Basic mathematics for nano-engineers

Horia D. Cornean

18/09/2007

Dept. of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, 9220 Aalborg, Denmark.

1 Taylor approximation in one variable

We begin with a fundamental theorem in analysis. Assume that we have a "nice" (continuously differentiable) function $f : \mathbb{R} \to \mathbb{R}$, for which we know its derivative f' at every point. Assume also that we know the value of f at a point x_0. Then we know the function everywhere, and it is given by the formula

$$f(x) = f(x_0) + \int_{x_0}^{x} f'(x_1)dx_1. \tag{1.1}$$

If we know the second derivative, too, then we can write

$$f'(x_1) = f'(x_0) + \int_{x_0}^{x_1} f''(x_2)dx_2,$$

and by inserting it into (1.1) we obtain:

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \int_{x_0}^{x} \left\{ \int_{x_0}^{x_1} f''(x_2)dx_2 \right\} dx_1. \tag{1.2}$$

This double integral must be interpreted in the following way: first one integrates with respect to the x_2 variable, thus obtaining a function of x_1, then we integrate with respect to x_1. For example, if c is a constant, we have:

$$\int_{x_0}^{x} \left\{ \int_{x_0}^{x_1} c \cdot dx_2 \right\} dx_1 = c \int_{x_0}^{x} (x_1 - x_0)dx_1 = \frac{c}{2}(x - x_0)^2.$$

It is now clear that we can continue this procedure to higher orders, if we know higher and higher order derivatives of f. We obtain

$$\begin{aligned} f(x) &= f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2} f''(x_0) + ... + \frac{(x - x_0)^n}{n!} f^{(n)}(x_0) \\ &+ \int_{x_0}^{x} \left\{ \int_{x_0}^{x_1} \left\{ \int_{x_0}^{x_2} ... \left\{ \int_{x_0}^{x_n} f^{(n+1)}(x_{n+1})dx_{n+1} \right\} ...dx_3 \right\} dx_2 \right\} dx_1. \end{aligned} \tag{1.3}$$

The notation $f^{(n)}(x_0)$ means the n'th order derivative of f at x_0. The nth order polynomial

$$P_n(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2} f''(x_0) + ... + \frac{(x - x_0)^n}{n!} f^{(n)}(x_0)$$

is called the nth order Taylor polynomial at x_0, of variable x. Now if n is large and $|x - x_0|$ is small, we see that the Taylor polynomial is a good approximation to the actual value of $f(x)$. From (1.3), we can estimate the error we make when we replace f by P_n; its absolute value is less than

$$\frac{|x - x_0|^{n+1}}{(n+1)!} \sup\{|f^{(n+1)}(t)| : t \text{ between } x \text{ and } x_0\}. \tag{1.4}$$

For example, if $f(x) = \sin(x)$, $x_0 = 0$, $x \in [0, \pi/4]$, and $n = 4$, we have $|f^{(5)}(t)| = |\cos(t)| \leq 1$ for all t between 0 and x, hence the error we make is less than

$$\frac{|x - x_0|^5}{120} \leq \frac{\pi^5}{4^5 \cdot 120}.$$

Exercise 1.1. *Estimate the error we make for the same function as above, but for $n = 6$.*

Exercise 1.2. *Assume $f(x) = e^x$, $x \in [0, 2]$, $x_0 = 1$ and $n = 2$. Show that we have*

$$\left| f(x) - \frac{x^2 + 1}{2} e \right| \leq \frac{e^2}{6}$$

for all $x \in [0, 2]$. Is this a good error? Show that for any n, the error is less than $e^2/(n+1)!$.

2 Local minima and maxima in one variable

A point of local minimum for a function f is a point x_0 around which we have the inequality $f(x) \geq f(x_0)$. If f has a derivative at x_0, then the tangent to its graph must be parallel to the $0x$ axis at this point. This is because the angle is negative on the left of x_0, and is positive on its right. So one equation is $f'(x_0) = 0$. But the same equation applies for local maxima, too, so we need more information.

If we use the Taylor expansion of order 2 around x_0, we get that

$$f(x) \sim f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 \tag{2.1}$$

for x close enought to x_0. But since $f'(x_0)$ has to be zero, we have

$$f(x) \sim f(x_0) + \frac{f''(x_0)}{2}(x - x_0)^2. \tag{2.2}$$

It is now clear that x_0 is a minimum if $f''(x_0) > 0$ and is a maximum if $f''(x_0) < 0$. If the second derivatives are zero, we cannot conlude anything, and we must look at higher order derivatives.

Exercise 2.1. *Consider the function $f(x) = e^{-x^2}$ defined on R. Solve the equation $f'(x) = 0$. Is the solution a minimum?*

3 Real functions of two variables

This section will mostly contain hand-waiving arguments, and no rigorous statements. The point is to get some fundamental results down to earth, and to understand why they hold true.

Let us start with partial derivatives. Assume that we have a function $f(x_1, x_2)$ where x_1 and x_2 are real variables. We can also say that f is defined on the vector space $\mathbb{R}^2$, where the variable is a two dimensional vector $\mathbf{x} = [x_1, x_2]$.

If one variable is kept fixed, for example x_2, then the function $g_{x_2}(x) = f(x, x_2)$ is a "normal" one-variable function. If g is differentiable at the point a, then f has a partial derivative with respect to the first variable and we have

$$g'_{x_2}(a) = \lim_{\Delta x \to 0} \frac{f(a + \Delta x, , x_2) - f(a, x_2)}{\Delta x} =: (\partial_1 f)(a, x_2). \tag{3.1}$$

In a similar way, if we fix the first variable we obtain a function $h_{x_1}(y) = f(x_1, y)$ and we can write

$$h'_{x_1}(b) = \lim_{\Delta y \to 0} \frac{f(x_1, b + \Delta y) - f(x_1, b)}{\Delta y} =: (\partial_2 f)(x_1, b). \tag{3.2}$$

Then we can define higher order partial derivatives by induction. For example, let us consider the meaning of $[\partial_1(\partial_1 f)](x_1, x_2)$. It means that we first differentiate f with respect to x_1, obtaining another function $\tilde{f}(x_1, x_2) := (\partial_1 f)(x_1, x_2)$. So the second order partial derivative of f with respect to x_1 is the first order partial derivative of $\tilde{f}$.

In order to simplify notation, we write

$$\partial_1(\partial_1 f) := \partial_1^2 f,\ \partial_2(\partial_2 f) := \partial_2^2 f,\ \partial_1(\partial_2 f) := \partial_{1,2}^2 f,\ \partial_2(\partial_1 f) := \partial_{2,1}^2 f$$

Let us now investigate a very important result on the mixed partial derivatives, widely and directly used in thermodynamics for example (see page 251 in McQuarrie).

Exercise 3.1. *Assume that f has continuous second order partial derivatives. Then we have*

$$(\partial_{1,2}^2 f)(a, b) = (\partial_{2,1}^2 f)(a, b).$$

This means that the order of performing partial derivatives is not important.

Hint to the solution. Use the fundamental formula (1.1) for the function $g_y(x) = f(x, y)$. We then obtain:

$$f(x, y) = g_y(x) = g_y(a) + \int_a^x g_y'(t)dt = f(a, y) + \int_a^x (\partial_1 f)(t, y)dt.$$

Hence

$$f(x, y) = f(a, y) + \int_a^x (\partial_1 f)(t, y)dt. \tag{3.3}$$

If we differentiate this equality with respect to y we get:

$$(\partial_2 f)(x, y) = (\partial_2 f)(a, y) + \int_a^x (\partial_2 \partial_1 f)(t, y)dt. \tag{3.4}$$

Then we differentiate both sides with respect to x. The left hand side will give us $(\partial_1 \partial_2 f)(x, y)$. The right hand side has a first term which does not depend on x so its partial derivative is zero, while the other one gives (according to (1.1)) $(\partial_2 \partial_1 f)(x, y)$. And we are done.
Example. Consider $f(x, y) = x^2 y^3 + e^x$. We have

$$(\partial_1 f)(x, y) = 2xy^3 + e^x,\ (\partial_2 f)(x, y) = 3x^2 y^2,\ (\partial_{1,2}^2 f)(x, y) = (\partial_{2,1}^2 f)(x, y) = 6xy^2.$$

For more details, read Examples 5, 6, and 7 in McQuarrie, at page 248 and 249.

4 Derivatives of vector functions

Let us fix a point in the two dimensional plane $X_0 := [x_0, y_0]$. We say that another point $X := [x, y]$ is close to X_0 if the euclidian distance

$$||X - X_0|| := \sqrt{(x - x_0)^2 + (y - y_0)^2}$$

is small.

Exercise 4.1. *Show that* $|x - x_0| \leq ||X - X_0||$ *and* $|y - y_0| \leq ||X - X_0||$.

Hint to the solution. Just use the obvious inequality

$$\max\{(x - x_0)^2,\ (y - y_0)^2\} \leq ||X - X_0||^2.$$

If one is interested in the difference $f(X) - f(X_0)$ when X is close to X_0, then one has to generalize the notion of derivative, since the variable is a vector.

Define the function $g_y(x) = f(x, y)$. The Taylor formula approximates g in the following way around x_0:

$$g_y(x) \sim g_y(x_0) + g_y'(x_0)(x - x_0) + \frac{g_y''(x_0)}{2}(x - x_0)^2$$

or using f:

$$f(x,y) \sim f(x_0,y) + (\partial_1 f)(x_0,y)(x-x_0) + \frac{(\partial_1^2 f)(x_0,y)}{2}(x-x_0)^2. \tag{4.1}$$

If we develop $f(x_0,y)$ around y_0 up to second order we get:

$$f(x_0,y) \sim f(x_0,y_0) + (\partial_2 f)(x_0,y_0)(y-y_0) + \frac{(\partial_2^2 f)(x_0,y_0)}{2}(y-y_0)^2. \tag{4.2}$$

Now we develop $(\partial_1 f)(x_0,y)$ in the first order around y_0:

$$(\partial_1 f)(x_0,y) \sim (\partial_1 f)(x_0,y_0) + (\partial_{2,1}^2 f)(x_0,y_0)(y-y_0). \tag{4.3}$$

If we are only interested in expressing $f(x,y)$ with an error of up to $||X-X_0||^2$, then we can insert (4.3) and (4.2) into (4.1) and write

$$\begin{aligned} f(x,y) &= f(x_0,y_0) + (\partial_1 f)(x_0,y_0)(x-x_0) + (\partial_2 f)(x_0,y_0)(y-y_0) \\ &+ \frac{1}{2}\Big\{(\partial_1^2 f)(x_0,y_0)(x-x_0)^2 + 2(\partial_{2,1}^2 f)(x_0,y_0)(x-x_0)(y-y_0) \\ &+ \ (\partial_2^2 f)(x_0,y_0)(y-y_0)^2\Big\} + \mathcal{O}(||X-X_0||^3). \end{aligned} \tag{4.4}$$

The expression $\mathcal{O}(||X-X_0||^3)$ denotes all terms which are as small as $||X-X_0||^3$. It means that (4.4) gives the quadratic approximation of f near X_0.

The linear approximation of f near X_0 is now given by two numbers: $(\partial_1 f)(x_0,y_0)$ and $(\partial_2 f)(x_0,y_0)$. These two numbers form a vector, called the gradient of f at X_0, and denoted by

$$(\nabla f)(X_0) := [(\partial_1 f)(x_0,y_0), (\partial_2 f)(x_0,y_0)]. \tag{4.5}$$

This vector is the derivative of f. We can write (using (4.4) and the definition of the scalar product of two vectors)

$$f(X) = f(X_0) + (\nabla f)(X_0)\cdot(X-X_0) + \mathcal{O}(||X-X_0||^2). \tag{4.6}$$

Exercise 4.2. *Consider $f(x,y) = \sin(x-y)\cos(xy)$. Develop f around $X_0 = [0,0]$ up to the second order in $||X-X_0||$.*

5 Local extrema for functions of two variables

We say that X_0 is a point of local minimum for f, if for every X close to X_0 we have $f(X) \geq f(X_0)$. Here the variable X is located in a small disk centred at X_0. It is clear that $f(x,y_0)$ is a one-variable function, which must have a minimum at x_0. Therefore, we must have $(\partial_1 f)(x_0,y_0) = 0$ and $(\partial_1^2 f)(x_0,y_0) \geq 0$. Similarly, because $f(x_0,y)$ has a minimum at y_0, we must have $(\partial_2 f)(x_0,y_0) = 0$ and $(\partial_2^2 f)(x_0,y_0) \geq 0$.

Now let us find some sufficient conditions for a point X_0 to be a local minimum of f. The key formula is (4.4). We have just showed that $(\nabla f)(X_0)$ must be zero, hence:

$$
\begin{aligned}
& f(x,y) - f(x_0,y_0) \\
& \quad = \frac{1}{2}\Big\{(\partial_1^2 f)(x_0,y_0)(x-x_0)^2 + 2(\partial_{2,1}^2 f)(x_0,y_0)(x-x_0)(y-y_0) \\
& \quad + \ (\partial_2^2 f)(x_0,y_0)(y-y_0)^2\Big\} + \mathcal{O}(||X-X_0||^3). \qquad (5.1)
\end{aligned}
$$

How can we be sure that the right hand side of this equality is non-negative?

1. Assume that both $(\partial_1^2 f)(x_0,y_0) > 0$ and $(\partial_2^2 f)(x_0,y_0) > 0$.
2. We have the obvious double identity:

$$
\begin{aligned}
& (\partial_1^2 f)(x_0,y_0)(x-x_0)^2 + 2(\partial_{2,1}^2 f)(x_0,y_0)(x-x_0)(y-y_0) \\
& + \ (\partial_2^2 f)(x_0,y_0)(y-y_0)^2 \\
& = \left\{\sqrt{(\partial_1^2 f)(x_0,y_0)}(x-x_0) + \frac{(\partial_{2,1}^2 f)(x_0,y_0)}{\sqrt{(\partial_1^2 f)(x_0,y_0)}}(y-y_0)\right\}^2 \\
& + \left\{(\partial_2^2 f)(x_0,y_0) - \frac{[(\partial_{2,1}^2 f)(x_0,y_0)]^2}{(\partial_1^2 f)(x_0,y_0)}\right\}(y-y_0)^2 \qquad (5.2) \\
& = \left\{\sqrt{(\partial_2^2 f)(x_0,y_0)}(y-y_0) + \frac{(\partial_{2,1}^2 f)(x_0,y_0)}{\sqrt{(\partial_2^2 f)(x_0,y_0)}}(x-x_0)\right\}^2 \\
& + \left\{(\partial_1^2 f)(x_0,y_0) - \frac{[(\partial_{2,1}^2 f)(x_0,y_0)]^2}{(\partial_2^2 f)(x_0,y_0)}\right\}(x-x_0)^2. \qquad (5.3)
\end{aligned}
$$

Denote by $a := \max\{(\partial_1^2 f)(x_0,y_0), (\partial_2^2 f)(x_0,y_0)\}$ and by

$$\Delta = (\partial_1^2 f)(x_0,y_0)(\partial_2^2 f)(x_0,y_0) - [(\partial_{1,2}^2 f)(x_0,y_0)]^2.$$

Assume that $\Delta > 0$. Then (5.2) and (5.3) imply the double inequality:

$$
\begin{aligned}
& (\partial_1^2 f)(x_0,y_0)(x-x_0)^2 + 2(\partial_{2,1}^2 f)(x_0,y_0)(x-x_0)(y-y_0) \\
& + \ (\partial_2^2 f)(x_0,y_0)(y-y_0)^2 \\
& \geq \frac{\Delta}{a}(y-y_0)^2 \qquad (5.4)
\end{aligned}
$$

and

$$
\begin{aligned}
& (\partial_1^2 f)(x_0,y_0)(x-x_0)^2 + 2(\partial_{2,1}^2 f)(x_0,y_0)(x-x_0)(y-y_0) \\
& + \ (\partial_2^2 f)(x_0,y_0)(y-y_0)^2 \\
& \geq \frac{\Delta}{a}(x-x_0)^2. \qquad (5.5)
\end{aligned}
$$

If we sum up (5.4) and (5.5), we get:

$$\begin{aligned}
&(\partial_1^2 f)(x_0, y_0)(x - x_0)^2 + 2(\partial_{2,1}^2 f)(x_0, y_0)(x - x_0)(y - y_0) \\
&+\ (\partial_2^2 f)(x_0, y_0)(y - y_0)^2 \\
&\geq \frac{\Delta}{2a}||X - X_0||^2. \qquad (5.6)
\end{aligned}$$

Use this in (5.1) and obtain the key inequality:

$$f(X) - f(X_0) \geq \frac{\Delta}{2a}||X - X_0||^2 + \mathcal{O}(||X - X_0||^3). \qquad (5.7)$$

This means that if $||X - X_0||$ is amll enough, the right hand side will be non-negative, therefore $f(X) \geq f(X_0)$ for every X in a small disk around X_0.

CONCLUSION: in order for a point X_0 to be a local minimum for f, the following two conditions are sufficient:

I. $(\partial_1^2 f)(x_0, y_0) > 0$ and $(\partial_2^2 f)(x_0, y_0) > 0$;

II. $\Delta = (\partial_1^2 f)(x_0, y_0)(\partial_2^2 f)(x_0, y_0) - [(\partial_{1,2}^2 f)(x_0, y_0)]^2 > 0$.

In a similar manner, we can establish two sufficient conditions for a point X_0 to be a local maximum, i.e. $f(X) \leq f(X_0)$ for every X in a small disk around X_0:

I. $(\partial_1^2 f)(x_0, y_0) < 0$ and $(\partial_2^2 f)(x_0, y_0) < 0$;

II. $\Delta = (\partial_1^2 f)(x_0, y_0)(\partial_2^2 f)(x_0, y_0) - [(\partial_{1,2}^2 f)(x_0, y_0)]^2 > 0$. We see that the second condition is the same.

If $(\partial_1^2 f)(x_0, y_0)$ and $(\partial_2^2 f)(x_0, y_0)$ are non-zero but have opposite signs, then X_0 is a saddle-point (i.e. a local minimum for one variable and a local maximum for the other one).

Exercise 5.1. *Investigate Example 3, page 283 in McQuarrie.*

6 Scalar functions with n variables

Let $n \geq 1$. A vector X in $\mathbb{R}^n$ will have the components $[x_1, \ldots, x_n]$. Consider the scalar field $\phi : \mathbb{R}^n \to \mathbb{R}$, which we want to approximate near the point $X_0 = [x_{1,0}, \ldots, x_{n,0}]$. Here the smallness parameter will be

$$||X - X_0|| = \sqrt{\sum_{i=1}^{n} (x_i - x_{i,0})^2}.$$

Reasoning in the same way as we did in order to obtain (4.4), we can generalize that identity in the following way:

$$\begin{aligned}
\phi(X) = \phi(X_0) + \sum_{i=1}^{n} (\partial_i \phi)(X_0)\ (x_i - x_{i,0}) + \frac{1}{2}\sum_{i=1}^{n} (\partial_{i,i}^2 \phi)(X_0)\ (x_i - x_{i,0})^2 \\
+ \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (\partial_{i,j}^2 \phi)(X_0)\ (x_i - x_{i,0})(x_j - x_{j,0}) + \mathcal{O}(||X - X_0||^3). \qquad (6.1)
\end{aligned}$$

Here we used the symmetry property of second order partial derivatives $\partial^2_{i,j}\phi(X_0) = \partial^2_{j,i}\phi(X_0)$.

In order to write (6.1) in a more compact way, we introduce the Hessian matrix associated to ϕ by

$$H_\phi(X) = \{\partial^2_{i,j}\phi(X)\}_{1\leq i\leq j\leq n}. \tag{6.2}$$

Then (6.1) can be rewritten as:

$$\phi(X) \approx \phi(X_0) + \nabla\phi(X_0)\cdot(X-X_0) + \frac{1}{2}\sum_{i=1}^{n}(x_i - x_{i,0})[H_\phi(X_0)]_{i,j}(x_j - x_{j,0})$$
$$= \phi(X_0) + \nabla\phi(X_0)\cdot(X-X_0) + \frac{1}{2}(X-X_0)\cdot\{[H_\phi(X_0)](X-X_0)\}. \tag{6.3}$$

6.1 Extremum points for functions of several variables

We will now give sufficient conditions such that X_0 to be a maximum (minimum) point for a smooth function ϕ. First, $\nabla\phi(X_0) = [0,\dots,0]$ is a necessary condition. Solving this equation gives us all possible interior extremum points. Now around such a point, (6.3) becomes:

$$\phi(X) = \phi(X_0) + \frac{1}{2}(X-X_0)\cdot\{[H_\phi(X_0)](X-X_0)\} + \mathcal{O}(||X-X_0||^3). \tag{6.4}$$

The Hessian matrix is symmetric (real and equal to its transposed matrix), therefore the Complex Spectral Theorem says that it has exactly n real eigenvalues. Denote these eigenvalues (they can be degenerate) as $\{\lambda_i\}_{i=1}^n$, in increasing order: $\lambda_1 \leq ... \leq \lambda_n$. Then a result in linear algebra (given without proof) states that for every X:

$$\lambda_1||X-X_0||^2 \leq (X-X_0)\cdot\{[H_\phi(X_0)](X-X_0)\} \leq \lambda_n||X-X_0||^2.$$

Thus if $0 < \lambda_1$, then for $||X-X_0||$ small enough we have $\phi(X) \geq \phi(X_0)$, hence X_0 is a local minimum. If $\lambda_n < 0$, then for $||X-X_0||$ small enough we have $\phi(X) \leq \phi(X_0)$, hence X_0 is in this case a local maximum.

7 General vector fields

IN THIS SECTION, ALL VECTORS ARE COLUMN VECTORS!!!! Their transposed * vectors are line vectors.

The most general real vector field is a mapping

$$\vec{F}:\mathbb{R}^n \to \mathbb{R}^m, \quad X = [x_1,\dots,x_n]^*, \quad \vec{F}(X) = [F_1(X),\dots,F_m(X)]^*.$$

Each component of $\vec{F}$ is in fact a scalar field, of the type we studied in the previous section. Let us try to approximate $\vec{F}(X)$ in a small neighborhood of X_0; we only look at the linear approximation.

Using (6.1), we can replace ϕ by any of F_j's and write m equations:

$$F_j(X) = F_j(X_0) + \sum_{i=1}^{n} (\partial_i F_j)(X_0)\ (x_i - x_{i,0}) + \mathcal{O}(||X - X_0||^2). \tag{7.1}$$

In order to write this in a shorter way, we introduce the Jacobi matrix:

$$[J_F(X)]_{j,i} = \partial_i F_j(X),\ 1 \le i \le n,\ 1 \le j \le m, \quad m \times n \text{ type matrix.} \tag{7.2}$$

Then the above m scalar equations can be written as one vector equation:

$$\vec{F}(X) = \vec{F}(X_0) + [J_F(X_0)](X - X_0) + \mathcal{O}(||X - X_0||^2). \tag{7.3}$$

Here all vectors are column vectors. The Jacobi matrix can be considered as the natural generalization of the derivative of a function. Note that if $m = 1$ then

$$[J_\phi(X)] = [\nabla\phi(\mathbf{X})]^*.$$

7.1 The chain rule for vector fields

The chain rule is about computing the Jacobi matrix of a vector field which is defined as the composition of two others. Let $\vec{F} : \mathbb{R}^n \to \mathbb{R}^m$, $\vec{G} : \mathbb{R}^m \to \mathbb{R}^p$ and $\vec{H} : \mathbb{R}^n \to \mathbb{R}^p$ where

$$\vec{H}(X) = \vec{G}(\vec{F}(X)). \tag{7.4}$$

Denote by $\vec{F}(X) = Y$ and $\vec{F}(X_0) = Y_0$. We then have

$$\begin{aligned}\vec{H}(X) &= \vec{H}(X_0) + [J_G(Y_0)](Y - Y_0) + \mathcal{O}(||X - X_0||^2)\\ &= \vec{H}(X_0) + [J_G(Y_0)][J_F(X_0)](X - X_0) + \mathcal{O}(||X - X_0||^2).\end{aligned} \tag{7.5}$$

It follows the identity:

$$[J_H(X_0)] = [J_G(\vec{F}(X_0))][J_F(X_0)], \tag{7.6}$$

which is nothing but the chain rule in the matrix form.

Let us use this in an important example, where $n = p = 1$ and $m > 1$. Consider a vector field $\vec{F} : \mathbb{R} \to \mathbb{R}^m$, a scalar field $\phi : \mathbb{R}^m \to \mathbb{R}$, and their composition:

$$h : \mathbb{R} \to \mathbb{R}, \quad h(t) = \phi(\vec{F}(t)).$$

The Jacobi matrix of h is a 1×1 type, it is just a number which coincides with its usual derivative. Moreover, $[J_F(t)] = [F_1'(t), \dots, F_m'(t)]^*$ is a column vector, i.e. a $m \times 1$ matrix. And $[J_\phi(\vec{F}(t))] = [\nabla\phi(\vec{F}(t))]^*$ is a line vector, i.e. a $1 \times m$ matrix. Thus

$$\begin{aligned}J_h(t) = h'(t) &= [J_\phi(\vec{F}(t))][J_F(t)] = [\nabla\phi(\vec{F}(t))]^*[F_1'(t), \dots, F_m'(t)]^*\\ &= \sum_{i=1}^{m} \{\partial_i \phi(\vec{F}(t))\}\ F_i'(t).\end{aligned} \tag{7.7}$$