شبکه‌های عصبی مصنوعی

درس ۸

رُویه‌های کارایی و نقاط بهینه

Performance Surfaces and Optimum Points

کاظم فولادی قلعه
دانشکده مهندسی، پردیس فارابی
دانشگاه تهران

http://courses.fouladi.ir/nn

8

Performance Surfaces and Optimum Points

رویه‌های کارآیی و نقاط بهینه

PERFORMANCE SURFACES AND OPTIMUM POINTS

یکی از تکنیک‌های پایه برای آموزش شبکه‌ی عصبی: تکنیک یادگیری کارآیی
(پارامترهای شبکه تنظیم می‌شوند تا کارآیی شبکه بهینه شود)

سایر تکنیک‌ها: یادگیری انجمنی، یادگیری رقابتی

هدف این فصل: بررسی رُویه‌های کارآیی و تعیین شرایط وجود ماکزیمم‌ها و می‌نیمم‌های آن

دو گام برای **فرآیند بهنیه‌سازی**

i. تعریف کارآیی: شاخص کارآیی: **تابع هزینه**
ii. **جستجو** در فضای پارامتر: تنظیم وزن‌ها و بایاس‌ها به‌منظور بهبود شاخص کارآیی

شاخص کارآیی مورد نظر برای می‌نیمم‌سازی را با $F(x)$ نمایش می‌دهیم.
x: اسکالری که باید تنظیم شود.
فرض: تابع F تحلیلی است (یعنی تمام مشتقات آن موجود است)

Prepared by Kazim Fouladi | Fall 2017 | 2nd Edition
کاظم فولادی

شبکه‌های عصبی مصنوعی

رُویه‌های کارایی و نقاط بهینه

۱

سری تیلور

Taylor Series Expansion

$$F(x) = F(x^*) + \frac{d}{dx}F(x)\Big|_{x = x^*}(x - x^*)$$

$$+ \frac{1}{2}\frac{d^2}{dx^2}F(x)\Bigg|_{x = x^*}(x - x^*)^2 + \cdots$$

$$+ \frac{1}{n!}\frac{d^n}{dx^n}F(x)\Bigg|_{x = x^*}(x - x^*)^n + \cdots$$

بسط سری تیلور

TAYLOR SERIES EXPANSION

$$F(x) = F(x^*) + \frac{d}{dx}F(x)\Big|_{x = x^*}(x - x^*)$$
$$+ \frac{1}{2}\frac{d^2}{dx^2}F(x)\Big|_{x = x^*}(x - x^*)^2 + \cdots$$
$$+ \frac{1}{n!}\frac{d^n}{dx^n}F(x)\Big|_{x = x^*}(x - x^*)^n + \cdots$$

استفاده از بسط سری تیلور برای تقریب زدن شاخص کارآیی :
با محدود کردن تعداد جملات

Example

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

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

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

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

Taylor series approximations:

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

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

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

بسط سری تیلور

مثال

Taylor Series Expansion

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

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:

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

تقریب مرتبه صفر
Zeroth-Order Approximation

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

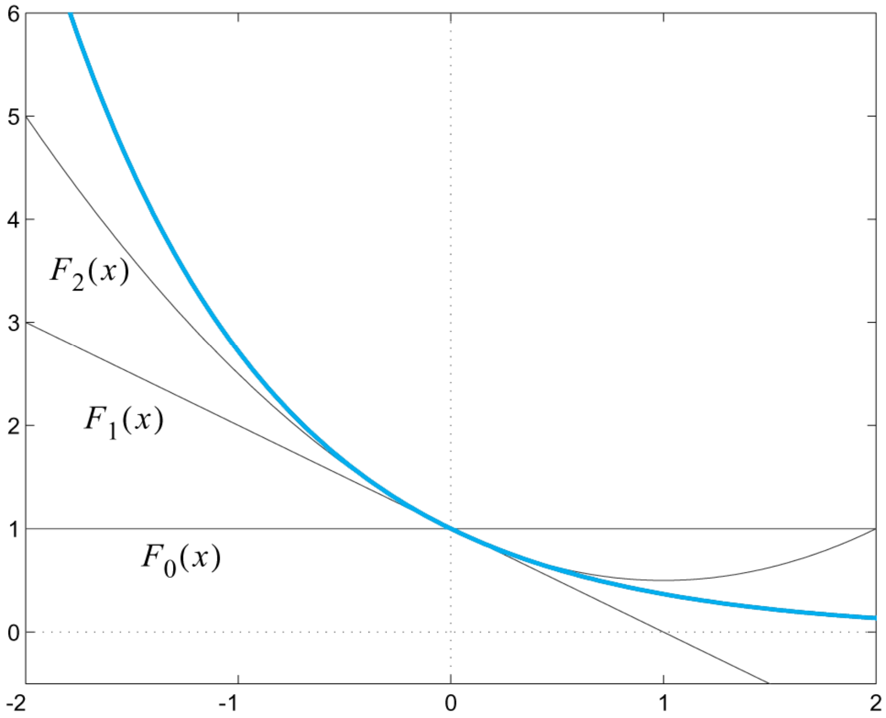
تقریب مرتبه یک
First-Order Approximation

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

تقریب مرتبه دو
Second-Order Approximation

8

Plot of Approximations

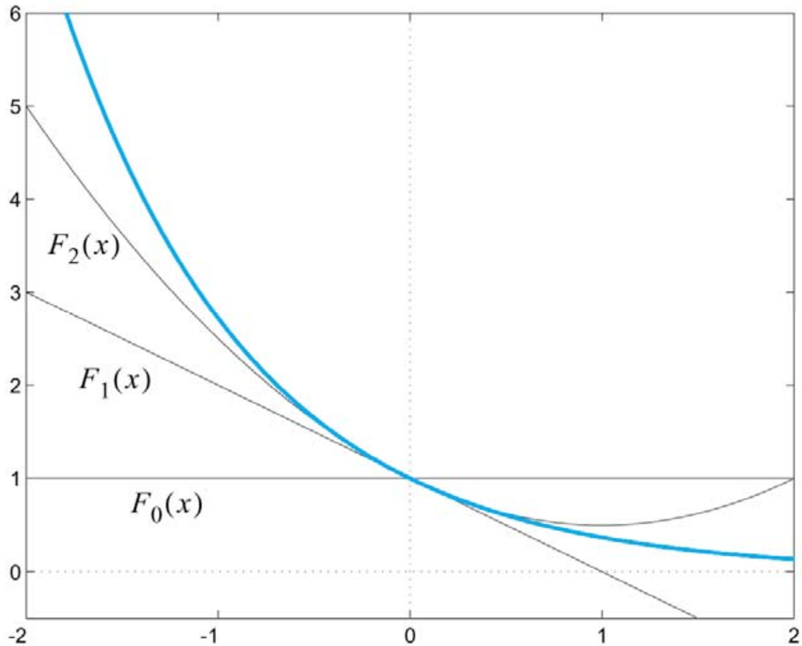


بسط سری تیلور

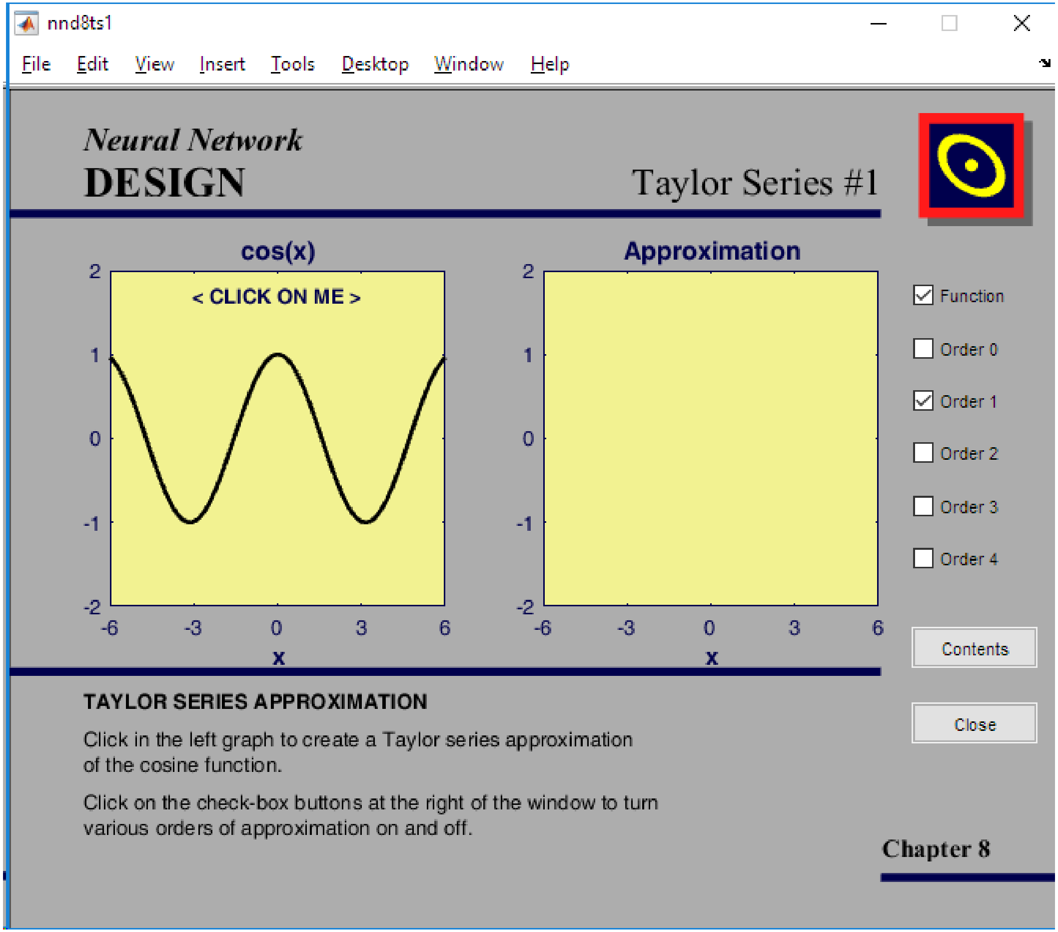
مثال: ترسیم تقریب‌ها

Taylor Series Expansion

همه‌ی تقریب‌ها دقیق هستند اگر x بسیار نزدیک به صفر باشد.
با دور شدن از صفر، تقریب‌های مرتبه بالاتر دقیق‌تر هستند.



از تقریب‌های سری تیلور شاخص کارآیی (تابع هزینه) برای بررسی شکل تابع در همسایگی نقاط بهینه‌ی ممکن استفاده می‌کنیم.


nnd8ts1
File Edit View Insert Tools Desktop Window Help
Neural Network
DESIGN
Taylor Series #1
cos(x)
< CLICK ON ME >
Approximation
2
1
0
-1
-2
-6 -3 0 3 6
x
Function
Order 0
Order 1
Order 2
Order 3
Order 4
Contents
Close
TAYLOR SERIES APPROXIMATION
Click in the left graph to create a Taylor series approximation of the cosine function.
Click on the check-box buttons at the right of the window to turn various orders of approximation on and off.
Chapter 8

```
>> nnd8ts1
```

Vector Case

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

$$F(\mathbf{x}) = F(\mathbf{x}^*) + \frac{\partial}{\partial x_1} F(\mathbf{x})\Big|_{\mathbf{x} = \mathbf{x}^*} (x_1 - x_1^*) + \frac{\partial}{\partial x_2} F(\mathbf{x})\Big|_{\mathbf{x} = \mathbf{x}^*} (x_2 - x_2^*)$$

$$+ \cdots + \frac{\partial}{\partial x_n} F(\mathbf{x})\Big|_{\mathbf{x} = \mathbf{x}^*} (x_n - x_n^*) + \frac{1}{2} \frac{\partial^2}{\partial x_1^2} F(\mathbf{x})\Big|_{\mathbf{x} = \mathbf{x}^*} (x_1 - x_1^*)^2$$

$$+ \frac{1}{2} \frac{\partial^2}{\partial x_1 \partial x_2} F(\mathbf{x})\Big|_{\mathbf{x} = \mathbf{x}^*} (x_1 - x_1^*)(x_2 - x_2^*) + \cdots$$

بسط سری تیلور

حالت برداری (چند متغیره)

Taylor Series Expansion: Vector Case

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

$$F(\mathbf{x}) = F(\mathbf{x}^*) + \frac{\partial}{\partial x_1}F(\mathbf{x})\Big|_{\mathbf{x}=\mathbf{x}^*}(x_1 - x_1^*) + \frac{\partial}{\partial x_2}F(\mathbf{x})\Big|_{\mathbf{x}=\mathbf{x}^*}(x_2 - x_2^*)$$

$$+ \cdots + \frac{\partial}{\partial x_n}F(\mathbf{x})\Big|_{\mathbf{x}=\mathbf{x}^*}(x_n - x_n^*) + \frac{1}{2}\frac{\partial^2}{\partial x_1^2}F(\mathbf{x})\Big|_{\mathbf{x}=\mathbf{x}^*}(x_1 - x_1^*)^2$$

$$+ \frac{1}{2}\frac{\partial^2}{\partial x_1 \partial x_2}F(\mathbf{x})\Big|_{\mathbf{x}=\mathbf{x}^*}(x_1 - x_1^*)(x_2 - x_2^*) + \cdots$$

$F(\mathbf{x})$ تابعی از بردار $\mathbf{x}$ شامل همه‌ی پارامترهای شبکه است.

Matrix Form

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

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

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

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

بسط سری تیلور

حالت برداری (چند متغیره): فرم ماتریسی

Taylor Series Expansion: Vector Case: Matrix Form

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

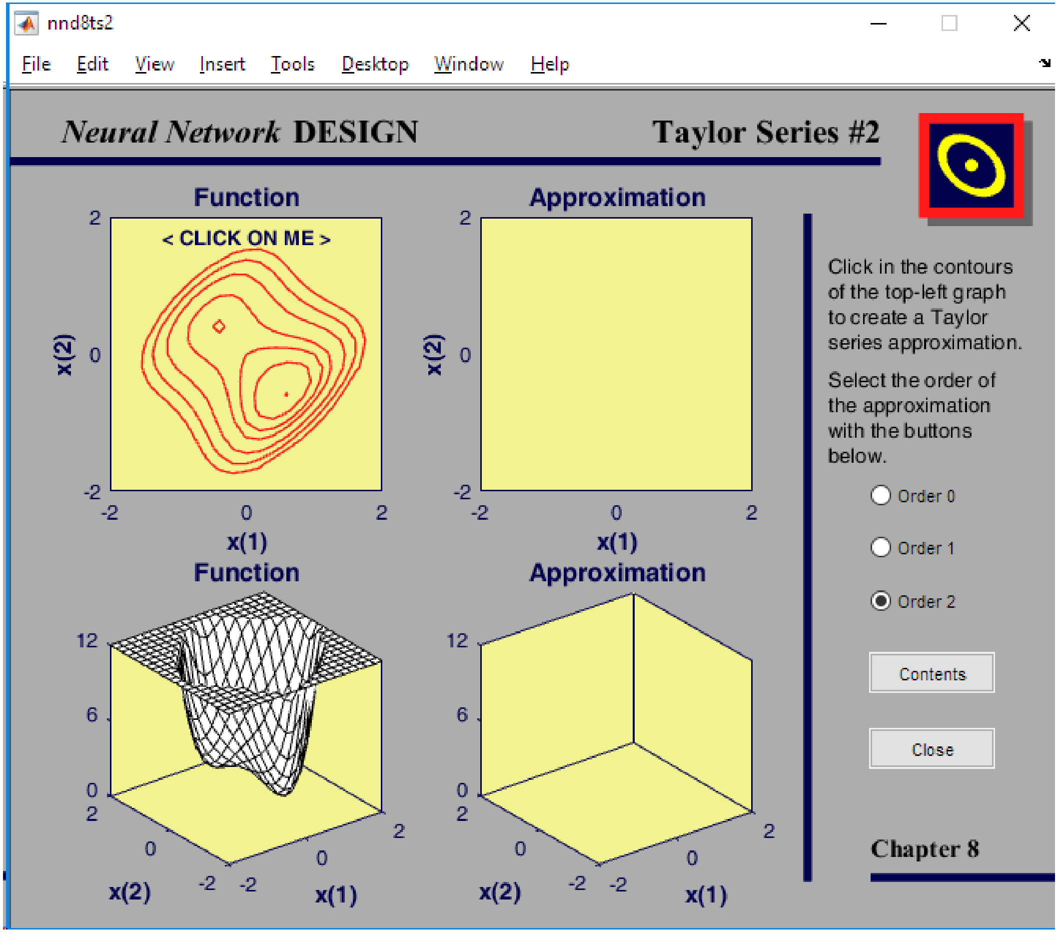
$$+ \frac{1}{2}(\mathbf{x}-\mathbf{x}^*)^T \nabla^2 F(\mathbf{x}) \Big|_{\mathbf{x}=\mathbf{x}^*} (\mathbf{x}-\mathbf{x}^*) + \cdots$$

مشتقات اول
گرادیان
Gradient

Hessian
مشتقات دوم
هسی

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

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


nnd8ts2
File Edit View Insert Tools Desktop Window Help
Neural Network DESIGN
Taylor Series #2
Function
< CLICK ON ME >
Approximation
Click in the contours of the top-left graph to create a Taylor series approximation.
Select the order of the approximation with the buttons below.
Order 0
Order 1
Order 2
Contents
Close
Chapter 8
x(1)
x(2)


MATLAB

```
>> nnd8ts2
```

رُویه‌های کارایی و نقاط بهینه

۲

مشتقات جهتی

8

Directional Derivatives

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

مشتقات جهتی

Directional Derivatives

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

(ith 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 $\mathbf{p}$: $\dfrac{\mathbf{p}^T \nabla F(\mathbf{x})}{\|\mathbf{p}\|}$ مشتق در یک راستای دلخواه

باید گرادیان در بردار جهت یکه‌ی **p** ضرب داخلی شود.

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

8

Example

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

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

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

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

مشتقات جهتی

مثال

Directional Derivatives

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

نقطه‌ی کار : $\mathbf{x}^* = \begin{bmatrix} 0.5 \\ 0 \end{bmatrix}$

$\mathbf{p} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$:جهت دلخواه

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

نکته:

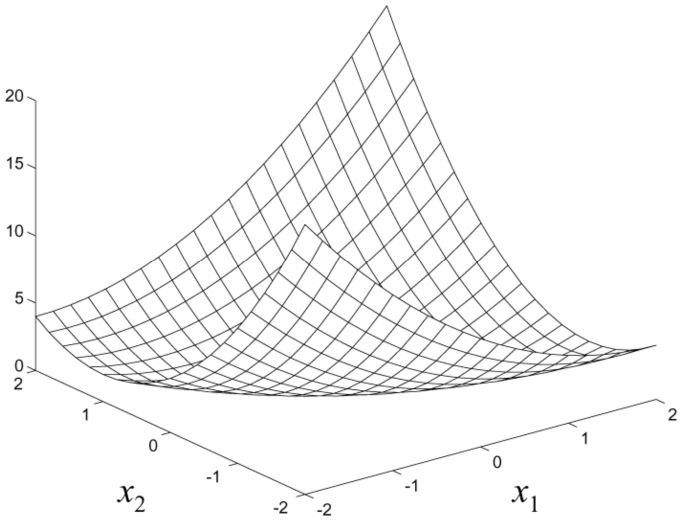
- هر جهت عمود بر گرادیان، شیب صفر دارد.
- بیشترین شیب در جهت گرادیان است.
  (طول آن مهم نیست)

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

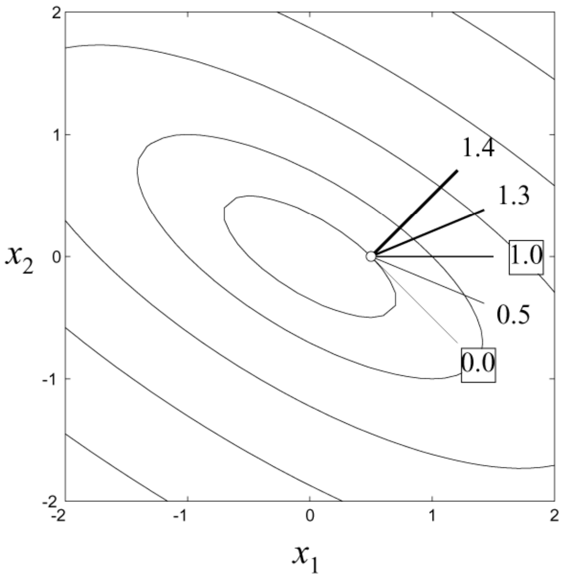
شیب تابع در راستای $\mathbf{p}$ در نقطه‌ی $\mathbf{x}^*$ صفر است.

Prepared by Kazim Fouladi | Fall 2017 | 2nd Edition
کاظم فولادی

Plots



Directional
Derivatives



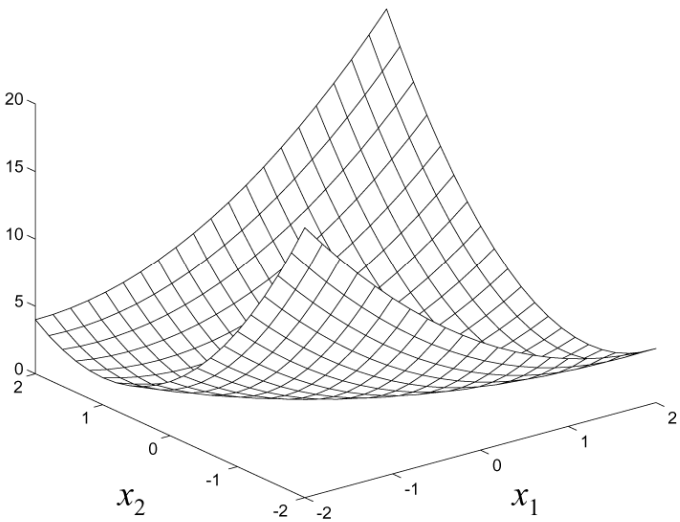
مشتقات جهتی

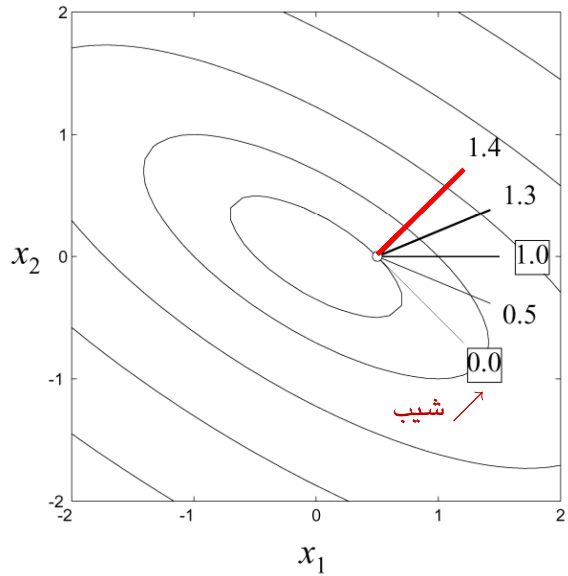
مثال: نمودار

DIRECTIONAL DERIVATIVES

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

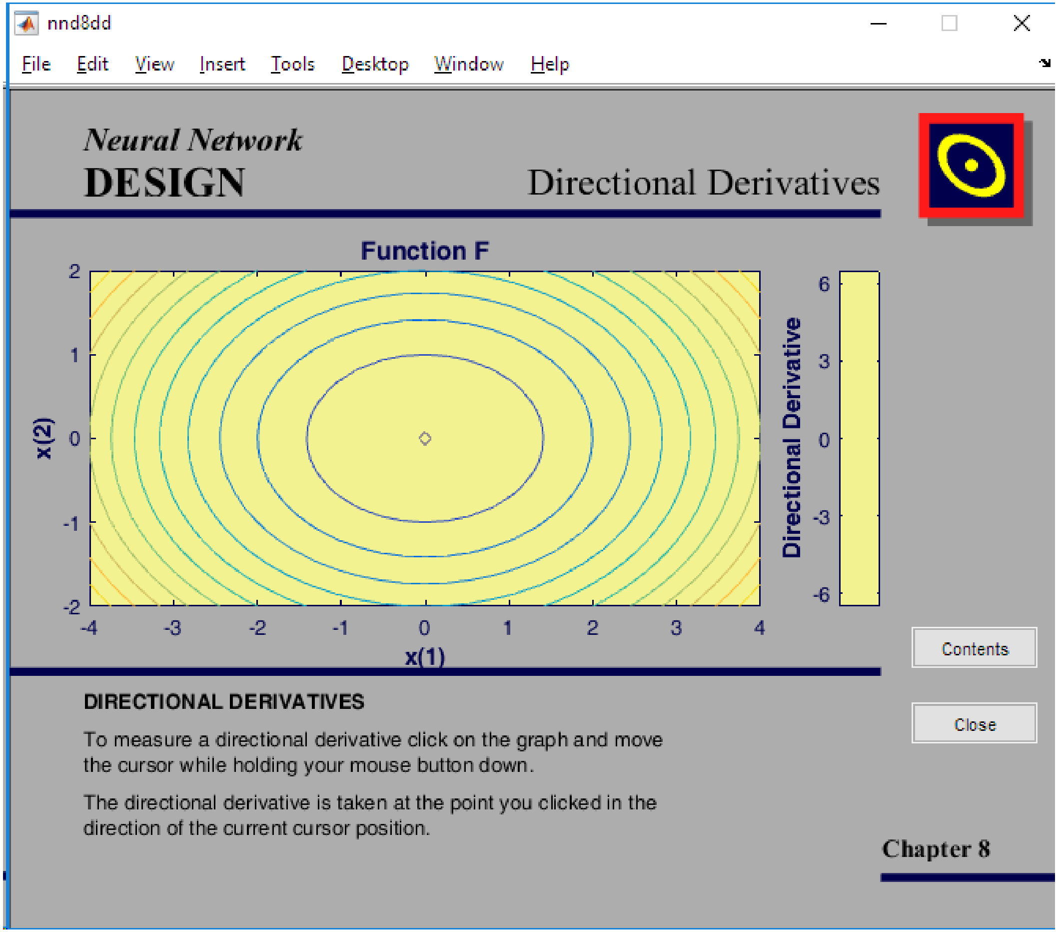
Directional Derivatives





نمودار کانتوری
contour plot

(مقدار تابع بر روی هر کانتور ثابت است.)



MATLAB

>> nnd8dd

رُویه‌های کارایی و نقاط بهینه

۳

نقاط می‌نیمم

Minima

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 $\delta > \|\Delta\mathbf{x}\| > 0$.

می‌نیمم‌ها

MINIMA

می‌نیمم محلی — Strong Minimum — می‌نیمم قوی

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 $\delta > \|\Delta\mathbf{x}\| > 0$.

اگر از می‌نیمم قوی (محلی) در هر جهتی به میزان فاصله‌ی کمی دور شویم، مقدار تابع افزایش می‌یابد.

می‌نیمم مطلق — Global Minimum — می‌نیمم سراسری

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 — می‌نیمم ضعیف

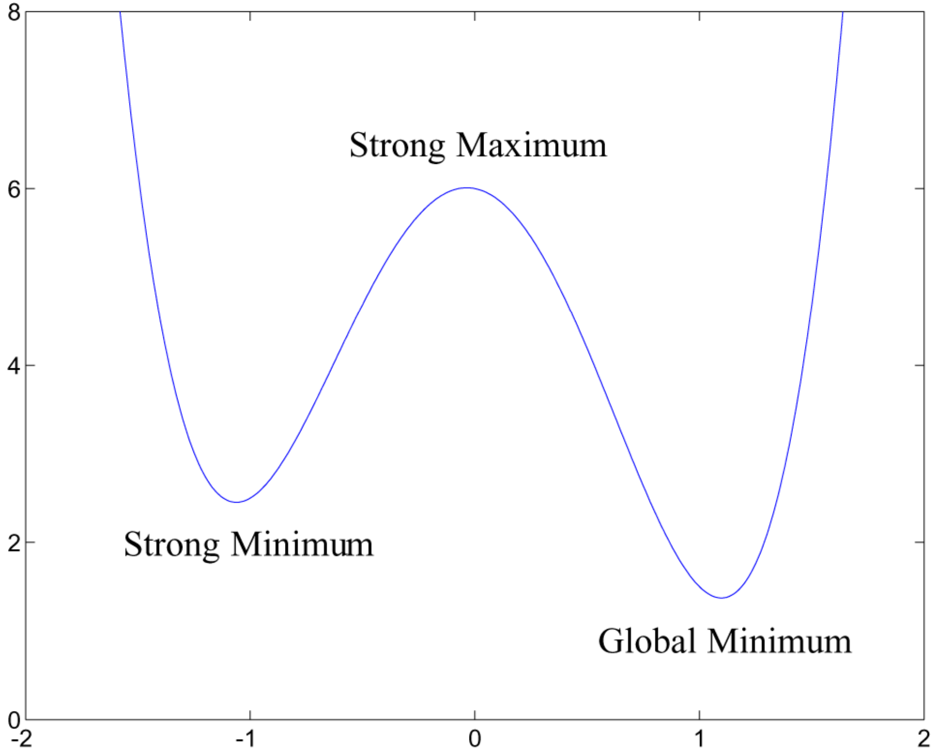
The point $\mathbf{x}^*$ is a weak minimum of $F(\mathbf{x})$ if it is not a strong minimum, and a scalar $\delta > 0$ exists, such that $F(\mathbf{x}^*) \le F(\mathbf{x}^* + \Delta\mathbf{x})$ for all $\Delta\mathbf{x}$ such that $\delta > \|\Delta\mathbf{x}\| > 0$.

می‌نیمم ضعیف مانند می‌نیمم قوی است اما مقدار تابع در بعضی جهت‌ها ممکن است ثابت بماند.

8

Scalar Example

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

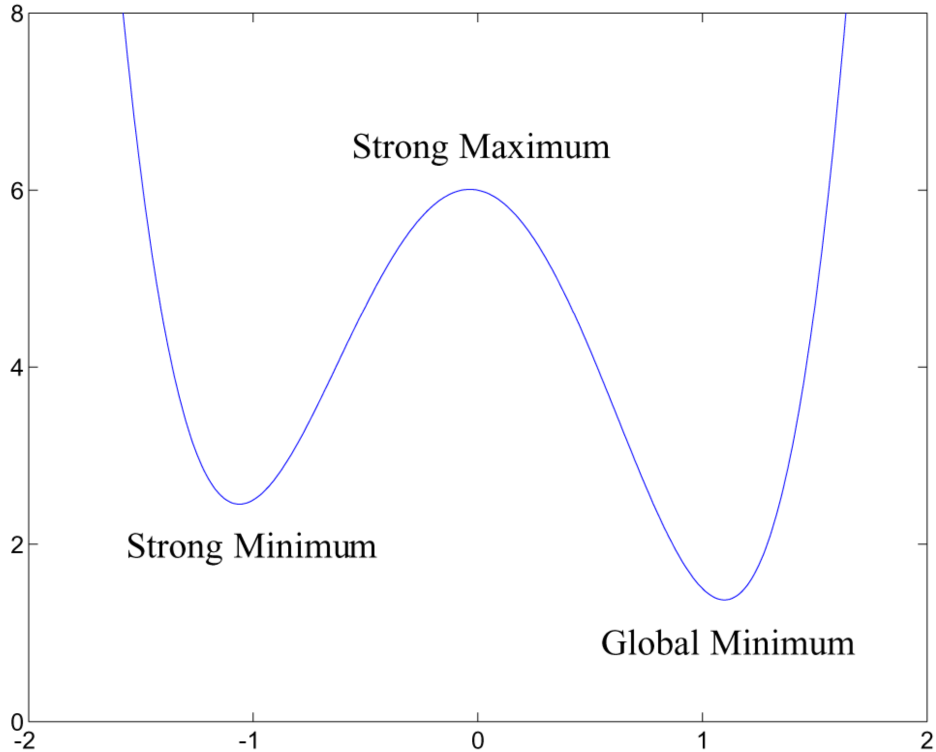


می‌نیمم‌ها

مثال اسکالر

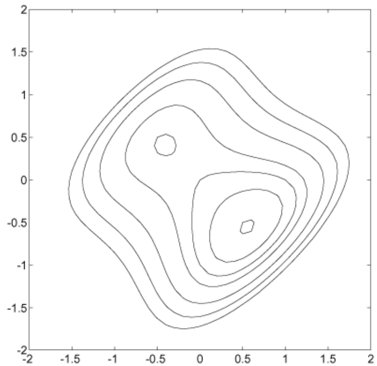
MINIMA

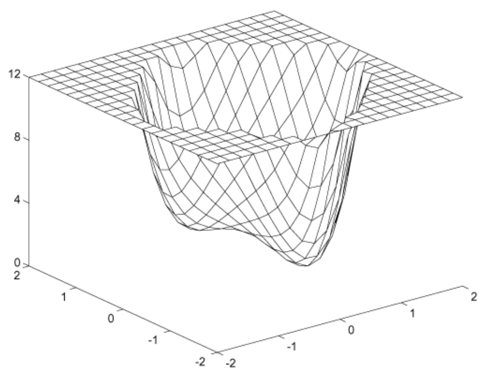
$$F(x) = 3x^4 - 7x^2 - \frac{1}{2}x + 6$$



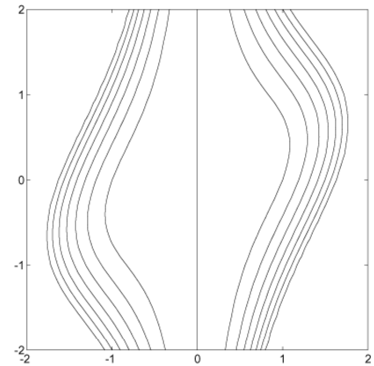
Vector Example

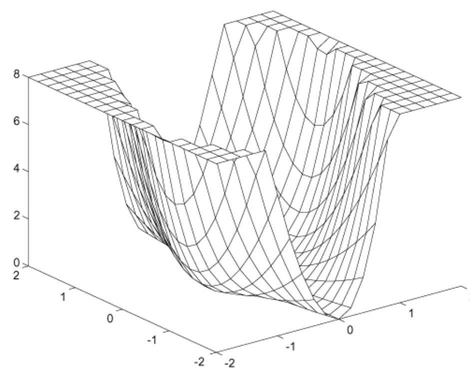
$$F(\mathbf{x}) = (x_2 - x_1)^4 + 8x_1x_2 - x_1 + x_2 + 3$$





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





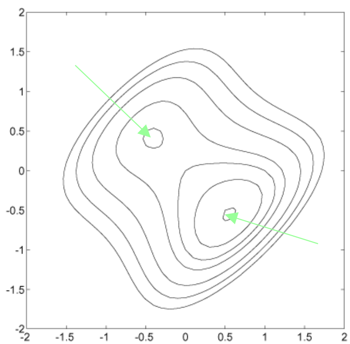
می‌نیمم‌ها

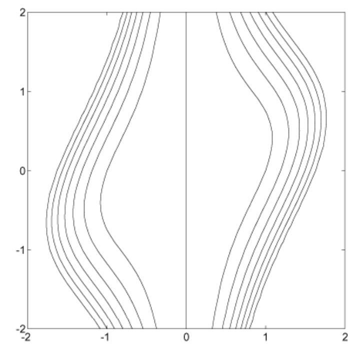
مثال برداری

MINIMA

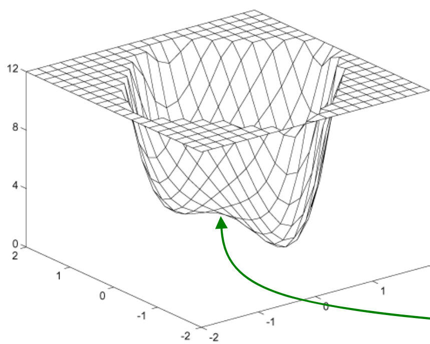
$$F(\mathbf{x}) = (x_2 - x_1)^4 + 8x_1x_2 - x_1 + x_2 + 3$$

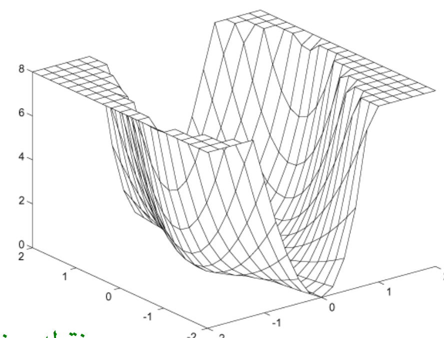
$$F(\mathbf{x}) = (x_1^2 - 1.5x_1x_2 + 2x_2^2)x_1^2$$





$x_1 = 0$
می‌نیمم ضعیف





نقطه‌ی زینی
saddle point

(در یک راستا ماکزیمم محلی و در راستای عمود بر آن می‌نیمم محلی است.)

رُویه‌های کارایی و نقاط بهینه

۴

شرایط لازم برای نقطه‌ی بهینه

8 First-Order Optimality Condition

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

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

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

شرایط لازم برای بهینگی

شرایط مرتبه اول

First-Order Optimality Condition

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

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

For small $\Delta\mathbf{x}$: جملات مرتبه بالاتر قابل چشم‌پوشی است

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

برای می‌نیمم لازم است که:

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

Since this must be true for every $\Delta\mathbf{x}$, $\boxed{\nabla F(\mathbf{x})\Big|_{\mathbf{x}=\mathbf{x}^*} = \mathbf{0}}$ برای هر $\Delta\mathbf{x}$ داریم:

stationary point نقطه‌ی ایستان

هر نقطه‌ای که این شرط (گرادیان صفر) را دارا باشد، **نقطه‌ی ایستان** نام دارد (شرط لازم و نه کافی برای می‌نیمم بودن).

8 Second-Order Condition

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

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

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:

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

This is a **<u>sufficient</u>** condition for optimality.

A **<u>necessary</u>** condition is that the Hessian matrix be positive semidefinite. A matrix A is positive semidefinite if:

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

شرایط لازم برای بهینگی

شرایط مرتبه دوم

SECOND-ORDER CONDITION

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

برای نقطه‌ی ایستان $\mathbf{x}^*$ داریم:

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

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 **A** is positive definite if:

شرط کافی برای بهینگی:
ماتریس هسی باید معین مثبت باشد

$$\boxed{\mathbf{z}^T \mathbf{A} \mathbf{z} > 0}$$

for any $\mathbf{z} \neq 0$.

همه‌ی مقادیر ویژه مثبت باشد

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

A **necessary** condition is that the Hessian matrix be positive semidefinite. A matrix A is positive semidefinite if:

شرط لازم برای بهینگی:
ماتریس هسی باید نیمه معین مثبت باشد

$$\boxed{\mathbf{z}^T \mathbf{A} \mathbf{z} \geq 0}$$

for any $\mathbf{z}$.

همه‌ی مقادیر ویژه نامنفی باشد

اگر جمله‌ی مرتبه دوم صفر باشد، اما جمله‌ی مرتبه سوم مثبت باشد، باز هم می‌توان می‌نیمم محلی داشت.

Example

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

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

$$\nabla^2 F(\mathbf{x}) = \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix}$$

(Not a function of $\mathbf{x}$ 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.

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

$\lambda = 0.76, 5.24$ Both eigenvalues are positive, therefore <u>strong minimum</u>.

شرایط لازم برای بهینگی

مثال

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

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

$$\nabla^2 F(\mathbf{x}) = \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix}$$

(Not a function of $\mathbf{x}$ 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.

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

$\lambda = 0.76, 5.24$ Both eigenvalues are positive, therefore <u>strong minimum</u>.

شبکه‌های عصبی مصنوعی

رُویه‌های کارایی و نقاط بهینه

۵

توابع مربعی

Quadratic Functions

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

Gradient and Hessian:

Useful properties of gradients:

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

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

Gradient of Quadratic Function:

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

Hessian of Quadratic Function:

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

تابع درجه دوم

تابع مربعی

Quadratic Functions

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

Gradient and Hessian:

Useful properties of gradients:

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

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

Gradient of Quadratic Function:

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

Hessian of Quadratic Function:

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

همه‌ی مشتقات مرتبه بالاتر آن صفر است.

8 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.

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

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

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

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

سیستم ویژه‌ی هسی

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.

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

استفاده از
بردارهای ویژه‌ی ماتریس هسی به‌عنوان پایه

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

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

ماتری هسی متقارن است،
پس پردارهای ویژه‌ی آن متعامد هستند.

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

Second Directional Derivative

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

Represent **p** with respect to the eigenvectors (new basis):

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

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

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

سیستم ویژه‌ی هسی

مشتقات دوم جهتی

SECOND DIRECTIONAL DERIVATIVE

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

مشتق دوم در راستای **p**:

Represent **p** with respect to the eigenvectors (new basis):

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

تبدیل پایه‌ی **p** به **B**:

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

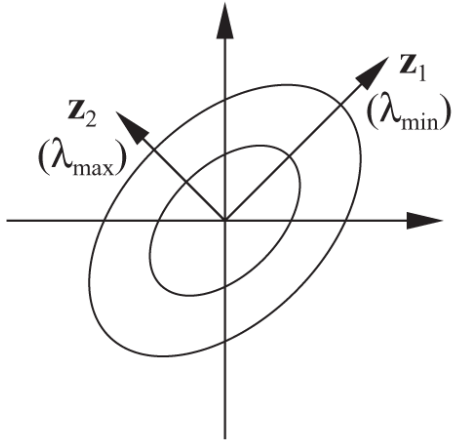
میانگین وزن‌دار
مقادیر ویژه
$\Downarrow$
بین بزرگ‌ترین و
کوچک‌ترین مقدار ویژه

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

Eigenvector (Largest Eigenvalue)

$$\mathbf{p} = \mathbf{z}_{max} \qquad \mathbf{c} = \mathbf{B}^T\mathbf{p} = \mathbf{B}^T\mathbf{z}_{max} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

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



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

سیستم ویژه‌ی هسی

بردار ویژه (متناظر با بزرگترین مقدار ویژه)

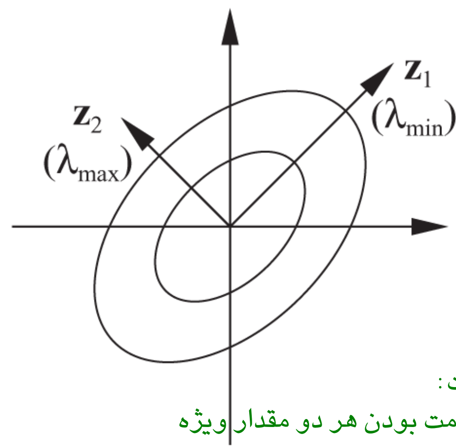
EIGENVECTOR (LARGEST EIGENVALUE)

مشتق دوم = بزرگترین مقدار ویژه

$$\mathbf{p} = \mathbf{z}_{max} \qquad \mathbf{c} = \mathbf{B}^T\mathbf{p} = \mathbf{B}^T\mathbf{z}_{max} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

فقط در یک نقطه،
1 می‌شود.
(زیرا بردارهای یکه،
متعامد یکه هستند)

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



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

مقادیر ویژه، انحنا در راستای بردارهای ویژه را بازنمایی می‌کنند.

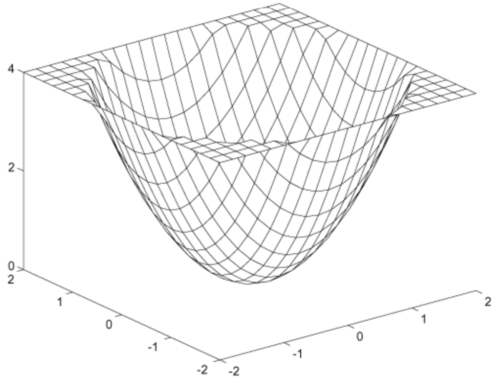
❄ حالت:
هم علامت بودن هر دو مقدار ویژه

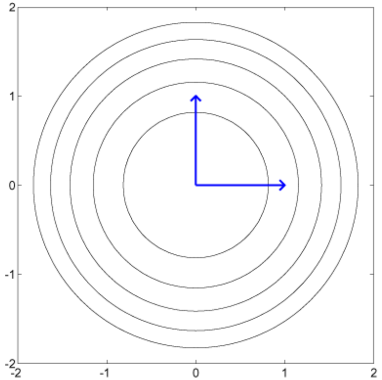
Circular Hollow

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

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

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





سیستم ویژه‌ی هسی

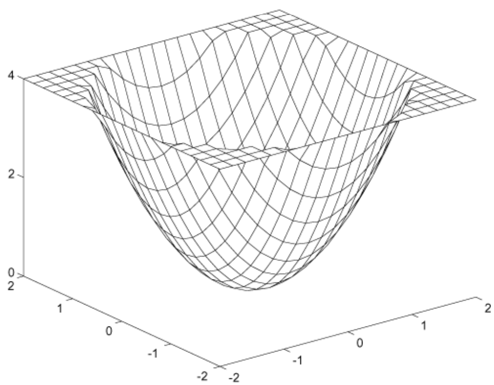
تباهیدگی دایره‌ای

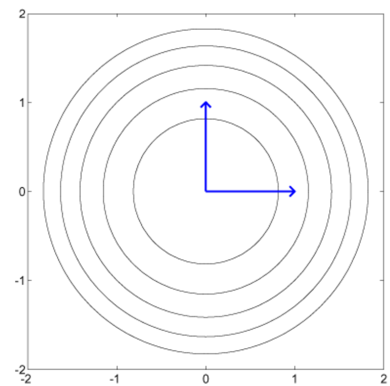
Circular Hollow

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

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

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





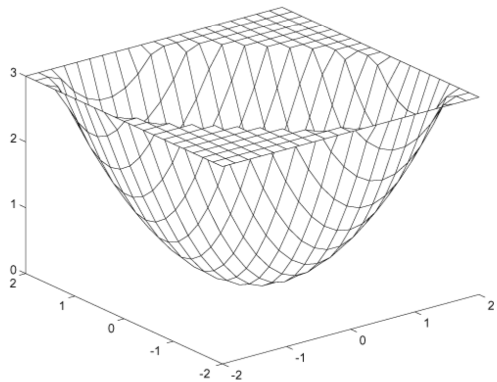
❊ حالت :

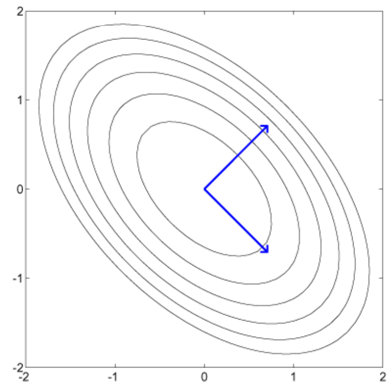
تساوی مقادیر ویژه ⇐ انحنا در هر دو راستا مساوی است ⇐ کانتورهای دایره‌ای

Elliptical Hollow

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

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





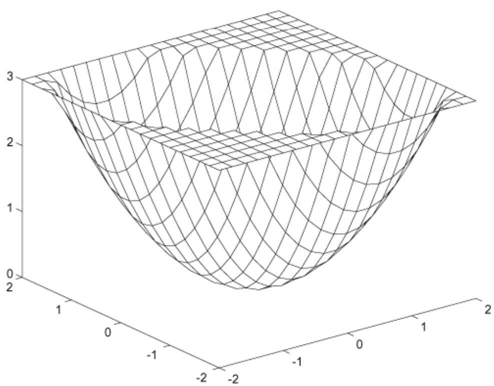
سیستم ویژه‌ی هسی

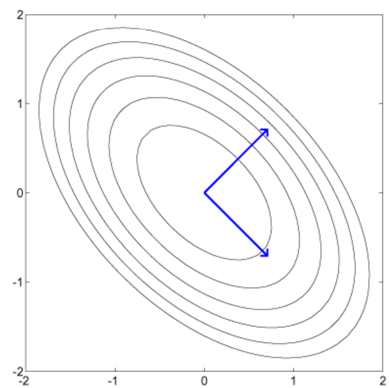
تباهیدگی بیضوی

ELLIPTICAL HOLLOW

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

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





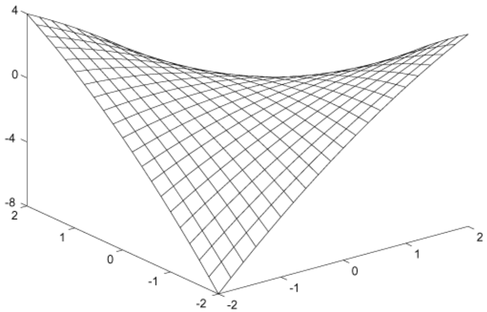
❋ حالت :

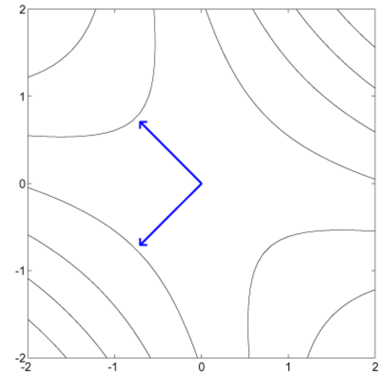
حداکثر انحنا در جهت بزرگ‌ترین مقدار ویژه است ⇐ کانتورها در این جهت به هم نزدیک هستند.

Elongated Saddle

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

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





سیستم ویژه‌ی هسی

زین ممتد کشیده

ELONGATED SADDLE

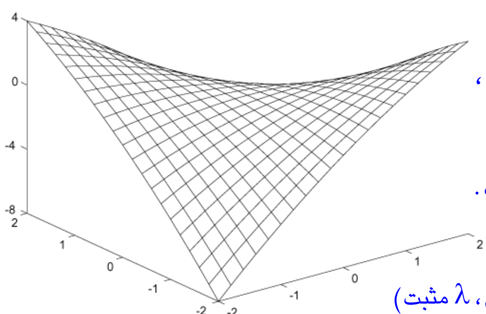
$$F(\mathbf{x}) = -\frac{1}{4}x_1^2 - \frac{3}{2}x_1x_2 - \frac{1}{4}x_2^2 = \frac{1}{2}\mathbf{x}^T\begin{bmatrix}-0.5 & -1.5\\ -1.5 & -0.5\end{bmatrix}\mathbf{x}$$

❋ حالت:
مقادیر ویژه‌ی ناهم‌علامت

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

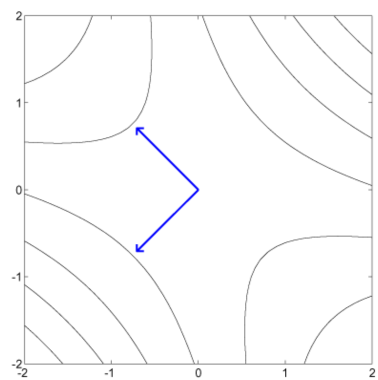
انحنای مثبت در جهت $\mathbf{z}_1$
(اندازه‌ی کوچک‌تر، انحنای کمتر)

انحنای منفی در جهت $\mathbf{z}_2$
(اندازه‌ی بزرگ‌تر، انحنای بیشتر)



چون هسی
معین مثبت نیست،
پس
$\mathbf{x}^* = 0$
نقطه‌ی زینی است.

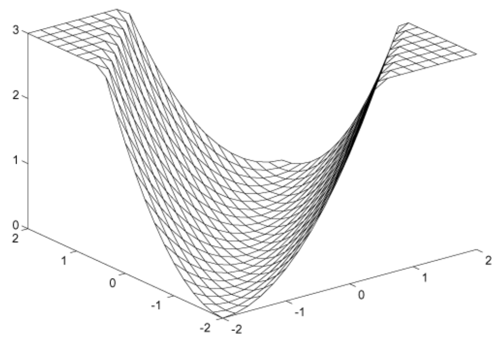
در جهت $\mathbf{z}_1$ (می‌نیمم، λ مثبت)
در جهت $\mathbf{z}_2$ (ماکزیمم، λ منفی)

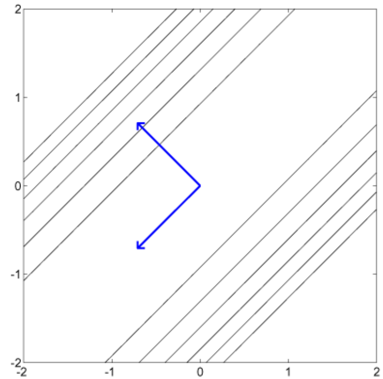


Stationary Valley

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

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





سیستم ویژه‌ی هسی

دره‌ی ایستان

STATIONARY VALLEY

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

❄ حالت :

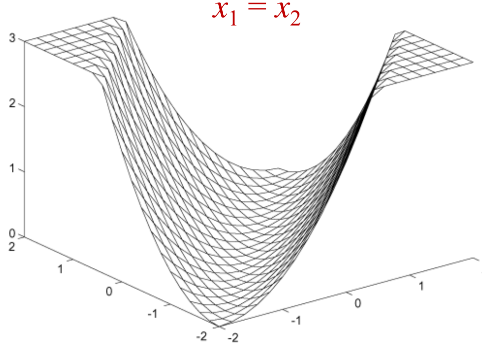
یک مقدار ویژه صفر است.

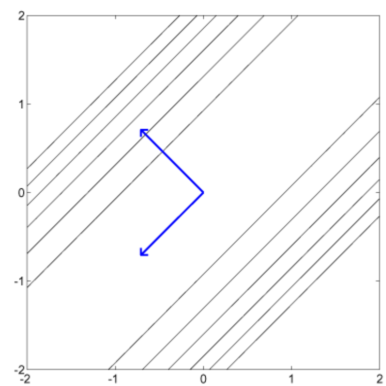
$$\nabla^2 F(\mathbf{x}) = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \qquad \lambda_1 = 1 \qquad \mathbf{z}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \qquad \lambda_2 = 0 \qquad \mathbf{z}_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

درراستای $\mathbf{z}_2$ انحنا صفر است

می‌نیمم ضعیف در راستای $\mathbf{z}_2$

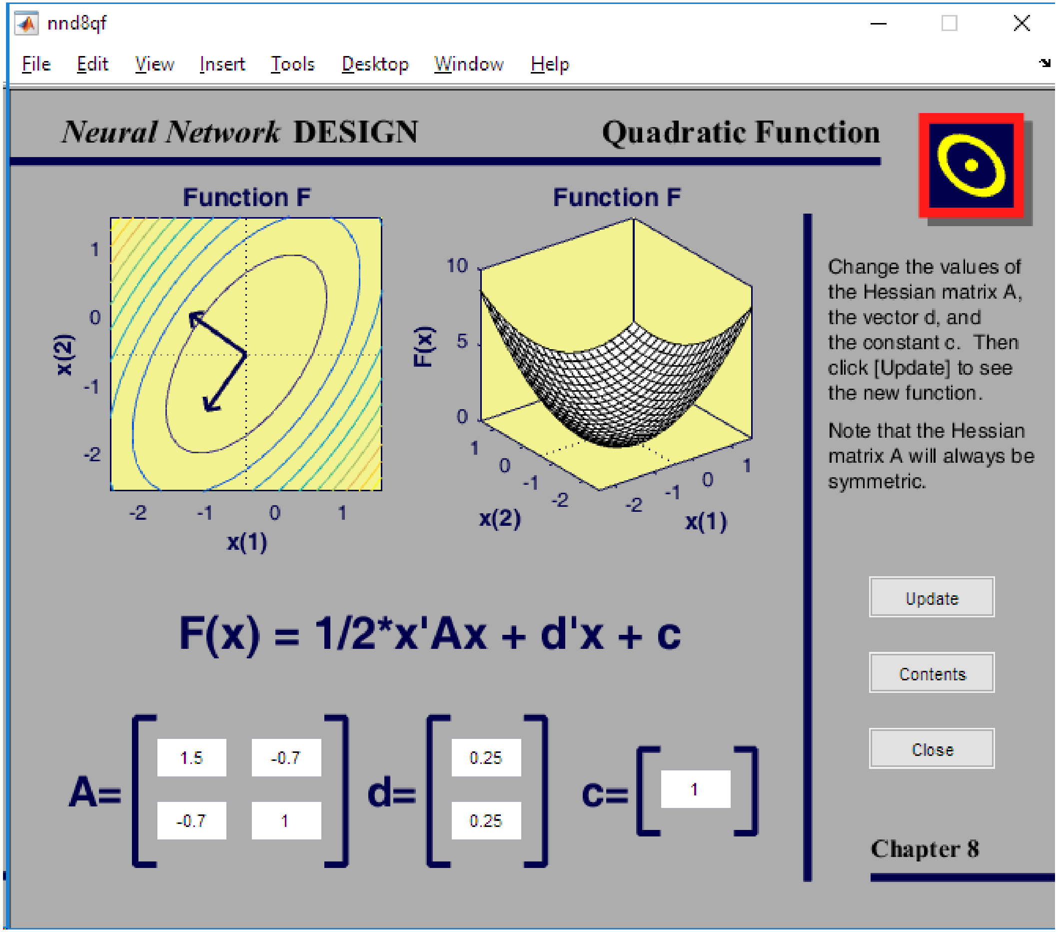
$x_1 = x_2$





نکته: توابع مرتبه بالاتر از دو می‌توانند با هسی نیمه معین مثبت، می‌نیمم قوی داشته باشند.

کاظم فولادی


nnd8qf
File Edit View Insert Tools Desktop Window Help
Neural Network DESIGN
Quadratic Function
Function F
Function F
x(2)
x(1)
F(x)
x(2)
x(1)
Change the values of the Hessian matrix A, the vector d, and the constant c. Then click [Update] to see the new function.
Note that the Hessian matrix A will always be symmetric.
F(x) = 1/2*x'Ax + d'x + c
A=
1.5
-0.7
-0.7
1
d=
0.25
0.25
c=
1
Update
Contents
Close
Chapter 8

```
>> nnd8qf
```

Quadratic Function Summary

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

خلاصه ویژگی‌های تابع درجه دوم

Quadratic Function Summary

مقادیر ویژه‌ی هسی:

همه مثبت ⇐ یک min قوی

- If the eigenvalues of the Hessian matrix are all positive, the function will have a single strong minimum.

همه منفی ⇐ یک max قوی

- If the eigenvalues are all negative, the function will have a single strong maximum.

بعضی مثبت، بعضی منفی ⇐ وجود نقطه‌ی زینی

- If some eigenvalues are positive and other eigenvalues are negative, the function will have a single saddle point.

همه نامنفی، بعضی صفر ⇐ وجود نقطه‌ی min ضعیف یا عدم وجود نقطه‌ی زینی

- 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.

همه نامثبت، بعضی صفر ⇐ وجود نقطه‌ی max ضعیف یا عدم وجود نقطه‌ی زینی

- 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.

نقطه‌ی ایستان

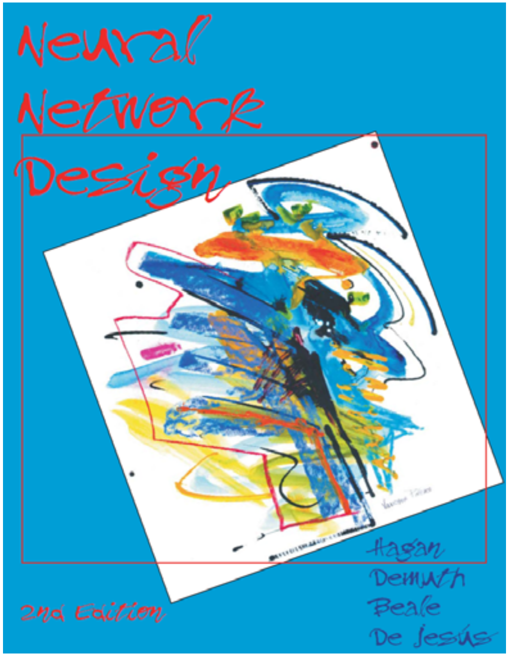
Stationary Point: $\mathbf{x}^* = -\mathbf{A}^{-1}\mathbf{d}$

رُویه‌های کارایی و نقاط بهینه

۶

منابع

منبع اصلی



8 Performance Surfaces and Optimum Points

Objectives

This chapter lays the foundation for a type of neural network training technique called performance learning. There are several different classes of network learning laws, including associative learning (as in the Hebbian learning of Chapter 7) and competitive learning (which we will discuss in Chapter 16). Performance learning is another important class of learning law, in which the network parameters are adjusted to optimize the performance of the network. In the next two chapters we will lay the groundwork for the development of performance learning, which will then be presented in detail in Chapters 10–14. The main objective of the present chapter is to investigate performance surfaces and to determine conditions for the existence of minima and maxima of the performance surface. Chapter 9 will follow this up with a discussion of procedures to locate the minima or maxima.

8-1

Martin T. Hagan, Howard B. Demuth, Mark H. Beale, Orlando De Jesus,
Neural Network Design,
2nd Edition, Martin Hagan, 2014.
Chapter 8

Online version can be downloaded from: http://hagan.okstate.edu/nnd.html