In vector calculus for Euclidean space, people often use the word "gradient" and the symbol ∇f to mean the *vector* whose components are $\partial_\mu f$. Then the directional derivative of f along a vector v is given by:

$(\nabla f) \cdot v$

This notion of "gradient" involves two uses of the metric, which cancel each other out, in a sense.

1. To construct a vector from the components $\partial_\mu f$, you have to use the metric: $(\nabla f)^\nu = g^{\mu\nu} \partial_\mu f$
2. To construct the dot-product, you use the metric again: $(\nabla f) \cdot v = g_{\mu\nu} (\nabla f)^\mu v^\nu$

But the metric drops out of the final result:
$(\nabla f) \cdot v = g_{\mu\nu} (\nabla f)^\mu v^\nu = g_{\mu\nu} (g^{\mu\lambda} \partial_\lambda f) v^\nu = (g_{\mu\nu} g^{\mu\lambda}) \partial_\lambda f v^\nu = \delta^\lambda_\nu \partial_\lambda v^\nu = \partial_\nu v^\nu$

As a vector, ∇f is perpendicular to the surfaces of constant f, the "level curves" of f.

The gradient in a general coordinate system depends on the metric tensor but the Jacobian matrix consists of only the partial derivatives.

The gradient of a vector field is given by:

$$\nabla \mathbf{f} = g^{jk} \frac{\partial f^i}{\partial x^j} \mathbf{e}_i \otimes \mathbf{e}_j,$$

where the Einstein summation notation is implied, g^{jk} are the metric tensor elements evaluated from the Jacobian matrix consisting of the partial derivatives of the coordinate transformation from the Cartesian coordinate system. In Cartesian coordinate system, this equals to exactly the transpose of the Jacobian matrix, which is given by, regardless of the metric tensor,

$$\{\boldsymbol{J}\mathbf{f}\}_{i,j} = \frac{\partial f^i}{\partial x^j}.$$

For example, for the spherical coordinate system with the coordinate $\mathbf{z}$, we have

$x_0 = z_0 \sin z_1 \sin z_2,\ x_1 = z_0 \cos z_1 \sin z_2,\ x_2 = z_0 \cos z_2$, where $\mathbf{x}$ is the coordinate in the Cartesian coordinate system.

Given a vector function f, Each column of gradient is given by

$$\{\nabla \mathbf{f}\}_{:,i} = \frac{\partial f^i}{\partial \mathbf{z}_0} \hat{\mathbf{z}}_0 + \frac{1}{\mathbf{z}_0} \frac{\partial f^i}{\partial \mathbf{z}_1} \hat{\mathbf{z}}_1 + \frac{1}{\mathbf{z}_0 \sin \mathbf{z}_2} \frac{\partial f^i}{\partial \mathbf{z}_2} \hat{\mathbf{z}}_2$$

or

$$\begin{bmatrix} \frac{\partial f^i}{\partial \mathbf{z}_0} & \frac{1}{\mathbf{z}_0} \frac{\partial f^i}{\partial \mathbf{z}_1} & \frac{1}{\mathbf{z}_0 \sin \mathbf{z}_2} \frac{\partial f^i}{\partial \mathbf{z}_2} \end{bmatrix}^{\mathrm{T}}$$

but the each row of Jacobian is still given by

$$\{\boldsymbol{J}\mathbf{f}\}_{i,:} = \begin{bmatrix} \frac{\partial f^i}{\partial \mathbf{z}_0} & \frac{\partial f^i}{\partial \mathbf{z}_1} & \frac{\partial f^i}{\partial \mathbf{z}_2} \end{bmatrix}$$

The reason for this is because the Jacobian matrix is applied to solve integrals by substitution where the determinant of the Jacobian matrix is needed. It is also used to transform partial derivatives into partial derivatives of another coordinate system. Another application is to evaluate the metric tensor as mentioned before.

I will use the notation $f_{,i}$ for the ith partial derivative of a function $f:\mathbb{R}^n\to\mathbb{R}$. In differential geometry, partial derivatives are defined using coordinate systems (charts). If M is a smooth manifold, $p\in U\subseteq M$, and $x:U\to\mathbb{R}^n$ is a coordinate system, then we define $\frac{\partial}{\partial x^i}\big|_p$ by

$$\frac{\partial}{\partial x^i}\bigg|_p f=(f\circ x^{-1})_{,i}(x(p))$$

for all smooth functions $f:M\to\mathbb{R}$. The notation $\left(\frac{\partial}{\partial x^i}\right)_p$ can be used instead of $\frac{\partial}{\partial x^i}\big|_p$, and the notation $\frac{\partial f(p)}{\partial x^i}$ can be used instead of $\frac{\partial}{\partial x^i}\big|_p f$.

The gradient of a differentiable function $f:\mathbb{R}^n\to\mathbb{R}$ is the function $\nabla f:\mathbb{R}^n\to\mathbb{R}^n$ defined by $\nabla f(x)=(f_{,1}(x),\dots,f_{,n}(x))$ for all $x\in\mathbb{R}^n$. The right-hand side can be rewritten using the differential geometry definition of partial derivative, if we use the fact that the identity map on $\mathbb{R}^n$ is a coordinate system. I will denote the identity map by I. We have

$$\nabla f(x)=(f_{,1}(x),\dots,f_{,n}(x))=\big((f\circ I^{-1})_{,1}(I(x)),\dots,(f\circ I^{-1})_{,n}(I$$

This is the n-tuple of components of the cotangent vector $(df)_x$ in the coordinate system I, since

$$(\mathrm{d}f)_x=\left((df)_x\frac{\partial}{\partial I^i}\bigg|_x\right)(dI^i)_x=\left(\frac{\partial}{\partial I^i}\bigg|_x f\right)(dI^i)_x.$$

This is one reason to think of the gradient of f as a cotangent vector, specifically as the cotangent vector $(\mathrm{d}f)_x$.

Another reason is that we can interpret the formula for $\nabla f(x)$ above as associating an n-tuple *with each coordinate system*. The right-hand side is the n-tuple associated with the coordinate system I. To get the n-tuple associated with an arbitrary coordinate system J, just make the substitution $I\to J$. Now that we have an n-tuple associated with each coordinate system, we can investigate how they're related to each other, i.e. we can investigate how an n-tuple "transforms" under a change of coordinates. If you know that $\left(\frac{\partial}{\partial J^1}\big|_x,\dots,\frac{\partial}{\partial J^1}\big|_x\right)$ is an ordered basis for the tangent space at x, and that "transforms covariantly" means "transforms in the same way as the ordered basis", then you should see that it follows almost immediately that our n-tuple $\left(\frac{\partial}{\partial J^1}\big|_x f,\dots,\frac{\partial}{\partial J^1}\big|_x f\right)$ "transforms covariantly" (unlike n-tuples of components of tangent vectors, which "transform contravariantly").

Suppose you have some weird coordinate system with two coordinates, u and v. You want to know the distance between point A and point B is. If you were using cartesian coordinates, then the distance would be given by: $D = \sqrt{\delta u^2 + \delta v^2}$, where δu is the change in the u coordinate in going from A to B, and δv is the change in the v coordinate. But what about polar coordinates? In terms of r and θ, the distance between A and B is given approximately (when A and B are very close together) by:

$$D = \sqrt{\delta r^2 + r^2 \delta\theta^2}$$

In general, for points that are close together, the distance will be given by: $D^2 = g_{uu}\delta u^2 + g_{uv}\delta u\delta v + g_{vu}\delta v\delta u + g_{vv}\delta v^2$. Those four numbers, $g_{uu}, g_{uv}, g_{vu}, g_{vv}$ are the components of the "metric tensor" in the $u - v$ coordinate system.

In cartesian coordinates x, y, it's trivial: $g_{xx} = 1, g_{xy} = 0, g_{yx} = 0, g_{yy} = 0$. But in polar coordinates, it's a little more interesting: $g_{rr} = 1, g_{r\theta} = 0, g_{\theta r} = 0, g_{\theta\theta} = r^2$.

The metric tensor is how you compute dot-products of two vectors: $\vec{A} \cdot \vec{B} = (A^u)(B^u)g_{uu} + (A^u)(B^v)g_{uv} + (A^v)(B^u)g_{vu} + (A^v)(B^v)g_{vv}$. For cartesian coordinates, since the components of g are pretty trivial, then it simplifies a lot: $\vec{A} \cdot \vec{B} = A^x B^x + A^y B^y$.

Viewed as a 2x2 matrix, the metric tensor g has an inverse. It's components are denoted by raised indices: g^{ij}. You can use the inverse metric tensor to take a "dot" product of two covectors, or to convert a covector into a vector.

So let's take the case of the gradient in polar coordinates. The covector form is: $\nabla\phi$ with components $(\nabla\phi)_r = \frac{\partial\phi}{\partial r}$ and $(\nabla\phi)_\theta = \frac{\partial\phi}{\partial\theta}$. To convert it into a vector, you use the inverse of the metric tensor. In this case,

$$g^{rr} = \frac{1}{g_{rr}} = 1$$
$$g^{\theta\theta} = \frac{1}{g_{\theta\theta}} = \frac{1}{r^2}$$

(the other two components are zero).

So the vector form of the gradient is: $\overrightarrow{\nabla\phi}$ with components

So the vector form of the gradient is: $\overrightarrow{\nabla\phi}$ with components

$$(\overrightarrow{\nabla\phi})^r = g^{rr}\frac{\partial\phi}{\partial r} = \frac{\partial\phi}{\partial r}$$
$$(\overrightarrow{\nabla\phi})^\theta = g^{\theta\theta}\frac{\partial\phi}{\partial\theta} = \frac{1}{r^2}\frac{\partial\phi}{\partial\theta}$$

To get the directional derivative, you take the dot-product with a direction vector $\overrightarrow{V}$:
$\overrightarrow{\nabla\phi}\cdot\overrightarrow{V}$

but computing the dot-product in curvilinear coordinates involves the metric tensor again:

$$\overrightarrow{\nabla\phi}\cdot\overrightarrow{V} = g_{rr}(\overrightarrow{\nabla\phi})^r(\overrightarrow{V})^r + g_{\theta\theta}(\overrightarrow{\nabla\phi})^\theta(\overrightarrow{V})^\theta$$

The metric tensor just cancels out the use of the inverse metric tensor in forming the vector gradient, so the result is just:

$$\overrightarrow{\nabla\phi}\cdot\overrightarrow{V} = \frac{\partial\phi}{\partial r}V^r + \frac{\partial\phi}{\partial\theta}V^\theta$$

This shows that the covariant form of the gradient is more natural; the vector form uses the inverse metric tensor to create the vector, and then uses the metric tensor again to get the result. In the final result, the metric tensor components drop out.