. It is clear that $\mathrm{Ri}_{\mathbf{L}}$ is invertible if and only if $\mathbf{L}$ is invertible. (We have $(\mathrm{Ri}_{\mathbf{L}})^{-1} = \mathrm{Ri}_{\mathbf{L}^{-1}}$ if this is the case.) Thus, given $\mathbf{L}_o \in \mathrm{Lis}(\mathcal{V},\mathcal{V}')$, we can apply Lemma 3 with $\mathbf{L}_o, \mathbf{L}_o^{-1}$, and $\mathbf{1}_{\mathcal{V}}$ playing the roles of x_o, y_o, and z_o, respectively. We conclude that there is an open neighborhood $\mathcal{D}$ of $\mathbf{L}_o$ in $\mathrm{Lin}(\mathcal{V}, \mathcal{V}')$ such that the equation

$$? \, \mathbf{M} \in \mathrm{Lin}(\mathcal{V}', \mathcal{V}), \quad \mathbf{LM} = \mathbf{1}_{\mathcal{V}}$$

has a solution for every $\mathbf{L} \in \mathcal{D}$. Since this solution can only be $\mathbf{L}^{-1} = \mathrm{inv}(\mathbf{L})$, it follows that the role of the mapping φ of Lemma 3 is played by $\mathrm{inv}|_{\mathcal{D}}$. Therefore, we must have $\mathcal{D} \subset \mathrm{Lis}(\mathcal{V}, \mathcal{V}')$ and inv must be differentiable at $\mathbf{L}_o$. Since $\mathbf{L}_o \in \mathrm{Lis}(\mathcal{V}, \mathcal{V}')$ was arbitrary, it follows that $\mathrm{Lis}(\mathcal{V}, \mathcal{V}')$ is open and that inv is differentiable. ■

Completion of Proofs: We assume that hypotheses of the Local Inversion Theorem are satisfied. The assumption that φ has an invertible tangent at $x \in \mathcal{D}$ is equivalent to $\nabla_x \varphi \in \mathrm{Lis}(\mathcal{V}, \mathcal{V}')$. Since $\nabla\varphi$ is continuous and since $\mathrm{Lis}(\mathcal{V}, \mathcal{V}')$ is an open subset of $\mathrm{Lin}(\mathcal{V}, \mathcal{V}')$ by Lemma 4, it follows that $\mathcal{G} := (\nabla\varphi)^{<}(\mathrm{Lis}(\mathcal{V}, \mathcal{V}'))$ is an open subset of $\mathcal{E}$. By Lemma 2, we can choose a local inverse of φ near x which is differentiable and hence continuous at $\varphi(x)$. By Prop.2, (ii), we can adjust this local inverse such that its range is included in $\mathcal{G}$. Let ψ be the local inverse so adjusted. Then φ has an invertible tangent at every $z \in \mathrm{Rng}\,\psi$. Let $z \in \mathrm{Rng}\,\psi$ be given. Applying Lemma 2 to z, we see that φ must have a local inverse near z that is differentiable at $\varphi(z)$. By Prop.1, it follows that ψ must also be differentiable at $\varphi(z)$. Since $z \in \mathrm{Rng}\,\psi$ was arbitrary, it follows that ψ is differentiable. The formula (68.4) follows from Prop.1. Since $\mathrm{inv} : \mathrm{Lis}(\mathcal{V}, \mathcal{V}') \to \mathrm{Lin}(\mathcal{V}', \mathcal{V})$ is differentiable and hence continuous by Lemma 4, since $\nabla\varphi$ is continuous by assumption, and since ψ is continuous because it is differentiable, it follows that the composite $\mathrm{inv} \circ \nabla\varphi|_{\mathrm{Rng}\,\psi}^{\mathrm{Lis}(\mathcal{V},\mathcal{V}')} \circ \psi$ is continuous. By (68.4), this composite is none other than $\nabla\psi$, and hence the proof of the Local Inversion Theorem is complete.

Assume, now, that the hypotheses of the Implicit Mapping Theorem are satisfied. By the Local Inversion Theorem, we can then choose a local inverse ρ of $\tilde{\omega}$, as defined by (68.18), that is of class C^1. It then follows that the function $\varphi : \mathcal{D} \to \mathcal{E}'$ defined in the proof of Lemma 3 is also of class C^1. The formula (68.6) is obtained easily by differentiating the constant $x \mapsto \omega(x, \varphi(x))$, using Prop.1 of Sect.65 and the Chain Rule.

To prove the Differentiation Theorem for Inversion Mappings, one applies the Implicit Mapping Theorem in the same way as Lemma 3 was applied to obtain Lemma 4. ■

Pitfall: Let I and I' be intervals and let $f : I \to I'$ be of class C^1 and surjective. If f has an invertible tangent at every $t \in I$, then f is not only locally invertible near every $t \in I$ but (globally) invertible, as is easily seen. This conclusion does *not* generalize to the case when I and I' are replaced by connected open subsets of flat spaces of higher dimension. The curvilinear coordinate systems discussed in Sect.74 give rise to counterexamples. One can even give counterexamples in which I and I' are replaced by *convex* open subsets of flat spaces. ■

Notes 68

(1) The Local Inversion Theorem is often called the "Inverse Function Theorem". Our term is more descriptive.

(2) The Implicit Mapping Theorem is most often called the "Implicit Function Theorem".

(3) If the sets $\mathcal{D}$ and $\mathcal{D}'$ of the Local Inversion Theorem are both subsets of $\mathbb{R}^n$ for some $n \in \mathbb{N}$, then $\nabla_x \varphi$ can be identified with an n-by-n matrix (see (65.12)). This matrix is often called the "Jacobian matrix" and its determinant the "Jacobian" of φ at x. Some textbook authors replace the condition that $\nabla_x \varphi$ be invertible by the condition that the Jacobian be non-zero. I believe it is a red herring to drag in determinants here. The Local Inversion Theorem has an extension to infinite-dimensional spaces, where it makes no sense to talk about determinants.

69 Extreme Values, Constraints

In this section $\mathcal{E}$ and $\mathcal{F}$ denote flat spaces with translation spaces $\mathcal{V}$ and $\mathcal{W}$, respectively, and $\mathcal{D}$ denotes a subset of $\mathcal{E}$.

The following definition is "local" variant of the definition of an extremum given in Sect.08.

Definition 1: *We say that a function* $f : \mathcal{D} \to \mathbb{R}$ *attains a* **local maximum [local minimum]** *at* $x \in \mathcal{D}$ *if there is* $\mathcal{N} \in \operatorname{Nhd}_x(\mathcal{D})$ (see (56.1)) *such that* $f|_{\mathcal{N}}$ *attains a maximum [minimum] at* x. *We say that* f *attains a* **local extremum** *at* x *if it attains a local maximum or local minimum at* x.

The following result is a direct generalization of the Extremum Theorem of elementary calculus (see Sect.08).

Extremum Theorem: *If* $f : \mathcal{D} \to \mathbb{R}$ *attains a local extremum at* $x \in \operatorname{Int}\mathcal{D}$ *and if* f *is differentiable at* x, *then* $\nabla_x f = \mathbf{0}$.

Proof: Let $\mathbf{v} \in \mathcal{V}$ be given. It is clear that $S_{x,\mathbf{v}} := \{s \in \mathbb{R} \mid x + s\mathbf{v} \in \operatorname{Int}\mathcal{D}\}$ is an open subset of $\mathbb{R}$ and the function $(s \mapsto f(x + s\mathbf{v})) : S_{x,\mathbf{v}} \to \mathbb{R}$ attains a local extremum at $0 \in \mathbb{R}$ and is differentiable at $0 \in \mathbb{R}$. Hence, by the Extremum Theorem of elementary calculus, and by (65.13) and (65.14), we have

$$0 = \partial_0(s \mapsto f(x + s\mathbf{v})) = (\mathrm{dd}_{\mathbf{v}} f)(x) = (\nabla_x f)\mathbf{v}.$$

Since $\mathbf{v} \in \mathcal{V}$ was arbitrary, the desired result $\nabla_x f = \mathbf{0}$ follows. ■

From now on we assume that $\mathcal{D}$ is an open subset of $\mathcal{E}$.

The next result deals with the case when a restriction of $f : \mathcal{D} \to \mathbb{R}$ to a suitable subset of $\mathcal{D}$, but not necessarily f itself, has an extremum at a given point $x \in \mathcal{D}$. This subset of $\mathcal{D}$ is assumed by the set on which a given mapping $\varphi : \mathcal{D} \to \mathcal{F}$ has a constant value, i.e. the set $\varphi^{<}(\{\varphi(x)\})$. The assertion that $f|_{\varphi^{<}(\{\varphi(x)\})}$ has an extremum at x is often expressed by saying that f attains an extremum at x subject to the *constraint* that φ be constant.

Constrained-Extremum Theorem: *Assume that $f : \mathcal{D} \to \mathbb{R}$ and $\varphi : \mathcal{D} \to \mathcal{F}$ are both of class C^1 and that $\nabla_x \varphi \in \mathrm{Lin}(\mathcal{V}, \mathcal{W})$ is surjective for a given $x \in \mathcal{D}$. If $f|_{\varphi^{<}(\{\varphi(x)\})}$ attains a local extremum at x, then*

$$\nabla_x f \in \mathrm{Rng}\,(\nabla_x \varphi)^{\top}. \tag{69.1}$$

Proof: We put $\mathcal{U} := \mathrm{Null}\ \nabla_x \varphi$ and choose a supplement $\mathcal{Z}$ of $\mathcal{U}$ in $\mathcal{V}$. Every neighborhood of $\mathbf{0}$ in $\mathcal{V}$ includes a set of the form $\mathcal{N} + \mathcal{M}$, where $\mathcal{N}$ is an open neighborhood of $\mathbf{0}$ in $\mathcal{U}$ and $\mathcal{M}$ is an open neighborhood of $\mathbf{0}$ in $\mathcal{Z}$. Since $\mathcal{D}$ is open, we may select $\mathcal{N}$ and $\mathcal{M}$ such that $x + \mathcal{N} + \mathcal{M} \subset \mathcal{D}$. We now define $\omega : \mathcal{N} \times \mathcal{M} \to \mathcal{F}$ by

$$\omega(\mathbf{u}, \mathbf{z}) := \varphi(x + \mathbf{u} + \mathbf{z}) \quad \text{for all} \quad \mathbf{u} \in \mathcal{N}, \mathbf{z} \in \mathcal{M}. \tag{69.2}$$

It is clear that ω is of class C^1 and we have

$$\nabla_{\mathbf{0}} \omega(\mathbf{0}, \cdot) = \nabla_x \varphi|_{\mathcal{Z}}. \tag{69.3}$$

Since $\nabla_x \varphi$ is surjective, it follows from Prop.5 of Sect.13 that $\nabla_x \varphi|_{\mathcal{Z}}$ is invertible. Hence we can apply the Implicit Mapping Theorem of Sect.68 to the case when $\mathcal{A}$ there is replaced by $\mathcal{N} \times \mathcal{M}$, the points x_o, y_o by $\mathbf{0}$, and z_o by $\varphi(x)$. We obtain an open neighborhood $\mathcal{G}$ of $\mathbf{0}$ in $\mathcal{U}$ with $\mathcal{G} \subset \mathcal{N}$ and a mapping $\mathbf{h} : \mathcal{G} \to \mathcal{Z}$, of class C^1, such that $\mathbf{h}(\mathbf{0}) = \mathbf{0}$ and

$$\omega(\mathbf{u}, \mathbf{h}(\mathbf{u})) = \varphi(x) \quad \text{for all} \quad \mathbf{u} \in \mathcal{G}. \tag{69.4}$$

Since $\nabla_{(1)} \omega(\mathbf{0}, \mathbf{0}) = \nabla_{\mathbf{0}} \omega(\cdot, \mathbf{0}) = \nabla_x \varphi|_{\mathcal{U}} = \mathbf{0}$ by the definition $\mathcal{U} := \mathrm{Null}\ \nabla_x \varphi$, the formula (68.6) yields in our case that $\nabla_{\mathbf{0}} \mathbf{h} = \mathbf{0}$, i.e. we have

$$\mathbf{h}(\mathbf{0}) = \mathbf{0}, \qquad \nabla_{\mathbf{0}} \mathbf{h} = \mathbf{0}. \tag{69.5}$$

It follows from (69.2) and (69.4) that $x + \mathbf{u} + \mathbf{h}(\mathbf{u}) \in \varphi^<(\{\varphi(x)\})$ for all $\mathbf{u} \in \mathcal{G}$ and hence that

$$(\mathbf{u} \mapsto f(x + \mathbf{u} + \mathbf{h}(\mathbf{u}))) : \mathcal{G} \to \mathbb{R}$$

has a local extremum at $\mathbf{0}$. By the Extremum Theorem and the Chain Rule, it follows that

$$0 = \nabla_{\mathbf{0}}(\mathbf{u} \mapsto f(x + \mathbf{u} + \mathbf{h}(\mathbf{u}))) = \nabla_x f(\mathbf{1}_{\mathcal{U}\subset\mathcal{V}} + \nabla_{\mathbf{0}}\mathbf{h}|^{\mathcal{V}}),$$

and therefore, by (69.5), that $\nabla_x f|_{\mathcal{U}} = (\nabla_x f)\mathbf{1}_{\mathcal{U}\subset\mathcal{V}} = \mathbf{0}$, which means that $\nabla_x f \in \mathcal{U}^\perp = (\text{Null } \nabla_x\varphi)^\perp$. The desired result (69.1) is obtained by applying (22.9). ∎

In the case when $\mathcal{F} := \mathbb{R}$, the Constrained-Extremum Theorem reduces to

Corollary 1: *Assume that $f, g : \mathcal{D} \to \mathbb{R}$ are both of class C^1 and that $\nabla_x g \neq \mathbf{0}$ for a given $x \in \mathcal{D}$. If $f|_{g^<(\{g(x)\})}$ attains a local extremum at x, then there is $\lambda \in \mathbb{R}$ such that*

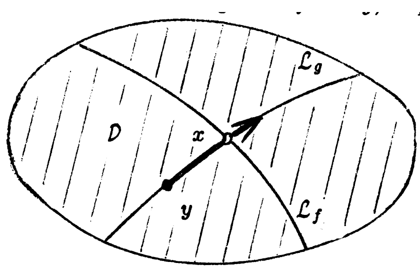
$$\nabla_x f = \lambda \nabla_x g. \tag{69.6}$$

In the case when $\mathcal{F} := \mathbb{R}^I$, we can use (23.19) and obtain

Corollary 2: *Assume that $f : \mathcal{D} \to \mathbb{R}$ and all terms $g_i : \mathcal{D} \to \mathbb{R}$ in a finite family $g := (g_i \mid i \in I) : \mathcal{D} \to \mathbb{R}^I$ are of class C^1 and that $(\nabla_x g_i \mid i \in I)$ is linearly independent for a given $x \in \mathcal{D}$. If $f|_{g^<(\{g(x)\})}$ attains a local extremum at x, then there is $\lambda \in \mathbb{R}^I$ such that*

$$\nabla_x f = \sum_{i\in I} \lambda_i \nabla_x g_i. \tag{69.7}$$

Remark 1: If $\mathcal{E}$ is two-dimensional then Cor.1 can be given a geometrical interpretation as follows: The sets $\mathcal{L}_g := g^<(\{g(x)\})$ and $\mathcal{L}_f := f^<(\{f(x)\})$ are the "level-lines" through x of f and g, respectively (see Figure).



If these lines cross at x, then the value $f(y)$ of f at $y \in \mathcal{L}_g$ strictly increases or decreases as y moves along $\mathcal{L}_g$ from one side of the line $\mathcal{L}_f$ to the other and hence $f|_{\mathcal{L}_g}$ cannot have an extremum at x. Hence, if $f|_{\mathcal{L}_g}$ has an extremum at x, the level-lines must be tangent. The assertion (69.6) expresses this tangency. The condition $\nabla_x g \neq \mathbf{0}$ insures that the level-line $\mathcal{L}_g$ does not degenerate to a point. ∎

Notes 69

(1) The number λ of (69.6) or the terms λ_i occuring in (69.7) are often called "Lagrange multipliers".

610 Integral Representations

Let $I \in \operatorname{Sub}\mathbb{R}$ be some genuine interval and let $\mathbf{h} : I \to \mathcal{V}$ be a continuous process with values in a given linear space $\mathcal{V}$. For every $\boldsymbol{\lambda} \in \mathcal{V}^*$ the composite $\boldsymbol{\lambda}\mathbf{h} : I \to \mathbb{R}$ is then continuous. Given $a, b \in I$ one can therefore form the integral $\int_a^b \boldsymbol{\lambda}\mathbf{h}$ in the sense of elementary integral calculus (see Sect.08). It is clear that the mapping

$$(\boldsymbol{\lambda} \mapsto \int_a^b \boldsymbol{\lambda}\mathbf{h}) : \mathcal{V}^* \to \mathbb{R}$$

is linear and hence an element of $\mathcal{V}^{**} \cong \mathcal{V}$.

Definition 1: *Let $\mathbf{h} : I \to \mathcal{V}$ be a continuous process with values in the linear space $\mathcal{V}$. Given any $a, b \in I$ the* **integral of h from** *a* **to** *b is defined to be the unique element $\int_a^b \mathbf{h}$ of $\mathcal{V}$ which satisfies*

$$\boldsymbol{\lambda}\int_a^b \mathbf{h} = \int_a^b \boldsymbol{\lambda}\mathbf{h} \quad \text{for all} \quad \boldsymbol{\lambda} \in \mathcal{V}^*. \tag{610.1}$$

Proposition 1: *Let $\mathbf{h} : I \to \mathcal{V}$ be a continuous process and let $\mathbf{L}$ be a linear mapping from $\mathcal{V}$ to a given linear space $\mathcal{W}$. Then $\mathbf{L}\mathbf{h} : I \to \mathcal{W}$ is a continuous process and*

$$\mathbf{L}\int_a^b \mathbf{h} = \int_a^b \mathbf{L}\mathbf{h} \quad \text{for all} \quad a, b \in I. \tag{610.2}$$

Proof: Let $\boldsymbol{\omega} \in \mathcal{W}^*$ and $a, b \in I$ be given. Using Def.1 twice, we obtain

$$\boldsymbol{\omega}(\mathbf{L}\int_a^b \mathbf{h}) = (\boldsymbol{\omega}\mathbf{L})\int_a^b \mathbf{h} = \int_a^b (\boldsymbol{\omega}\mathbf{L})\mathbf{h} = \int_a^b \boldsymbol{\omega}(\mathbf{L}\mathbf{h}) = \boldsymbol{\omega}\int_a^b \mathbf{L}\mathbf{h}.$$

Since $\boldsymbol{\omega} \in \mathcal{W}^*$ was arbitrary, (610.2) follows. ■

Proposition 2: *Let ν be a norm on the linear space $\mathcal{V}$ and let $\mathbf{h} : I \to \mathcal{V}$ be a continuous process. Then $\nu \circ \mathbf{h} : I \to \mathbb{R}$ is continuous and, for every $a, b \in I$, we have*

$$\nu\Big(\int_a^b \mathbf{h}\Big) \leq \left| \int_a^b \nu \circ \mathbf{h} \right|. \tag{610.3}$$

Proof: The continuity of $\nu \circ \mathbf{h}$ follows from the continuity of ν (Prop.7 of Sect.56) and the Composition Theorem for Continuity of Sect.56.

It follows from (52.14) that

$$|\boldsymbol{\lambda}\mathbf{h}|(t) = |\boldsymbol{\lambda}\mathbf{h}(t)| \leq \nu^*(\boldsymbol{\lambda})\nu(\mathbf{h}(t)) = \nu^*(\boldsymbol{\lambda})(\nu \circ \mathbf{h})(t)$$

holds for all $\boldsymbol{\lambda} \in \mathcal{V}^*$ and all $t \in I$, and hence that

$$|\boldsymbol{\lambda}\mathbf{h}| \leq \nu \circ \mathbf{h} \quad \text{for all} \quad \boldsymbol{\lambda} \in \mathrm{Ce}(\nu^*).$$

Using this result and (610.1), (08.40), and (08.42), we obtain

$$\left|\boldsymbol{\lambda} \int_a^b \mathbf{h}\right| = \left|\int_a^b \boldsymbol{\lambda}\mathbf{h}\right| \leq \left|\int_b^a |\boldsymbol{\lambda}\mathbf{h}|\;\right| \leq \left|\int_a^b \nu \circ \mathbf{h}\right|$$

for all $\boldsymbol{\lambda} \in \mathrm{Ce}(\nu^*)$ and all $a, b \in I$. The desired result now follows from the Norm-Duality Theorem, i.e. (52.15). ■

Most of the rules of elementary integral calculus extend directly to integrals of processes with values in a linear space. For example, if $\mathbf{h} : I \to \mathcal{V}$ is continuous and if $a, b, c \in I$, then

$$\int_a^b \mathbf{h} = \int_a^c \mathbf{h} + \int_c^b \mathbf{h}. \tag{610.4}$$

The following result is another example.

Fundamental Theorem of Calculus: *Let $\mathcal{E}$ be a flat space with translation space $\mathcal{V}$. If $\mathbf{h} : I \to \mathcal{V}$ is a continuous process and if $x \in \mathcal{E}$ and $a \in I$ are given, then the process $p : I \to \mathcal{E}$ defined by*

$$p(t) := x + \int_a^t \mathbf{h} \quad \text{for all} \quad t \in I \tag{610.5}$$

is of class C^1 and $p^\bullet = \mathbf{h}$.

Proof: Let $\boldsymbol{\lambda} \in \mathcal{V}^*$ be given. By (610.5) and (610.1) we have

$$\boldsymbol{\lambda}(p(t) - x) = \boldsymbol{\lambda} \int_a^t \mathbf{h} = \int_a^t \boldsymbol{\lambda}\mathbf{h} \quad \text{for all} \quad t \in I.$$

By Prop.1 of Sect.61 p is differentiable and we have $(\boldsymbol{\lambda}(p-x))^{\bullet} = \boldsymbol{\lambda}p^{\bullet}$. Hence, the elementary Fundamental Theorem of Calculus (see Sect.08) gives $\boldsymbol{\lambda}p^{\bullet} = \boldsymbol{\lambda}\mathbf{h}$. Since $\boldsymbol{\lambda} \in \mathcal{V}^*$ was arbitrary, we conclude that $p^{\bullet} = \mathbf{h}$, which is continuous by assumption. ■

In the remainder of this section, we assume that the following items are given: (i) a flat space $\mathcal{E}$ with translation space $\mathcal{V}$, (ii) An open subset $\mathcal{D}$ of $\mathcal{E}$, (iii) An open subset $\overline{\mathcal{D}}$ of $\mathbb{R} \times \mathcal{E}$ such that, for each $x \in \mathcal{D}$,

$$\overline{\mathcal{D}}_{(\cdot,x)} := \{s \in \mathbb{R} \mid (s,x) \in \overline{\mathcal{D}}\}$$

is a non-empty open inteval, (iv) a linear space $\mathcal{W}$.

Consider now a mapping $\mathbf{h} : \overline{\mathcal{D}} \to \mathcal{W}$ such that $\mathbf{h}(\cdot,x) : \overline{\mathcal{D}}_{(\cdot,x)} \to \mathcal{W}$ is continuous for all $x \in \mathcal{D}$. Given $a, b \in \mathbb{R}$ such that $a, b \in \overline{\mathcal{D}}_{(\cdot,x)}$ for all $x \in \mathcal{D}$, we can then define $\mathbf{k} : \mathcal{D} \to \mathcal{W}$ by

$$\mathbf{k}(x) := \int_a^b \mathbf{h}(\cdot,x) \quad \text{for all} \quad x \in \mathcal{D}. \tag{610.6}$$

We say that $\mathbf{k}$ is defined by an **integral representation**.

Proposition 3: *If $\mathbf{h} : \overline{\mathcal{D}} \to \mathcal{W}$ is continuous, so is the mapping $\mathbf{k} : \mathcal{D} \to \mathcal{W}$ defined by the integral representation* (610.6).

Proof: Let $x \in \mathcal{D}$ be given. Since $\mathcal{D}$ is open, we can choose a norm ν on $\mathcal{V}$ such that $x + \overline{\mathrm{Ce}}(\nu) \subset \mathcal{D}$. We may assume, without loss of generality, that $a < b$. Then $[a,b]$ is a compact interval and hence, in view of Prop.4 of Sect.58, $[a,b] \times (x + \overline{\mathrm{Ce}}(\nu))$ is a compact subset of $\overline{\mathcal{D}}$. By the Uniform Continuity Theorem of Sect. 58, it follows that the restriction of $\mathbf{h}$ to $[a,b] \times (x + \overline{\mathrm{Ce}}(\nu))$ is uniformly continuous. Let μ be a norm on $\mathcal{W}$ and let $\varepsilon \in \mathbb{P}^{\times}$ be given. By Prop.4 of Sect.56, we can determine $\delta \in]0,1]$ such that

$$\mu(\mathbf{h}(s,y) - \mathbf{h}(s,x)) < \frac{\varepsilon}{b-a}$$

for all $s \in [a,b]$ and all $y \in x + \delta\mathrm{Ce}(\nu)$. Hence, by (610.6) and Prop.2, we have

$$\begin{aligned} \mu(\mathbf{k}(y) - \mathbf{k}(x)) &= \mu\left(\int_a^b (\mathbf{h}(\cdot,y) - \mathbf{h}(\cdot,x))\right) \le \int_a^b \mu \circ (\mathbf{h}(\cdot,y) - \mathbf{h}(\cdot,x)) \\ &\le (b-a)\frac{\varepsilon}{b-a} = \varepsilon \end{aligned}$$

whenever $\nu(y - x) < \delta$. Since $\varepsilon \in \mathbb{P}^{\times}$ was arbitrary, the continuity of $\mathbf{k}$ at x follows by Prop.1 of Sect.56. Since $x \in \mathcal{D}$ was arbitrary, the assertion follows. ■

The following is a stronger version of Prop.3.

Proposition 4: *Let* $\overline{\overline{\mathcal{D}}} \subset \mathbb{R} \times \mathbb{R} \times \mathcal{E}$ *be defined by*

$$\overline{\overline{\mathcal{D}}} := \{(a, b, x) \mid x \in \mathcal{D},\ (a, x) \in \overline{\mathcal{D}}, (b, x) \in \overline{\mathcal{D}}\}. \tag{610.7}$$

Then, if $\mathbf{h} : \overline{\mathcal{D}} \to \mathcal{W}$ *is continuous, so is the mapping* $\Big((a, b, x) \mapsto \int_a^b \mathbf{h}(\cdot, x)\Big) : \overline{\overline{\mathcal{D}}} \to \mathcal{W}$.

Proof: Choose a norm μ on $\mathcal{W}$. Let $(a, b, x) \in \overline{\overline{\mathcal{D}}}$ be given. Since $\mu \circ \mathbf{h}$ is continuous at (a, x) and at (b, x) and since $\overline{\mathcal{D}}$ is open, we can find $\mathcal{N} \in \mathrm{Nhd}_x(\mathcal{D})$ and $\sigma \in \mathbb{P}^\times$ such that

$$\overline{\mathcal{N}} := ((a+]-\sigma, \sigma[\,) \cup (b+]-\sigma, \sigma[\,)) \times \mathcal{N}$$

is a subset of $\overline{\mathcal{D}}$ and such that the restriction of $\mu \circ \mathbf{h}$ to $\overline{\mathcal{N}}$ is bounded by some $\beta \in \mathbb{P}^\times$. By Prop.2, it follows that

$$\mu(\int_s^a \mathbf{h}(\cdot, y) + \int_b^t \mathbf{h}(\cdot, y)) \leq (\,|s - a| + |t - b|\,)\beta \tag{610.8}$$

for all $y \in \mathcal{N}$, all $s \in a +]-\sigma, \sigma[$, and all $t \in b +]-\sigma, \sigma]$.

Now let $\varepsilon \in \mathbb{P}^\times$ be given. If we put $\delta := \min\{\sigma, \frac{\varepsilon}{4\beta}\}$, it follows from (610.8) that

$$\mu\left(\int_s^a \mathbf{h}(\cdot, y) + \int_b^t \mathbf{h}(\cdot, y)\right) < \frac{\varepsilon}{2} \tag{610.9}$$

for all $y \in \mathcal{N}$, all $s \in a +]-\delta, \delta]$, and all $t \in b +]-\delta, \delta]$. On the other hand, by Prop.3, we can determine $\mathcal{M} \in \mathrm{Nhd}_x(\mathcal{D})$ such that

$$\mu\left(\int_a^b \mathbf{h}(\cdot, y) - \int_a^b \mathbf{h}(\cdot, x)\right) < \frac{\varepsilon}{2} \tag{610.10}$$

for all $y \in \mathcal{M}$. By (610.4) we have

$$\begin{aligned} \int_s^t \mathbf{h}(\cdot, y) - \int_a^b \mathbf{h}(\cdot, x) \quad &= \quad \int_a^b \mathbf{h}(\cdot, y) - \int_a^b \mathbf{h}(\cdot, x) \\ &\quad + \int_s^a \mathbf{h}(\cdot, y) + \int_b^t \mathbf{h}(\cdot, y). \end{aligned}$$

Hence it follows from (610.9) and (610.10) that

$$\mu\left(\int_s^t \mathbf{h}(\cdot, y) - \int_a^b \mathbf{h}(\cdot, x)\right) < \varepsilon$$

for all $y \in \mathcal{M} \cap \mathcal{N} \in \mathrm{Nhd}_x(\mathcal{D})$, all $s \in a+]-\delta, \delta]$, and all $t \in b+]-\delta, \delta[$. Since $\varepsilon \in \mathbb{P}^\times$ was arbitrary, it follows from Prop.1 of Sect.56 that the mapping under consideration is continuous at (a, b, x). ■

Differentiation Theorem for Integral Representations: *Assume that* $\mathbf{h} : \overline{\mathcal{D}} \to \mathcal{W}$ *satisfies the following conditions:*

(i) $\mathbf{h}(\bullet, x)$ *is continuous for every* $x \in \mathcal{D}$.

(ii) $\mathbf{h}(s, \bullet)$ *is differentiable at* x *for all* $(s, x) \in \overline{\mathcal{D}}$ *and the partial 2-gradient* $\nabla_{(2)}\mathbf{h} \;:\; \overline{\mathcal{D}} \to \mathrm{Lin}(\mathcal{V}, \mathcal{W})$ *is continuous.*

Then the mapping $\mathbf{k} : \mathcal{D} \to \mathcal{W}$ *defined by the integral representation* (610.6) *is of class* C^1 *and its gradient is given by the integral representation*

$$\nabla_x \mathbf{k} = \int_a^b \nabla_{(2)}\mathbf{h}(\bullet, x) \quad \text{for all} \quad x \in \mathcal{D}. \tag{610.11}$$

Roughly, this theorem states that if $\mathbf{h}$ satisfies the conditions (i) and (ii), one can differentiate (610.6) with respect to x by "differentiating under the integral sign".

Proof: Let $x \in \mathcal{D}$ be given. As in the proof of Prop.3, we can choose a norm ν on $\mathcal{V}$ such that $x + \overline{\mathrm{Ce}}(\nu) \subset \mathcal{D}$, and we may assume that $a < b$. Then $[a, b] \times \big(x + \overline{\mathrm{Ce}}(\nu)\big)$ is a compact subset of $\overline{\mathcal{D}}$. By the Uniform Continuity Theorem of Sect.58, the restriction of $\nabla_{(2)}\mathbf{h}$ to $[a, b] \times \big(x + \overline{\mathrm{Ce}}(\nu)\big)$ is uniformly continuous. Hence, if a norm μ on $\mathcal{W}$ and $\varepsilon \in \mathbb{P}^\times$ are given, we can determine $\delta \in\ ]0, 1\,]$ such that

$$\|\nabla_{(2)}\mathbf{h}(s, x + \mathbf{u}) - \nabla_{(2)}\mathbf{h}(s, x)\|_{\nu,\mu} < \frac{\varepsilon}{b - a} \tag{610.12}$$

holds for all $s \in [a, b]$ and $\mathbf{u} \in \delta\mathrm{Ce}(\nu)$.

We now define $\mathbf{n} \;:\; \big(\overline{\mathcal{D}} - (0, x)\big) \to \mathcal{W}$ by

$$\mathbf{n}(s, \mathbf{u}) := \mathbf{h}(s, x + \mathbf{u}) - \mathbf{h}(s, x) - \big(\nabla_{(2)}\mathbf{h}(s, x)\big)\,\mathbf{u}. \tag{610.13}$$

It is clear that $\nabla_{(2)}\mathbf{n}$ exists and is given by

$$\nabla_{(2)}\mathbf{n}(s, \mathbf{u}) = \nabla_{(2)}\mathbf{h}(s, x + \mathbf{u}) - \nabla_{(2)}\mathbf{h}(s, x).$$

By (610.12), we hence have

$$\|\nabla_{(2)}\mathbf{n}(s, \mathbf{u})\|_{\nu,\mu} < \frac{\varepsilon}{b - a}$$

for all $s \in [a,b]$ and $\mathbf{u} \in \delta\mathrm{Ce}(\nu)$. By the Striction Estimate for Differentiable Mapping of Sect.64, it follows that $\mathbf{n}(s,\bullet)|_{\delta\mathrm{Ce}(\nu)}$ is constricted for all $s \in [a,b]$ and that

$$\mathrm{str}(\mathbf{n}(s,\bullet)|_{\delta\mathrm{Ce}(\nu)}\ ;\ \nu,\mu) \leq \frac{\varepsilon}{b-a} \quad \text{for all} \quad s \in [a,b].$$

Since $\mathbf{n}(s,\mathbf{0}) = \mathbf{0}$ for all $s \in [a,b]$, the definition (64.1) shows that

$$\mu\left(\mathbf{n}(s,\mathbf{u})\right) \leq \frac{\varepsilon}{b-a}\nu(\mathbf{u}) \quad \text{for all} \quad s \in [a,b]$$

and all $\mathbf{u} \in \delta\mathrm{Ce}(\nu)$. Using Prop.2, we conclude that

$$\mu\left(\int_a^b \mathbf{n}(\bullet,\mathbf{u})\right) \leq \int_a^b \mu \circ (\mathbf{n}(\bullet,\mathbf{u})) \leq \varepsilon\nu(\mathbf{u})$$

whenever $\mathbf{u} \in \delta\mathrm{Ce}(\nu)$. Since $\varepsilon \in \mathbb{P}^\times$ was arbitrary, it follows that the mapping $\left(\mathbf{u} \mapsto \int_a^b \mathbf{n}(\bullet,\mathbf{u})\right)$ is small near $\mathbf{0} \in \mathcal{V}$ (see Sect.62). Now, integrating (610.13) with respect to $s \in [a,b]$ and observing the representation (610.6) of $\mathbf{k}$, we obtain

$$\int_a^b \mathbf{n}(\bullet,\mathbf{u}) = \mathbf{k}(x+\mathbf{u}) - \mathbf{k}(\mathbf{u}) - \left(\int_a^b \nabla_{(2)}\mathbf{h}(\bullet,x)\right)\mathbf{u}$$

for all $\mathbf{u} \in \mathcal{D}-x$. Therefore, by the Characterization of Gradients of Sect.63, $\mathbf{k}$ is differentiable at x and its gradient is given by (610.11). The continuity of $\nabla\mathbf{k}$ follows from Prop.3. ∎

The following corollary deals with generalizations of integral representations of the type (610.6).

Corollary: *Assume that* $\mathbf{h} : \overline{\mathcal{D}} \to \mathcal{W}$ *satisfies the conditions* (i) *and* (ii) *of the Theorem. Assume, further, that* $f : \mathcal{D} \to \mathbb{R}$ *and* $g : \mathcal{D} \to \mathbb{R}$ *are differentiable and satisfy*

$$f(x), g(x) \in \overline{\mathcal{D}}_{(\bullet,x)} \quad \text{for all} \quad x \in \mathcal{D}.$$

Then $\mathbf{k} : \mathcal{D} \to \mathcal{W}$*, defined by*

$$\mathbf{k}(x) := \int_{f(x)}^{g(x)} \mathbf{h}(\bullet,x) \quad \text{for all} \quad x \in \mathcal{D}, \tag{610.14}$$

is of class C^1 *and its gradient is given by*

$$\begin{aligned} \nabla_x\mathbf{k} &= \mathbf{h}(g(x),x) \otimes \nabla_x g - \mathbf{h}(f(x),x) \otimes \nabla_x f \\ &\quad + \int_{f(x)}^{g(x)} \nabla_{(2)}\mathbf{h}(\bullet,x) \quad \text{for all} \quad x \in \mathcal{D}. \end{aligned} \tag{610.15}$$

Proof: Consider the set $\overline{\overline{\mathcal{D}}} \subset \mathbb{R} \times \mathbb{R} \times \mathcal{E}$ defined by (610.7), so that $a, b \in \overline{\mathcal{D}}_{(\bullet,x)}$ for all $(a, b, x) \in \overline{\overline{\mathcal{D}}}$. Since $\overline{\mathcal{D}}_{(\bullet,x)}$ is assumed to be an interval for each $x \in \mathcal{D}$, we can define $\mathbf{m} \ : \ \overline{\overline{\mathcal{D}}} \to \mathcal{W}$ by

$$\mathbf{m}(a, b, x) := \int_a^b \mathbf{h}(\bullet, x). \tag{610.16}$$

By the Theorem, $\nabla_{(3)}\mathbf{m} \ : \ \overline{\overline{\mathcal{D}}} \to \mathrm{Lin}(\mathcal{V}, \mathcal{W})$ exists and is given by

$$\nabla_{(3)}\mathbf{m}(a, b, x) = \int_a^b \nabla_{(2)}\mathbf{h}(\bullet, x). \tag{610.17}$$

It follows from Prop.4 that $\nabla_{(3)}\mathbf{m}$ is continuous. By the Fundamental Theorem of Calculus and (610.16), the partial derivatives $\mathbf{m},_1$ and $\mathbf{m},_2$ exist, are continuous, and are given by

$$\mathbf{m},_1 (a, b, x) = -\mathbf{h}(a, x), \quad \mathbf{m},_2 (a, b, x) = \mathbf{h}(b, x). \tag{610.18}$$

By the Partial Gradient Theorem of Sect.65, it follows that $\mathbf{m}$ is of class C^1. Since $\nabla_{(1)}\mathbf{m} = \mathbf{m},_1 \otimes$ and $\nabla_{(2)}\mathbf{m} = \mathbf{m},_2 \otimes$, it follows from (65.9), (610.17), and (610.18) that

$$\begin{aligned}(\nabla_{(a,b,x)}\mathbf{m})(\alpha, \beta, \mathbf{u}) \quad &= \quad \mathbf{h}(b, x) \otimes \beta - \mathbf{h}(a, x) \otimes \alpha \\ + \left(\int_a^b \nabla_{(2)}\mathbf{h}(\bullet, x) \right) \mathbf{u} & \end{aligned} \tag{610.19}$$

for all $(a, b, x) \in \overline{\overline{\mathcal{D}}}$ and all $(\alpha, \beta, \mathbf{u}) \in \mathbb{R} \times \mathbb{R} \times \mathcal{V}$.

Now, since f and g are differentiable, so is $(f, g, 1_{\mathcal{D}\subset\mathcal{E}}) \ : \ \mathcal{D} \to \mathbb{R}\times\mathbb{R}\times\mathcal{E}$ and we have

$$\nabla_x(f, g, 1_{\mathcal{D}\subset\mathcal{E}}) = (\nabla_x f, \nabla_x g, \mathbf{1}_{\mathcal{V}}) \tag{610.20}$$

for all $x \in \mathcal{D}$. (See Prop.2 of Sect.63). By (610.14) and (610.16) we have $\mathbf{k} = \mathbf{m} \circ (f, g, \mathbf{1}_{\mathcal{D}\subset\mathcal{E}})|^{\overline{\overline{\mathcal{D}}}}$. Therefore, by the General Chain Rule of Sect.63, $\mathbf{k}$ is of class C^1 and we have

$$\nabla_x\mathbf{k} = (\nabla_{(f(x),g(x),x)}\mathbf{m})\nabla_x(f, g, \mathbf{1}_{\mathcal{D}\subset\mathcal{E}})$$

for all $x \in \mathcal{D}$. Using (610.19) and (610.20), we obtain (610.15). ∎

611 Curl, Symmetry of Second Gradients

In this section, we assume that flat spaces $\mathcal{E}$ and $\mathcal{E}'$, with translation spaces $\mathcal{V}$ and $\mathcal{V}'$, and an open subset $\mathcal{D}$ of $\mathcal{E}$ are given.

If $\mathbf{H} : \mathcal{D} \to \mathrm{Lin}(\mathcal{V}, \mathcal{V}')$ is differentiable, we can apply the identification (24.1) to the codomain of the gradient

$$\nabla\mathbf{H} \ : \ \mathcal{D} \to \mathrm{Lin}\big(\mathcal{V}, \mathrm{Lin}(\mathcal{V}, \mathcal{V}')\big) \cong \mathrm{Lin}_2(\mathcal{V}^2, \mathcal{V}'),$$

and hence we can apply the switching defined by (24.4) to the values of $\nabla\mathbf{H}$.

Definition 1: *The* **curl** *of a differentiable mapping* $\mathbf{H} : \mathcal{D} \to \mathrm{Lin}(\mathcal{V}, \mathcal{V}')$ *is the mapping* $\mathrm{Curl}\,\mathbf{H} : \mathcal{D} \to \mathrm{Lin}_2(\mathcal{V}^2, \mathcal{V}')$ *defined by*

$$(\mathrm{Curl}\,\mathbf{H})(x) := \nabla_x\mathbf{H} - (\nabla_x\mathbf{H})^{\sim} \quad \text{for all} \quad x \in \mathcal{D}. \tag{611.1}$$

The values of $\mathrm{Curl}\,\mathbf{H}$ are skew, i.e. $\mathrm{Rng}\,\mathrm{Curl}\,\mathbf{H} \subset \mathrm{Skew}_2(\mathcal{V}^2, \mathcal{V}')$.

The following result deals with conditions that are sufficient or necessary for the curl to be zero.

Curl-Gradient Theorem: *Assume that* $\mathbf{H} : \mathcal{D} \to \mathrm{Lin}(\mathcal{V}, \mathcal{V}')$ *is of class* C^1. *If* $\mathbf{H} = \nabla\varphi$ *for some* $\varphi : \mathcal{D} \to \mathcal{E}'$ *then* $\mathrm{Curl}\,\mathbf{H} = \mathbf{0}$. *Conversely, if* $\mathcal{D}$ *is convex and* $\mathrm{Curl}\,\mathbf{H} = \mathbf{0}$ *then* $\mathbf{H} = \nabla\varphi$ *for some* $\varphi \ : \ \mathcal{D} \to \mathcal{E}'$.

The proof will be based on the following

Lemma: *Assume that* $\mathcal{D}$ *is convex. Let* $q \in \mathcal{D}$ *and* $q' \in \mathcal{E}'$ *be given and let* $\varphi \ : \ \mathcal{D} \to \mathcal{E}'$ *be defined by*

$$\varphi(q + \mathbf{v}) := q' + \left(\int_0^1 \mathbf{H}(q + s\mathbf{v})\mathrm{d}s\right)\mathbf{v} \quad \text{for all} \quad \mathbf{v} \in \mathcal{D} - q. \tag{611.2}$$

Then φ *is of class* C^1 *and we have*

$$\nabla\varphi(q + \mathbf{v}) = \mathbf{H}(q + \mathbf{v}) - \left(\int_0^1 s(\mathrm{Curl}\,\mathbf{H})(q + s\mathbf{v})\mathrm{d}s\right)\mathbf{v} \tag{611.3}$$

for all $\mathbf{v} \in \mathcal{D} - q$.

Proof: We define $\overline{\mathcal{D}} \subset \mathbb{R} \times \mathcal{V}$ by

$$\overline{\mathcal{D}} := \{(s, \mathbf{u}) \mid \mathbf{u} \in \mathcal{D} - q, \ q + s\mathbf{u} \in \mathcal{D}\}.$$

Since $\mathcal{D}$ is open and convex, it follows that, for each $\mathbf{u} \in \mathcal{D} - q$, $\overline{\mathcal{D}}_{(\bullet,\mathbf{u})} := \{s \in \mathbb{R} \mid q + s\mathbf{u} \in \mathcal{D}\}$ is an open interval that includes $[0, 1]$. Since $\mathbf{H}$ is of class C^1, so is

$$((s, \mathbf{u}) \mapsto \mathbf{H}(q + s\mathbf{u})) \ : \ \overline{\mathcal{D}} \to \mathrm{Lin}(\mathcal{V}, \mathcal{V}');$$

hence we can apply the Differentiation Theorem for Integral Representations of Sect.610 to conclude that $\mathbf{u} \mapsto \int_0^1 \mathbf{H}(q+s\mathbf{u})\mathrm{d}s$ is of class C^1 and that its gradient is given by

$$\nabla_{\mathbf{v}}\left(\mathbf{u} \mapsto \int_0^1 \mathbf{H}(q+s\mathbf{u})\mathrm{d}s\right) = \int_0^1 s\nabla\mathbf{H}(q+s\mathbf{v})\mathrm{d}s$$

for all $\mathbf{v} \in \mathcal{D} - q$. By the Product Rule (66.7), it follows that the mapping φ defined by (611.2) is of class C^1 and that its gradient is given by

$$(\nabla_{q+\mathbf{v}}\varphi)\mathbf{u} = \left(\int_0^1 s\nabla\mathbf{H}(q+s\mathbf{v})\mathbf{u}\mathrm{d}s\right)\mathbf{v} + \left(\int_0^1 \mathbf{H}(q+s\mathbf{v})\mathrm{d}s\right)\mathbf{u}$$

for all $\mathbf{v} \in \mathcal{D} - q$ and all $\mathbf{u} \in \mathcal{V}$. Applying the definitions (24.2), (24.4), and (611.1), and using the linearity of the switching and Prop.1 of Sect.610, we obtain

$$\begin{aligned}(\nabla_{q+\mathbf{v}}\varphi)\mathbf{u} \quad &= \quad \left(\int_0^1 s(\mathrm{Curl}\,\mathbf{H})(q+s\mathbf{v})\mathrm{d}s\right)(\mathbf{u},\mathbf{v}) \\ &\quad + \left(\int_0^1 (s(\nabla\mathbf{H}(q+s\mathbf{v}))\mathbf{v} + \mathbf{H}(q+s\mathbf{v}))\mathrm{d}s\right)\mathbf{u}\end{aligned} \tag{611.4}$$

for all $\mathbf{v} \in \mathcal{D} - q$ and all $\mathbf{u} \in \mathcal{V}$. By the Product Rule (66.10) and the Chain Rule we have, for all $s \in [0,1]$,

$$\partial_s(t \mapsto t\mathbf{H}(q+t\mathbf{v})) = \mathbf{H}(q+s\mathbf{v}) + s(\nabla\mathbf{H}(q+s\mathbf{v}))\mathbf{v}.$$

Hence, by the Fundamental Theorem of Calculus, the second integral on the right side of (611.4) reduces to $\mathbf{H}(q+\mathbf{v})$. Therefore, since $\mathrm{Curl}\,\mathbf{H}$ has skew values and since $\mathbf{u} \in \mathcal{V}$ was arbitrary, (611.4) reduces to (611.3). ■

Proof of the Theorem: Assume, first, that $\mathcal{D}$ is convex and that $\mathrm{Curl}\,\mathbf{H} = \mathbf{0}$. Choose $q \in \mathcal{D}$, $q' \in \mathcal{E}'$ and define $\varphi : \mathcal{D} \to \mathcal{E}'$ by (611.2). Then $\mathbf{H} = \nabla\varphi$ by (611.3).

Conversely, assume that $\mathbf{H} = \nabla\varphi$ for some $\varphi : \mathcal{D} \to \mathcal{E}'$. Let $q \in \mathcal{D}$ be given. Since $\mathcal{D}$ is open, we can choose a convex open neighborhood $\mathcal{N}$ of q with $\mathcal{N} \subset \mathcal{D}$. Let $\mathbf{v} \in \mathcal{N} - q$ by given. Since $\mathcal{N}$ is convex, we have $q + s\mathbf{v} \in \mathcal{N}$ for all $s \in [0,1]$, and hence we can apply the Fundamental Theorem of Calculus to the derivative of $s \mapsto \varphi(q+s\mathbf{v})$, with the result

$$\varphi(q+\mathbf{v}) = \varphi(q) + \left(\int_0^1 \mathbf{H}(q+s\mathbf{v})\mathrm{d}s\right)\mathbf{v}.$$

Since $\mathbf{v} \in \mathcal{N} - q$ was arbitrary, we see that the hypotheses of the Lemma are satisfied for $\mathcal{N}$ instead of $\mathcal{D}$, $q' := \varphi(q)$, $\varphi|_{\mathcal{N}}$ instead of φ, and $\mathbf{H}|_{\mathcal{N}} = \nabla\varphi|_{\mathcal{N}}$ instead of $\mathbf{H}$. Hence, by (611.3), we have

$$\left(\int_0^1 s(\operatorname{Curl}\mathbf{H})(q + s\mathbf{v})\mathrm{d}s\right)\mathbf{v} = \mathbf{0} \tag{611.5}$$

for all $\mathbf{v} \in \mathcal{N} - q$. Now let $\mathbf{u} \in \mathcal{V}$ be given. Since $\mathcal{N} - q$ is a neighborhood of $\mathbf{0} \in \mathcal{V}$, there is a $\delta \in \mathbb{P}^\times$ such that $t\mathbf{u} \in \mathcal{N} - q$ for all $t \in]0, \delta]$. Hence, substituting $t\mathbf{u}$ for $\mathbf{v}$ in (611.5) and dividing by t gives

$$\left(\int_0^1 s(\operatorname{Curl}\mathbf{H})(q + st\mathbf{u})\mathrm{d}s\right)\mathbf{u} = \mathbf{0} \quad \text{for all} \quad t \in]0, \delta].$$

since $\operatorname{Curl}\mathbf{H}$ is continuous, we can apply Prop.3 of Sect.610 and, in the limit $t \to 0$, obtain $((\operatorname{Curl}\mathbf{H})(q))\mathbf{u} = \mathbf{0}$. Since $\mathbf{u} \in \mathcal{V}$ and $q \in \mathcal{D}$ were arbitrary, we conclude that $\operatorname{Curl}\mathbf{H} = \mathbf{0}$. ■

Remark 1: In the second part of the Theorem, the condition that $\mathcal{D}$ be convex can be replaced by the weaker one that $\mathcal{D}$ be "simply connected". This means, intuitively, that every closed curve in $\mathcal{D}$ can be continuously shrunk entirely within $\mathcal{D}$ to a point. ■

Since $\operatorname{Curl}\nabla\varphi = \nabla^{(2)}\varphi - (\nabla^{(2)}\varphi)^\sim$ by (611.1), we can restate the first part of the Curl-Gradient Theorem as follows:

Theorem on Symmetry of Second Gradients: *If* $\varphi : \mathcal{D} \to \mathcal{E}'$ *is of class* C^2*, then its second gradient* $\nabla^{(2)}\varphi : \mathcal{D} \to \operatorname{Lin}_2(\mathcal{V}^2, \mathcal{V}')$ *has symmetric values, i.e.* $\operatorname{Rng}\nabla^{(2)}\varphi \subset \operatorname{Sym}_2(\mathcal{V}^2, \mathcal{V}')$.

Remark 2: The assertion that $\nabla_x(\nabla\varphi)$ is symmetric for a given $x \in \mathcal{D}$ remains valid if one merely assumes that φ is differentiable and that $\nabla\varphi$ is differentiable at x. A direct proof of this fact, based on the results of Sect.64, is straightforward although somewhat tedious. ■

We assume now that $\mathcal{E} := \mathcal{E}_1 \times \mathcal{E}_2$ is the set-product of flat spaces $\mathcal{E}_1, \mathcal{E}_2$ with translation spaces $\mathcal{V}_1, \mathcal{V}_2$, respectively. Assume that $\varphi : \mathcal{D} \to \mathcal{E}'$ is twice differentiable. Using Prop.1 of Sect.65 repeatedly, it is easily seen that

$$\begin{aligned}\nabla^{(2)}\varphi(x) = &\\ &(\nabla_{(1)}\nabla_{(1)}\varphi)(x)(\mathrm{ev}_1, \mathrm{ev}_1) + (\nabla_{(1)}\nabla_{(2)}\varphi)(x)(\mathrm{ev}_1, \mathrm{ev}_2) + \\ &(\nabla_{(2)}\nabla_{(1)}\varphi)(x)(\mathrm{ev}_2, \mathrm{ev}_1) + (\nabla_{(2)}\nabla_{(2)}\varphi)(x)(\mathrm{ev}_2, \mathrm{ev}_2)\end{aligned} \tag{611.6}$$

for all $x \in \mathcal{D}$, where ev_1 and ev_2 are the evaluation mappings from $\mathcal{V} := \mathcal{V}_1 \times \mathcal{V}_2$ to $\mathcal{V}_1$ and $\mathcal{V}_2$, respectively (see Sect.04). Evaluation of (611.6) at

$(\mathbf{u},\mathbf{v}) = ((\mathbf{u}_1,\mathbf{u}_2),(\mathbf{v}_1,\mathbf{v}_2)) \in \mathcal{V}^2$ gives

$$\begin{aligned}\nabla^{(2)}\varphi(x)(\mathbf{u},\mathbf{v}) = \\ (\nabla_{(1)}\nabla_{(1)}\varphi)(x)(\mathbf{u}_1,\mathbf{v}_1) + (\nabla_{(1)}\nabla_{(2)}\varphi)(x)(\mathbf{u}_1,\mathbf{v}_2) + \\ (\nabla_{(2)}\nabla_{(1)}\varphi)(x)(\mathbf{u}_2,\mathbf{v}_1) + (\nabla_{(2)}\nabla_{(2)}\varphi)(x)(\mathbf{u}_2,\mathbf{v}_2).\end{aligned} \tag{611.7}$$

The following is a corollary to the preceding theorem.

Theorem on the Interchange of Partial Gradients: *Assume that $\mathcal{D}$ is an open subset of a product space $\mathcal{E} := \mathcal{E}_1 \times \mathcal{E}_2$ and that $\varphi : \mathcal{D} \to \mathcal{E}'$ is of class C^2. Then*

$$\nabla_{(1)}\nabla_{(2)}\varphi = (\nabla_{(2)}\nabla_{(1)}\varphi)^{\sim}, \tag{611.8}$$

where the operation $\sim$ is to be understood as value-wise switching.

Proof: Let $x \in \mathcal{D}$ and $\mathbf{w} \in \mathcal{V}_1 \times \mathcal{V}_2$ be given. If we apply (611.7) to the case when $\mathbf{u} := (\mathbf{w}_1, \mathbf{0})$, $\mathbf{v} := (\mathbf{0}, \mathbf{w}_2)$ and then with $\mathbf{u}$ and $\mathbf{v}$ interchanged we obtain

$$\nabla^{(2)}\varphi(x)(\mathbf{u},\mathbf{v}) = (\nabla_{(1)}\nabla_{(2)}\varphi)(x)(\mathbf{w}_1,\mathbf{w}_2),$$

$$\nabla^{(2)}\varphi(x)(\mathbf{v},\mathbf{u}) = (\nabla_{(2)}\nabla_{(1)}\varphi)(x)(\mathbf{w}_2,\mathbf{w}_1).$$

Since $\mathbf{w} \in \mathcal{V}_1 \times \mathcal{V}_2$ was arbitrary, the symmetry of $\nabla^{(2)}\varphi(x)$ gives (611.8). ■

Corollary: *Let I be a finite index set and let $\mathcal{D}$ be an open subset of $\mathbb{R}^I$. If $\varphi : \mathcal{D} \to \mathcal{E}'$ is of class C^2, then the second partial derivatives $\varphi_{,i\,,k} : \mathcal{D} \to \mathcal{V}'$ satisfy*

$$\varphi_{,i\,,k} = \varphi_{,k\,,i} \quad \text{for all} \quad i,k \in I. \tag{611.9}$$

Remark 3: Assume $\mathcal{D}$ is an open subset of a Euclidean space $\mathcal{E}$ with translation space $\mathcal{V}$. A mapping $\mathbf{h} : \mathcal{D} \to \mathcal{V} \cong \mathcal{V}^*$ is then called a vector-field (see Sect.71). If $\mathbf{h}$ is differentiable, then the range of $\operatorname{Curl}\mathbf{h}$ is included in $\operatorname{Skew}_2(\mathcal{V}^2,\mathbb{R}) \cong \operatorname{Skew}\mathcal{V}$. If $\mathcal{V}$ is 3-dimensional, there is a natural doubleton of orthogonal isomorphisms from $\operatorname{Skew}\mathcal{V}$ to $\mathcal{V}$, as will be explained in Vol.II. If one of these two isomorphisms, say $\mathbf{V} \in \operatorname{Orth}(\operatorname{Skew}\mathcal{V},\mathcal{V})$, is singled out, we may consider the **vector-curl** $\operatorname{curl}\mathbf{h} := \mathbf{V}(\operatorname{Curl}\mathbf{h})|^{\operatorname{Skew}\mathcal{V}}$ of the vector field $\mathbf{h}$ (Note the lower-case "c"). This vector-curl rather than the curl of Def.1 is used in much of the literature. ■

Notes 611

(1) In some of the literature on "Vector Analysis", the notation $\nabla \times \mathbf{h}$ instead of $\operatorname{curl}\mathbf{h}$ is used for the vector-curl of $\mathbf{h}$, which is explained in the Remark above. This notation should be avoided for the reason mentioned in Note (1) to Sect.67.

612 Lineonic Exponentials

We note the following generalization of a familiar result of elementary calculus.

Proposition 1: *Let* $\mathcal{E}$ *be a flat space with translation space* $\mathcal{V}$ *and let* I *be an interval. Let* $\mathbf{g}$ *be a sequence of continuous processes in* $\text{Map}\,(I, \mathcal{V})$ *that converges locally uniformly to* $\mathbf{h} \in \text{Map}\,(I, \mathcal{V})$*. Let* $a \in I$ *and* $q \in \mathcal{E}$ *be given and define the sequence* z *in* $\text{Map}\,(I, \mathcal{E})$ *by*

$$z_n(t) := q + \int_a^t \mathbf{g}_n \quad \text{for all} \quad n \in \mathbb{N}^\times \quad \text{and} \quad t \in I. \tag{612.1}$$

Then $\mathbf{h}$ *is continuous and* z *converges locally uniformly to the process* $p : I \to \mathcal{E}$ *defined by*

$$p(t) := q + \int_a^t \mathbf{h} \quad \text{for all} \quad t \in I. \tag{612.2}$$

Proof: The continuity of $\mathbf{h}$ follows from the Theorem on Continuity of Uniform Limits of Sect.56. Hence (612.2) is meaningful. Let ν be a norm on $\mathcal{V}$. By Prop.2 of Sect.610 and (612.1) and (612.2) we have

$$\nu(z_n(t) - p(t)) = \nu\left(\int_a^t (\mathbf{g}_n - \mathbf{h})\right) \le \left|\int_a^t \nu \circ (\mathbf{g}_n - \mathbf{h})\right| \tag{612.3}$$

for all $n \in \mathbb{N}^\times$ and all $t \in I$. Now let $s \in I$ be given. We can choose a compact interval $[b, c] \subset I$ such that $a \in [b, c]$ and such that $[b, c]$ is a neightborhood of s relative to I (see Sect.56). By Prop.7 of Sect.58, $\mathbf{g}|_{[b,c]}$ converges uniformly to $\mathbf{h}|_{[b,c]}$. Now let $\epsilon \in \mathbb{P}^\times$ be given. By Prop.7 of Sect.55 we can determine $m \in \mathbb{N}^\times$ such that

$$\nu \circ (\mathbf{g}_n - \mathbf{h})|_{[b,c]} < \frac{\epsilon}{c - b} \quad \text{for all} \quad n \in m + \mathbb{N}.$$

Hence, by (612.3), we have

$$\nu(z_n(t) - p(t)) < |t - a| \frac{\epsilon}{c - b} < \epsilon \quad \text{for all} \quad n \in m + \mathbb{N}$$

and all $t \in [b, c]$. Since $\varepsilon \in \mathbb{P}^\times$ was arbitrary, we can use Prop.7 of Sect.55 again to conclude that $z|_{[b,c]}$ converges uniformly to $p|_{[b,c]}$. Since $s \in I$ was arbitrary and since $[b, c] \in \text{Nhd}_s(I)$, the conclusion follows. ■

From now on we assume that a linear space $\mathcal{V}$ is given and we consider the algebra $\text{Lin}\mathcal{V}$ of lineons on $\mathcal{V}$ (see Sect.18).

Proposition 2: *The sequence* $\mathbf{S}$ *in* Map (Lin$\mathcal{V}$, Lin$\mathcal{V}$) *defined by*

$$\mathbf{S}_n(\mathbf{L}) := \sum_{k \in n^{[}} \frac{1}{k!}\mathbf{L}^k \tag{612.4}$$

for all $\mathbf{L} \in \text{Lin}\mathcal{V}$ *and all* $n \in \mathbb{N}^\times$ *converges locally uniformly. Its limit is called the* **(lineonic) exponential** *for* $\mathcal{V}$ *and is denoted by* $\exp_{\mathcal{V}} : \text{Lin}\mathcal{V} \to \text{Lin}\mathcal{V}$. *The exponential* $\exp_{\mathcal{V}}$ *is continuous.*

Proof: We choose a norm ν on $\mathcal{V}$. Using Prop.6 of Sect.52 and induction, we see that $\|\mathbf{L}^k\|_\nu \leq \|\mathbf{L}\|_\nu{}^k$ for all $k \in \mathbb{N}$ and all $\mathbf{L} \in \text{Lin}\mathcal{V}$ (see (52.9)). Hence, given $\sigma \in \mathbb{P}^\times$, we have

$$\|\frac{1}{k!}\mathbf{L}^k\|_\nu \leq \frac{\sigma^k}{k!}$$

for all $k \in \mathbb{N}$ and all $\mathbf{L} \in \sigma\text{Ce}(\| \bullet \|_\nu)$. Since the sum-sequence of $(\frac{\sigma^k}{k!} \mid k \in \mathbb{N})$ converges (to e^σ), we can use Prop.9 of Sect.55 to conclude that the restriction of the sum-sequence $\mathbf{S}$ of $(\frac{1}{k!}\mathbf{L}^k \mid k \in \mathbb{N})$ to $\sigma\text{Ce}(\| \bullet \|_\nu)$ converges uniformly. Now, given $\mathbf{L} \in \text{Lin}\mathcal{V}$, $\sigma\text{Ce}(\| \bullet \|_\nu)$ is a neighborhood of $\mathbf{L}$ if $\sigma > \|\mathbf{L}\|_\nu$. Therefore $\mathbf{S}$ converges locally uniformly. The continuity of its limit $\exp_{\mathcal{V}}$ follows from the Continuity Theorem for Uniform Limits. ■

For later use we need the following:

Proposition 3: *Let* $\mathbf{L} \in \text{Lin}\mathcal{V}$ *be given. Then the constant zero is the only differentiable process* $\mathbf{D} : \mathbb{R} \to \text{Lin}\mathcal{V}$ *that satisfies*

$$\mathbf{D}^\bullet = \mathbf{L}\mathbf{D} \quad \text{and} \quad \mathbf{D}(0) = \mathbf{0}. \tag{612.5}$$

Proof: By the Fundamental Theorem of Calculus, (612.5) is equivalent to

$$\mathbf{D}(t) = \int_0^t (\mathbf{L}\mathbf{D}) \quad \text{for all} \quad t \in \mathbb{R}. \tag{612.6}$$

Choose a norm ν on $\mathcal{V}$. Using Prop.2 of Sect.610 and Prop.6 of Sect.52, we conclude from (612.6) that

$$\|\mathbf{D}(t)\|_\nu \leq \int_0^t \|\mathbf{L}\mathbf{D}(s)\|_\nu \text{d}s \leq \|\mathbf{L}\|_\nu \int_0^t \|\mathbf{D}(s)\|_\nu \text{d}s,$$

i.e., with the abbreviations

$$\sigma(t) := \|\mathbf{D}(t)\|_\nu \quad \text{for all} \quad t \in \mathbb{P}, \quad \kappa := \|\mathbf{L}\|_\nu, \tag{612.7}$$

that

$$0 \leq \sigma(t) \leq \kappa \int_0^t \sigma \quad \text{for all} \quad t \in \mathbb{P}. \tag{612.8}$$

We define $\varphi : \mathbb{P} \to \mathbb{R}$ by

$$\varphi(t) := \mathrm{e}^{-\kappa t} \int_0^t \sigma \quad \text{for all} \quad t \in \mathbb{P}. \tag{612.9}$$

Using elementary calculus, we infer from (612.9) and (612.8) that

$$\varphi^\bullet(t) = -\kappa \mathrm{e}^{-\kappa t} \int_0^t \sigma + \mathrm{e}^{-\kappa t}\sigma(t) \leq 0$$

for all $t \in \mathbb{P}$ and hence that φ is antitone. On the other hand, it is evident from (612.9) that $\varphi(0) = 0$ and $\varphi \geq 0$. This can happen only if $\varphi = 0$, which, in turn, can happen only if $\sigma(t) = 0$ for all $t \in \mathbb{P}$ and hence, by (612.7), if $\mathbf{D}(t) = \mathbf{0}$ for all $t \in \mathbb{P}$. To show that $\mathbf{D}(t) = \mathbf{0}$ for all $t \in -\mathbb{P}$, one need only replace $\mathbf{D}$ by $\mathbf{D} \circ (-\iota)$ and $\mathbf{L}$ by $-\mathbf{L}$ in (612.5) and then use the result just proved. ∎

Proposition 4: *Let* $\mathbf{L} \in \mathrm{Lin}\mathcal{V}$ *be given. Then the only differentiable process* $\mathbf{E} : \mathbb{R} \to \mathrm{Lin}\mathcal{V}$ *that satisfies*

$$\mathbf{E}^\bullet = \mathbf{L}\mathbf{E} \quad \text{and} \quad \mathbf{E}(0) = \mathbf{1}_\mathcal{V} \tag{612.10}$$

is the one given by

$$\mathbf{E} := \exp_\mathcal{V} \circ (\iota \mathbf{L}). \tag{612.11}$$

Proof: We consider the sequence $\mathbf{G}$ in $\mathrm{Map}\,(\mathbb{R}, \mathrm{Lin}\mathcal{V})$ defined by

$$\mathbf{G}_n(t) := \sum_{k \in n^[} \frac{t^k}{k!} \mathbf{L}^{k+1} \tag{612.12}$$

for all $n \in \mathbb{N}^\times$ and all $t \in \mathbb{R}$. Since, by (612.4), $\mathbf{G}_n(t) = \mathbf{L}\mathbf{S}_n(t\mathbf{L})$ for all $n \in \mathbb{N}^\times$ and all $t \in \mathbb{R}$, it follows from Prop.2 that $\mathbf{G}$ converges locally uniformly to $\mathbf{L}\exp \circ (\iota \mathbf{L}) : \mathbb{R} \to \mathrm{Lin}\mathcal{V}$. On the other hand, it follows from (612.12) that

$$\mathbf{1}_\mathcal{V} + \int_0^t \mathbf{G}_n = \mathbf{1}_\mathcal{V} + \sum_{k \in n^[} \frac{t^{k+1}}{(k+1)!} \mathbf{L}^{k+1} = \mathbf{S}_{n+1}(t\mathbf{L})$$

for all $n \in \mathbb{N}^\times$ and all $t \in \mathbb{R}$. Applying Prop.1 to the case when $\mathcal{E}$ and $\mathcal{V}$ are replaced by $\mathrm{Lin}\mathcal{V}$, I by $\mathbb{R}$, $\mathbf{g}$ by $\mathbf{G}$, a by 0, and q by $\mathbf{1}_\mathcal{V}$, we conclude that

$$\exp_\mathcal{V}(t\mathbf{L}) = \mathbf{1}_\mathcal{V} + \int_0^t (\mathbf{L} \exp_\mathcal{V} \circ (\iota \mathbf{L})) \quad \text{for all} \quad t \in \mathbb{R}, \tag{612.13}$$

which, by the Fundamental Theorem of Calculus, is equivalent to the assertion that (612.10) holds when $\mathbf{E}$ is defined by (612.11).

The uniqueness of $\mathbf{E}$ is an immediate consequence of Prop.3. ∎

Proposition 5: *Let* $\mathbf{L} \in \mathrm{Lin}\mathcal{V}$ *and a continuous process* $\mathbf{F} : \mathbb{R} \to \mathrm{Lin}\mathcal{V}$ *be given. Then the only differentiable process* $\mathbf{D} : \mathbb{R} \to \mathrm{Lin}\mathcal{V}$ *that satisfies*

$$\mathbf{D}^\bullet = \mathbf{L}\mathbf{D} + \mathbf{F} \quad \text{and} \quad \mathbf{D}(0) = \mathbf{0} \tag{612.14}$$

is the one given by

$$\mathbf{D}(t) = \int_0^t \mathbf{E}(t-s)\mathbf{F}(s)\mathrm{d}s \quad \text{for all} \quad t \in \mathbb{R}, \tag{612.15}$$

where $\mathbf{E} : \mathbb{R} \to \mathrm{Lin}\mathcal{V}$ *is given by* (612.11).

Proof: Let $\mathbf{D}$ be defined by (612.15). Using the Corollary to the Differentiation Theorem for Integral Representations of Sect.610, we see that $\mathbf{D}$ is of class C^1 and that $\mathbf{D}^\bullet$ is given by

$$\mathbf{D}^\bullet(t) = \mathbf{E}(0)\mathbf{F}(t) + \int_0^t \mathbf{E}^\bullet(t-s)\mathbf{F}(s)\mathrm{d}s \quad \text{for all} \quad t \in \mathbb{R}.$$

Using (612.10) and Prop.1 of Sect.610, and then (612.15), we find that (612.14) is satisfied.

The uniqueness of $\mathbf{D}$ is again an immediate consequence of Prop.3. ∎

Differentiation Theorem for Lineonic Exponentials: *Let* $\mathcal{V}$ *be a linear space. The exponential* $\exp_{\mathcal{V}} : \mathrm{Lin}\mathcal{V} \to \mathrm{Lin}\mathcal{V}$ *is of class* C^1 *and its gradient is given by*

$$(\nabla_{\mathbf{L}} \exp_{\mathcal{V}})\mathbf{M} = \int_0^1 \exp_{\mathcal{V}}(s\mathbf{L})\mathbf{M}\exp_{\mathcal{V}}((1-s)\mathbf{L})\mathrm{d}s \tag{612.16}$$

for all $\mathbf{L}, \mathbf{M} \in \mathrm{Lin}\mathcal{V}$.

Proof: Let $\mathbf{L}, \mathbf{M} \in \mathrm{Lin}\mathcal{V}$ be given. Also, let $r \in \mathbb{R}^\times$ be given and put

$$\mathbf{K} := \mathbf{L} + r\mathbf{M}, \quad \mathbf{E}_1 := \exp_{\mathcal{V}} \circ (\iota\mathbf{L}), \quad \mathbf{E}_2 := \exp_{\mathcal{V}} \circ (\iota\mathbf{K}). \tag{612.17}$$

By Prop.4 we have $\mathbf{E}_1{}^\bullet = \mathbf{L}\mathbf{E}_1$, $\mathbf{E}_2{}^\bullet = \mathbf{K}\mathbf{E}_2$, and $\mathbf{E}_1(0) = \mathbf{E}_2(0) = \mathbf{1}_{\mathcal{V}}$. Taking the difference and observing $(612.17)_1$ we see that $\mathbf{D} := \mathbf{E}_2 - \mathbf{E}_1$ satisfies

$$\mathbf{D}^\bullet = \mathbf{K}\mathbf{E}_2 - \mathbf{L}\mathbf{E}_1 = \mathbf{L}\mathbf{D} + r\mathbf{M}\mathbf{E}_2, \quad \mathbf{D}(0) = \mathbf{0}.$$

Hence, if we apply Prop.5 with the choice $\mathbf{F} := r\mathbf{M}\mathbf{E}_2$ we obtain

$$(\mathbf{E}_2 - \mathbf{E}_1)(t) = r\int_0^t \mathbf{E}_1(t-s)\mathbf{M}\mathbf{E}_2(s)\mathrm{d}s \quad \text{for all} \quad t \in \mathbb{R}.$$

For $t := 1$ we obtain, in view of (612.17),

$$\frac{1}{r}(\exp_{\mathcal{V}}(\mathbf{L}+r\mathbf{M}) - \exp_{\mathcal{V}}(\mathbf{L})) = \int_0^1 (\exp_{\mathcal{V}}((1-s)\mathbf{L}))\mathbf{M}\exp_{\mathcal{V}}(s(\mathbf{L}+r\mathbf{M}))\mathrm{d}s.$$

Since $\exp_{\mathcal{V}}$ is continuous by Prop.2, and since bilinear mappings are continuous (see Sect.66), we see that the integrand on the right depends continuously on $(s,r) \in \mathbb{R}^2$. Hence, by Prop.3 of Sect.610 and by (65.13), in the limit $r \to 0$ we find

$$(\mathrm{dd}_{\mathbf{M}}\exp_{\mathcal{V}})(\mathbf{L}) = \int_0^1 (\exp_{\mathcal{V}}((1-s)\mathbf{L}))\mathbf{M}\exp_{\mathcal{V}}(s\mathbf{L})\mathrm{d}s. \tag{612.18}$$

Again, we see that the integrand on the right depends continuously on $(s,\mathbf{L}) \in \mathbb{R}\times\mathrm{Lin}\mathcal{V}$. Hence, using Prop.3 of Sect.610 again, we conclude that the mapping $\mathrm{dd}_{\mathbf{M}}\exp_{\mathcal{V}} : \mathrm{Lin}\mathcal{V} \to \mathrm{Lin}\mathcal{V}$ is continuous. Since $\mathbf{M} \in \mathrm{Lin}\mathcal{V}$ was arbitrary, it follows from Prop.7 of Sect.65 that $\exp_{\mathcal{V}}$ is of class C^1. The formula (612.16) is the result of combining (65.14) with (612.18). ■

Proposition 6: *Assume that* $\mathbf{L},\mathbf{M} \in \mathrm{Lin}\mathcal{V}$ *commute, i.e. that* $\mathbf{LM} = \mathbf{ML}$. *Then:*

(i) $\mathbf{M}$ *and* $\exp_{\mathcal{V}}(\mathbf{L})$ *commute.*

(ii) *We have*

$$\exp_{\mathcal{V}}(\mathbf{L}+\mathbf{M}) = (\exp_{\mathcal{V}}(\mathbf{L}))(\exp_{\mathcal{V}}(\mathbf{M})). \tag{612.19}$$

(iii) *We have*

$$(\nabla_{\mathbf{L}}\exp_{\mathcal{V}})\mathbf{M} = (\exp_{\mathcal{V}}(\mathbf{L}))\mathbf{M}. \tag{612.20}$$

Proof: Put $\mathbf{E} := \exp_v \circ(\iota\mathbf{L})$ and $\mathbf{D} := \mathbf{EM} - \mathbf{ME}$. By Prop.4, we find $\mathbf{D}^\bullet = \mathbf{E}^\bullet\mathbf{M} - \mathbf{ME}^\bullet = \mathbf{LEM} - \mathbf{MLE}$. Hence, since $\mathbf{LM} = \mathbf{ML}$, we obtain $\mathbf{D}^\bullet = \mathbf{LD}$. Since $\mathbf{D}(0) = \mathbf{1}_{\mathcal{V}}\mathbf{M} - \mathbf{M}\mathbf{1}_{\mathcal{V}} = \mathbf{0}$, it follows from Prop.3 that $\mathbf{D} = \mathbf{0}$ and hence $\mathbf{D}(1) = \mathbf{0}$, which proves (i).

Now put $\mathbf{F} := \exp_v \circ(\iota\mathbf{M})$. By the Product Rule (66.13) and by Prop.4, we find

$$(\mathbf{EF})^\bullet = \mathbf{E}^\bullet\mathbf{F} + \mathbf{EF}^\bullet = \mathbf{LEF} + \mathbf{EMF}.$$

Since $\mathbf{EM} = \mathbf{ME}$ by part (i), we obtain $(\mathbf{EF})^\bullet = (\mathbf{L}+\mathbf{M})\mathbf{EF}$. Since $(\mathbf{EF})(0) = \mathbf{E}(0)\mathbf{F}(0) = \mathbf{1}_{\mathcal{V}}$, by Prop.4, $\mathbf{EF} = \exp_v \circ(\iota(\mathbf{L}+\mathbf{M}))$. Evaluation at 1 gives (612.19), which proves (ii).

Part (iii) is an immediate consequence of (612.16) and of (i) and (ii). ■

Pitfalls: Of course, when $\mathcal{V} = \mathbb{R}$, then $\mathrm{Lin}\mathcal{V} = \mathrm{Lin}\mathbb{R} \cong \mathbb{R}$ and the lineonic exponential reduces to the ordinary exponential exp. Prop.6 shows how the rule $\exp(s+t) = \exp(s)\exp(t)$ for all $s, t \in \mathbb{R}$ and the rule $\exp^\bullet = \exp$ generalize to the case when $\mathcal{V} \neq \mathbb{R}$. The assumption, in Prop.6, that $\mathbf{L}$ and $\mathbf{M}$ commute cannot be omitted. The formulas (612.19) and (612.20) need not be valid when $\mathbf{LM} \neq \mathbf{ML}$.

If the codomain of the ordinary exponential is restricted to $\mathbb{P}^\times$ it becomes invertible and its inverse $\log : \mathbb{P}^\times \to \mathbb{R}$ is of class C^1. If $\dim\mathcal{V} > 1$, then $\exp_\mathcal{V}$ does not have differentiable local inverses near certain values of $\mathbf{L} \in \mathrm{Lin}\mathcal{V}$ because for these values of $\mathbf{L}$, the gradient $\nabla_\mathbf{L} \exp_\mathcal{V}$ fails to be injective (see Problem 10). In fact, one can prove that $\exp_\mathcal{V}$ is not locally invertible near the values of $\mathbf{L}$ in question. Therefore, there is no general lineonic analogue of the logarithm. See, however, Sect.85. ■

613 Problems for Chapter 6

(1) Let I be a genuine interval, let $\mathcal{E}$ be a flat space with translation space $\mathcal{V}$, and let $p : I \to \mathcal{E}$ be a differentiable process. Define $\mathbf{h} : I^2 \to \mathcal{V}$ by

$$\mathbf{h}(s,t) := \left\{ \begin{array}{ccc} \frac{p(s)-p(t)}{s-t} & \text{if} & s \neq t \\ p^\bullet(s) & \text{if} & s = t \end{array} \right\}. \tag{P6.1}$$

(a) Prove: If $p^\bullet$ is continuous at $t \in I$, then $\mathbf{h}$ is continuous at $(t,t) \in I^2$.

(b) Find a counterexample which shows that $\mathbf{h}$ need not be continuous if p is merely differentiable and not of class C^1.

(2) Let $f : \mathcal{D} \to \mathbb{R}$ be a function whose domain $\mathcal{D}$ is an open convex subset of a flat space $\mathcal{E}$ with translation space $\mathcal{V}$.

(a) Show: If f is differentiable and ∇f is constant, then $f = a|_\mathcal{D}$ for some flat function $a \in \mathrm{Flf}(\mathcal{E})$ (see Sect.36).

(b) Show: If f is twice differentiable and $\nabla^{(2)} f$ is constant, then f has the form

$$f = a|_\mathcal{D} + \mathbf{Q} \circ (1_{\mathcal{D}\subset\mathcal{E}} - q_{\mathcal{D}\to\mathcal{E}}), \tag{P6.2}$$

where $a \in \mathrm{Flf}(\mathcal{E})$, $q \in \mathcal{E}$, and $\mathbf{Q} \in \mathrm{Qu}(\mathcal{V})$ (see Sect. 27).

(3) Let $\mathcal{D}$ be an open subset of a flat space $\mathcal{E}$ and let $\mathcal{W}$ be a genuine inner-product space. Let $\mathbf{k} : \mathcal{D} \to \mathcal{W}^\times$ be a mapping that is differentiable at a given $x \in \mathcal{D}$.

(a) Show that the value-wise magnitude $|\mathbf{k}| : \mathcal{D} \to \mathbb{P}^\times$, defined by $|\mathbf{k}|(y) := |\mathbf{k}(y)|$ for all $y \in \mathcal{D}$, is differentiable at x and that

$$\nabla_x|\mathbf{k}| = \tfrac{1}{|\mathbf{k}(x)|}(\nabla_x\mathbf{k})^\top\mathbf{k}(x). \tag{P6.3}$$

(b) Show that $\mathbf{k}/|\mathbf{k}| : \mathcal{D} \to \mathcal{W}^\times$ is differentiable at x and that

$$\nabla_x\left(\tfrac{\mathbf{k}}{|\mathbf{k}|}\right) = \tfrac{1}{|\mathbf{k}(x)|^3}\left(|\mathbf{k}(x)|^2\nabla_x\mathbf{k} - \mathbf{k}(x)\otimes(\nabla_x\mathbf{k})^\top\mathbf{k}(x)\right). \tag{P6.4}$$

(4) Let $\mathcal{E}$ be a genuine inner-product space with translation space $\mathcal{V}$, let $q \in \mathcal{E}$ be given, and define $\mathbf{r} : \mathcal{E}\setminus\{q\} \to \mathcal{V}^\times$ by

$$\mathbf{r}(x) := x - q \quad \text{for all} \quad x \in \mathcal{E}\setminus\{q\} \tag{P6.5}$$

and put $r := |\mathbf{r}|$ (see Part (a) of Problem 3).

(a) Show that $\mathbf{r}/r : \mathcal{E}\setminus\{q\} \to \mathcal{V}^\times$ is of class C^1 and that

$$\nabla\left(\tfrac{\mathbf{r}}{r}\right) = \tfrac{1}{r^3}(r^2\mathbf{1}_\mathcal{V} - \mathbf{r}\otimes\mathbf{r}). \tag{P6.6}$$

(Hint: Use Part (b) of Problem 3.)

(b) Let the function $h : \mathbb{P}^\times \to \mathbb{R}$ be twice differentiable. Show that $h \circ r : \mathcal{E}\setminus\{q\} \to \mathbb{R}$ is twice differentiable and that

$$\nabla^{(2)}(h \circ r) = \tfrac{1}{r}\left(\tfrac{1}{r^2}(r(h^{\bullet\bullet}\circ r) - (h^\bullet\circ r))\mathbf{r}\otimes\mathbf{r} + (h^\bullet\circ r)\mathbf{1}_\mathcal{V}\right). \tag{P6.7}$$

(c) Evaluate the Laplacian $\Delta(h \circ r)$ and reconcile your result with (67.17).

(5) A Euclidean space $\mathcal{E}$ of dimension n with $n \geq 2$, a point $q \in \mathcal{E}$, and a linear space $\mathcal{W}$ are assumed given.

(a) Let I be an open interval, let $\mathcal{D}$ be an open subset of $\mathcal{E}$, let $f : \mathcal{D} \to I$ be given by (67.16), let a be a flat function on $\mathcal{E}$, let $\mathbf{g} : I \to \mathcal{W}$ be twice differentiable, and put $\mathbf{h} := a|_\mathcal{D}(\mathbf{g}\circ f) : \mathcal{D} \to \mathcal{W}$. Show that

$$\Delta\mathbf{h} = 2a|_\mathcal{D}\left((2\iota\mathbf{g}^{\bullet\bullet} + (n+2)\mathbf{g}^\bullet)\circ f\right) - 4a(q)(\mathbf{g}^\bullet\circ f). \tag{P6.8}$$

(b) Assuming that the Euclidean space $\mathcal{E}$ is genuine, show that the function $\mathbf{h} : \mathcal{E} \setminus \{q\} \to \mathcal{W}$ given by

$$\mathbf{h}(x) := ((\mathbf{e} \cdot (x - q))|x - q|^{-n})\mathbf{a} + \mathbf{b} \tag{P6.9}$$

is harmonic for all $\mathbf{e} \in \mathcal{V} := \mathcal{E} - \mathcal{E}$ and all $\mathbf{a}, \mathbf{b} \in \mathcal{W}$. (Hint: Use Part (a) with a determined by $\nabla a = \mathbf{e},\ a(q) = 0$.)

(6) Let $\mathcal{E}$ be a flat space, let $q \in \mathcal{E}$, let $\mathbf{Q}$ be a non-degenerate quadratic form (see Sect.27) on $\mathcal{V} := \mathcal{E} - \mathcal{E}$, and define $f : \mathcal{E} \to \mathbb{R}$ by

$$f(y) := \mathbf{Q}(y - q) \quad \text{for all} \quad y \in \mathcal{E}. \tag{P6.10}$$

Let a be a non-constant flat function on $\mathcal{E}$ (see Sect.36).

(a) Prove: If the restriction of f to the hyperplane $\mathcal{F} := a^{<}(\{0\})$ attains an extremum at $x \in \mathcal{F}$, then x must be given by

$$x = q - \lambda \ \overset{\sqcup}{Q}{}^{-1}(\nabla a), \quad \text{where} \quad \lambda := \frac{a(q)}{\overset{\sqcup}{\mathbf{Q}}{}^{-1}(\nabla a, \nabla a)}. \tag{P6.11}$$

(Hint: Use the Constrained-Extremum Theorem.)

(b) Under what condition does $f|_{\mathcal{F}}$ actually attain a maximum or a minimum at the point x given by $(P6.11)$?

(7) Let $\mathcal{E}$ be a 2-dimensional Euclidean space with translation-space $\mathcal{V}$ and let $\mathbf{J} \in \mathrm{Orth}\mathcal{V} \cap \mathrm{Skew}\mathcal{V}$ be given (see Problem 2 of Chapt.4). Let $\mathcal{D}$ be an open subset of $\mathcal{E}$ and let $\mathbf{h} : \mathcal{D} \to \mathcal{V}$ be a vector-field of class C^1.

(a) Prove that

$$\mathrm{Curl}\,\mathbf{h} = -(\mathrm{div}(\mathbf{J}\mathbf{h}))\mathbf{J}. \tag{P6.12}$$

(b) Assuming that $\mathcal{D}$ is convex and that $\mathrm{div}\ \mathbf{h} = 0$, prove that $\mathbf{h} = \mathbf{J}\nabla f$ for some function $f : \mathcal{D} \to \mathbb{R}$ of class C^2.

Note: If $\mathbf{h}$ is interpreted as the velocity field of a volume-preserving flow, then $\mathrm{div}\ \mathbf{h} = 0$ is valid and a function f as described in Part (b) is called a "stream-function" of the flow. ■

(8) Let a linear space $\mathcal{V}$, a lineon $\mathbf{L} \in \mathrm{Lin}\mathcal{V}$, and $\mathbf{u} \in \mathcal{V}$ be given. Prove: The only differentiable process $\mathbf{h} : \mathbb{R} \to \mathcal{V}$ that satisfies

$$\mathbf{h}^{\bullet} = \mathbf{L}\mathbf{h} \quad \text{and} \quad \mathbf{h}(0) = \mathbf{u} \tag{P6.13}$$

is the one given by

$$\mathbf{h} := (\exp_{\mathcal{V}} \circ (\iota\mathbf{L}))\mathbf{u}. \tag{P6.14}$$

(Hint: To prove existence, use Prop. 4 of Sect. 612. To prove uniqueness, use Prop.3 of Sect.612 with the choice $\mathbf{D} := \mathbf{h} \otimes \boldsymbol{\lambda}$ (value-wise), where $\boldsymbol{\lambda} \in \mathcal{V}^*$ is arbitrary.)

(9) Let a linear space $\mathcal{V}$ and a lineon $\mathbf{J}$ on $\mathcal{V}$ that satisfies $\mathbf{J}^2 = -\mathbf{1}_{\mathcal{V}}$ be given. (There are such $\mathbf{J}$ if $\dim \mathcal{V}$ is even; see Sect.89.)

(a) Show that there are functions $c : \mathbb{R} \to \mathbb{R}$ and $d : \mathbb{R} \to \mathbb{R}$, of class C^1, such that

$$\exp_{\mathcal{V}} \circ (\iota\mathbf{J}) = c\mathbf{1}_{\mathcal{V}} + d\mathbf{J}. \tag{P6.15}$$

(Hint: Apply the result of Problem 8 to the case when $\mathcal{V}$ is replaced by $\mathcal{C} := \mathrm{Lsp}(\mathbf{1}_{\mathcal{V}}, \mathbf{J}) \subset \mathrm{Lin}\mathcal{V}$ and when $\mathbf{L}$ is replaced by $\mathrm{Le}_{\mathbf{J}} \in \mathrm{Lin}\mathcal{C}$, defined by $\mathrm{Le}_{\mathbf{J}}\mathbf{U} = \mathbf{J}\mathbf{U}$ for all $\mathbf{U} \in \mathcal{C}$.)

(b) Show that the functions c and d of Part (a) satisfy

$$d^{\bullet} = c, \;\; c^{\bullet} = -d, \;\; c(0) = 1, \;\; d(0) = 0, \tag{P6.16}$$

and

$$\left.\begin{array}{l} c(t+s) = c(t)c(s) - d(t)d(s) \\ d(t+s) = c(t)d(s) + d(t)c(s) \end{array}\right\} \quad \text{for all} \quad s, t \in \mathbb{R}. \tag{P6.17}$$

(c) Show that $c = \cos$ and $d = \sin$.

Remark: One could, in fact, use Part (a) to *define* sin and cos. ■

(10) Let a linear space $\mathcal{V}$ and $\mathbf{J} \in \mathrm{Lin}\mathcal{V}$ satisfying $\mathbf{J}^2 = -\mathbf{1}_{\mathcal{V}}$ be given and put

$$\mathcal{A} := \{\mathbf{L} \in \mathrm{Lin}\mathcal{V} \mid \mathbf{L}\mathbf{J} = -\mathbf{J}\mathbf{L}\}. \tag{P6.18}$$

(a) Show that $\mathcal{A} = \text{Null}\ (\nabla_{\pi\mathbf{J}} \exp_{\mathcal{V}})$ and conclude that $\nabla_{\pi\mathbf{J}} \exp_{\mathcal{V}}$ fails to be invertible when $\dim \mathcal{V} > 0$. (Hint: Use the Differentiation Theorem for Lineonic Exponentials, Part (a) of Problem 9, and Part (d) of Problem 12 of Chap. 1).

(b) Prove that $-\mathbf{1}_{\mathcal{V}} \in \text{Bdy}\,\text{Rng}\,\exp_{\mathcal{V}}$ and hence that $\exp_{\mathcal{V}}$ fails to be locally invertible near $\pi\mathbf{J}$.

(11) Let $\mathcal{D}$ be an open subset of a flat space $\mathcal{E}$ with translation space $\mathcal{V}$ and let $\mathcal{V}'$ be a linear space.

(a) Let $\mathbf{H} : \mathcal{D} \to \text{Lin}(\mathcal{V}, \mathcal{V}')$ be of class C^2. Let $x \in \mathcal{D}$ and $\mathbf{v} \in \mathcal{V}$ be given. Note that

$$\nabla_x \text{Curl}\,\mathbf{H} \in \text{Lin}(\mathcal{V}, \text{Lin}(\mathcal{V}, \text{Lin}(\mathcal{V}, \mathcal{V}'))) \cong \text{Lin}_2(\mathcal{V}^2, \text{Lin}(\mathcal{V}, \mathcal{V}'))$$

and hence

$$(\nabla_x \text{Curl}\,\mathbf{H})^\sim \in \text{Lin}_2(\mathcal{V}^2, \text{Lin}(\mathcal{V}, \mathcal{V}')) \cong \text{Lin}(\mathcal{V}, \text{Lin}(\mathcal{V}, \text{Lin}(\mathcal{V}, \mathcal{V}'))).$$

Therefore we have

$$\mathbf{G} := (\nabla_x \text{Curl}\,\mathbf{H})^\sim \mathbf{v} \in \text{Lin}(\mathcal{V}, \text{Lin}(\mathcal{V}, \mathcal{V}')) \cong \text{Lin}_2(\mathcal{V}^2, \mathcal{V}').$$

Show that

$$\mathbf{G} - \mathbf{G}^\sim = (\nabla_x \text{Curl}\,\mathbf{H})\mathbf{v}. \tag{P6.19}$$

(Hint: Use the Symmetry Theorem for Second Gradients.)

(b) Let $\boldsymbol{\eta} : \mathcal{D} \to \mathcal{V}^*$ be of class C^2 and put

$$\mathbf{W} := \text{Curl}\,\boldsymbol{\eta} : \mathcal{D} \to \text{Lin}(\mathcal{V}, \mathcal{V}^*) \cong \text{Lin}_2(\mathcal{V}^2, \mathbb{R}).$$

Prove: If $\mathbf{h} : \mathcal{D} \to \mathcal{V}$ is of class C^1, then

$$\text{Curl}\,(\mathbf{W}\mathbf{h}) = (\nabla \mathbf{W})\mathbf{h} + \mathbf{W}\nabla\mathbf{h} + (\nabla\mathbf{h})^\top \mathbf{W}, \tag{P6.20}$$

where value-wise evaluation and composition are understood.

(12) Let $\mathcal{E}$ and $\mathcal{E}'$ be flat spaces with $\dim \mathcal{E} = \dim \mathcal{E}' \geq 2$, let $\varphi : \mathcal{E} \to \mathcal{E}'$ be a mapping of class C^1 and put

$$\mathcal{C} := \{x \in \mathcal{E} \mid \nabla_x \varphi \quad \text{is not invertible}\}.$$

(a) Show that $\varphi_>(\mathcal{C}) \supset \operatorname{Rng}\varphi \cap \operatorname{Bdy}(\operatorname{Rng}\varphi)$.

(b) Prove: If the pre-image under φ of every bounded subset of $\mathcal{E}'$ is bounded and if $\operatorname{Acc}\varphi_>(\mathcal{C}) = \emptyset$ (see Def.1 of Sect.57), then φ is surjective. (Hint: Use Problem 13 of Chap.5.)

(13) By a **complex polynomial** p we mean an element of $\mathbb{C}^{(\mathbb{N})}$, i.e. a sequence in $\mathbb{C}$ indexed on $\mathbb{N}$ and with finite support (see (07.10)). If $p \neq 0$, we define the **degree** of p by

$$\deg p := \max \operatorname{Supt} p = \max\{k \in \mathbb{N} \mid p_k \neq 0\}.$$

By the **derivative** p' of p we mean the complex polynomial $p' \in \mathbb{C}^{(\mathbb{N})}$ defined by

$$(p')_k := (k+1)p_{k+1} \quad \text{for all} \quad k \in \mathbb{N}. \tag{P6.21}$$

The **polynomial function** $\tilde{p} : \mathbb{C} \to \mathbb{C}$ of p is defined by

$$\tilde{p}(z) := \textstyle\sum_{k\in\mathbb{N}} p_k z^k \quad \text{for all} \quad z \in \mathbb{C}. \tag{P6.22}$$

Let $p \in (\mathbb{C}^{(\mathbb{N})})^\times$ be given.

(a) Show that $\tilde{p}^<(\{0\})$ is finite and $\sharp\tilde{p}^<(\{0\}) \leq \deg p$.

(b) Regarding $\mathbb{C}$ as a two-dimensional linear space over $\mathbb{R}$, show that $\tilde{p}$ is of class C^2 and that

$$(\nabla_z\tilde{p})w = \tilde{p}'(z)w \quad \text{for all} \quad z, w \in \mathbb{C}. \tag{P6.23}$$

(c) Prove that $\tilde{p}$ is surjective if $\deg p > 0$. (Hint: Use Part (b) of Problem 12.)

(d) Show: If $\deg p > 0$, then the equation ? $z \in \mathbb{C}$, $\tilde{p}(z) = 0$ has at least one and no more than $\deg p$ solutions.

Note: The assertion of Part (d) is usually called the "Fundamental Theorem of Algebra". It is really a theorem of analysis, not algebra, and it is not all that fundamental. ■