GEOMETRY OF QUADRATIC FUNCTIONS / 2

- $\boldsymbol{v}_{\max}$ ($\boldsymbol{v}_{\min}$) direction of highest (lowest) curvature

  Proof: With $\boldsymbol{v} = \mathbf{V}^T\mathbf{x}$:

  $$\mathbf{x}^T\mathbf{H}\mathbf{x} = \mathbf{x}^T\mathbf{V}\Lambda\mathbf{V}^T\mathbf{x} = \boldsymbol{v}^T\Lambda\boldsymbol{v} = \sum_{i=1}^{d}\lambda_i v_i^2 \leq \lambda_{\max}\sum_{i=1}^{d} v_i^2 = \lambda_{\max}\|\boldsymbol{v}\|^2$$

  Since $\|\boldsymbol{v}\| = \|\mathbf{x}\|$ ($\mathbf{V}$ orthogonal): $\max_{\|\mathbf{x}\|=1}\mathbf{x}^T\mathbf{H}\mathbf{x} \leq \lambda_{\max}$
  Additional: $\boldsymbol{v}_{\max}{}^T\mathbf{H}\boldsymbol{v}_{\max} = \mathbf{e}_1^T\Lambda\mathbf{e}_1 = \lambda_{\max}$
  Analogous: $\min_{\|\mathbf{x}\|=1}\mathbf{x}^T\mathbf{H}\mathbf{x} \geq \lambda_{\min}$ and $\boldsymbol{v}_{\min}{}^T\mathbf{H}\boldsymbol{v}_{\min} = \lambda_{\min}$

- Contour lines of any quadratic form are ellipses
  (with eigenvectors of A as principal axes, principal axis theorem)

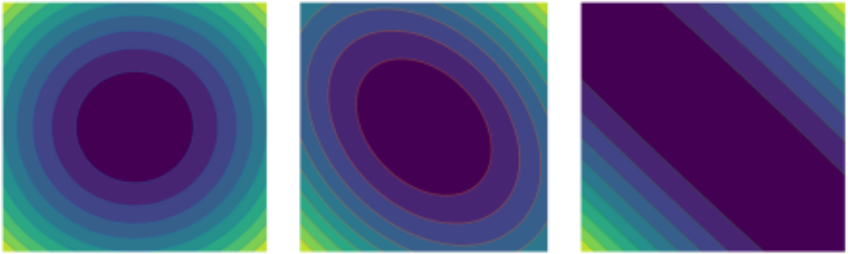
  Look at $q(\mathbf{x}) = \mathbf{x}^T\boldsymbol{A}\mathbf{x} + \boldsymbol{b}^T\mathbf{x} + c$
  Now use $\boldsymbol{y} = \mathbf{x} - \boldsymbol{w} = \mathbf{x} + \frac{1}{2}\boldsymbol{A}^{-1}\boldsymbol{b}$
  This already gives us the general form of an ellipse:
  $\boldsymbol{y}^T\boldsymbol{A}\boldsymbol{y} = (\mathbf{x} - \boldsymbol{w})^T\boldsymbol{A}(\mathbf{x} - \boldsymbol{w}) = q(\mathbf{x}) + const$
  If we use $\boldsymbol{z} = \boldsymbol{V}^T y$ we obtain it in standard form
  $\sum_{i=1}^{n}\lambda_i z_i^2 = \boldsymbol{z}^T\Lambda\boldsymbol{z} = y^T\boldsymbol{V}\Lambda\boldsymbol{V}^T y = \boldsymbol{y}^T\boldsymbol{A}\boldsymbol{y} = q(\mathbf{x}) + const$

CONDITION AND CURVATURE

Condition of $\mathbf{H} = 2\mathbf{A}$ is given by $\kappa(\mathbf{H}) = \kappa(\mathbf{A}) = |\lambda_{max}|/|\lambda_{min}|$.

High condition means:

- $|\lambda_{max}| \gg |\lambda_{min}|$
- Curvature along $\mathbf{v}_{max} \gg$ curvature along $\mathbf{v}_{min}$
- **Problem** for optimization algorithms like **gradient descent** (later)

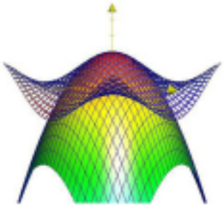


Left: Excellent condition. **Middle:** Good condition. **Right:** Bad condition.

APPROXIMATION OF SMOOTH FUNCTIONS

Any function $f \in \mathcal{C}^2$ can be locally approximated by a quadratic function via second order Taylor approximation:

$$f(\mathbf{x}) \approx f(\tilde{\mathbf{x}}) + \nabla f(\tilde{\mathbf{x}})^T(\mathbf{x} - \tilde{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \tilde{\mathbf{x}})^T \nabla^2 f(\tilde{\mathbf{x}})(\mathbf{x} - \tilde{\mathbf{x}})$$



f and its second order approximation is shown by the dark and bright grid, respectively. (Source: `daniloroccatano.blog`)

$\Longrightarrow$ Hessians provide information about **local** geometry of a function.