Understanding the Hessian.

We now know that

\[ f(\vec{x}_0 + \vec{v}) \approx f(\vec{x}_0) + \langle \vec{v}, \nabla f(\vec{x}_0) \rangle + \frac{1}{2} \langle \vec{v}, \nabla^2 f(\vec{x}_0) \vec{v} \rangle \]

The gradient vector is easy to interpret: it points "straight uphill" and its norm is the slope of the hill.

What does \( \nabla f(\vec{x}_0) \) tell us? We'll need some more linear algebra.

Definition. If \( A \) is an \( n \times n \) matrix, then \( \vec{v} \in \mathbb{R}^n \) is called an eigenvector of \( A \) with eigenvalue \( \lambda \) if

\[ A\vec{v} = \lambda \vec{v}. \]
Example. \[ A = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \quad \text{eigenvectors} \quad e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \]

\[ \lambda_1 \quad \text{and} \quad \lambda_2 \quad \text{eigenvalues} \]

Proposition. The eigenvalues of \( A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \) are the roots of the polynomial

\[ \det(A - \lambda I) = (a - \lambda)(d - \lambda) - bc. \]

Proof. \( \lambda \) is an eigenvalue of \( A \leftrightarrow \) there is some \( \vec{v} \) with \( A \vec{v} = \lambda \vec{v} \) or \( (A - \lambda I) \vec{v} = 0 \). The matrix \( (A - \lambda I) \) has zero determinant \( \leftrightarrow \) there is some such \( \vec{v} \).
Here is a useful fact:

Spectral Theorem. If \( A \) is a symmetric, real \( n \times n \) matrix then there exists an orthonormal basis \( \vec{v}_1, \ldots, \vec{v}_n \) for \( \mathbb{R}^n \) so that \( A \vec{v}_i = \lambda_i \vec{v}_i \) for all \( i \in 1, \ldots, n \). All \( \vec{v}_i \) and \( \lambda_i \) are real.

This is often summarized as "every symmetric matrix can be diagonalized."

Homework. Suppose \( f : \mathbb{R}^n \to \mathbb{R} \) and \( B : \mathbb{R}^n \to \mathbb{R}^n \) is a linear map. Then

\[
\nabla (f(B(x))) = B \nabla f
\]

and

\[
\nabla (f(Bx)) = B^T \nabla f B
\]

\( a \) Equivalently, \( B \) is an \( n \times n \) matrix.
Now let's return to Taylor's theorem.

Proposition. If \( f: \mathbb{R}^n \to \mathbb{R} \) and \( \bar{x}_0 \in \mathbb{R}^n \) there is an orthonormal basis for \( \mathbb{R}^n \) so that (in these coordinates) \( \nabla f(\bar{x}_0) \) is a diagonal matrix.

Proof. Combine homework with Spectral theorem.

Proposition. If \( f: \mathbb{R}^n \to \mathbb{R} \) and \( \bar{x}_0 \in \mathbb{R}^n \) there is an orthonormal basis for \( \mathbb{R}^n \) so that

\[
f(\bar{x}_0 + \tilde{v}) \approx f(\bar{x}_0) + \left< \nabla f(\bar{x}_0), \tilde{v} \right> + \frac{1}{2} \left< \tilde{v}, \nabla^2 f(\bar{x}_0) \tilde{v} \right>
\]

\[
\approx C + b_1 v_1 + \ldots + b_nv_n + a_1 v_1^2 + \ldots + a_nv_n^2
\]
Here the $a_i$ are the eigenvalues of $H_f(x_0)$ (which are the same in any orthogonal basis for $\mathbb{R}^n$).

Definition. A surface in the form $z = ax^2 + 2bxy + cy^2$ is called a paraboloid.

Definition. If the eigenvalues of $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$ have the same sign, the surface is called an elliptic paraboloid.

The level curves of an elliptic paraboloid are ellipses, with major and minor axes in the directions of the eigenvectors.

\[ \text{Equivalently, } \det \begin{bmatrix} a & b \\ b & c \end{bmatrix} > 0. \]
\( \vec{v}_1 \) and \( \vec{v}_2 \) of \( \begin{bmatrix} a & b \\ b & c \end{bmatrix} \).

\[
ax^2 + 2bxy + cy^2 = 1 = \lambda_1 \vec{v}_1 \cdot \vec{v}_1 + \lambda_2 \vec{v}_2 \cdot \vec{v}_2
\]

Definition. If the eigenvalues \( \lambda_1, \lambda_2 \) of \( \begin{bmatrix} a & b \\ b & c \end{bmatrix} \) have opposite signs, the surface is called a hyperbolic paraboloid.

The level curves of a hyperbolic paraboloid are hyperbolas. The hyperbolas are perpendicular to the eigenvectors \( \vec{v}_1 \) and \( \vec{v}_2 \) of \( \begin{bmatrix} a & b \\ b & c \end{bmatrix} \).

\( a \) Equivalently, \( \det \begin{bmatrix} a & b \\ b & c \end{bmatrix} < 0. \)
ax^2 + 2bxy + cy^2 = 1 = \lambda_1 V_1^2 + \lambda_2 V_2^2

the asymptotic lines have equation \( V_2 = \pm \sqrt{\frac{\lambda_2}{\lambda_1}} V_1 \).

the individual hyperbolas are orthogonal to \( V_1 \) and \( V_2 \) axes where they cross.

Definition. If exactly one of the eigenvalues \( \lambda_1, \lambda_2 \) of \( \begin{bmatrix} a & b \\ b & c \end{bmatrix} \) is zero, the surface is called a cylindrical paraboloid.

The level curves of a cylindrical paraboloid are lines parallel to the direction of the eigenvector with 0 eigenvalue.
Note: If both eigenvalues of \( \begin{bmatrix} a & b \\ b & c \end{bmatrix} \) are zero then \( \begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \) and the surface is the x-y plane.

Let's summarize what we've learned. For scalar functions, \( f: \mathbb{R}^n \to \mathbb{R} \),

if \( n = 1 \), \( f(x_0 + v) \approx f(x_0) + f'(x_0)v + \frac{1}{2} f''(x_0)v^2 \)

so \( f(x_0 + v) \) is approximated by the parabola

\[ f(x_0) + f'(x_0)v + \frac{1}{2}f''(x_0)v^2 = p(v) \]
The vertex of the parabola is found by solving \( p'(v) = 0 \) (or completing the square).

\[
2 \cdot \frac{1}{2} f''(x_0) v + f'(x_0) = p'(v)
\]

so

\[ p'(v) = 0 \iff v = -\frac{f'(x_0)}{f''(x_0)} \]

So we could write

\[
p(v) = \frac{1}{2} f''(x_0) \left( v + \frac{f'(x_0)}{f''(x_0)} \right)^2 + (f(x_0) - \frac{f'(x_0)^2}{2f''(x_0)})
\]

\[ = \alpha (v - \beta)^2 + \gamma \]

if we wanted to.
If \( n > 1 \), \( f(\tilde{x}_0 + \tilde{v}) \approx f(\tilde{x}_0) + \langle \tilde{v}, \nabla f(\tilde{x}_0) \rangle + \frac{1}{2} \langle \tilde{v}, Hf(\tilde{x}_0) \tilde{v} \rangle \)

so \( f(\tilde{x}_0 + \tilde{v}) \) is approximated by the paraboloid

\[
f(\tilde{x}_0) + \langle \tilde{v}, \nabla f(\tilde{x}_0) \rangle + \frac{1}{2} \langle \tilde{v}, Hf(\tilde{x}_0) \tilde{v} \rangle = p(\tilde{v})
\]

If we write \( \tilde{v} \) in the basis of eigenvectors of the \( nxn \) symmetric matrix \( Hf(\tilde{x}_0) \), then (in this basis) \( Hf(\tilde{x}_0) \) is a diagonal matrix, and

\[
p(\tilde{v}) = f(\tilde{x}_0) + \sum_{i=1}^{n} v_i \frac{\partial f(\tilde{x}_0)}{\partial v_i} \frac{1}{2} v_i^2 \frac{\partial^2 f(\tilde{x}_0)}{\partial v_i^2}
\]

So the center of the paraboloid \( p(\tilde{v}) \) is located at

\[
\tilde{V}_0 = \left( -\frac{\partial f(\tilde{x}_0)}{\partial v_1(\tilde{x}_0)}, \ldots, -\frac{\partial f(\tilde{x}_0)}{\partial v_n(\tilde{x}_0)} \right)
\]

(in eigenvector coordinates for \( Hf(\tilde{x}_0) \)).