Suggested Problems:

- pages 168-169 numbers 1,3,5a-d,9

REMEMBER? If f and g are functions of one variable x, then $d/dx f(g(x)) = f'(g(x))g'(x)$.

A GENERALISATION. If r is curve (a function of one variable which has values in space) and F is a function of three variables, we get a function of one variables $t \mapsto F(r(t))$. The derivative is $d/dt F(r(t)) = \nabla F(r(t)) \cdot r'(t)$.

INTUITIVE EXAMPLE. If F is the temperature distribution in a room and $r(t)$ is the path of a spider (it is soon Halloween), then $F(r(t))$ is the temperature, the spider experiences at time t. The rate of change depends on the velocity $r'(t)$ of the spider as well as the temperature gradient ∇F and the angle between gradient and velocity. If the spider moves orthogonal to the gradient it moves tangent to a level curve and the rate of change is 0.

WRITING IT OUT. Writing the scalar product out gives

$$\frac{d}{dt} F(r(t)) = F_x(x(t),y(t),z(t))x'(t) + F_y(x(t),y(t),z(t))y'(t) + F_z(x(t),y(t),z(t))z'(t) \ .$$

SPIDER. Assume a spider moves along a circle $r(t) = (\cos(t), \sin(t))$ on a table with temperature distribution $T(x,y) = x^2 - y^3$. Find the rate of change of the temperature, the spider experiences in time.

SOLUTION. $\nabla T(x,y) = (2x, -3y^2)$, $r'(t) = (-\sin(t), \cos(t))$ $d/dt T(r(t)) = \nabla T(r(t)) \cdot r'(t) = (2\cos(t), -3\sin(t)^2) \cdot (-\sin(t), \cos(t)) = -2\cos(t)\sin(t) - 3\sin^2(t)\cos(t)$.

DERIVATIVE. If $f : \mathbf{R}^n \to \mathbf{R}^m$ is a map, its **derivative** f' is defined to be the $m \times n$ matrix $[f']_{ij} = \frac{\partial}{\partial x_j} f_i$.

EXAMPLES OF DERIVATIVES.

$f : \mathbf{R} \to \mathbf{R}^3$	$f : \mathbf{R}^3 \to \mathbf{R}$	$f : \mathbf{R}^2 \to \mathbf{R}^3$	$f : \mathbf{R}^2 \to \mathbf{R}^2$	$f : \mathbf{R}^3 \to \mathbf{R}^3$
curve	scalar function	surface	coordinate change	gradient field
f' velocity vector.	f' gradient vector.	f' tangent matrix	f' Jacobean matrix.	f' Hessian matrix.

THE GENERAL CHAIN RULE.
If $f : \mathbf{R}^n \to \mathbf{R}^m$ and $g : \mathbf{R}^k \to \mathbf{R}^n$, we can compose $f * g$, which is a map from $\mathbf{R}^k$ to $\mathbf{R}^m$. The chain rule expresses the derivative of $f \circ g(x) = f(g(x))$ in terms of the derivatives of f and g.

$$\frac{\partial}{\partial x_j} f(g(x))_i = \sum_n \frac{\partial}{\partial x_n} f_i(g(x)) \frac{\partial}{\partial x_j} g_n(x)$$

We can write this as

$$f'(g(x))g'(x)$$

where $f'(g(x))$ and $g'(x)$ are matrices. The chain rule in higher dimensions looks exactly as the chain rule in one dimension, only that the objects are matrices and the multiplication is matrix multiplication.

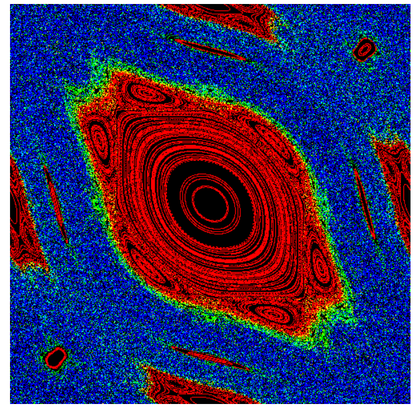
REMARK. An other notation is used used especially in general relativity: if f^i_j denotes the derivative of f^i with respect to x_j, where $f = (f^1, \dots, f^n)$. The chain rule is

$$(f \circ g)^i_j(x) = \sum_n f^i_n(g(x)) g^n_j(x) = f^i_n(g(x)) g^n_j(x) \ .$$

The expression to the right uses automatic summation over the index n which appears in upper and lower places (Einstein convention).

EXAMPLE. GRADIENT IN POLAR COORDINATES. In Polar coordinates, the gradient is $\nabla = (\partial_r, \partial_\theta/r)$. Writing it out: $d/dr f(x(r,\theta), y(r,\theta)) = f_x(x,y)\cos(\theta) + f_y(x,y)\sin(\theta)$ and $d/(rd\theta) f(x(r,\theta), y(r,\theta)) = -f_x(x,y)\sin(\theta) + f_y(x,y)\cos(\theta)$ means (check it!) that the length of ∇f is the same in both coordinate systems.

CHAOS. Consider the map $T : (x,y) \mapsto (2x - y + 2\sin(x), y) = (x_1, y_1)$ in the plane. Because adding a multiple of 2π to x or y has as a result again an addition of a multiple of 2π, we can look at the map on $(0, 2\pi) \times (0 \times 2\pi$. What happens, if we iterate this map? Start with a point (x_0, y_0) and call (x_1, y_1) the image, (x_2, y_2) the image of the image etc. In the picture to the right you see some orbits. Sometimes, the points stay on curves, sometimes, they behave erratically, chaotically. The derivative $T'(x,y)$ is the matrix: $T'(x,y) = \begin{bmatrix} 2+2\cos(x) & -1 \\ 0 & 1 \end{bmatrix}$. By the chain rule, the derivative of $T \circ T$ is $(T \circ T)'(x_0, y_0) = \begin{bmatrix} 2+2\cos(x_1) & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2+2\cos(x_0) & -1 \\ 0 & 1 \end{bmatrix}$ and the derivative of the nth iterate is the product of n matrices $\begin{bmatrix} 2+2\cos(x_{n-1}) & -1 \\ 0 & 1 \end{bmatrix} \dots \begin{bmatrix} 2+2\cos(x_0) & -1 \\ 0 & 1 \end{bmatrix}$.



Experimentally one finds that the first entry A^n_{11} of this matrix grows exponentially in the number of iterations. If this happens on a large set of initial conditions, then one calls the map **chaotic**.

WHERE DO WE NEED THE CHAIN RULE.
While the chain rule is useful in calculations using the composition of functions, the iteration of maps or in doing change of variables, it is also useful for **understanding** some theoretical aspects. Examples (both practical and theoretical):

- The chain rule is used in **change of variable** formulas. For example, if F is a function in the plane and $x = r\cos(\theta), y = r\sin(\theta)$ is a change of coordinates into polar coordinates, then calculating the gradient of F in the new coordinates requires the chain rule.
- Remember the **fundamental theorem for line integrals**: $\int_a^b \nabla U(r(t)) \cdot r'(t)\ dt = U(b) - U(a)$. The chain rule said that inside the integral, we have $d/dt U(r(t))$.
- Gradients are orthogonal to level surfaces: assume we have a curve $r(t)$ on a surface $G(x,y,z) = c$. Because we move on the surface where G is constant, we have $d/dt G(r(t)) = 0$. The chain rule says $\nabla G(r(t)) \cdot r'(t) = 0$. In other words, the gradient of G is orthogonal to the surface $G = const$. (We had argued earlier using the linear approximation of G.)
- The chain rule illustrates also the **Lagrange multiplier** method. If we want to extremize $F(x,y)$ on the constraint $G(x,y)$ and $r(t)$ is a curve on $G(x,y) = c$, then at a critical point, we must have $d/dt F(r(t)) = 0$. This means with the chain rule that the velocity vector $r'(t)$ is orthogonal to $\nabla F(r(t))$. Because the velocity vector is orthogonal to ∇G The vectors ∇G and ∇F have to be parallel.
- In **fluid dynamics**, you often see partial differential equations starting with $u_t + u\nabla u$, where u is the velocity of the fluid. For example, the conservation of momentum of an ideal fluid in the presence of an external force F is $u_t + u \cdot \nabla u = (1/\rho)\nabla p + F$, where p is the pressure and ρ is the density. What does this mean? The term $u_t + u \cdot \nabla u$ is the change of velocity in the coordinate frame of a particle in the fluid. $(u(t,x,y,z) = (u_1(t,x,y,z), u_2(x,y,z), u_3(x,y,z)$ is a vector and ∇u means applying the gradient to each coordinate). For any quantity like pressure, vorticity, density etc. , $Df/Dt = f_t + u \cdot \nabla f$ means the time derivative of f **moving with the fluid**.
- In **chaos theory**, one wants to understand what happens after iterating a map f. If $f^{(n)}(x) = ff^{(n-1)}(x)$, then $(f^{(n)})'(x) = f'(f^{(n-1)}x)f'(f^{(n-2)}(x))...f'(x)$ is a product of matrices. Chaos means that $(f^n)'(x)$ grows exponentially for a large set of x. A measure of chaos is the Lyapunov exponent $\lambda = n^{-1}\log|(f^n)'(x)|$ in the limit $n \to \infty$. If this number is positive, one has sensitive dependence on initial conditions. The map $x \mapsto 4x(1-x)$ on the interval $[0,1]$ for example has this property. Small chaos of order $\lambda = 10^{-15}$ is present in the solar system. Large chaos $\lambda = 10^{15}$ is present in a small volume of argon gas at room temperature.