

8



Performance Surfaces

Taylor Series Expansion



$$\begin{aligned} F(x) = & F(x^*) + \frac{d}{dx}F(x) \Big|_{x=x^*} (x - x^*) \\ & + \frac{1}{2} \frac{d^2}{dx^2} F(x) \Big|_{x=x^*} (x - x^*)^2 + \dots \\ & + \frac{1}{n!} \frac{d^n}{dx^n} F(x) \Big|_{x=x^*} (x - x^*)^n + \dots \end{aligned}$$

Example

$$F(x) = e^{-x}$$

Taylor series of $F(x)$ about $x^* = 0$:

$$F(x) = e^{-x} = e^{-0} - e^{-0}(x-0) + \frac{1}{2}e^{-0}(x-0)^2 - \frac{1}{6}e^{-0}(x-0)^3 + \dots$$

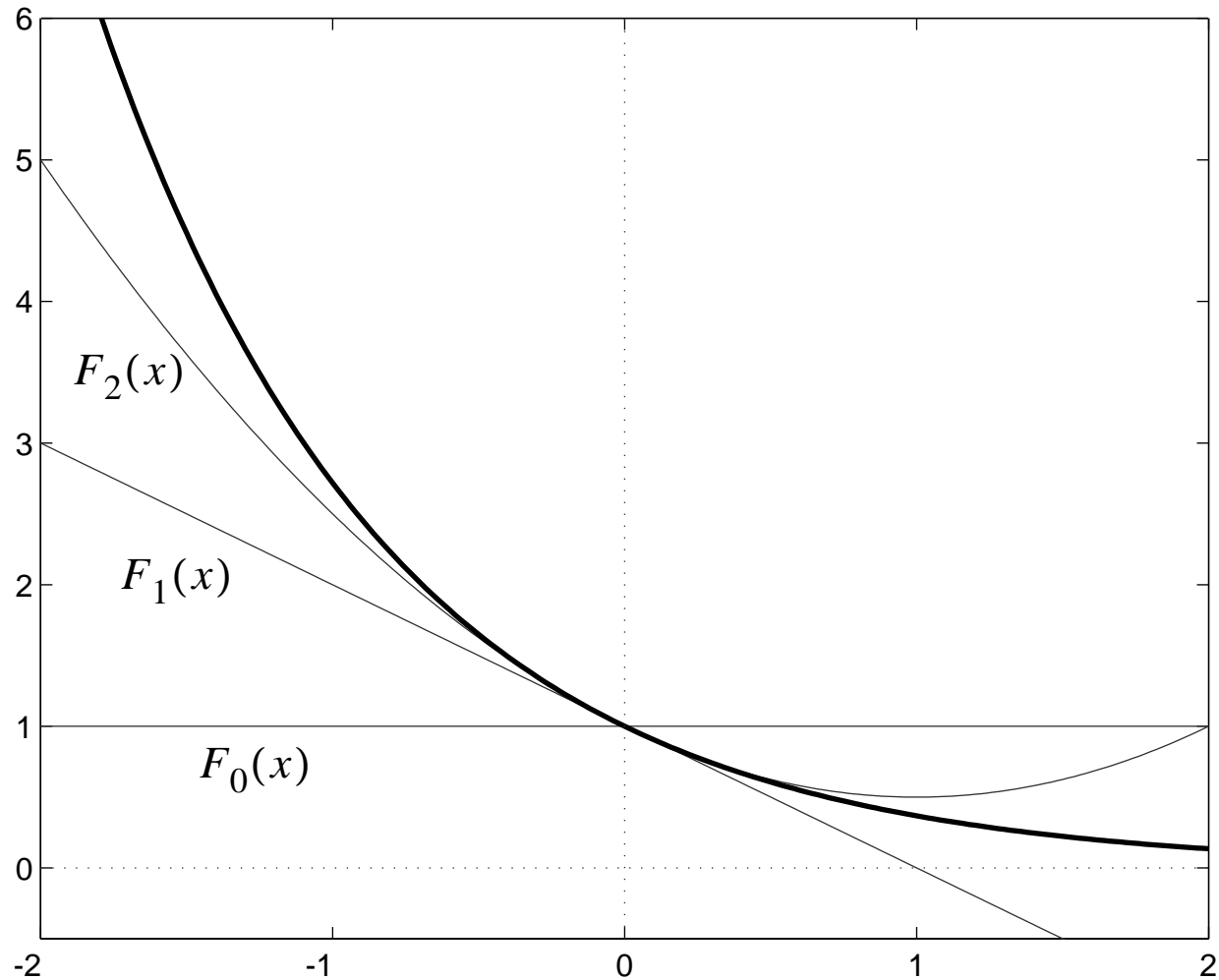
$$F(x) = 1 - x + \frac{1}{2}x^2 - \frac{1}{6}x^3 + \dots$$

Taylor series approximations:

$$F(x) \approx F_0(x) = 1$$

$$F(x) \approx F_1(x) = 1 - x$$

$$F(x) \approx F_2(x) = 1 - x + \frac{1}{2}x^2$$





$$F(\mathbf{x}) = F(x_1, x_2, \dots, x_n)$$

$$\begin{aligned} F(\mathbf{x}) = & F(\mathbf{x}^*) + \frac{\partial}{\partial x_1} F(\mathbf{x}) \Big|_{\mathbf{x} = \mathbf{x}^*} (x_1 - x_1^*) + \frac{\partial}{\partial x_2} F(\mathbf{x}) \Big|_{\mathbf{x} = \mathbf{x}^*} (x_2 - x_2^*) \\ & + \dots + \frac{\partial}{\partial x_n} F(\mathbf{x}) \Big|_{\mathbf{x} = \mathbf{x}^*} (x_n - x_n^*) + \frac{1}{2} \frac{\partial^2}{\partial x_1^2} F(\mathbf{x}) \Big|_{\mathbf{x} = \mathbf{x}^*} (x_1 - x_1^*)^2 \\ & + \frac{1}{2} \frac{\partial^2}{\partial x_1 \partial x_2} F(\mathbf{x}) \Big|_{\mathbf{x} = \mathbf{x}^*} (x_1 - x_1^*) (x_2 - x_2^*) + \dots \end{aligned}$$



$$F(\mathbf{x}) = F(\mathbf{x}^*) + \nabla F(\mathbf{x})^T \Big|_{\mathbf{x}=\mathbf{x}^*} (\mathbf{x} - \mathbf{x}^*) \\ + \frac{1}{2} (\mathbf{x} - \mathbf{x}^*)^T \nabla^2 F(\mathbf{x}) \Big|_{\mathbf{x}=\mathbf{x}^*} (\mathbf{x} - \mathbf{x}^*) + \dots$$

Gradient

$$\nabla F(\mathbf{x}) = \begin{bmatrix} \frac{\partial}{\partial x_1} F(\mathbf{x}) \\ \frac{\partial}{\partial x_2} F(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial x_n} F(\mathbf{x}) \end{bmatrix}$$

Hessian

$$\nabla^2 F(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} F(\mathbf{x}) & \frac{\partial^2}{\partial x_1 \partial x_2} F(\mathbf{x}) & \dots & \frac{\partial^2}{\partial x_1 \partial x_n} F(\mathbf{x}) \\ \frac{\partial^2}{\partial x_2 \partial x_1} F(\mathbf{x}) & \frac{\partial^2}{\partial x_2^2} F(\mathbf{x}) & \dots & \frac{\partial^2}{\partial x_2 \partial x_n} F(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} F(\mathbf{x}) & \frac{\partial^2}{\partial x_n \partial x_2} F(\mathbf{x}) & \dots & \frac{\partial^2}{\partial x_n^2} F(\mathbf{x}) \end{bmatrix}$$



First derivative (slope) of $F(\mathbf{x})$ along x_i axis: $\partial F(\mathbf{x})/\partial x_i$

(i th element of gradient)

Second derivative (curvature) of $F(\mathbf{x})$ along x_i axis: $\partial^2 F(\mathbf{x})/\partial x_i^2$

(i,i element of Hessian)

First derivative (slope) of $F(\mathbf{x})$ along vector \mathbf{p} : $\frac{\mathbf{p}^T \nabla F(\mathbf{x})}{\|\mathbf{p}\|}$

Second derivative (curvature) of $F(\mathbf{x})$ along vector \mathbf{p} : $\frac{\mathbf{p}^T \nabla^2 F(\mathbf{x}) \mathbf{p}}{\|\mathbf{p}\|^2}$

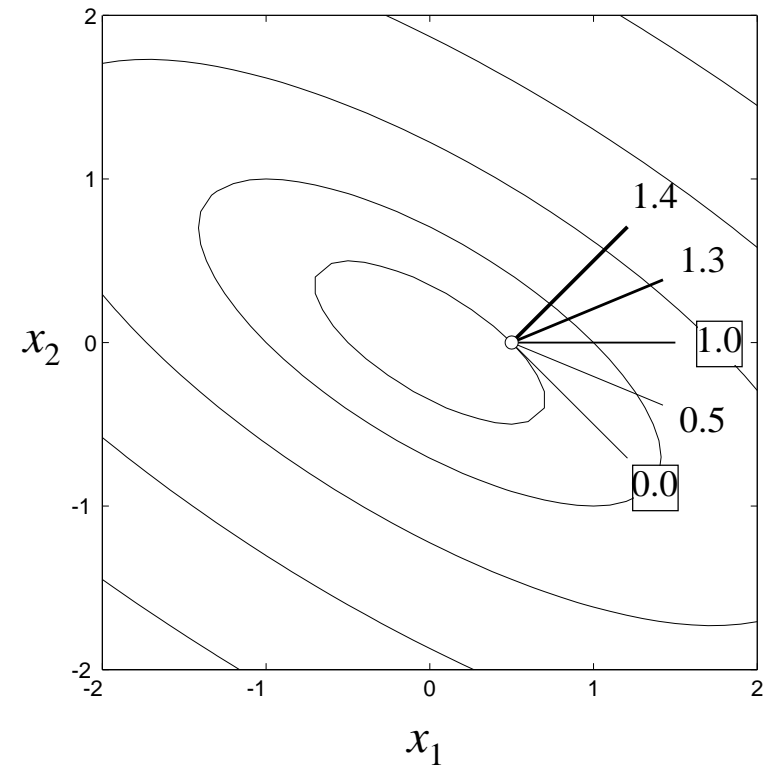
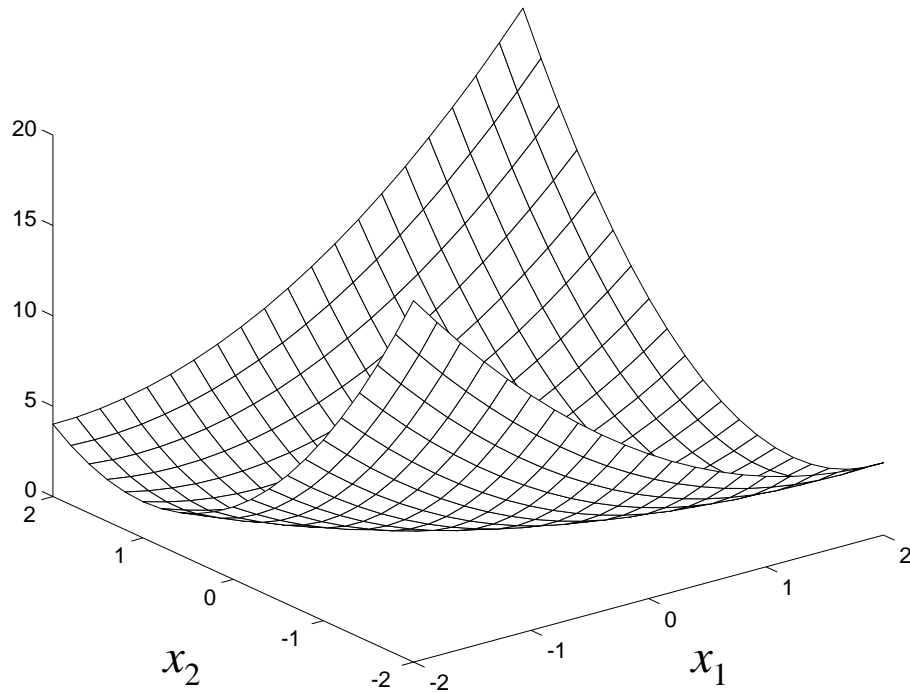
Example

$$F(\mathbf{x}) = x_1^2 + 2x_1x_2 + 2x_2^2$$

$$\mathbf{x}^* = \begin{bmatrix} 0.5 \\ 0 \end{bmatrix} \quad \mathbf{p} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\nabla F(\mathbf{x}) \Big|_{\mathbf{x} = \mathbf{x}^*} = \begin{bmatrix} \frac{\partial}{\partial x_1} F(\mathbf{x}) \\ \frac{\partial}{\partial x_2} F(\mathbf{x}) \end{bmatrix} \Big|_{\mathbf{x} = \mathbf{x}^*} = \begin{bmatrix} 2x_1 + 2x_2 \\ 2x_1 + 4x_2 \end{bmatrix} \Big|_{\mathbf{x} = \mathbf{x}^*} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\frac{\mathbf{p}^T \nabla F(\mathbf{x})}{\|\mathbf{p}\|} = \frac{\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}}{\left\| \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\|} = \frac{0}{\sqrt{2}} = 0$$

Directional
Derivatives



Strong Minimum

The point \mathbf{x}^* is a strong minimum of $F(\mathbf{x})$ if a scalar $\delta > 0$ exists, such that $F(\mathbf{x}^*) < F(\mathbf{x}^* + \Delta\mathbf{x})$ for all $\Delta\mathbf{x}$ such that $\delta > \|\Delta\mathbf{x}\| > 0$.

Global Minimum

The point \mathbf{x}^* is a unique global minimum of $F(\mathbf{x})$ if $F(\mathbf{x}^*) < F(\mathbf{x}^* + \Delta\mathbf{x})$ for all $\Delta\mathbf{x} \neq 0$.

Weak Minimum

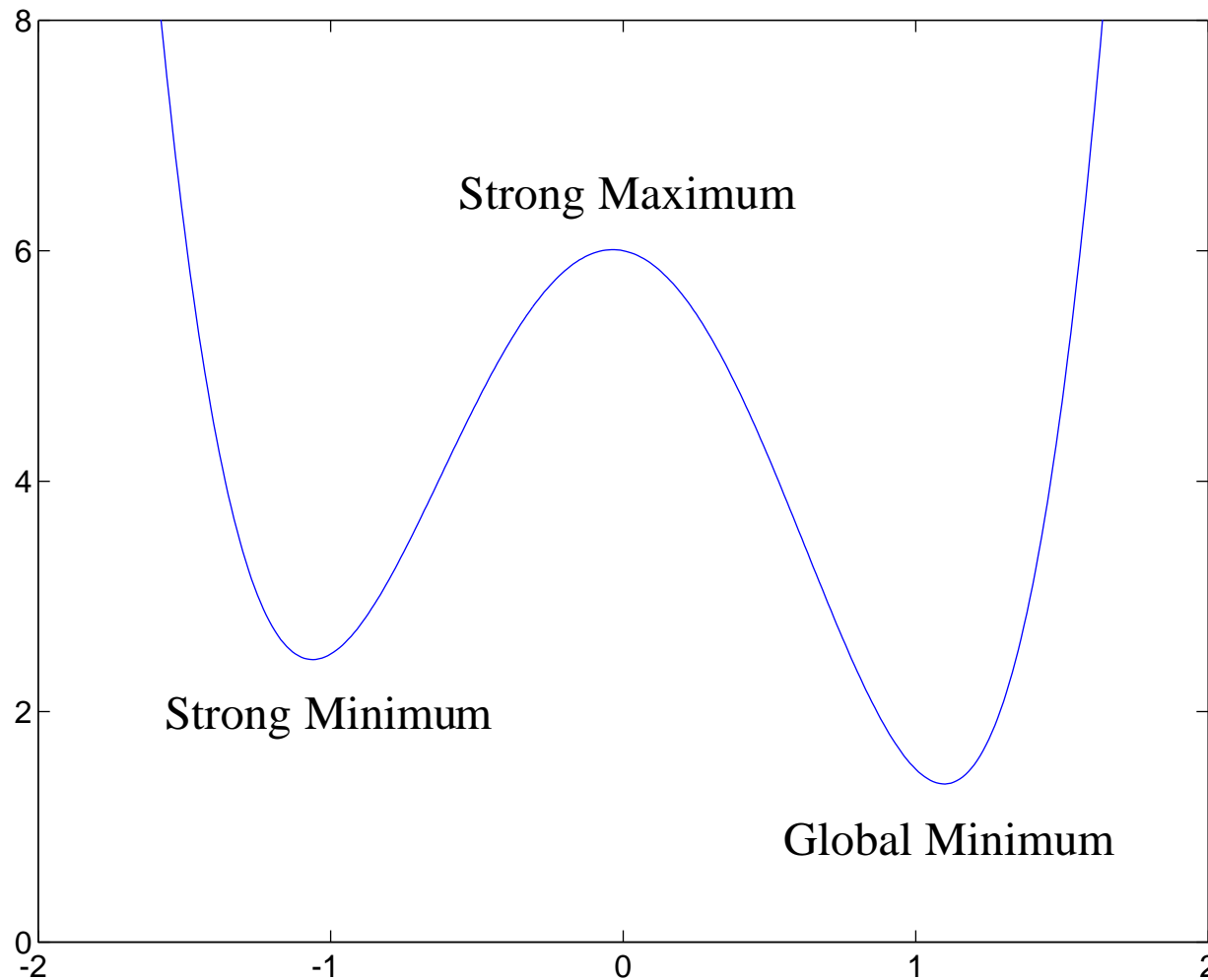
The point \mathbf{x}^* is a weak minimum of $F(\mathbf{x})$ if it is not a strong minimum, and a scalar $\delta > 0$ exists, such that $F(\mathbf{x}^*) \leq F(\mathbf{x}^* + \Delta\mathbf{x})$ for all $\Delta\mathbf{x}$ such that $\delta > \|\Delta\mathbf{x}\| > 0$.

8

Scalar Example

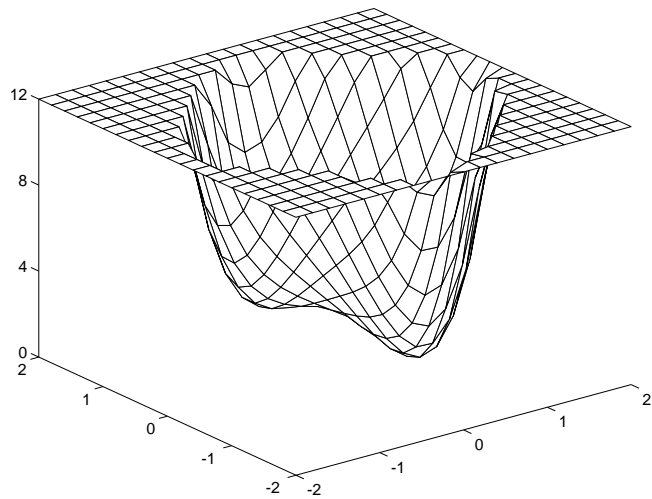
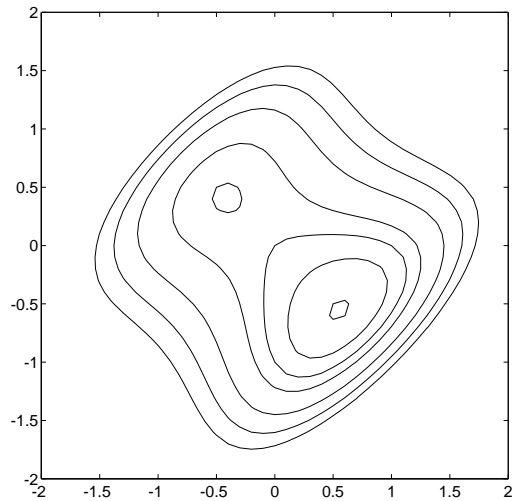


$$F(x) = 3x^4 - 7x^2 - \frac{1}{2}x + 6$$

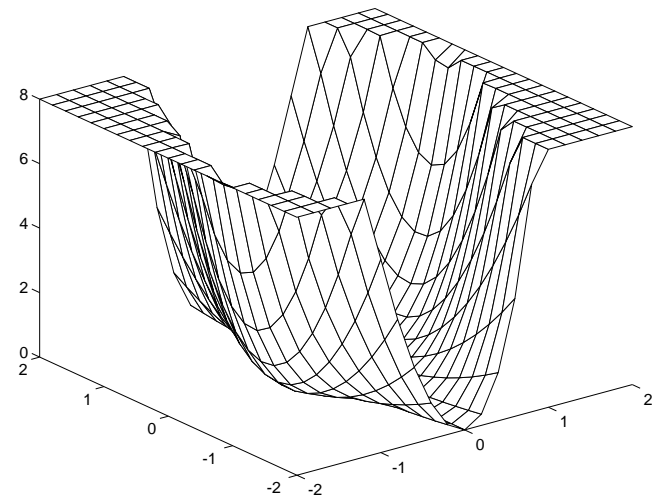
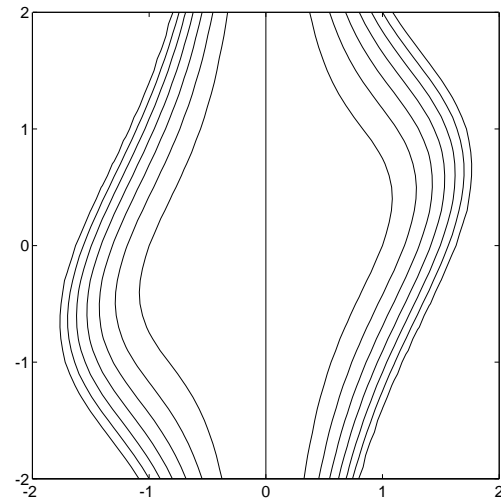




$$F(\mathbf{x}) = (x_2 - x_1)^4 + 8x_1x_2 - x_1 + x_2 + 3$$



$$F(\mathbf{x}) = (x_1^2 - 1.5x_1x_2 + 2x_2^2)x_1^2$$



8

First-Order Optimality Condition

$$F(\mathbf{x}) = F(\mathbf{x}^* + \Delta\mathbf{x}) = F(\mathbf{x}^*) + \nabla F(\mathbf{x})^T \Big|_{\mathbf{x}=\mathbf{x}^*} \Delta\mathbf{x} + \frac{1}{2} \Delta\mathbf{x}^T \nabla^2 F(\mathbf{x}) \Big|_{\mathbf{x}=\mathbf{x}^*} \Delta\mathbf{x} + \dots$$

$$\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}^*$$

For small $\Delta\mathbf{x}$:

$$F(\mathbf{x}^* + \Delta\mathbf{x}) \cong F(\mathbf{x}^*) + \nabla F(\mathbf{x})^T \Big|_{\mathbf{x}=\mathbf{x}^*} \Delta\mathbf{x}$$

If \mathbf{x}^* is a minimum, this implies:

$$\nabla F(\mathbf{x})^T \Big|_{\mathbf{x}=\mathbf{x}^*} \Delta\mathbf{x} \geq 0$$

If $\nabla F(\mathbf{x})^T \Big|_{\mathbf{x}=\mathbf{x}^*} \Delta\mathbf{x} > 0$ then $F(\mathbf{x}^* - \Delta\mathbf{x}) \cong F(\mathbf{x}^*) - \nabla F(\mathbf{x})^T \Big|_{\mathbf{x}=\mathbf{x}^*} \Delta\mathbf{x} < F(\mathbf{x}^*)$

But this would imply that \mathbf{x}^* is not a minimum. Therefore $\nabla F(\mathbf{x})^T \Big|_{\mathbf{x}=\mathbf{x}^*} \Delta\mathbf{x} = 0$

Since this must be true for every $\Delta\mathbf{x}$,

$$\nabla F(\mathbf{x}) \Big|_{\mathbf{x}=\mathbf{x}^*} = \mathbf{0}$$



If the first-order condition is satisfied (zero gradient), then

$$F(\mathbf{x}^* + \Delta\mathbf{x}) = F(\mathbf{x}^*) + \frac{1}{2}\Delta\mathbf{x}^T \nabla^2 F(\mathbf{x}) \Big|_{\mathbf{x} = \mathbf{x}^*} \Delta\mathbf{x} + \dots$$

A strong minimum will exist at \mathbf{x}^* if $\Delta\mathbf{x}^T \nabla^2 F(\mathbf{x}) \Big|_{\mathbf{x} = \mathbf{x}^*} \Delta\mathbf{x} > 0$ for any $\Delta\mathbf{x} \neq \mathbf{0}$.

Therefore the Hessian matrix must be positive definite. A matrix \mathbf{A} is positive definite if:

$$\mathbf{z}^T \mathbf{A} \mathbf{z} > 0 \quad \text{for any } \mathbf{z} \neq \mathbf{0}.$$

This is a **sufficient** condition for optimality.

A **necessary** condition is that the Hessian matrix be positive semidefinite. A matrix \mathbf{A} is positive semidefinite if:

$$\mathbf{z}^T \mathbf{A} \mathbf{z} \geq 0 \quad \text{for any } \mathbf{z}.$$

Example

$$F(\mathbf{x}) = x_1^2 + 2x_1x_2 + 2x_2^2 + x_1$$

$$\nabla F(\mathbf{x}) = \begin{bmatrix} 2x_1 + 2x_2 + 1 \\ 2x_1 + 4x_2 \end{bmatrix} = \mathbf{0} \quad \Rightarrow \quad \mathbf{x}^* = \begin{bmatrix} -1 \\ 0.5 \end{bmatrix}$$

$$\nabla^2 F(\mathbf{x}) = \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix} \quad (\text{Not a function of } \mathbf{x} \text{ in this case.})$$

To test the definiteness, check the eigenvalues of the Hessian. If the eigenvalues are all greater than zero, the Hessian is positive definite.

$$|\nabla^2 F(\mathbf{x}) - \lambda \mathbf{I}| = \left| \begin{bmatrix} 2 - \lambda & 2 \\ 2 & 4 - \lambda \end{bmatrix} \right| = \lambda^2 - 6\lambda + 4 = (\lambda - 0.76)(\lambda - 5.24)$$

$$\lambda = 0.76, 5.24$$

Both eigenvalues are positive, therefore strong minimum.

Quadratic Functions



$$F(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + \mathbf{d}^T \mathbf{x} + c \quad (\text{Symmetric } \mathbf{A})$$

Gradient and Hessian:

Useful properties of gradients:

$$\nabla(\mathbf{h}^T \mathbf{x}) = \nabla(\mathbf{x}^T \mathbf{h}) = \mathbf{h}$$

$$\nabla_{\mathbf{x}}^T \mathbf{Q}\mathbf{x} = \mathbf{Q}\mathbf{x} + \mathbf{Q}^T \mathbf{x} = 2\mathbf{Q}\mathbf{x} \quad (\text{for symmetric } \mathbf{Q})$$

Gradient of Quadratic Function:

$$\nabla F(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{d}$$

Hessian of Quadratic Function:

$$\nabla^2 F(\mathbf{x}) = \mathbf{A}$$

Eigensystem of the Hessian



Consider a quadratic function which has a stationary point at the origin, and whose value there is zero.

$$F(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x}$$

Perform a similarity transform on the Hessian matrix, using the eigenvalues as the new basis vectors.

$$\mathbf{B} = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \dots & \mathbf{z}_n \end{bmatrix}$$

Since the Hessian matrix is symmetric, its eigenvectors are orthogonal.

$$\mathbf{B}^{-1} = \mathbf{B}^T$$

$$\mathbf{A}' = [\mathbf{B}^T \mathbf{A} \mathbf{B}] = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} = \mathbf{\Lambda} \quad \mathbf{A} = \mathbf{B} \mathbf{\Lambda} \mathbf{B}^T$$



$$\frac{\mathbf{p}^T \nabla^2 F(\mathbf{x}) \mathbf{p}}{\|\mathbf{p}\|^2} = \frac{\mathbf{p}^T \mathbf{A} \mathbf{p}}{\|\mathbf{p}\|^2}$$

Represent \mathbf{p} with respect to the eigenvectors (new basis):

$$\mathbf{p} = \mathbf{B} \mathbf{c}$$

$$\frac{\mathbf{p}^T \mathbf{A} \mathbf{p}}{\|\mathbf{p}\|^2} = \frac{\mathbf{c}^T \mathbf{B}^T (\mathbf{B} \mathbf{A} \mathbf{B}^T) \mathbf{B} \mathbf{c}}{\mathbf{c}^T \mathbf{B}^T \mathbf{B} \mathbf{c}} = \frac{\mathbf{c}^T \mathbf{\Lambda} \mathbf{c}}{\mathbf{c}^T \mathbf{c}} = \frac{\sum_{i=1}^n \lambda_i c_i^2}{\sum_{i=1}^n c_i^2}$$

$$\lambda_{min} \leq \frac{\mathbf{p}^T \mathbf{A} \mathbf{p}}{\|\mathbf{p}\|^2} \leq \lambda_{max}$$

8

Eigenvector (Largest Eigenvalue)

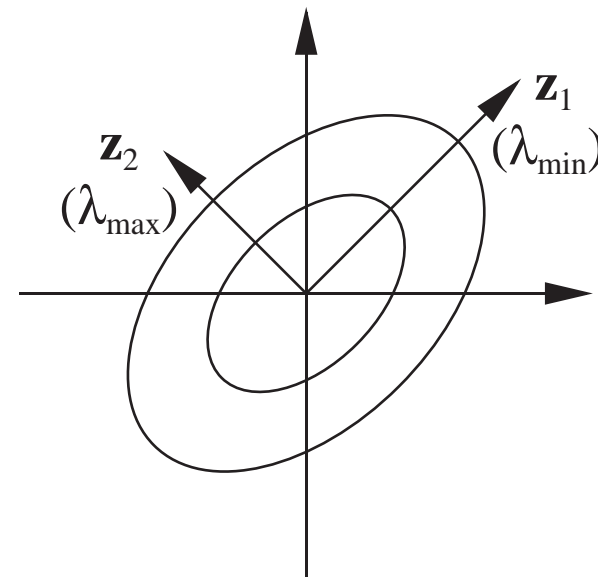


$$\mathbf{p} = \mathbf{z}_{max} \quad \mathbf{c} = \mathbf{B}^T \mathbf{p} = \mathbf{B}^T \mathbf{z}_{max} =$$

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\frac{\mathbf{z}_{max}^T \mathbf{A} \mathbf{z}_{max}}{\|\mathbf{z}_{max}\|^2} = \frac{\sum_{i=1}^n \lambda_i c_i^2}{\sum_{i=1}^n c_i^2} = \lambda_{max}$$

The eigenvalues represent curvature
(second derivatives) along the eigenvectors
(the principal axes).



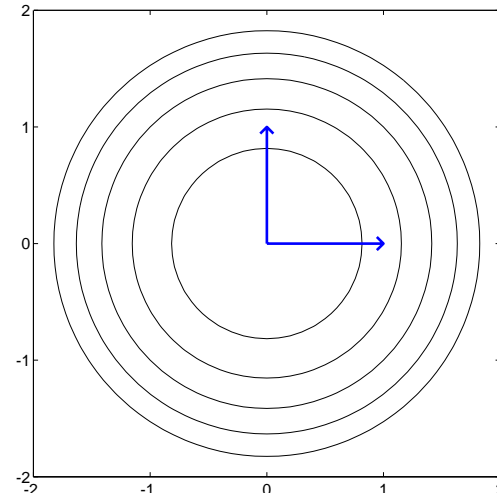
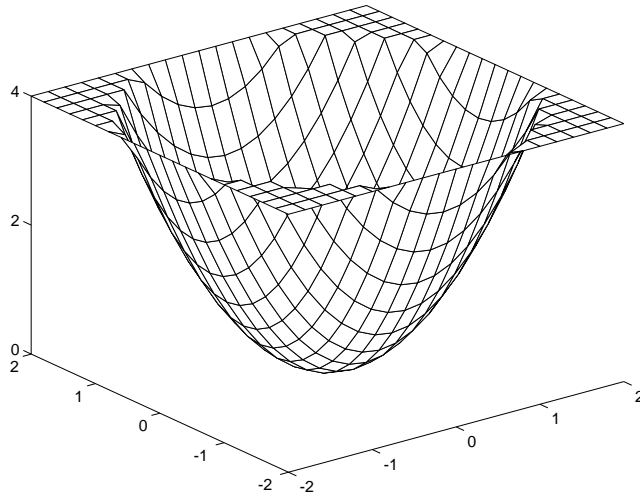
Circular Hollow



$$F(\mathbf{x}) = x_1^2 + x_2^2 = \frac{1}{2} \mathbf{x}^T \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \mathbf{x}$$

$$\nabla^2 F(\mathbf{x}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad \lambda_1 = 2 \quad \mathbf{z}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \lambda_2 = 2 \quad \mathbf{z}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

(Any two independent vectors in the plane would work.)

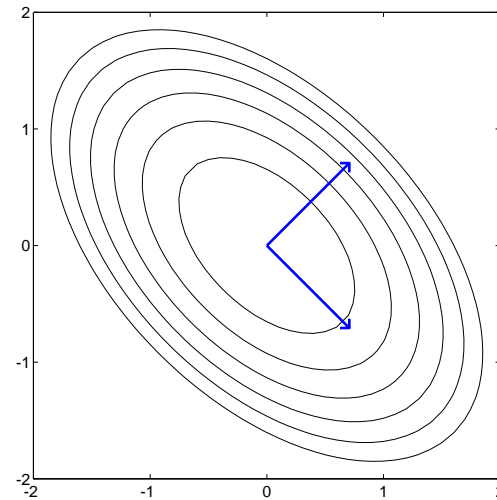
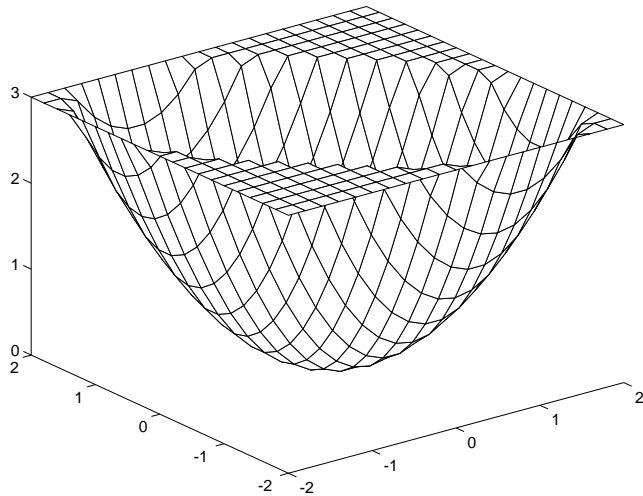


Elliptical Hollow



$$F(\mathbf{x}) = x_1^2 + x_1x_2 + x_2^2 = \frac{1}{2}\mathbf{x}^T \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \mathbf{x}$$

$$\nabla^2 F(\mathbf{x}) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad \lambda_1 = 1 \quad \mathbf{z}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \lambda_2 = 3 \quad \mathbf{z}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

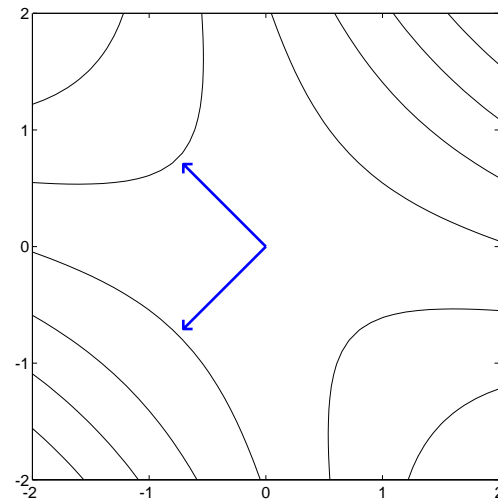
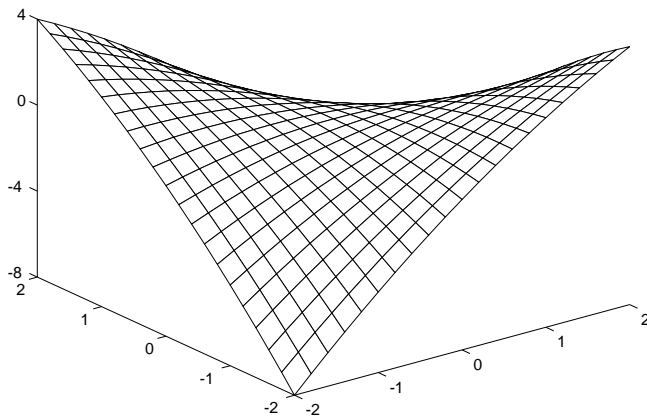


Elongated Saddle



$$F(\mathbf{x}) = -\frac{1}{4}x_1^2 - \frac{3}{2}x_1x_2 - \frac{1}{4}x_2^2 = \frac{1}{2}\mathbf{x}^T \begin{bmatrix} -0.5 & -1.5 \\ -1.5 & -0.5 \end{bmatrix} \mathbf{x}$$

$$\nabla^2 F(\mathbf{x}) = \begin{bmatrix} -0.5 & -1.5 \\ -1.5 & -0.5 \end{bmatrix} \quad \lambda_1 = 1 \quad \mathbf{z}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \lambda_2 = -2 \quad \mathbf{z}_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

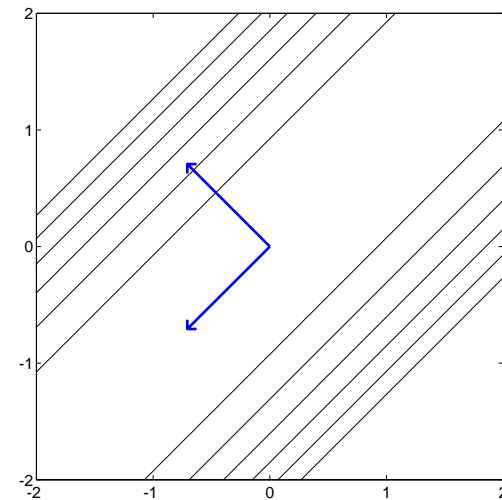
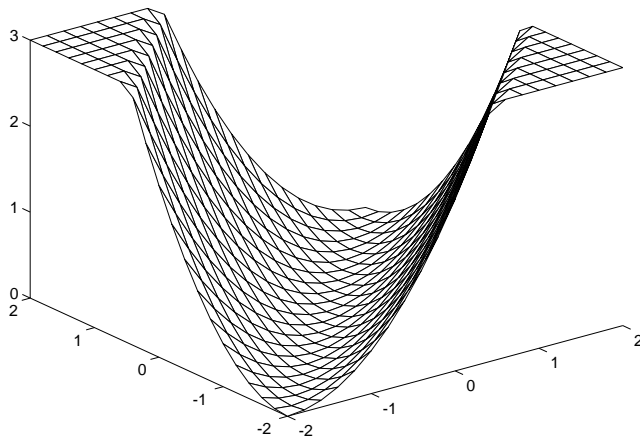


Stationary Valley



$$F(\mathbf{x}) = \frac{1}{2}x_1^2 - x_1x_2 + \frac{1}{2}x_2^2 = \frac{1}{2}\mathbf{x}^T \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \mathbf{x}$$

$$\nabla^2 F(\mathbf{x}) = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad \lambda_1 = 1 \quad \mathbf{z}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \lambda_2 = 0 \quad \mathbf{z}_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$





- If the eigenvalues of the Hessian matrix are all positive, the function will have a single strong minimum.
- If the eigenvalues are all negative, the function will have a single strong maximum.
- If some eigenvalues are positive and other eigenvalues are negative, the function will have a single saddle point.
- If the eigenvalues are all nonnegative, but some eigenvalues are zero, then the function will either have a weak minimum or will have no stationary point.
- If the eigenvalues are all nonpositive, but some eigenvalues are zero, then the function will either have a weak maximum or will have no stationary point.

Stationary Point: $\mathbf{x}^* = -\mathbf{A}^{-1} \mathbf{d}$