Vectors, directional derivatives and the chain rule

Sean Fitzpatrick

June 13, 2018

1 Conventions on "free" and "bound" vectors

In this course we meet vectors in a number of different contexts: we've identified points $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ with the corresponding vector $\vec{x} = \langle x_1, \ldots, x_n \rangle$, which we also view as an element of $\mathbb{R}^n$, with the agreement that such vectors are viewed as "bound" to the origin, with their tail at $(0, \ldots, 0)$ and their tip at $\mathbf{x}$. It also becomes necessary to consider "free" vectors; for example, when we want to discuss tangent vectors to a curve or a surface at a point other than the origin. To distinguish, we introduced the notation V^n for the set of all free vectors up to equivalence, meaning that an element of V^n is an arrow in $\mathbb{R}^n$ with a specified magnitude and direction, but no specified location, and the notation $V_{\mathbf{x}}^n$ to denote the space of vectors bound to the point $\mathbf{x} \in \mathbb{R}^n$. Note that for each vector $\vec{v} \in V^n$ and each $\mathbf{x} \in \mathbb{R}^n$ there is a "representative" vector $\vec{v}_{\mathbf{x}} \in \mathbb{R}^n$. We also use the notation $T_{\mathbf{x}}\mathbb{R}^n$ for $V_{\mathbf{x}}^n$, when we want to emphasize the role of bound vectors as tangent vectors. (The T stands for tangent.) Since we identify vectors bound at the origin with points, we have the equality $V_{\mathbf{0}}^n = \mathbb{R}^n$. Note that each space $V_{\mathbf{x}}^n$ is *isomorphic* to $\mathbb{R}^n$ (it's another copy of the same set of vectors, bound at a different point), but that in most cases we only identify points with vectors bound to the origin.

Usually we can rely on context and get away with a fair amount of abuse of notation when it comes to points and vectors in $\mathbb{R}^n$. For example, we make no distinction between parametric curves and vector-valued functions of one variable. If we wanted to be very specific, we might point out the distinction between a parametric curve and a vector field as follows: although both are expressed in terms of vector-valued functions, we require

$$\mathbf{r}(t) \in V_{\mathbf{0}}^n = \mathbb{R}^n$$

for each t in the parameter domain $[a, b]$, while for a vector field we require

$$\vec{F}(\mathbf{x}) \in V_{\mathbf{x}}^n$$

for each $\mathbf{x} \in \mathbb{R}^n$.

Recall that when we want to discuss things like linear transformations or the derivative matrix, we should think of a vector $\vec{v} \in V_{\mathbf{x}}^n$ as a column vector. (See the handout on linear algebra for a review of matrix multiplication and linear transformations.) Given a map $f : D \subset \mathbb{R}^n \to E \subset \mathbb{R}^m$, and a point $\mathbf{x} \in D$, the derivative of f gives us a matrix $Df(\mathbf{x})$ that defines a map

$$Df(\mathbf{x}) : T_{\mathbf{x}}\mathbb{R}^n \to T_{\mathbf{y}}\mathbb{R}^m,$$

1

where $\mathbf{y} = f(\mathbf{x})$. Thus, if f maps points $\mathbf{x} \in \mathbb{R}^n$ to points $\mathbf{y} \in \mathbb{R}^m$, the derivative of f maps tangent vectors at $\mathbf{x}$ to tangent vectors at $\mathbf{y}$.

We've briefly discussed the fact that, in linear algebra, one can prove that any linear transformation T can be expressed in terms of matrix multiplication, and that conversely, multiplication by a matrix determines a linear transformation. The constants in the matrix defining a linear transformation depend on choices of *basis vectors* for $\mathbb{R}^n$ and $\mathbb{R}^m$, but if we agree to stick to the standard unit basis vectors corresponding to the coordinate systems we're using, there's no ambiguity. A change of basis is roughly the same thing as a change of coordinate system, and as we'll see below, changes in coordinate systems can be handled via the chain rule. (We've already seen that a change of coordinates, like rectangular to polar, is not necessarily linear. However, this is at the level of *points*. The chain rule, interpreted in terms of the derivative matrix as above, tells us that a change of coordinates does determine a linear transformation on each of the vector spaces $V_{\mathbf{x}}^n = T_{\mathbf{x}}\mathbb{R}^n$.)

2 Derivatives

Recall that one of the reasons for moving to column vectors is that the definition of the derivative looks more natural that way. We declared that if $U \subseteq \mathbb{R}^n$ is open and $f : U \to \mathbb{R}^m$, then f is differentiable at $\mathbf{x}_0 \in U$ if there exists a linear function $T : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\mathbf{h} \to 0} \frac{\|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - T(\mathbf{h})\|}{\|\mathbf{h}\|} = 0.$$

We proved that the matrix A such that $T(\mathbf{h}) = A\mathbf{h}$ is the *derivative* of f at $\mathbf{x}_0$, which is the matrix of partial derivatives of f evaluated at $\mathbf{x}_0$:

$$\mathbf{D}f(\mathbf{x}_0) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1}(\mathbf{x}_0) & \cdots & \dfrac{\partial f_1}{\partial x_n}(\mathbf{x}_0) \\ \vdots & & \vdots \\ \dfrac{\partial f_m}{\partial x_1}(\mathbf{x}_0) & \cdots & \dfrac{\partial f_m}{\partial x_n}(\mathbf{x}_0) \end{bmatrix},$$

where the $f_i(\mathbf{x})$ are the components of $f(\mathbf{x}) \in \mathbb{R}^m$. In the special case that f is real-valued ($m = 1$, so $f : U \subset \mathbb{R}^n \to \mathbb{R}$), the derivative $\mathbf{D}f(\mathbf{x}_0)$ gives us the linear function $T : \mathbb{R}^n \to \mathbb{R}$, so $\mathbf{D}f(\mathbf{x}_0)$ is a $1 \times n$ matrix (which we might call a "row vector"). If we want to insist that the gradient of f at $\mathbf{x}_0$ be a vector — that is, a column vector — then we should define

$$\nabla f(\mathbf{x}_0) = (\mathbf{D}f(\mathbf{x}_0))^T,$$

where for any $m \times n$ matrix A, the *transpose* A^T of an $m \times n$ matrix A is the $n \times m$ matrix whose columns are the rows of A (and conversely, the rows of A^T are the columns of A).

We should remark that in more advanced mathematics, it may not be desirable (or even possible) to have a fixed choice of coordinates or a dot product. (This is the case in the study of non-Euclidean geometries, and with General Relativity, where the analogue of the dot product is determined locally by the physics.) In such settings, the function $T : \mathbb{R}^n \to \mathbb{R}$ giving the derivative of f at a point is still well-defined, even though the gradient will depend on choices we make. We'll say a bit more about this below.
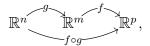
3   The Chain Rule

Given differentiable functions $g : \mathbb{R}^n \to \mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R}^p$ such that the composition $f \circ g : \mathbb{R}^n \to \mathbb{R}^p$ is defined, the chain rule tells us that $f \circ g$ is also differentiable, with derivative given by

$$\mathbf{D}(f \circ g)(\mathbf{x}) = \mathbf{D}f(g(\mathbf{x}))\mathbf{D}g(\mathbf{x}),$$
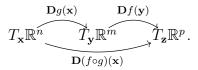
where the product on the right-hand side is matrix multiplication. On way to think about this is as follows: let $\mathbf{x} \in \mathbb{R}^n$ be a point in the domain of g. We know that the function g will take $\mathbf{x}$ to a point $\mathbf{y}$ in $\mathbb{R}^m$, and then f in turn will take $\mathbf{y}$ to some point $\mathbf{z}$ in $\mathbb{R}^p$. Now at each such point $\mathbf{x} \in \mathbb{R}^n$ we have the tangent space $T_{\mathbf{x}}\mathbb{R}^n$ at $\mathbf{x}$ as discussed above. (Soon, we will want to think of the tangent space as the space of all vectors along which we could compute a directional derivative.) Let $T_f : \mathbb{R}^n \to \mathbb{R}^m$ be the linear function given by the derivative of f, and let $T_g : \mathbb{R}^m \to \mathbb{R}^p$ be the linear function given by the derivative of g. The chain rule is then essentially telling us that *the derivative of a composition is the composition of the derivatives*: we have

$$T_f \circ T_g(\vec{v}) = T_f(T_g(\vec{v})) = \mathbf{D}f(\mathbf{y})(\mathbf{D}g(\mathbf{x})\vec{v}) = (\mathbf{D}f(g(\mathbf{x}))\mathbf{D}g(\mathbf{x}))\vec{v} = T_{f \circ g}(\vec{v}).$$

In other words, given the composition

$$\mathbb{R}^n \xrightarrow{\;g\;} \mathbb{R}^m \xrightarrow{\;f\;} \mathbb{R}^p \underset{f \circ g}{},$$

we have the corresponding composition

$$T_{\mathbf{x}}\mathbb{R}^n \xrightarrow{\mathbf{D}g(\mathbf{x})} T_{\mathbf{y}}\mathbb{R}^m \xrightarrow{\mathbf{D}f(\mathbf{y})} T_{\mathbf{z}}\mathbb{R}^p \underset{\mathbf{D}(f \circ g)(\mathbf{x})}{}.$$

We also know that this even makes sense if we restrict to subsets of each space. For example, if $S \subset \mathbb{R}^3$ is a surface and $f : \mathbb{R}^3 \to \mathbb{R}^m$ is defined on some open set U containing S, then for each $\mathbf{x} \in S$, $\mathbf{D}f(\mathbf{x})$ maps the tangent plane to S at $\mathbf{x}$ to the tangent space of $f(S) = \{f(\mathbf{s}) \in \mathbb{R}^m | \mathbf{s} \in S\}$ at $f(\mathbf{x})$ (you proved this for curves on an assignment).

The chain rule is also necessary to handle changes of coordinates. In the physical world there is no natural coordinate system, so (for example) two people setting up the same experiment might choose different coordinates with which to describe their results. Let's say the goal of the experiment is to measure a certain quantity given by some function of the coordinates chosen. Then our two experimenters may come up with very different looking functions to describe the same measurement, and they will need a way to compare their results. Of course, if all they need to do is calculate some quantity, they can compare their answers. But they may also be interested in things such as how this quantity is changing with time, or varies in space. In order to guarantee that they will get the same answers, the chain rule is needed.

Let's suppose that we have two coordinate systems on $\mathbb{R}^n$, and $F : \mathbb{R}^n \to \mathbb{R}^n$ is a mapping that converts one coordinate system into the other. Given a function f defined in terms of

the second coordinate system, we obtain a new function g defined in terms of the first[1], given by $g(\mathbf{x}) = f(\mathbf{y})$, where $\mathbf{y} = F(\mathbf{x})$. We can sum up the situation with the following diagram:

$$\begin{array}{ccc} \mathbf{x} \in \mathbb{R}^n & \xrightarrow{\;\;F\;\;} & \mathbf{y} \in \mathbb{R}^n \\ \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle f} \\ g(\mathbf{x}) \in \mathbb{R} & =\!\!=\!\!= & f(\mathbf{y}) \in \mathbb{R} \end{array} \quad .$$

For example, let's suppose we are interested in the function $f(x,y) = x^2 + y^2 - 2x + 1$ defined on $\mathbb{R}^2$ in the usual Cartesian coordinates, but our friend wishes to work in terms of polar coordinates. Then they will be interested in the function $g(r, \theta) = r^2 - 2r\cos\theta + 1$, which looks very different but computes the same thing. (One could even say that as functions on the *points* in $\mathbb{R}^2$, they are the same function!) Now, let's say we each want to compute a result which requires finding the gradient. We let $F(r, \theta) = (r\cos\theta, r\sin\theta)$, which is a map from $\mathbb{R}^2 \to \mathbb{R}^2$, such that $g(r, \theta) = f(F(r, \theta))$. The chain rule then tells us that

$$\begin{aligned} \mathbf{D}g(r, \theta) &= \mathbf{D}f(x, y)\mathbf{D}F(r, \theta) \\ &= \begin{bmatrix} 2x - 2 & 2y \end{bmatrix} \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{bmatrix} \\ &= \begin{bmatrix} 2r\cos\theta - 2 & 2r\sin\theta \end{bmatrix} \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix} \\ &= \begin{bmatrix} 2r\cos^2\theta - 2\cos\theta + 2r\sin^2\theta & -2r^2\sin\theta\cos\theta + 2r\sin\theta + 2r^2\sin\theta\cos\theta \end{bmatrix} \\ &= \begin{bmatrix} 2r - 2\cos\theta & 2r\sin\theta \end{bmatrix}, \end{aligned}$$

which of course is the same as if we simply computed the gradient of g directly. Although computing the gradient of g directly is less work, the importance of the chain rule here is that it guarantees that if we choose to work with f and Cartesian coordinates, we have a way of converting all of our results to the new coordinate system for comparison.

4 Directional derivatives, gradients, and differentials

Recall that the directional derivative of a function $f : U \subseteq \mathbb{R}^n$ in the direction of a vector $\vec{v} \in \mathbb{R}^n$ is given by

$$D_{\vec{v}}f(\mathbf{x}) = \frac{1}{\|\vec{v}\|} \lim_{h \to 0} \frac{f(\mathbf{x} + t\vec{v}) - f(\mathbf{x})}{h},$$

and given by the formula $D_{\vec{v}}f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \dfrac{\vec{v}}{\|\vec{v}\|}$, where $\nabla f(\mathbf{x})$ is the gradient of f. Notice that geometrically, we identify $\mathbf{x} \in \mathbb{R}^n$ with the vector pointing from the origin to $\mathbf{x}$ (which we could denote by $\vec{x}$ if we wanted to distinguish between the two, but we'll stick with $\mathbf{x}$ for simplicity), and think of $t\vec{v}$ as being added tip-to-tail with $\mathbf{x}$, so it's actually more natural to think of $\vec{v}$ as an element of the tangent space $T_{\mathbf{x}}\mathbb{R}^n$ of vectors with their tails at $\mathbf{x}$.

---

[1]The function g is sometimes called the *pullback of f by F, and denoted by $g = F^*f$.*

As mentioned above, the gradient and the dot product depend on the coordinate system we're using to describe our function f, so we might expect that the directional derivative depends on our choice of coordinate system as well. However, a change of coordinates induces a change of basis on the tangent space $T_{\mathbf{x}}\mathbb{R}^n$ consistent with the chain rule, and it turns out that the chain rule comes to the rescue, making all the changes fit together in such a way that the directional derivative remains unchanged.

Remark 4.1. We are in fact going to adjust our definition of the directional derivative slightly, and allow the length of the vector $\vec{v}$ to influence the result: we will take our directional derivative to be

$$D_{\vec{v}} f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \vec{v},$$

even if $\|\vec{v}\| \neq 1$. This is because we want to shift our point of view below and start thinking of vectors as *differential operators*, and we want vectors of different lengths to correspond to different operators. (Also, in general the length of a vector depends on the choice of dot product and may not even be defined in an abstract setting. If we want to work in a setting where we care about lengths, we would need to specify a dot product on each of the tangent spaces.)

Let f be a function defined on some open set $V \subset \mathbb{R}^n$, and let $F : U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^n$ be a continuously differentiable function defining a change of coordinates. Let f be a differentiable function defined on V. Given F and f, we can define a function g on U as follows: For each $\mathbf{y} = F(\mathbf{x})$ we define $g(\mathbf{x}) = f(F(\mathbf{x})) = f(\mathbf{y})$. For example, F could be the polar coordinate transformation $F(r, \theta) = (r \cos\theta, r \sin\theta)$, and given $f(x, y)$ we obtain $g(r, \theta) = f(r \cos\theta, r \sin\theta)$. Let $\vec{v} \in T_{\mathbf{x}}\mathbb{R}^n$ be a given vector. (Keep in mind that $\vec{v}$ is just a usual vector in $\mathbb{R}^n$, once we translate $\vec{v}$ to the origin.) Then we know that

$$D_{\vec{v}} g(\mathbf{x}) = \nabla g(\mathbf{x}) \cdot \vec{v} = \sum_{i=1}^{n} \frac{\partial g}{\partial x_i}(\mathbf{x}) v_i,$$

where $\vec{v} = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}^T$.[2] Under the change of coordinates F, the vector $\vec{v}$ transforms according to the *derivative* of F at $\mathbf{x}$, giving us the vector

$$\vec{w} \in T_{\vec{y}}\mathbb{R}^n, \quad \vec{w} = \mathbf{D}F(\mathbf{x})\vec{v}.$$

Note that the components of $\vec{w}$ are

$$w_i = \sum_{j=1}^{n} \frac{\partial F_i}{\partial x_j}(\mathbf{x}) v_j,$$

according to the rules for matrix multiplication. Therefore, we should expect that the directional derivative $D_{\vec{v}} g(\mathbf{x})$ is the same as the directional derivative $D_{\vec{w}} f(\mathbf{y})$ after our change

---

[2]Since we want to think of $\vec{v}$ as a column vector, but column vectors don't fit all that nicely into a line of text, while row vectors do, we write $\vec{v}$ as the transpose of a row vector.

of coordinates. Indeed, we have

$$\begin{aligned}
D_{\vec{v}}g(\mathbf{x}) &= \nabla g(\mathbf{x}) \cdot \vec{v} \\
&= \mathbf{D}g(\mathbf{x})\vec{v} \\
&= [\mathbf{D}f(\mathbf{y})\mathbf{D}F(\mathbf{x})]\vec{v} \\
&= \mathbf{D}f(\mathbf{y})[\mathbf{D}F(\mathbf{x})\vec{v}] \\
&= \mathbf{D}f(\mathbf{y})\vec{w} = D_{\vec{w}}f(\mathbf{y}),
\end{aligned}$$

where we have used the fact that for any two vectors $X, Y \in \mathbb{R}^n$, we have $X \cdot Y = X^T Y$, and recall that the gradient ∇g was defined as the transpose of the derivative $\mathbf{D}g$. In fact, it is common practice to *identify* the space $T_{\mathbf{x}}\mathbb{R}^n$ of tangent vectors at $\mathbf{x} \in \mathbb{R}^n$ with their corresponding directional derivatives! Thus, we would write $\vec{v}(f(\mathbf{x})) = D_{\vec{v}}f(\mathbf{x})$, and we could think of $\vec{v}$ as the "vector"

$$\vec{v} = \sum_{i=1}^{n} v_i \frac{\partial}{\partial x_i}.$$

Under a change of variables $\mathbf{x} \mapsto \mathbf{y}$, the components v_i change according to $v_i \mapsto w_i = \sum_{i=1}^{n} \frac{\partial y_i}{\partial x_j}v_j$, so that

$$\vec{v}(f(\mathbf{x})) = \sum_{i=1}^{n} v_i \frac{\partial f}{\partial x_i}(\mathbf{x}) = \sum_{i=1}^{n} v_i \sum_{j=1}^{n} \frac{\partial f}{\partial y_j}(\mathbf{y})\frac{\partial y_j}{\partial x_i}(\mathbf{x}) = \sum_{j=1}^{n} w_j \frac{\partial f}{\partial y_j}(\mathbf{y}) = \vec{w}(f(\vec{y})).$$

Remark 4.2. If we had specified a dot product on each of the tangent spaces $T_{\mathbf{x}}\mathbb{R}^n$, then it would be possible to use our definition of directional derivative from class, where we divide by the length of $\vec{v}$, since we can use the dot product to define length. In this case, the above argument would fail for a general transformation, since $\|\vec{w}\| \neq \|\vec{v}\|$, where $\vec{w} = DF(\mathbf{x})\vec{v}$, for most transformations F. If we insist that $DF(\mathbf{x})$ preserve the dot product as well, so that $\|\vec{w}\| = \|\vec{v}\|$, then we would call a transformation F with this property an **isometry**.

Remark 4.3. In more advanced texts (primarily in geometry or mathematical physics) a more abstract definition of a vector field is given, which relies on the interpretation of a vector field as a differential operator. A vector field $\vec{v}$ on a manifold M (think of M as a subset of $\mathbb{R}^n$ with a well-defined tangent space at each point) is *defined* as a function $\vec{v} : C^{\infty}(M) \to C^{\infty}(M)$ [3] such that

1. $\vec{v}(cf) = c\vec{v}(f)$ for all $c \in \mathbb{R}$ and $f \in C^{\infty}(M)$

2. $\vec{v}(f + g) = \vec{v}(f) + \vec{v}(g)$ for all $f, g \in C^{\infty}(M)$

3. $\vec{v}(fg) = f\vec{v}(g) + \vec{v}(f)g$ for all $f, g \in C^{\infty}(M)$.

Properties 1 and 2 tell us that $\vec{v}$ is a linear map, and property 3 is known as the Leibniz property (a.k.a. the product rule).

---

[3]The set $C^{\infty}(M)$ denotes the space of all functions with infinitely-many continuous partial derivatives. We can also define vector fields on the sets $C^k(M)$ of functions with k continuous partial derivatives, but in this case the range of $\vec{v}$ would be $C^{k-1}(M)$, since after taking one derivative there would be $k-1$ left.

Recall that we defined the *differential* of a function $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ to be the expression

$$df(\mathbf{x}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x})dx_i,$$

which can be interpreted as a way of telling us how small changes in each of the variables x_i results in a small change in f. In geometry, differentials are interpreted somewhat differently. First, we note that the differential contains the same information as the gradient; namely, the values of all the partial derivatives of f. Recall also that the gradient has the problem of depending on choices that we make. The differential, it turns out, does not have this problem. Suppose $\mathbf{y} = F(\mathbf{x})$ is a change of coordinates, and let $g(\mathbf{x}) = f(\mathbf{y})$. With respect to the $\mathbf{y}$ coordinates $(y_1, \ldots y_n)$, we would write

$$df(\mathbf{y}) = \sum_{i=1}^{n} \frac{\partial f}{\partial y_i}(\mathbf{y})dy_i.$$

Now each y_i is a function of $\mathbf{x}$, so we have

$$y_i = \sum_{j=1}^{n} \frac{\partial y_i}{\partial x_j}(\mathbf{x})dx_j,$$

and by the chain rule,

$$\frac{\partial f}{\partial y_i}(\mathbf{y}) = \sum_{k=1}^{n} \frac{\partial g}{\partial x_k}(\mathbf{x})\frac{\partial x_k}{\partial y_i}(\mathbf{y}).$$

(Notice that here we need to assume that F is invertible, since we're using the chain rule going in the opposite direction.) If we substitute these last two expressions into the expression for $df(\mathbf{y})$, we get (after a bit of rearranging)

$$df(\mathbf{y}) = \sum_{j=1}^{n} \left(\sum_{k=1}^{n} \frac{\partial g}{\partial x_k}(\mathbf{x}) \sum_{i=1}^{n} \frac{\partial x_k}{\partial y_i}\frac{\partial y_i}{\partial x_j} \right) dx_j.$$

But $\displaystyle\sum_{i=1}^{n} \frac{\partial x_k}{\partial y_i}\frac{\partial y_i}{\partial x_j} = \frac{\partial x_k}{\partial x_j} = \delta_{kj}$, where δ_{kj} is the *Kronecker delta*, which is zero if $k \neq j$ and one if $k = j$, and so the above sum collapses, giving

$$df(\mathbf{y}) = \sum_{j=1}^{n} \frac{\partial g}{\partial x_j}(\mathbf{x})dx_j = dg(\mathbf{x}).$$

So, we see that df has the same expression in either coordinate system. (We used $g = f \circ F$ to distinguish between the two sets of coordinates, but since differential looks the same in any such system, convention is to identify g with f, since they are really the same function on the underlying set of points.) Now, we will view df as a "dual vector", by making it into a linear function from the tangent space $T_{\mathbf{x}}\mathbb{R}^n$ to $\mathbb{R}$. To do this, we define

$$df(\mathbf{x})(\vec{v}) = \vec{v}f(\mathbf{x}) = D_{\vec{v}}f(\mathbf{x})$$

for any $\vec{v} \in T_{\mathbf{x}}\mathbb{R}^n$. Notice that if we think of the partial differential operators $\dfrac{\partial}{\partial x_i}$ as a basis for $T_{\mathbf{x}}\mathbb{R}^n$, then the differentials dx_j satisfy $dx_j(\dfrac{\partial}{\partial x_i}) = \dfrac{\partial x_j}{\partial x_i} = \delta_{ij}$. In the language of linear algebra, the dx_j form a *dual basis* to the basis for $T_{\mathbf{x}}\mathbb{R}^n$. In fact, we get a new vector space

$$T_{\mathbf{x}}^*\mathbb{R}^n = \{\alpha = \sum_{i=1}^{n} a_j dx_j | a_j \in \mathbb{R}, j = 1, \ldots, n\}$$

called the *cotangent space* at $\mathbf{x}$, whose elements are called cotangent vectors. For an abstract vector space V, recall that the dual space is defined as the space

$$V^* = \{\varphi : V \to \mathbb{R} | \varphi(a\vec{v} + b\vec{w}) = a\varphi(\vec{v}) + b\varphi(\vec{w})\}.$$

If V is equipped with a dot product, we showed on an earlier assignment that each $v \in V$ has a corresponding dual vector $\varphi_{\vec{v}}$ such that $\varphi_{\vec{v}}(\vec{w}) = \vec{v} \cdot \vec{w}$, and that this establishes an isomorphism between V and V^*. In particular, the spaces $T_{\mathbf{x}}\mathbb{R}^n$ and $T_{\mathbf{x}}^*\mathbb{R}^n$ are dual vector spaces, and the usual dot product on $\mathbb{R}^n$ gives us a correspondence between vectors and dual vectors: if $\vec{v} = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}^T$, we get the covector $\alpha_{\vec{v}} = \sum_{i=1}^{n} v_i dx_i$, such that $\alpha_{\vec{v}}(\vec{w}) = \vec{v} \cdot \vec{w}$.