Understanding the Hessian.

We now know that

$$f(\vec{x}_0+\vec{v}) \approx f(x_0) + \langle \vec{v}, \nabla f(x_0) \rangle + \frac{1}{2} \langle \vec{v}, Hf(x_0)\,\vec{v} \rangle$$

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

What does $Hf(x_0)$ 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~~ *eigenvector* of A with *eigenvalue* λ if

$$A\vec{v} = \lambda \vec{v}.$$

Example.

$$A = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

eigenvectors

$$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

eigen values

$$\lambda_1 \quad , \quad \lambda_2$$

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. λ is an eigenvalue of A <=> 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 <=> there is some such $\vec{v}$. $\square$

Here is a useful fact:

Spectral Theorem. If A is a symmetric, real $n \times n$ matrix then $\exists$ 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 λ_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[a]. Then

$$\nabla(f(B(\vec{x}))) = B \nabla f$$

and

$$H(f(B\vec{x})) = B^T Hf \, B$$

---

[a] Equivalently, B is an $n \times n$ matrix.

Now lets return to Taylor's theorem.

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

Proof. Combine homework with Spectral theorem.

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

$$f(\vec{x}_0 + \vec{v}) \approx f(\vec{x}_0) + \langle \nabla f(\vec{x}_0), \vec{v} \rangle + \frac{1}{2} \langle \vec{v}, Hf(\vec{x}_0)\vec{v} \rangle$$

$$\approx c + b_1 v_1 + \ldots + b_n v_n + a_1 v_1^2 + \ldots + a_n v_n^2$$

Here the a_i are the eigenvalues of $Hf(\vec{x}_0)$ (which are the same in any orthonormal basis for $\mathbb{R}^n$).

Definition. A surface in the form

$$z = ax^2 + 2bxy + cy^2$$ is called a <u>paraboloid</u>.

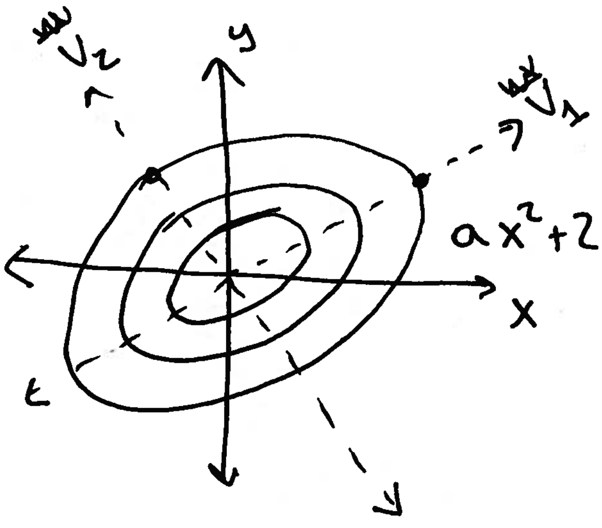
Definition. If the eigenvalues λ_1, λ_2 of $\begin{bmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{bmatrix}$ have the same sign[a], the surface is called an <u>elliptic</u> paraboloid.

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

---

[a] Equivalently, $\det \begin{bmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{bmatrix} > 0$.

$\vec{v}_1$ and $\vec{v}_2$ of $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$.



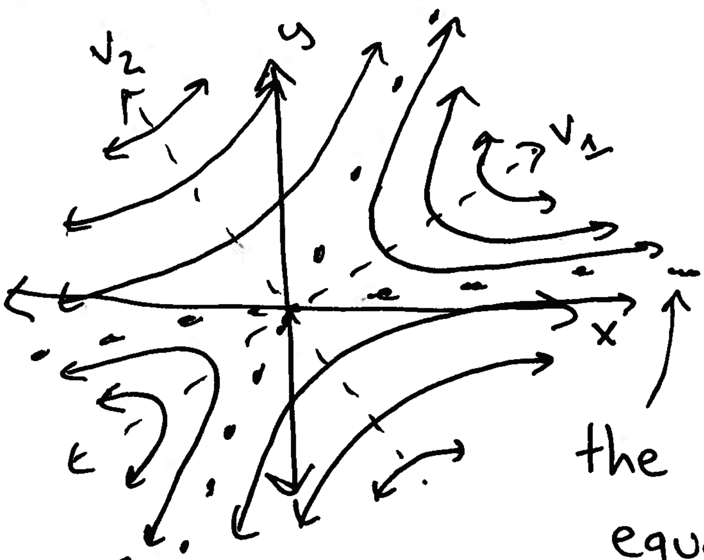
$ax^2 + 2bxy + cy^2 = 1 = \lambda_1 v_1^2 + \lambda_2 v_2^2$

Definition. If the eigenvalues λ_1, λ_2 of $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$ have opposite signs[a], 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, $\det \begin{bmatrix} a & b \\ b & c \end{bmatrix} < 0$.



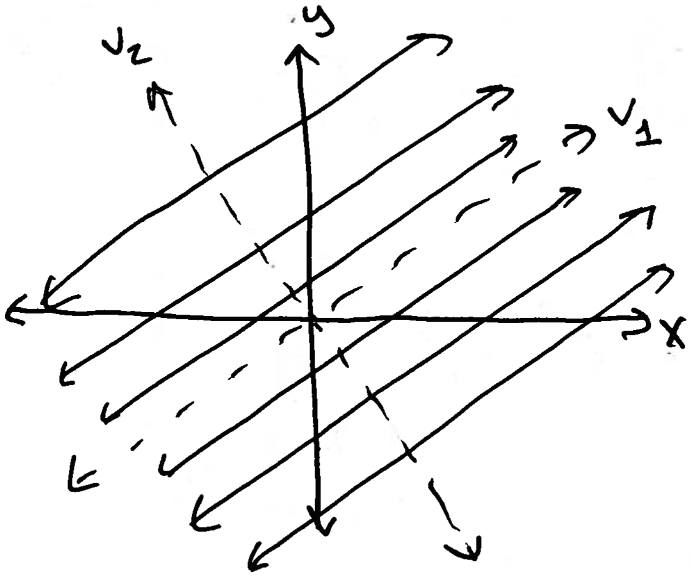
$ax^2 + 2bxy + cy^2 = 1 = \lambda_1 v_1^2 + \lambda_2 v_2^2$

the asymptotic lines have equation $v_2 = \pm\sqrt{\frac{\lambda_1}{\lambda_2}}\, v_1$.

the individual hyperbolas are orthogonal to v_1 and v_2 axes where they cross

Definition. If exactly one of the eigenvalues λ_1, λ_2 of $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$ ~~are~~ is zero, the surface is called a <u>cylindrical</u> <u>paraboloid</u>.

The level curves of a cylindrical paraboloid are lines parallel to the direction of the eigenvector with 0 eigenvalue.



$$ax^2 + 2bxy + cy^2 = 1 = 0v_1^2 + \lambda_2 v_2^2$$

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

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

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

if we wanted to.

If $n>1$, $f(\vec{x}_0+\vec{v}) \approx f(\vec{x}_0) + \langle \vec{v}, \nabla f(x_0)\rangle + \frac{1}{2}\langle \vec{v}, Hf(x_0)\vec{v}\rangle$

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

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

If we write $\vec{v}$ in the basis of eigenvectors of the $n \times n$ symmetric matrix $Hf(\vec{x}_0)$, then (in this basis) $Hf(x_0)$ is a diagonal matrix, and

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

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

$$\vec{v}_0 = \left(- \frac{\partial f}{\partial v_1}(\vec{x}_0) \Big/ \frac{\partial^2 f}{\partial v_1^2}(\vec{x}_0), \ldots, - \frac{\partial f}{\partial v_n}(\vec{x}_0) \Big/ \frac{\partial^2 f}{\partial v_n^2}(\vec{x}_0) \right)$$

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