Ralph M Kaufmann

# The Linear Algebra Version of the Chain Rule [1]

## Idea

The differential of a differentiable function at a point gives a good linear approximation of the function – by definition. This means that locally one can just regard linear functions. The algebra of linear functions is best described in terms of linear algebra, i.e. vectors and matrices. Now, in terms of matrices the concatenation of linear functions is the matrix product. Putting these observations together gives the formulation of the chain rule as the Theorem that the linearization of the concatenations of two functions at a point is given by the concatenation of the respective linearizations. Or in other words that matrix describing the linearization of the concatenation is the product of the two matrices describing the linearizations of the two functions.

## 1. Linear Maps

Let $V^n$ be the space of n–dimensional vectors.

1.1. **Definition.** A *linear map* $F : V^n \to V^m$ is a rule that associates to each n–dimensional vector $\vec{x} = \langle x_1, \dots x_n \rangle$ an m–dimensional vector $F(\vec{x}) = \vec{y} = \langle y_1, \dots, y_n \rangle = \langle f_1(\vec{x}), \dots, (f_m(\vec{x})) \rangle$ in such a way that:

1) For $c \in \mathbb{R} : F(c\vec{x}) = cF(\vec{x})$
2) For any two n–dimensional vectors $\vec{x}$ and $\vec{x}'$: $F(\vec{x} + \vec{x}') = F(\vec{x}) + F(\vec{x}')$

If $m = 1$ such a map is called a linear function. Note that the *component functions* $f_1, \dots, f_m$ are all linear functions.

1.2. **Examples.**

1) m=1, n=3: all linear functions are of the form
$$y = ax_1 + bx_2 + cx_3$$
for some $a, b, c \in \mathbb{R}$. E.g.: $y = 2x_1 + 15x_2 - 2\pi x_3$

2 m=2, n=3: The linear maps are of the form
$$\begin{aligned} y_1 &= ax_1 + bx_2 + cx_3 \\ y_2 &= dx_1 + ex_2 + fx_3 \end{aligned}$$
for some $a, b, c, d, e, f \in \mathbb{R}$. E.g.: $y_1 = 17x_1 + 15.6x_2 - 3x_3, y_2 = \sqrt{2}x_1 - 5x_2 - \frac{3}{4}x_3$

3) m=3, n=2:The linear maps are of the form
$$\begin{aligned} y_1 &= ax_1 + bx_2 \\ y_2 &= cx_1 + dx_2 \\ y_3 &= ex_1 + fx_2 \end{aligned}$$
for some $a, b, c, d, e, f \in \mathbb{R}$. E.g.: $y_1 = 17x_1 + 2x_2, y_2 = -5x_2, y_3 = -\frac{3}{4}x_1$

4) n=m=2:
$$\begin{aligned} y_1 &= ax_1 + bx_2 \\ y_2 &= bx_1 + cx_2 \end{aligned}$$
for some $a, b, c, d \in \mathbb{R}$. E.g.: $y_1 = 3x_1 + 2x_2, y_2 = x_1$

1.3. **Remark.** If $F : V^k \to V^n$ and $G : V^n \to V^k$ are linear maps the the concatenation $F \circ G$ given by $\vec{x} \mapsto F \circ G(\vec{x}) := F(G(\vec{x}))$ is also a linear map.

## 2. Matrices

2.1. **Definition.** A $m \times n$ matrix is an array of real numbers made up of m rows and n columns. It will be denoted as follows:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

Notice that $a_{ij}$ is the the entry in the i–th row and j–th column of $A$.

We call $m \times 1$ matrices *column vectors* and $1 \times n$ matrices *row vectors*.

2.2. **Examples.**

1) m=1, n=3: The $3 \times 1$ matrices have the following form:

$$\begin{pmatrix} a & b & c \end{pmatrix}$$

for some $a, b, c \in \mathbb{R}$

E.g.: $\begin{pmatrix} 2 & 15 & -2\pi \end{pmatrix}$

2 n= 3, m=2: The matrices have the following form

$$\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$$

for some $a, b, c, d, e, f \in \mathbb{R}$. E.g.:$\begin{pmatrix} 17 & 15.6 & -3 \\ \sqrt{2} & -5 & \frac{3}{4} \end{pmatrix}$

3) m=3, n=2: The matrices have the following form

$$\begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix}$$

for some $a, b, c, d, e, f \in \mathbb{R}$. E.g.: $\begin{pmatrix} 17 & 2 \\ 0 & -5 \\ -\frac{3}{4} & 0 \end{pmatrix}$

4) m=n=2: The matrices have the following form

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

for some $a, b, c, d \in \mathbb{R}$. E.g.: $\begin{pmatrix} 3 & 2 \\ 0 & 1 \end{pmatrix}$

2.3. **Remark.** We can think of a $n \times m$ matrix in two ways: either as a collection of n row vectors or a collection of m column vectors.

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{12} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = \begin{pmatrix} r_1(A) \\ r_2(A) \\ \vdots \\ r_m(A) \end{pmatrix} = \begin{pmatrix} c_1(A) & c_2(A) & \dots & c_n(A) \end{pmatrix}$$

where $r_i(A)$ is the i–th row of $A$ and $c_i(A)$ is the i–th column of $A$.

## 3. Matrix multiplication

**3.1. The dot product.** Given a row vector $u = (u_1 u_2 \dots u_n)$ and a column vector $v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$ we set

$$uv := u_1 v_1 + u_2 v_2 + \dots u_n v_n = \sum_{i=1}^{n} u_i v_i$$

**3.2. Definition.** Given an $m \times k$ matrix $A$ and and $k \times n$ matrix $B$ we define their product

$$AB = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mk} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{12} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \dots & b_{kn} \end{pmatrix}$$

to be the following $n \times m$ matrix

$$AB = \begin{pmatrix} r_1(A)c_1(B) & r_1(A)c_2(B) & \dots & r_1(A)c_n(B) \\ r_2(A)c_1(B) & r_2(A)c_2(B) & \dots & r_2(A)c_n(B) \\ \vdots & \vdots & \ddots & \vdots \\ r_m(A)c_1(B) & r_m(A)c_2(B) & \dots & r_n(A)c_m(B) \end{pmatrix}$$

where again $r_i(A)$ is the i–th row vector of $A$ and $c_j(B)$ is the j–th column vector of $B$.
In other words, if we denote by $(AB)_{ij}$ the entry in the i–th row and j–th column of $AB$ then

$$(AB)_{ij} = r_i(A)c_j(B) = \sum_{s=1}^{k} a_{is} b_{sj} = a_{i1} b_{1j} + a_{i2} b_{2j} + \dots a_{ik} b_{kj}$$

**3.3. Remarks.**

1) Remember that $n \times k$ and $k \times m$ yields $n \times m$. Thus one can think of plumbing pipes: you can plumb them together only if they fit. After fitting them together the ends in the middle are eliminated, leaving only the outer ends.
2) The matrix product is associative.
3) In general, if $AB$ makes sense, then $BA$ does not. If one restricts to square matrices, i.e. $n \times n$ matrices then $AB$ and $BA$ are also $n \times n$ matrices, but even then the matrix product is *not commutative*.

**3.4. Examples.**

1) $\begin{pmatrix} 2 & 3 \\ 1 & 4 \\ 4 & 5 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 3 \\ 1 & -1 & 0 \end{pmatrix} = \begin{pmatrix} 2 \cdot 0 + 3 \cdot 1 & 2 \cdot 1 + 3 \cdot -1 & 2 \cdot 3 + 3 \cdot 0 \\ 1 \cdot 0 + 4 \cdot 1 & 1 \cdot 1 + 4 \cdot -1 & 1 \cdot 3 + 4 \cdot 0 \\ 4 \cdot 0 + 5 \cdot 1 & 4 \cdot 1 + 5 \cdot -1 & 4 \cdot 3 + 5 \cdot 0 \\ 0 \cdot 0 + 0 \cdot 1 & 0 \cdot 1 + 0 \cdot -1 & 0 \cdot 3 + 0 \cdot 0 \end{pmatrix} = \begin{pmatrix} 3 & -1 & 6 \\ 4 & -3 & 3 \\ 5 & -1 & 12 \\ 0 & 0 & 0 \end{pmatrix}$

2) $\begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 3 & 1 & 0 \\ 2 & 1 & 3 \\ 7 & 4 & 0 \end{pmatrix} = \begin{pmatrix} 30 & 14 & 6 \end{pmatrix}$

3) $\begin{pmatrix} 3 & 1 & 0 \\ 2 & 1 & 3 \\ 7 & 4 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 5 & 13 & 15 \end{pmatrix}$

## 4. Linear maps given by matrices

In order to connect the matrix notation with linear maps we think of vectors $\vec{x} \in V^n$ as *column vectors*!

4.1. **Definition.** Given an $m \times n$ matrix $A$ we associate to it the following linear map:

$$F_A(\vec{x}) := A\vec{x} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{12} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{12}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Thus $y_i = \sum_{j=1}^n a_{ij}x_j$.

4.2. **Proposition.** If $F : V^k \to V^n$ is a linear map given by a matrix $A$ and $G : V^n \to V^k$ is a linear map given by a matrix $B$ then concatenation $F \circ G$ is given by the matrix $AB$.

**Proof.**

We set $\vec{y} = G(\vec{x})$ and $\vec{z} = F(\vec{y})$ then $y_s = \sum_{j=1}^n b_{sj}x_j$ and $z_i = \sum_{s=1}^k a_{is}y_s$ and thus $z_i = \sum_{s=1}^k a_{is}(\sum_{j=1}^n b_{sj}x_j) = \sum_{j=1}^n(\sum_{j=1}^k a_{is}b_{sj})x_j$

## 5. The chain rule for maps of several variables

5.1. **Definition.** A map $F$ from $D \subset \mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that associates to each point $\mathbf{x} \in D$ a point $F(\mathbf{x}) = \mathbf{y}$ in $\mathbb{R}^m$. It is given by its component functions: $F = (f_1(x_1, \dots x_n), \dots, f_m(x_1, \dots x_n))$ which are just functions of n variables.

We call a map continuous or differentiable if all of the component functions have this property.

5.2. **Examples.**

1) Polar coordinates: $F(r, \theta) = (r\cos\theta, r\sin\theta)$ with domain $\mathbb{R}^2$ mapping to $\mathbb{R}^2$
2) Spherical coordinates: $F(r, \theta, \phi) = (\rho\cos\theta\sin\phi, \rho\sin\theta\sin\phi, \rho\cos\phi)$ with domain $\mathbb{R}^3$ mapping to $\mathbb{R}^3$
3) Some arbitrary function: e.g. $F(x, y, z) = (x^2 + y, \tan(z)\mathrm{e}^{x+y}, \frac{xy}{z}, xyz)$ with the domain $D = \{(x, y, z) \in \mathbb{R}^3 : z \in (-\pi/2, \pi/2) \setminus \{0\}\}$ mapping to $\mathbb{R}^4$

5.3. **Definition.** Suppose $F = (f_1(x_1, \dots x_n), \dots, f_m(x_1, \dots x_n))$ is a map $D \subset \mathbb{R}^n$ to $\mathbb{R}^m$ from such that all of partial derivatives of its component function $\frac{\partial f_i}{\partial x_j}$ exist at a point $\mathbf{x}_0$. We define the Jacobian of $F$ at $\mathbf{x}_0$ to be the $m \times n$ matrix of all partial differentials at that point

$$J_F(\mathbf{x_0}) := \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}_0) & \frac{\partial f_1}{\partial x_2}(\mathbf{x}_0) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}_0) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{x}_0) & \frac{\partial f_2}{\partial x_2}(\mathbf{x}_0) & \dots & \frac{\partial f_2}{\partial x_n}(\mathbf{x}_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}_0) & \frac{\partial f_m}{\partial x_2}(\mathbf{x}_0) & \dots & \frac{\partial f_m}{\partial x_n}(\mathbf{x}_0) \end{pmatrix}$$

that is the ij-th entry is $(J_F)_{ij}(\mathbf{x}_0) = \frac{\partial f_i}{\partial x_j}(\mathbf{x}_0))$

5.4. **Definition.** The linear approximation $L_F$ of a map $F$ at a point $\mathbf{x}_0$ is given by

$$L_F(\mathbf{x}) = F(\mathbf{x}_0) + J_F(\mathbf{x} - \mathbf{x}_0)$$

5.5. **Examples.**

1) The Jacobian a function of three variables $f(x,y,z)$: $J_F = \nabla f = (f_x \ f_y \ f_z)$ and the linear approximation at $(x_0, y_0, z_0)$ is
$L_F(x,y,z) = f(x_0,y_0,z_0) + f_x(x_0,y_0,z_0)(x-x_0) + f_y(x_0,y_0,z_0)(y-y_0) + f_z(x_0,y_0,z_0)(z-z_0)$ – whose graph is the tangent plane.
2) The Jacobian and the linear approximation at $t_0$ of a vector function $\vec{r}(t) = \langle x(t), y(t), z(t)\rangle$ are $J(\vec{r})(t_0) = \begin{pmatrix} x'(t_0) \\ y'(t_0) \\ z'(t_0) \end{pmatrix}$ and $L(\vec{r})(t) = \vec{r}(t_0) + (t-t_0)\vec{r}'(t_0)$ – the tangent line.
3) The Jacobian of a map $F = (g(x,y), h(x,y))$ from $D \subset \mathbb{R}^2$ to $\mathbb{R}^2$ is: $J_F = \begin{pmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{pmatrix}$
4) The Jacobian of a function $f(x_1, \dots, x_n)$ is $J_f = \nabla f = (\frac{\partial f}{\partial x_1} \ \frac{\partial f}{\partial x_2} \ \cdots \ \frac{\partial f}{\partial x_n})$

5.6. **Theorem.** (The chain rule)
Given two differentiable maps $F : D \to \mathbb{R}^m$, in components $F = (f_1(y_1, \dots y_k), \dots, f_n(y_1, \dots y_k))$, and $G : E \to \mathbb{R}^k$, in components $G = (g_1(x_1, \dots, x_n), \dots, g_k(x_1, \dots, x_n))$, with $E \subset \mathbb{R}^n$ and $D \subset G(E) \subset \mathbb{R}^k$ then

$$J_{F \circ G} = J_F J_G$$

**Proof.** The ij–th entry of $J_{F\circ G}$ is $(J_{F\circ G})_{ij} = \frac{\partial}{\partial x_j}(f_i(g_1(x_1, \dots, x_n), \dots, g_k(x_1, \dots, x_n))$. Setting $\mathbf{y} = G(\mathbf{x})$ and $\mathbf{z} = F(\mathbf{y})$ the chain rule yields $\frac{\partial}{\partial x_j}(f_i(g_1(x_1, \dots, x_n), \dots, g_k(x_1, \dots, x_n)) = \frac{\partial z_i}{\partial y_1}\frac{\partial y_1}{\partial x_j} + \dots + \frac{\partial z_i}{\partial y_k}\frac{\partial y_k}{\partial x_j}$ and this is just the ij–th entry of $J_F J_G$

5.7. **Examples.**

1) $z = f(x,y,z) = f(\mathbf{x})$, $\mathbf{x} = \mathbf{r}(t)$

$$J_{f\circ \mathbf{r}}(x(t_0), y(t_0), z(t_0)) = \nabla f(x(t_0), y(t_0), z(t_0)) \begin{pmatrix} x'(t) \\ y'(t) \\ z'(t) \end{pmatrix}$$

$= f_x(x(t_0), y(t_0), z(t_0))x'(t_0) + f_y(x(t_0), y(t_0), z(t_0))y'(t_0) + f_z(x(t_0), y(t_0), z(t_0))z'(t_0)$
2) $z = f(x,y), x = g(s,t), y = h(s,t)$:

$$J_{f\circ(g,h)} = (\frac{\partial f}{\partial x} \ \frac{\partial f}{\partial y}) \begin{pmatrix} \frac{\partial g}{\partial s} & \frac{\partial g}{\partial t} \\ \frac{\partial h}{\partial s} & \frac{\partial h}{\partial t} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} \\ \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} \end{pmatrix}$$

3) $z = f(x_1, \dots x_n \ x_1 = g_1(t_1, \dots, t_m), x_2 = g_2(t_1, \dots, t_m), \dots, x_n = g_n(t_1, \dots, t_m)$. We set $G = (g_1, \dots g_n)$ and obtain:

$$J_{f\circ G} = \nabla f J_G = (\frac{\partial f}{\partial x_1} \ \frac{\partial f}{\partial x_2} \ \cdots \ \frac{\partial f}{\partial x_n}) \begin{pmatrix} \frac{\partial x_1}{\partial t_1} & \frac{\partial x_1}{\partial t_2} & \cdots & \frac{\partial x_1}{\partial t_n} \\ \frac{\partial x_2}{\partial t_1} & \frac{\partial x_2}{\partial t_2} & \cdots & \frac{\partial x_2}{\partial t_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_m}{\partial t_1} & \frac{\partial x_m}{\partial t_2} & \cdots & \frac{\partial x_m}{\partial t_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t_1} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t_1} + \dots + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial t_1} \\ \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t_2} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t_2} + \dots + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial t_2} \\ \vdots \\ \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t_m} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t_m} + \dots + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial t_m} \end{pmatrix}$$

## 6. Exercises

1) Show that the matrix multiplication is associative.

2) Show that the $n \times n$ matrix with 1s on the diagonal and all other entries 0: $E = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$ is a left and right unit. I.e. for any $n \times n$ matrix $A$ the following holds: $AE = EA = A$.
3) Show that the matrix multiplication of $2 \times 2$ is not commutative. Consider $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ and calculate $AB$ and $BA$.
4) Prove Remark 1.3!
5) Prove that a linear function of $n$ variables is of the form: $a_1x_1 + \ldots + a_nx_n$. (Hint either show that all partial derivatives are constant, or use the linearity and the fact that any vector $\vec{x}$ can be written as $\sum_{i=1}^{n} x_i e_i$ where the $e_i$ are the basis vectors that have all 0 entries except for the i–th one. (In three dimensions these are the vectors $e_1 = i, e_2 = j$ and $e_3 = k$))
6) Show that indeed the component functions of a linear map are linear.
7) Use 5) and 6) to show that any linear function can be written in the form $F(\vec{x}) = A\vec{x}$ for some matrix $A$ and $\vec{x}$ considered as a column vector.
8) Calculate the Jacobian of the functions in the Example 5.2
9) Calculate the Jacobian of the function in Example 5.2 3) written in polar coordinates. I.e. $f(x(r, \theta, z), y(r, \theta, z), z(r, \theta, z))$.
10) Do the same for spherical coordinates: calculate the Jacobian of the function in Example 5.2 3) in spherical coordinates $f(x(\rho, \theta, \phi), y(\rho, \theta, \phi), z(\rho, \theta, \phi))$.