A

Mathematical Background of Variational Calculus

TABLE OF CONTENTS

This Appendix outlines key portions of the mathematical apparatus that supports Chapters 1–5 of AVMM. The exposition omits proofs but gives references to where those can be found. A rich source for this kind of background material is Chapter 1 of Gelfand and Fomin [274].[1]

§A.1. Vector Spaces

A *vector space* $\mathcal{V}$ is a nonempty collection of objects called *vectors*, which may be added together and multiplied by numbers called *scalars*. Generic vectors are denoted by lower case bold symbols such as $\mathbf{x}$, $\mathbf{y}$, etc. This notation agrees with that used for the ordinary vectors of linear algebra.

If $\mathbf{x}$ belongs to $\mathcal{V}$, we write $\mathbf{x} \in \mathcal{V}$, and say that $\mathbf{x}$ is a *member*[2] of $\mathcal{V}$.

Scalars that may appear in the vector-scaling operation defined below are denoted by c, d, etc. In the ensuing description these will be real numbers: $c, d \in \mathcal{R}$, in which as usual $\mathcal{R}$ denotes the set of real numbers. (Generalizations to complex scalars, etc., are possible but unnecessary here.)

§A.1.1. Vector Space Properties

At a minimum, two operations are defined in a vector space: *vector addition*, denoted $\mathbf{x} + \mathbf{y}$, and *scalar multiplication*, denoted $c\,\mathbf{x}$. These obey the following properties, sometimes called *axioms*[3]

Closure Properties:

- **C1**. *Closure Under Addition*: $\mathbf{x} + \mathbf{y} \in \mathcal{V}$ whenever $\mathbf{x} \in \mathcal{V}$ and $\mathbf{y} \in \mathcal{V}$.
- **C2**. *Closure Under Scalar Multiplication*: $c\,\mathbf{x} \in \mathcal{V}$ whenever $\mathbf{x} \in \mathcal{V}$ and $c \in \mathcal{R}$.

Addition Properties:

- **A1**. *Additive Identity*: There is a *zero vector*, denoted $\mathbf{0} \in \mathcal{V}$, such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for all $x \in \mathcal{V}$.
- **A2**. *Additive Negation*: For all $x \in \mathcal{V}$ there is a vector $-\mathbf{x} \in \mathcal{V}$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$. This $-\mathbf{x}$ is called the *negative* of $\mathbf{x}$.
- **A3**. *Associativity*: $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{V}$.
- **A4**. *Commutativity*: $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{V}$.

Scalar Multiplication Properties:

- **M1**. *Scalar Multiplicative Identity*: $1\,\mathbf{x} = \mathbf{x}$ for each $\mathbf{x} \in \mathcal{V}$, in which 1 is the real number 1.
- **M2**. *First Distributive Property*: $c\,(\mathbf{x} + \mathbf{y}) = c\,\mathbf{x} + c\,\mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ and all $c \in \mathcal{R}$.
- **M3**. *Second Distributive Property*. $(c + d)\,\mathbf{x} = c\,\mathbf{x} + d\,\mathbf{x}$ for all $\mathbf{x} \in \mathcal{V}$ and all $c, d \in \mathcal{R}$.
- **M4**. *Associativity*: $c\,(d\,\mathbf{x}) = (c\,d)\,\mathbf{x}$ for all $\mathbf{x} \in \mathcal{V}$ and all $c, d \in \mathcal{R}$.

The two closure properties can be checked at once by verifying the following property:

- **C0**. *Closure Under Linear Combination*: $c\,\mathbf{x} + d\,\mathbf{y} \in \mathcal{V}$ whenever $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ and $c, d \in \mathcal{R}$.

A *vector subspace* $\mathcal{W}$ of a vector space $\mathcal{V}$ is a nonempty subset of $\mathcal{V}$ that is itself a vector space — a set of vectors that satisfy the foregoing 10 conditions.

---

[1] Unfortunately the Russian to English translation therein is occasionally outdated. E.g., "linear space" should be "vector space." (The former term is imprecise as it is presently used for other things.) Such infelicities are silently emended.

[2] Some texts, such as [274], use *element* of a vector space instead of member. That term will be avoided here to reduce the risk of confusion with finite elements.

[3] Transcribed from [746] with minor corrections and additions.

Table A.1. Vector Space Examples

Symbol	*Dim**	*Description*
$\mathcal{R}$ or $\mathcal{R}^1$	1	Space of all real numbers
$\mathcal{R}^2$	2	Space of all ordered real number pairs (a.k.a. 2-vectors)
$\mathcal{R}^3$	3	Space of all ordered real number triples (a.k.a. 3-vectors)
$\mathcal{R}^n$	n	Space of all ordered n-tuples of real numbers (a.k.a. n-vectors)
$\mathcal{C}$ or $\mathcal{C}^1$	1	Space of all complex numbers
$\mathcal{C}^n$	$n^\dagger$	Space of all ordered n-tuples of complex numbers
$\mathcal{P}$	∞	Space of all polynomials
$\mathcal{P}_n$	$n+1$	Space of all polynomials of degree $\leq n$
$\mathcal{M}_{mn}$	$m\,n$	Space of all $m \times n$ matrices
$\mathcal{F}$	∞	Generic function space
$C^0(I)$	∞	Space of all continuous functions on the interval I (which may be open or closed, finite or infinite)
$C^n(I)$	∞	Space of all functions on the interval I that have n continuous derivatives (I as above)

* Space dimensionality: cardinality of basis.
† Dimensionality of $\mathcal{C}^n$ is $2n$ if the basis consists of real numbers.

§A.1.2. Vector Space Examples

Some important examples of vector spaces are listed in Table A.1. Note that dimensionality may be finite, as in the ordinary n-vectors and $m \times n$ matrices, or infinite, as in function spaces.

§A.1.3. Normed Vector Spaces

A vector space $\mathcal{V}$ is said to ne *normed* if each member $\mathbf{x} \in \mathcal{V}$ can be assigned a non-negative number $||\mathbf{x}||$ called the *norm* of $\mathbf{x}$, which obeys the following rules:

N1. *Uniqueness of Zero Norm*: $||\mathbf{x}|| = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

N2. *Positive Homogeneity Under Scaling*: $||c\,\mathbf{x}|| = |c|\,||\mathbf{x}||$ for all $\mathbf{x} \in \mathcal{V}$ and all $c \in \mathcal{R}$.

N3. *Triangular Inequality*. $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{V}$.

If $\mathcal{V}$ is a normed vector space the *distance* between two vectors $\mathbf{x}$ and $\mathbf{y}$ that belong to $\mathcal{V}$ is defined as the nonnegative quantity $||\mathbf{x} - \mathbf{y}||$.

§A.1.4. One-Forms

The principal vector space studied in linear algebra is $\mathcal{R}_n$, the set of all ordered n-tuples. Its member objects are n-dimensional column vectors, which for $n = 3$ reduce to the ordinary vectors (visualizable as arrowed segments) of 3D space.[4]

A *linear map* from $\mathcal{R}_n$ to a scalar field is called a *one-form* (often written 1-form) or *covector*. See Figure A.1(a) for a“slot machine” representation. For exposition simplicity it is assumed that the

[4] The $\mathcal{R}_n$ space is actually the source of the vector space concept, which generalizes linear algebra rules and operations to more complex mathematical objects.

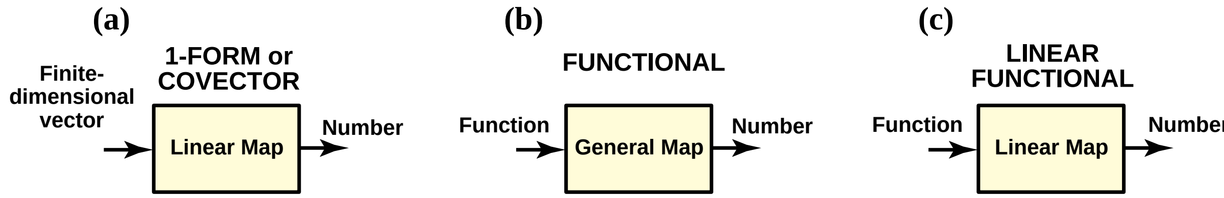

FIGURE A.1. "Slot machine" visualization of the mapping of members of a vector space onto numbers (scalars): (a) A *one-form*, also spelled 1-form, receives a finite dimensional vector and maps it linearly to a number; (b) A *functional* (without qualifier) receives a function and maps it to a number; (c) A *linear functional* receives a function and linearly maps it to a number.

finite-dimensional vector space is $\mathcal{R}_n$, the set of real n-vectors. Its members are denoted $\mathbf{x}$, $\mathbf{y}$, etc. A 1-form map from that space to a scalar must comply with the following linearity rules.

L1. *Degree-One Homogeneity*: $F[\alpha\,\mathbf{x}] = \alpha\,F[\mathbf{x}]$ for any $\mathbf{x} \in \mathcal{R}_n$ and $\alpha \in \mathcal{R}$.

L2. *Linear Superposition*: $F[\mathbf{x}+\mathbf{y}] = F[\mathbf{x}] + F[\mathbf{y}]$ for any $\mathbf{x}, \mathbf{y} \in \mathcal{R}_n$.

For the space $\mathcal{R}_n$ a 1-form can be expediciously built on premultiplying by a row n-vector, since the dot product produces a scalar. For example, suppose that $n = 3$ so $\mathbf{x} = [\,x_1 \;\; x_2 \;\; x_3\,]^T \in \mathcal{R}_3$ is a 3-vector. The vector entry sum is provided by the 1-form

$$S(\mathbf{x}) = \mathbf{e}_3^T\,\mathbf{x} = [\,1 \;\; 1 \;\; 1\,] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1 + x_2 + x_3. \tag{A.1}$$

in which $\mathbf{e}_k$ denotes the k-vector of ones. Here are additional 1-form examples:

- Algebraically largest entry of $\mathbf{x} \in \mathcal{R}_n$ is the 1-form $\max_{i=1}^n x_i$ (but not $\max_{i=1}^n |x_i|$, as rule **L2** would be violated).
- Projection length of $\mathbf{x}$ along the unit direction vector $\mathbf{d}$, in which $\mathbf{d}^T\mathbf{d} = 1$, is the 1-form $\mathbf{d}^T\mathbf{x}$, as long as $\mathbf{d}$ is independent of $\mathbf{x}$.
- Let $\mathbf{g} \in \mathcal{R}_3$ be the 3-vector gradient of a spatial field $\phi(x_1, x_2, x_3)$, that is, $\mathbf{g} = [\,\partial\phi/\partial x_1 \;\; \partial\phi/\partial x_2 \;\; \partial\phi/\partial x_3\,]^T$. The directional gradient value along unit direction $\mathbf{d}$ is the 1-form $g = \mathbf{d}^T\mathbf{g}$.
- The trace of a $n \times n$ matrix $\mathbf{A}$ with entries a_{ij}: $\mathbf{trace}(\mathbf{A}) = \sum_{i=1}^n a_{ii}$, is a 1-form.

On the other hand, the squared length of $\mathbf{x}$, given by the dot product

$$L^2(\mathbf{x}) = ||\mathbf{x}||_2^2 = \mathbf{x}^T\,\mathbf{x} = [\,x_1 \;\; x_2 \;\; x_3\,] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1^2 + x_2^2 + x_3^2. \tag{A.2}$$

is *not* a 1-form since it violates rules **L1** and **L2**. Actually (A.2) is an instance of a 2-form, also called bilinear or quadratic forms. Those are considered in the case of functionals in §A.4.1

What if the vector space becomes an function space? Its members are now functions and the space has infinite dimension. Then 1-forms become *linear functionals*. Those are covered in §A.2.2

§A.1.5. Function Spaces In 1D

Normed vector spaces whose members are functions are called *function vector spaces* or *function spaces* for short. Those are important in the study of variational methods. For description simplicity the members considered below are real-valued functions $y(x)$ of a real variable x, defined over a closed interval $x \in [a, b]$ with $a \leq b$, of the real axis. They will be classified on the bases of function continuity as follows.

Space $C^0(a, b)$ of continuous functions. This vector space consists of all continuous functions $y(x)$ defined in $[a, b]$. The addition and multiplication-by-scalar operations are defined by the ordinary addition of functions and multiplication of function by numbers, respectively. The norm is defined as the maximum of the absolute value:

$$||y||_0 = \max_{a \leq x \leq b} |y(x)|. \tag{A.3}$$

The zero subscript in $||y||$ indicates that taking the norm only involves the function itself. The number $||y||_0$ is usually called the *supremum norm* or *least upper bound* of $y(x)$ over $[a, b]$. Thus many authors call (A.3) the *supremum norm*.

Space $C^1(a, b)$ of continuously differentiable functions. This vector space consists of all functions $y(x)$ defined in $[a, b]$ that are continuous and possess continuous first derivatives. The addition and multiplication-by-scalar operations are the same as in $C^0(a, b)$. The norm is defined by the formula

$$||y||_1 = \max_{a \leq x \leq b} |y(x)| + \max_{a \leq x \leq b} |y'(x)|. \tag{A.4}$$

From this definition it follows that two functions in $C^1(a, b)$, say $y(x)$ and $z(x)$, are viewed as close if both the functions themselves and their first derivatives are close together, since if $||y - z|| < \epsilon$ implies $|y(x) - z(x)| < \epsilon$ and $||y'(x) - z'(x)|| < \epsilon$ for all $x \in [a, b]$. Consequently the norm (A.4) gives more weight to regularity than the supremum norm.

Space $C^n(a, b)$ of n-times continuously differentiable functions. This vector space consists of all functions $y(x)$ defined in $[a, b]$ that are continuous and possess continuous derivatives up to order n inclusive, in which n is a positive integer. The addition and multiplication-by-scalar operations are the same as the preding cases. The norm is defined by the formula

$$||y||_n = \sum_{k=0}^{n} \max_{a \leq x \leq b} |y^{(k)}(x)| \tag{A.5}$$

in which $y(k)(x)$ denotes the k^{th} derivative of $y(x)$.

The previously introduced vector spaces $C^0(a, b)$ and $C^1(a, b)$ correspond to the cases $n = 0$ and $n = 1$, respectively, of $C^n(a, b)$. Obviously $C^k(a, b)$ is included in $C^{k-1}(a, b)$ for $k \geq 1$.

Remark A.1. The case $n \to \infty$ is of special interest for some applications. It gives the class of functions $C^\infty(a, b)$ that are infinitely smooth and possess continuous derivatives of all orders. For example, polynomials. That space will not be used explicitly here.

Remark A.2. The foregoing norm choices do not exhaust all possibilities. One problem with (A.3)–(A.5) is their extreme sensitivity to tiny *local* chances in the functions, especially if derivatives are included. To sugarcoat that undesirable feature numerous *global* norms have been devised. These involve integration over the interval $[a, b]$. For example, the widely used L_p norms for the $C^0(a, b)$ space, with p a positive integer, are given by

$$||y||_{L_p} = \int_a^b |y(x)|^p \, dx, \quad p = 1, 2 \ldots, \quad y(x) \in C^0(a, b). \tag{A.6}$$

This is called the *Euclidean norm* for $p = 2$. The supremum norm (A.3) is recovered in the limit $p \to \infty$, whence the alternative name *infinity norm*.

Remark A.3. Another difficulty with norms that combine function values and derivatives, as in (A.5) for $k \geq 1$, is improper scaling when applied to physical problems. An example in dynamics: if $y(t)$ is a displacement response in time t, $\dot{y}(t) = dy(t)/dt$ is a velocity, so taking the norm $||y||_1$ of (A.4) would add apples to oranges. This nonsense can be resolved by applying appropriate weights that restore the correct physical dimensions (e.g., multiplying $\dot{y}(t)$ by a characteristic length), so one is effectively adding apples to apples.

§A.1.6. More General Function Spaces

The function spaces introduced above may be generalized in several directions, for example

Multiple Dimensions. The space of scalar functions that depend on multiple independent variables. For example $y = y(x_1, x_2, x_3)$ in 3D space or $y = y(x_1, x_2, x_3, t)$ in spacetime.

Multiple Functions. The space of m multiple functions that depend on n multiple independent variables. These are usually arranged as an ordinary m-vector that (as usual) is identified by a bold symbol. For example, if $m = 2$ and $n = 3$:

$$\mathbf{y}(\mathbf{x}) = \mathbf{y}(x_1, x_2, x_3) = \begin{bmatrix} y_1(x_1, x_2, x_3) \\ y_2(x_1, x_2, x_3) \end{bmatrix}, \quad \text{in which} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \tag{A.7}$$

Such generalizations are well covered in textbooks devoted to functional analysis, as well as advanced treatment of variational calculus.

§A.2. Functionals: Basic Concepts

The foregoing concepts are now applied to the study of *functionals*. As noted in Chapter 1 a functional is a function of a function. More precisely, it receives a function and produces a number. Figure A.1(b) gives a visual representation of a functional as a "slot machine."

In this Section we only consider functionals $J[y]$ that depend on a single function $y(x)$ in which x, y and J are real, and defined over the closed interval $x \in [a, b]$. The function $y(x)$ is viewed as a *point* of its vector space, which will be called $y \in \mathcal{V}_y$.

§A.2.1. Functional Continuity

The functional $J[y]$ is said to be *continuous* at the point $y^* \in \mathcal{R}$ if for any $\epsilon > 0$ there is a $\delta > 0$ such that

$$\big| J[y] - J[y^*] \big| < \epsilon, \tag{A.8}$$

provided that $||y - y^*|| < \delta$, in which $||$ is the norm defined for $\mathcal{V}_y$.

Remark A.4. The inequality (A.8) is equivalent to two inequalities:

$$|\, J[y] - J[y^*]\, | > -\epsilon, \tag{A.9}$$

and

$$|\, J[y] - J[y^*]\, | < \epsilon, \tag{A.10}$$

If in the definition of continuity, (A.8) is replaced by (A.9), $J[y]$ is said to be *lower semicontinuous* at y^*. If instead (A.8) is replaced by (A.10), $J[y]$ is said to be *upper semicontinuous* at y^*.

Remark A.5. Which space $\mathcal{V}_y$ is appropriate for $y(x)$? The decision depends primarily on the order of y-derivatives that appear in $J[y]$. For example, in the case of the simplest variational problem, which involves functionals of the form

$$J[y] = \int_a^b F(x, y, y')\, dx \tag{A.11}$$

the appropriate function space for $y(x)$ would be $C^1(a, b)$, the space of continuous differentiable functions. Selecting the larger space $C^0(a, b)$ of continuous functions may lead to problems since we could select two functions $y(x)$ and $y^*(x)$ arbitrarily close in the sense of the supremum norm (A.3), but whose derivatives differ by any given amount. The continuity test (A.8) would then fail. Bottom line: one should pick a function space in which the continuity of the functional under study holds.

Remark A.6. In many variational problems we deal with functionals defined on sets of *admissible functions* that do not form a vector space. For example, the admissible functions for the simplest functional (A.11) under the essential boundary conditions

$$y(a) = \hat{y}_a, \qquad y(b) = \hat{y}_b, \tag{A.12}$$

are the continuously differentiable plane curves $y = y(x)$ that pass through the points A$(a, \hat{y}_a)$ and B$(b, \hat{y}_b)$. But the sum of two admissible functions will not pass through those points unless $\hat{y}_a = \hat{y}_b = 0$. The concept of normed vector space as well as related notions of distance and continuity still play, however, an important role. It is generally possible to modify the given variational problem through various techniques so that the appropriate tools can be rigorously applied.

§A.2.2. Continuous Linear Functionals

The generalization of the 1-form discussed in §A.1.4 to function spaces is clear fits the definition of functional. But carrying over the linearity conditions **L1** and **L2** of §A.1.4 means that we get a restricted kind, called a *linear functional*. See Figure A.1(c). This can be further specialized by requiring that the continuity property (A.8) hold. The end product of these specializations is a *continuous linear functional*, or CLF. These figure prominently in variational calculus. For simplicity they are introduced next for the one-dimensional case.

Consider a one-dimensional, normed function vector space $\mathcal{F}(a, b)$. Its members are functions $h(x) \in \mathcal{F}$ with domain $x \in [a, b]$.[5] with domain $x \in [a, b]$. The functional $\phi[h]$ is called a *continuous linear functional*, or CLF for short, if it satisfies the following conditions.

CL1. $\phi[\alpha\, h] = \alpha\, \phi[h]$ for any $h \in \mathcal{F}$ and $\alpha \in \mathcal{R}$.

CL2. $\phi[h_1 + h_2] = \phi[h_1] + \phi[h_2]$ for any $h_1, h_2 \in \mathcal{F}$.

CL3. $\phi[h]$ is continuous for all $h \in \mathcal{F}$.

[5] We deviate here from the use of $y(x)$ as generic member function so as to merge seamlessly with **?**, which introduces the concept of *variation*. In that subsection, $h(x)$ is the *increment* of $y(x)$.

Example A.1. The function $h(x) \in C^0(a, b)$, the space of continuous functions in $x \in [a, b]$. Define the functional $\phi[h]$ as the value taken at a point $c \in [a, b]$:

$$\phi[h] = h(c). \tag{A.13}$$

This $\phi[h]$ is a CLF on $[a, b]$.

Example A.2. Again $h(x) \in C^0(a, b)$. The area integral

$$\phi[h] = \int_a^b h(x)\, dx, \tag{A.14}$$

defines a CLF in $C^0(a, b)$

Example A.3. Again $h(x) \in C^0(a, b)$ while $\alpha(x)$ denotes a given function also in $C^0(a, b)$. The integral

$$\phi[h] = \int_a^b \alpha(x)\, h(x)\, dx, \tag{A.15}$$

defines a CLF on the function space $C^0(a, b)$.

Example A.4. More generally, take $h(x) \in C^n(a, b)$, with n a positive integer whereas $\alpha_i(x)$, $i = 0, 1, \ldots n$ are fixed functions in $C(a, b)$. the integral

$$\phi[h] = \int_a^b \Big[\alpha_0(x)\, h(x) + \alpha_1(x)\, h'(x) + \ldots + \alpha_n(x)\, h^(n)(x)\Big]\, dx \tag{A.16}$$

defines a CLF on the function space $C^n(a, b)$.

§A.2.3. Fundamental Lemmas of Variational Calculus

Suppose that the CLF (A.16) stated in Example A.4 vanishes for all $h(x)$ pertaining to a certain function space. What can be said about the functions $\alpha_i(x)$? The following results are typical of the so-called *fundamental lemma of variational calculus*.[6]

LEMMA A.1. If $\alpha(x) \in C^0(a, b)$ and if

$$\int_a^b \alpha(x)\, h(x)\, dx = 0, \tag{A.17}$$

for every $h(x) \in C^0(a, b)$ such that $h(a) = h(b) = 0$, then $\alpha(x) = 0$ for all $x \in [a, b]$. Proof: see [274, p. 9]. It is shown there that the result also holds if $C^0(a, b)$ is replaced by $C^n(a, b)$.

LEMMA A.2. If $\alpha(x) \in C^0(a, b)$ and if

$$\int_a^b \alpha(x)\, h'(x)\, dx = 0, \tag{A.18}$$

[6] Strictly speaking, that name applies to LEMMA A.1, when $h(x)$ is taken as the admissible variation of the input to a functional. LEMMAS A.2–A.4 are directly deduced from it after minor gyrations, and so are largely variations on a theme.

for every $h(x) \in C^1(a,b)$ such that $h(a) = h(b) = 0$, then $\alpha(x) = c$ for all $x \in [a,b]$, in which c is a constant. Proof: see [274, p. 10].

LEMMA A.3. If $\alpha(x) \in C^0(a,b)$ and if

$$\int_a^b \alpha(x)\, h''(x)\, dx = 0, \tag{A.19}$$

for every $h(x) \in C^2(a,b)$ such that $h(a) = h(b) = 0$ and $h'(a) = h'(b) = 0$, then $\alpha(x) = c_0\, x + c_1$ for all $x \in [a,b]$, in which c_0 and c_1 are constants. Proof: see [274, p. 10].

LEMMA A.4. If $\alpha(x) \in C^0(a,b)$ and $\beta(x) \in C^0(a,b)$ and if

$$\int_a^b \left[\alpha(x)\, h(x) + \beta(x)\, h'(x)\right] dx = 0, \tag{A.20}$$

for every $h(x) \in C^1(a,b)$ such that $h(a) = h(b) = 0$, then $\beta(x)$ is differentiable and $\beta'(x) = \alpha(x)$ for all $x \in [a,b]$. Proof: see [274, p. 11].

§A.3. Functionals: First Variation

The concept of *variation* of a functional is analogous to that of a *differential* of a function of n variables. Its main application is to find *extrema* of functionals. Those are input function(s), known as *extremals*, for which the functional becomes stationary with respect to small variations. As usual we restrict the ensuing exposition to the one-dimensional case.

§A.3.1. Increments and Differentiability

Let $J[y]$ be a functional defined on some normed function vector space with domain $x \in [a,b]$. Suppose that an admissible input function $y = y(x)$ is incremented by $h = h(x)$, while taking care that $y + h = y(x) + h(x)$ *remains admissible*. (For example, if $y(x)$ has to satisfy the essential boundary conditions $y(a) = \hat{y}_a$ and $y(b) = \hat{y}_b$, plainly $h(a) = h(b) = 0$.) The *increment* of the functional is defined as

$$\Delta J[h] = J[y+h] - J[y]. \tag{A.21}$$

If y is fixed, $\Delta J[y+h] \equiv J[h]$ is a functional of $h = h(x)$. For arbitrary h this will be generally a nonlinear functional. If the increment ΔJ can be expressed as

$$\Delta J[h] = \phi[h] + \epsilon\, ||h|| \tag{A.22}$$

in which $\phi[h]$ is a CLF and $\epsilon \to 0$ as $||h|| \to 0$, the functional $J[y]$ is said to be *differentiable* at that fixed y. If (A.22) holds, the functional $\phi[h]$ is called the *principal linear part* of the increment $\Delta J[h]$.

If $J[y]$ is differentiable at each admissible y, the functional is generically called *differentiable*.

§A.3.2. First Variation Properties

For notational convenience, as well as smooth linkage to ordinary differential calculus, the increment $\phi[h]$ is renamed the *first variation* of $J[y]$, and denoted by $\delta J = \delta J[h]$.[7]

Two important first-variation theorems pertaining to differentiable functionals follow.

THEOREM A.1. The first variation of a differentiable functional is unique.

Proof: see [274, p. 12].

THEOREM A.2. A *necessary* condition for a differentiable functional $J[y]$ to have an extremum at $y = y^*$ is that its first variation vanish there; that is

$$\delta J[y] = 0, \tag{A.23}$$

for $y = y^*$ and all admissible h.

Proof: see [274, p. 13].

Functions $y = y(x)$ at which δJ vanishes are called *extremals*.

Remark A.7. THEOREM A.2 is the analog of a well known theorem of ordinary calculus. If the one-dimensional differentiable function $f(x)$ has an extremum at $x = x^*$, its differential must vanish there: $df(x^*) = 0$, which may also be expressed as $y'(x^*) = 0$. That is, the function must be stationary at x^*. Now in ordinary calculus an extremum can be a local maximum, a local minimum, or a local inflexion point (neither a maximum nor a minimum). If one is looking for actual (local) *maxima* or *minima*, the condition $df(x) = 0$ is only *necessary*. Sufficient conditions depend on examination of the sign of the *second differential* $d^2 f(x)$ — or, alternatively, $y''(x)$ — in the vicinity of $x = x^*$. The analogous sufficiency conditions for functionals are far more elaborated; in fact rigorous results took over a century (from Euler and Lagrange to Weierstrass) to be worked out.

§A.4. Functionals: Second Variation

THEOREM 2 above gives only *necessary* conditions for an extremum of $J[y]$. An extremum can be a maximum, minimum or just a stationary point. As hinted in Remark A.7, *sufficient* conditions for a maximum or a minimum require examination of the functional-level equivalent of the second differential of ordinary calculus. That equivalent object is called a *second variation*. Before defining it, some auxiliary concepts are introduced.

§A.4.1. Bilinear and Quadratic Functionals

An obvious generalization of the linear functionals introduced in §A.2.2 is the concept of *bilinear functional*. This becomes a *quadratic functional* as special case.

A functional $B[y, z]$ that depends on two input functions $y = y(x)$ and $z = z(x)$ that belong to a normed function vector space $\mathcal{F}$, is said to be *bilinear* if it is a linear functional of y for fixed z, and a linear functional of z for fixed y. More specifically, if it satifies the following "bilinearity" conditions

B1. *Homogeneity in y*: $B[\alpha\, y, z] = \alpha B[y, z]$ for all $y, z \in \mathcal{F}$ and $\alpha \in \mathcal{R}$.

[7] Of course the first variation is actually a functional of $y = y(x)$ (the"baseline" input function) and $h = h(x)$ (its increment). So in some contexts it may help to write $\Delta J[y, h] = \delta J[y, h] + \epsilon||h||$.

- **B2**. *Associativity in y*: $B[y_1 + y_2, z] = B[y_1, z] + B[y_2, z]$ for all $y_1, y_2, z \in \mathcal{F}$.
- **B3**. *Homogeneity in z*: $B[y, \alpha z] = \alpha B[y, z]$ for all $y, z \in \mathcal{F}$ and $\alpha \in \mathcal{R}$.
- **B4**. *Associativity in z*: $B[y, z_1 + z_2] = B[y, z_1] + B[y, z_2]$ for all $y, z_1, z_2 \in \mathcal{F}$

If we set $y = z$ in a bilinear functional, we obtain an expression called a *quadratic functional*

A quadratic functional $A[y] = B[y, y]$ is said to be *positive definite*

$$A[y] > 0 \quad \text{for all} \quad y \in \mathcal{F}, \quad y \neq 0. \tag{A.24}$$

If the $>$ is replaced by $\geq$ in (A.24) the functional is said to be *nonnegative*.

A quadratic functional $A[y] = B[y, y]$ is said to be *strongly positive definite* if there exists a constant $k > 0$ such that

$$A[y] \geq k\,|y||. \tag{A.25}$$

In all of the folloing examples, x, $y(x)$, $z(x)$, $\alpha(x)$, $\beta(x)$, $\gamma(x) \in \mathcal{R}$, and $x \in [a, b]$.

Example A.5. The expression

$$B[y, z] = \int_a^b y(x)\, z(x)\, dx, \tag{A.26}$$

in which $y, z \in C^0(a, b)$, is a bilinear functional. The corresponding quadratic functional is obtained by making $z(x) = y(x)$:

$$A[y] = \int_a^b (y(x))^2\, dx \int_a^b y(x) y(x)^2\, dx \tag{A.27}$$

The quadratic functional (A.27) is positive definite.

Example A.6. The expression

$$B[y, z] = \int_a^b \alpha(x)\, y(x)\, z(x)\, dx, \tag{A.28}$$

in which $y, z \in C^0(a, b)$ and $\alpha(x)$ is a given function, is a bilinear functional. The corresponding quadratic functional is

$$A[y] = \int_a^b \alpha(x)\, (y(x))^2\, dx = \int_a^b \alpha(x)\, y(x) y(x)\, dx \tag{A.29}$$

If $\alpha(x) > 0$ for all $x \in [a, b]$ the functional (A.29) is positive definite.

Example A.7. The expression

$$A[y] = \int_a^b \left[\alpha(x)\, y(x) y(x) + \beta(x)\, y(x)\, y'(x) + \gamma(x)\, y'(x) y'(x)\right] dx, \tag{A.30}$$

in which $y \in C^1[a, b]$ and $\alpha(x)$, $\beta(x)$, and $\gamma(x)$ are given functions, is a quadratic functional.

§A.4.2. Second Variation Definition

We retake the developments of §A.3.1 and **?** to proceed beyond the first variation. To briefly recapitulate, let $J[y]$ be a functional defined on some normed function vector space $\mathcal{F}$ with domain $x \in [a, b]$. Suppose that an admissible input function $y = y(x)$ is incremented by $h = h(x)$, while taking care that $y + h = y(x) + h(x)$ remains admissible. If the increment of the functional can be expressed as

$$\Delta J[h] = J[y + h] - J[y] = \delta J[h] + \epsilon \, ||h|| \tag{A.31}$$

in which $\phi[h]$ is a CLF and $\epsilon \to 0$ as $||h|| \to 0$, then: (i) δJ was called the first variation of $J[y]$ at the given y, and (ii) $J[y]$ was said to be differentiable at y. If (ii) holds for any y, the functional $J[y]$ was generically called *differentiable*.

Continuing on this theme, we will say that $J[y]$ is *twice differentiable* if its increment can be written as

$$\Delta J[h] = J[y + h] - J[y] = \phi_1[h] + \phi_2[h] + \epsilon \, ||h||^2, \tag{A.32}$$

in which $\phi_1[h]$ is a *linear* functional, $\phi_2[h]$ is a *quadratic* functional and $\epsilon \to 0$ as $||h|| \to 0$.

The $\phi_1[h]$ term that appears in (A.32) was called the first variation in **?**. Accordingly we will rename $\phi_2[h]$ the *second variation*, and denote it by $\delta^2 J[h]$.[8] On replacing $\phi_1 \to \delta J$ and $\phi_2 \to \delta^2 J$, the increment (A.32) may be recast as

$$\Delta J[h] = J[y + h] - J[y] = \delta J + \delta^2 J + \epsilon \, ||h||^2. \tag{A.33}$$

If (A.33) holds for any y and h, we will generically state that $J[y]$ is twice differentiable.

Two important second-variation theorems pertaining to twice differentiable functionals follow. (These correspond to those stated for the first variation in **?**.)

THEOREM A.3. The second variation of a twice differentiable functional is unique.

Proof: see [274, p. 99].

THEOREM A.4. A *necessary* condition for a twice differentiable functional $J[y]$ to have a *minimum* at $y = y^*$ is that

$$\delta J[y] = 0, \qquad \delta^2 J[y] \geq 0, \tag{A.34}$$

for $y = y^*$ and all admissible h. For a maximum, replace the sign in the second of (A.34) by $\leq$.

Proof: see [274, p. 99].

Note that condition (A.34), or its counterpart with $\leq$, is *not* sufficient to guarantee a minimum or a maximum. To achieve sufficiency we need to recall the concept of *strong positivity*, which was defined for a quadratic functional in (A.25).

THEOREM A.5. A *sufficient* condition for a twice differentiable functional $J[y]$ to have a *minimum* at $y = y^*$ is that is second variation $\delta^2 J[h]$ be strongly positive for $y = y^*$.

Proof: see [274, p. 100].

For a maximum, replace *strongly positive* by *strongly negative*, or change the sign of $\delta^2 J[h]$.

---

[8] The comment made in a previous footnote also applies here: $\delta^2 J$ is actually a function of y and h. In some contexts it might be convenient to denote it as $\delta^2 J[y, h]$.

Remark A.8. In a finite dimensional normed vector space, strong positivity of a quadratic form is equivalent to positive definiteness of that form. Therefore a twice differentiable function $f(\mathbf{x})$ of a finite number of variables collected in vector $\mathbf{x} = [\, x_1 \;\; x_2 \;\; \ldots \;\; x_n \,]^T$, has a minimum at a point $P^* = P(\mathbf{x}^*)$ where its first differential vanishes, if its second differential is positive definite at P^*. In more general case of a functional, however, strong positivity is a stronger condition than positive definiteness.

§A.4.3. Second Variation Differential Forms

We now find an expression for the second variation of the simplest variational problem, associated with functionals of the form (A.11), in which input functions $y = y(x)$ satisfy *a priori* the fixed-point conditions (A.12). As regards smoothness we assume that the integrand of (A.11), which has the form $F(x, y, z)$, has continuous partial derivatives up to the necessary order with respect to all its arguments.

Increment $y(x)$ by function $h(x)$ such that $y + h$ satisfies all admissibility conditions; in particular $h(a) = h(b) = 0$. The corresponding increment $\Delta J[h] = J[y + h] - J[y]$ of the functional is expanded in Taylor's series

$$\Delta J[h] = \int_a^b (F_y\, h + F_{y'}\, h')\, dx + \tfrac{1}{2} \int_a^b \left[\bar{F}_{yy}\, h^2 + \bar{F}_{yy'}\, h\, h' + \bar{F}_{y'y'}\, (h')^2 \right] dx. \tag{A.35}$$

in which the overbar indicated that the corresponding derivatives are evaluated along certain intermediate curves; for example

$$\bar{F}_{yy} = F_{yy}\left(x, y + \theta\, h, y' + \theta h'\right), \quad \theta \in [0, 1], \tag{A.36}$$

and similarly for $\bar{F}_{yy'}$ and $\bar{F}_{y'y'}$. If $\bar{F}_{yy}$, $\bar{F}_{yy'}$ and $\bar{F}_{y'y'}$ are replaced by the derivatives F_{yy}, $F_{yy'}$ and $F_{y'y'}$, respectivelyt, evaluated at $(x, y(x), y'(x)$ the expansion (A.35) can be rewritten with a remainder term as

$$\Delta J[h] = \int_a^b (F_y\, h + F_{y'}\, h')\, dx \int_a^b \left[\epsilon_1\, h^2 + \epsilon_2\, h\, h' + \epsilon_3\, (h')^2 \right] dx + \epsilon, \tag{A.37}$$

in which

$$\epsilon = \int_a^b \left[F_{yy}\, h^2 + F_{yy'}\, h\, h' + F_{y'y'}\, (h')^2 \right] dx + \tag{A.38}$$

Beacuse of the bassumed continuity of F_{yy}, $F_{yy'}$ and $F_{y'y'}$, it follows that ϵ_1, ϵ_2 and ϵ_3 tend to zero as $||h||_1|| \to 0$. Hence ϵ is an infinitesimal of order higher than 2 with respect to $||h||_1$. By definition the first and second terms of (A.37) are the first variation $\delta J[h]$ and $\delta^2 J[h]$, respectively, of (A.11). Therefore the latter can be expressed as

$$\delta^2 J[h] = \int_a^b \left[F_{yy}\, h^2 + F_{yy'}\, h\, h' + F_{y'y'}\, (h')^2 \right] dx. \tag{A.39}$$

Integrating the middle term by parts and noting that $h(a) = h(b) = 0$ we obtain $\int_a^b F_{yy'}\, h\, h'\, dx = -\int_a^b (dF_{yy'}/dx) h^2\, dx$, whence (A.39) can be transformed into the more compact form

$$\delta^2 J[h] = \int_a^b \left[Q\, h^2 + P\, (h')^2 \right] dx, \quad P = \tfrac{1}{2} F_{y'y'}, \quad Q = \tfrac{1}{2} \left(F_{yy'} - \frac{dF_{yy}}{dx} \right) dx. \tag{A.40}$$

Here $P = P(x)$ and $Q = Q(x)$. This is the most convenient expression for further investigations.

§A.4.4. Legendre's Condition

(TBC)