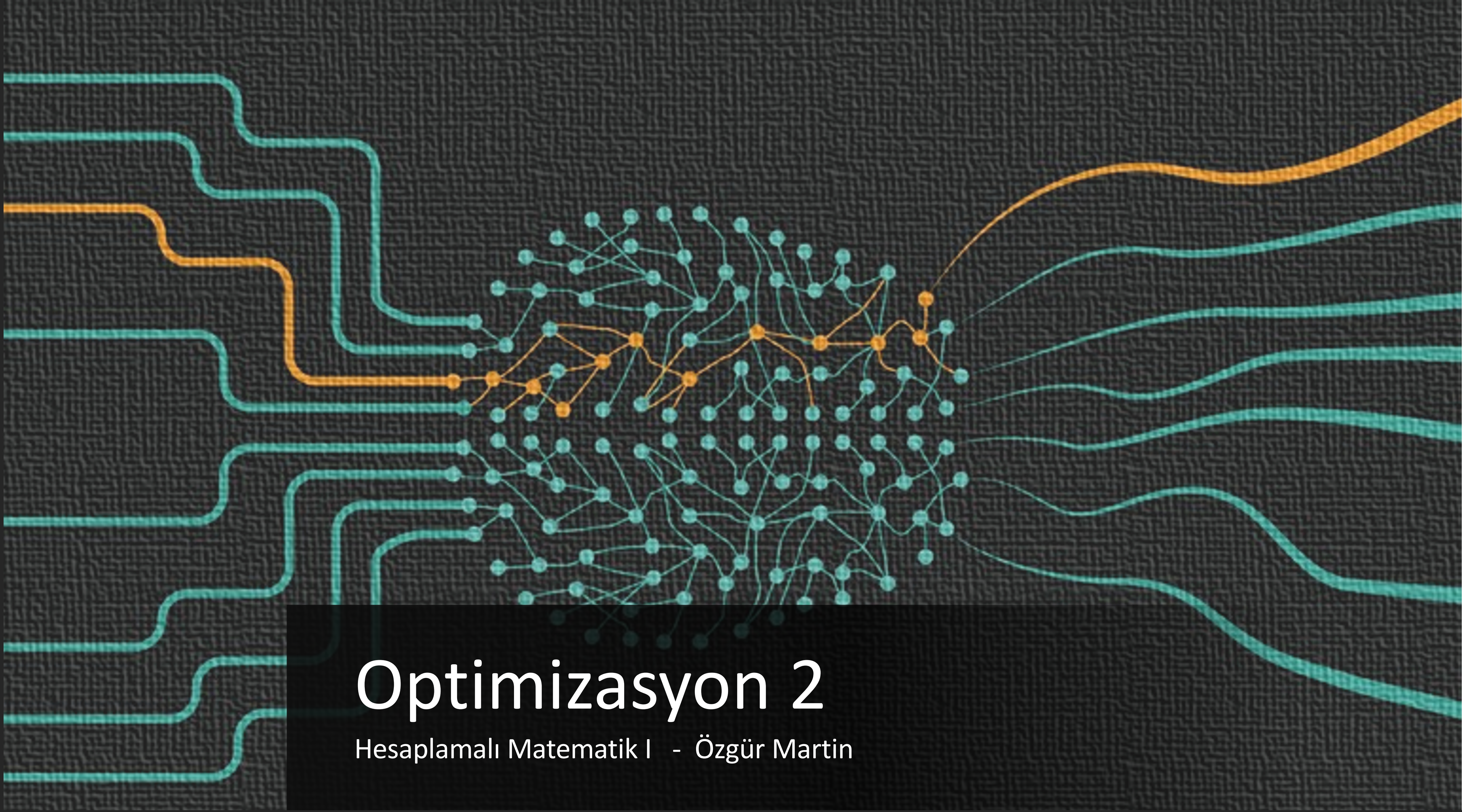


Optimizasyon 2

Hesaplamalı Matematik I - Özgür Martin

Kısıtsız Çok Değişkenli Optimizasyon

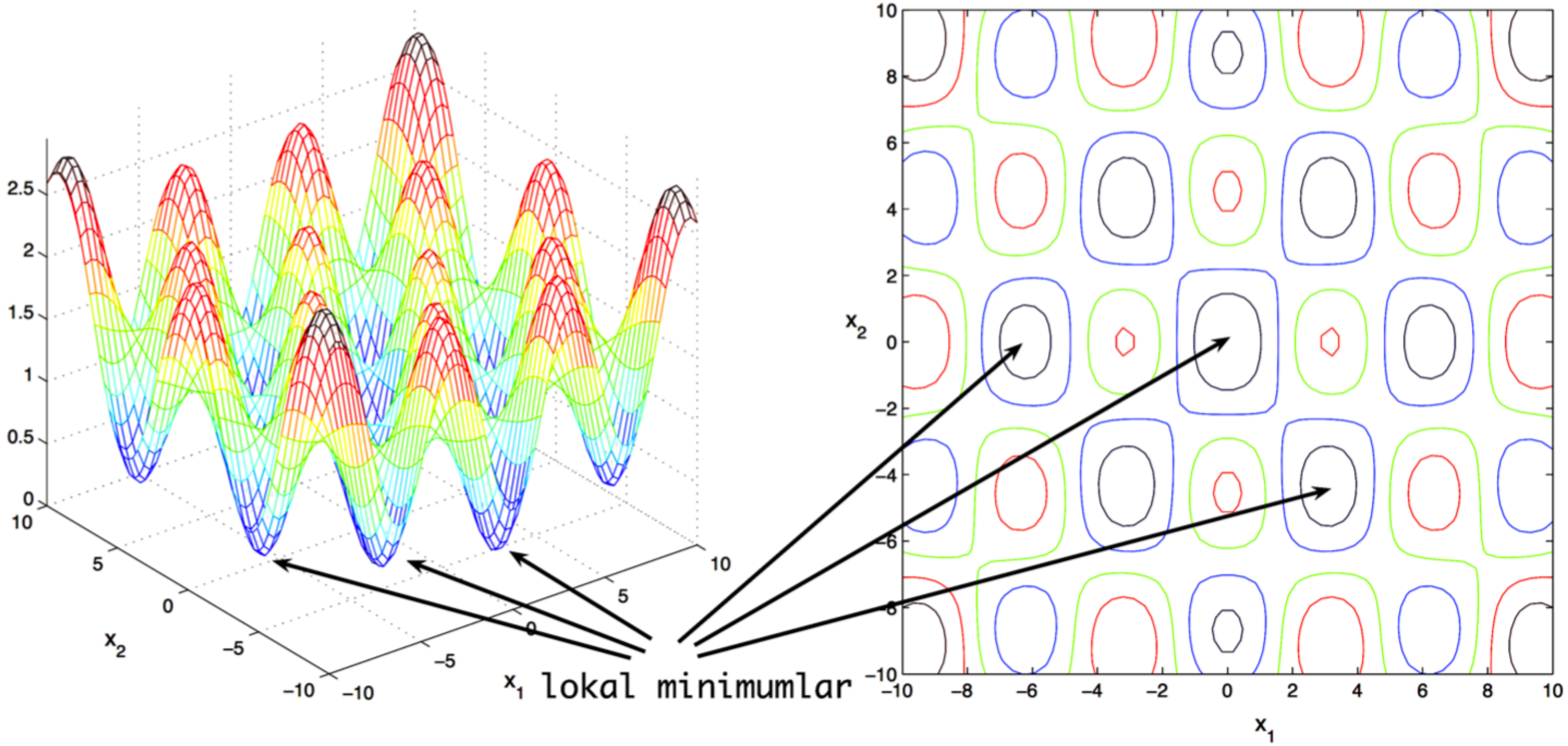
Dışbükey $f\colon \mathbb{R}^n \to \mathbb{R}$ fonksiyonunun birinci dereceden kısmi türevleri var ve sürekli olsun

Gradyan vektörü:

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

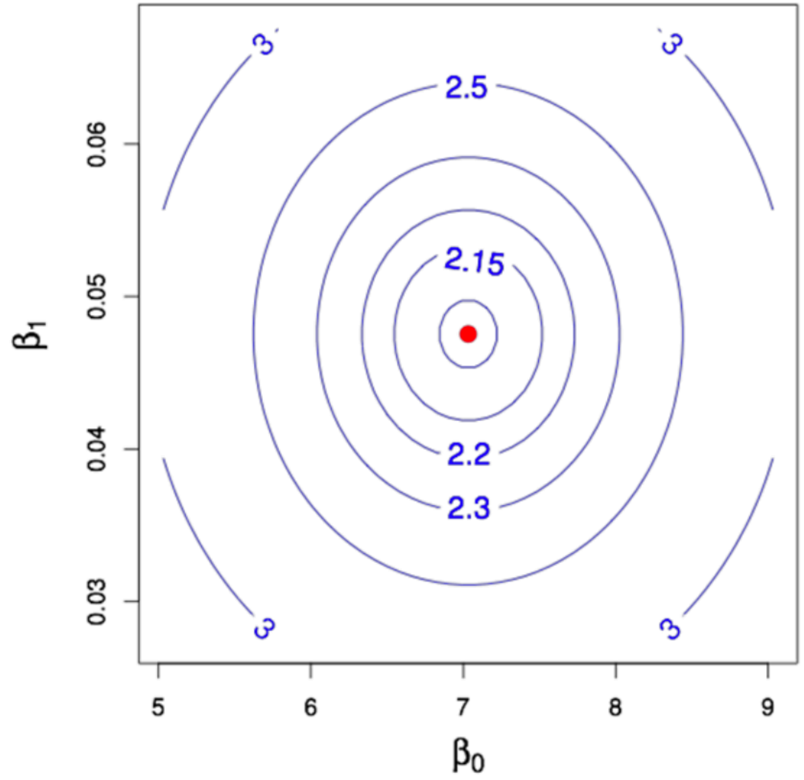
Hatırlatma: f fonksiyonun bir $x \in \mathbb{R}^n$ noktasında en hızlı azaldığı yön gradyanın ter yönüdür, yani $-\nabla f(x)$.

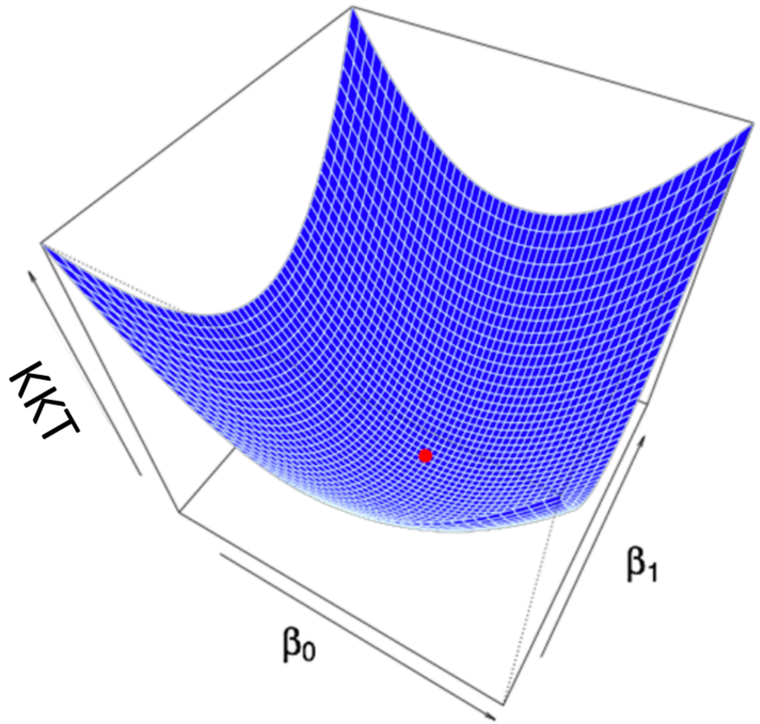
Amaç: Lokal minimumlardan bir tanesine ulaşmak. Bu noktalarda $\nabla f(x) = 0$.



Hatırlatma: Dışbükey fonksiyonlarda tek bir lokal minimum var. Bu aynı zamanda global minimum.

Dışbükey Fonksiyon



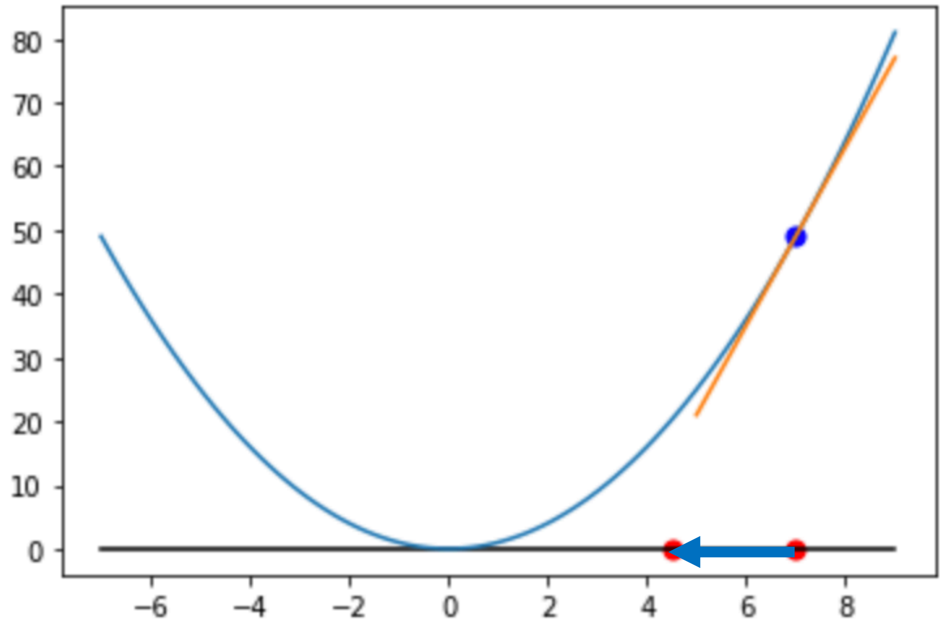


Gradyan İnişi Yöntemi (Tek Değişken)

$f\colon \mathbb{R} \to \mathbb{R}$ fonksiyonu için $f'(x) = 0$ olan x noktasını arıyoruz. Öğrenme hızı c'yi türevin mutlak değeri ile çarpıp kullanalım!

Gradyan İnişi Yöntemi

1. Herhangi bir x noktasından başla.
2. Bir ögrenme hızı c seç.
3. $x = x - c\, f'(x)$
4. $|f'(x)|$ yeterince küçük olana kadar devam et.

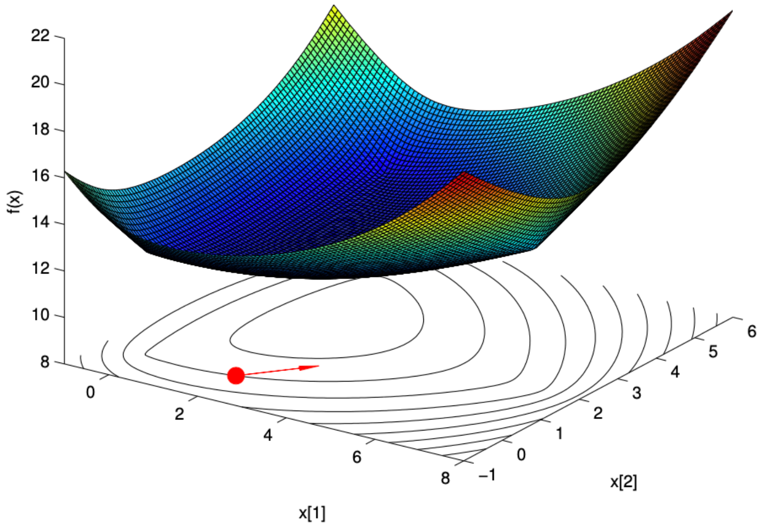


Gradyan İnişi Yöntemi (Çok Değişken)

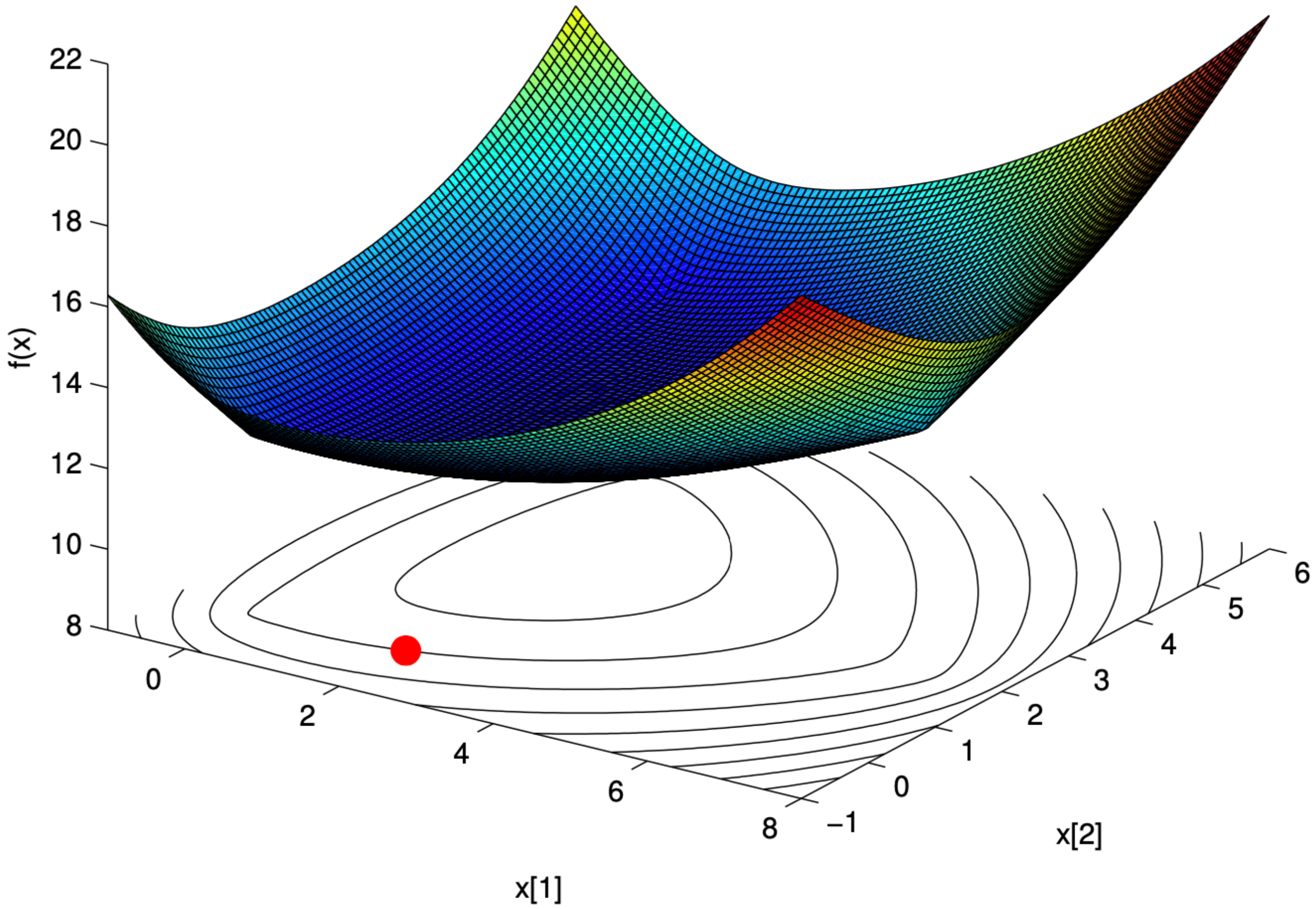
Dışbükey, bütün kısmi türevleri var ve sürekli olan bir $f\colon \mathbb{R}^n \to \mathbb{R}$ fonksiyonu için $\nabla f(x) = 0$ eşitliğini sağlayan $x \in \mathbb{R}^n$ noktasını arıyoruz.

Gradyan İnişi Yöntemi

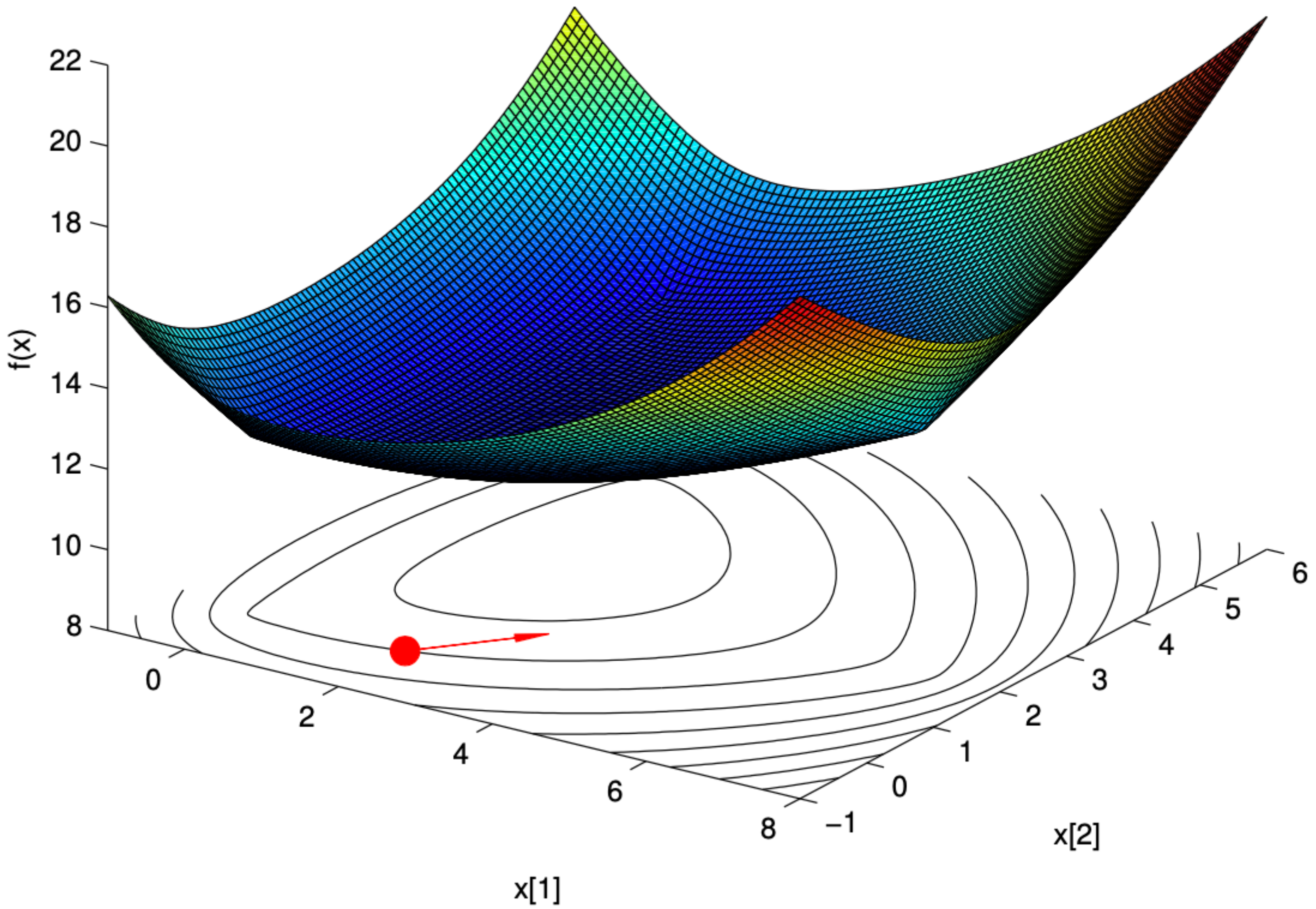
1. Herhangi bir $x \in \mathbb{R}^n$ noktasından başla.
2. Bir ögrenme hızı c seç.
3. $x = x - c\,\nabla f(x)$
4. $\| \nabla f(x) \|$ yeterince küçük olana kadar devam et.



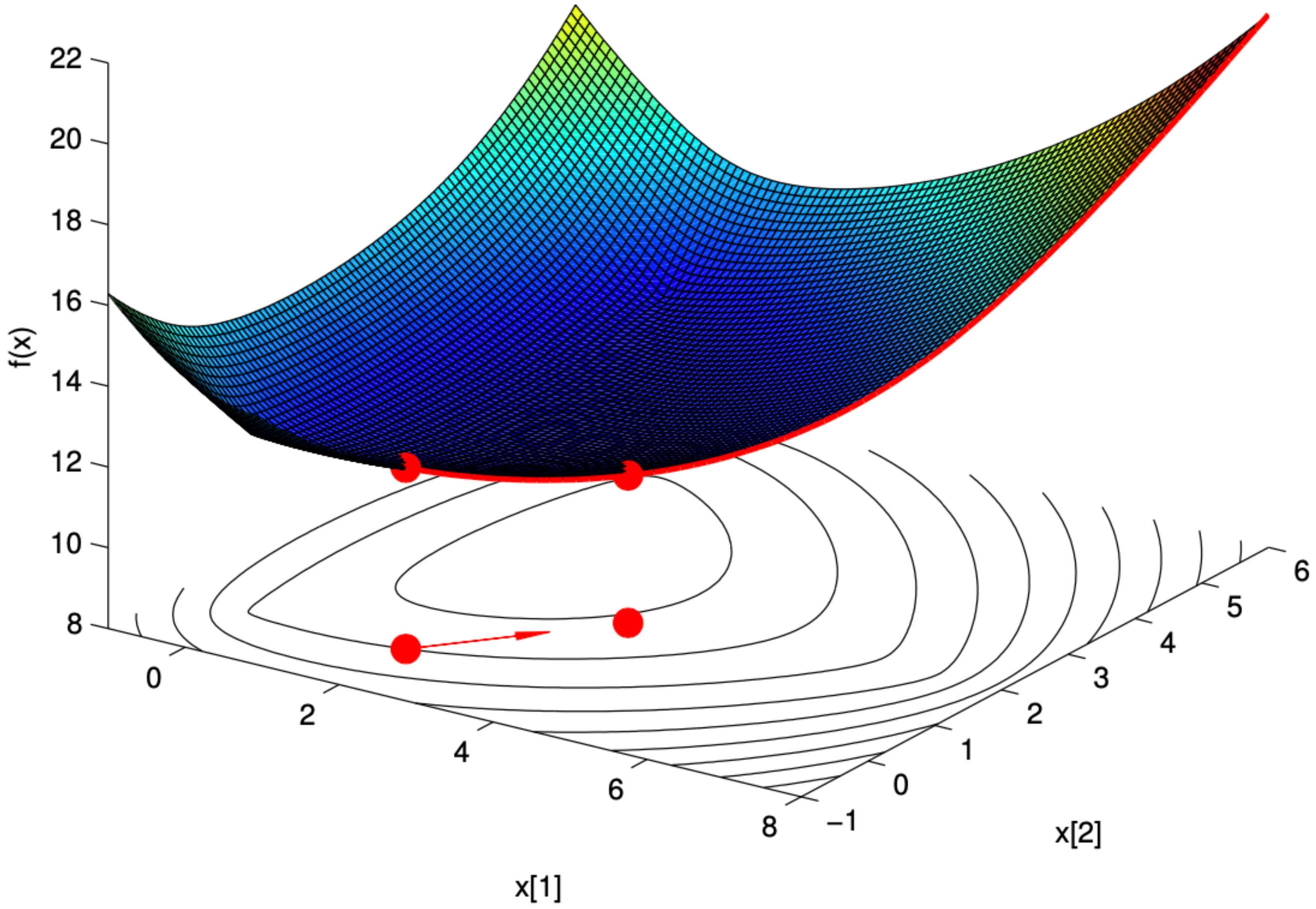
Gradyan İnişi Yöntemi (Çok Değişken)



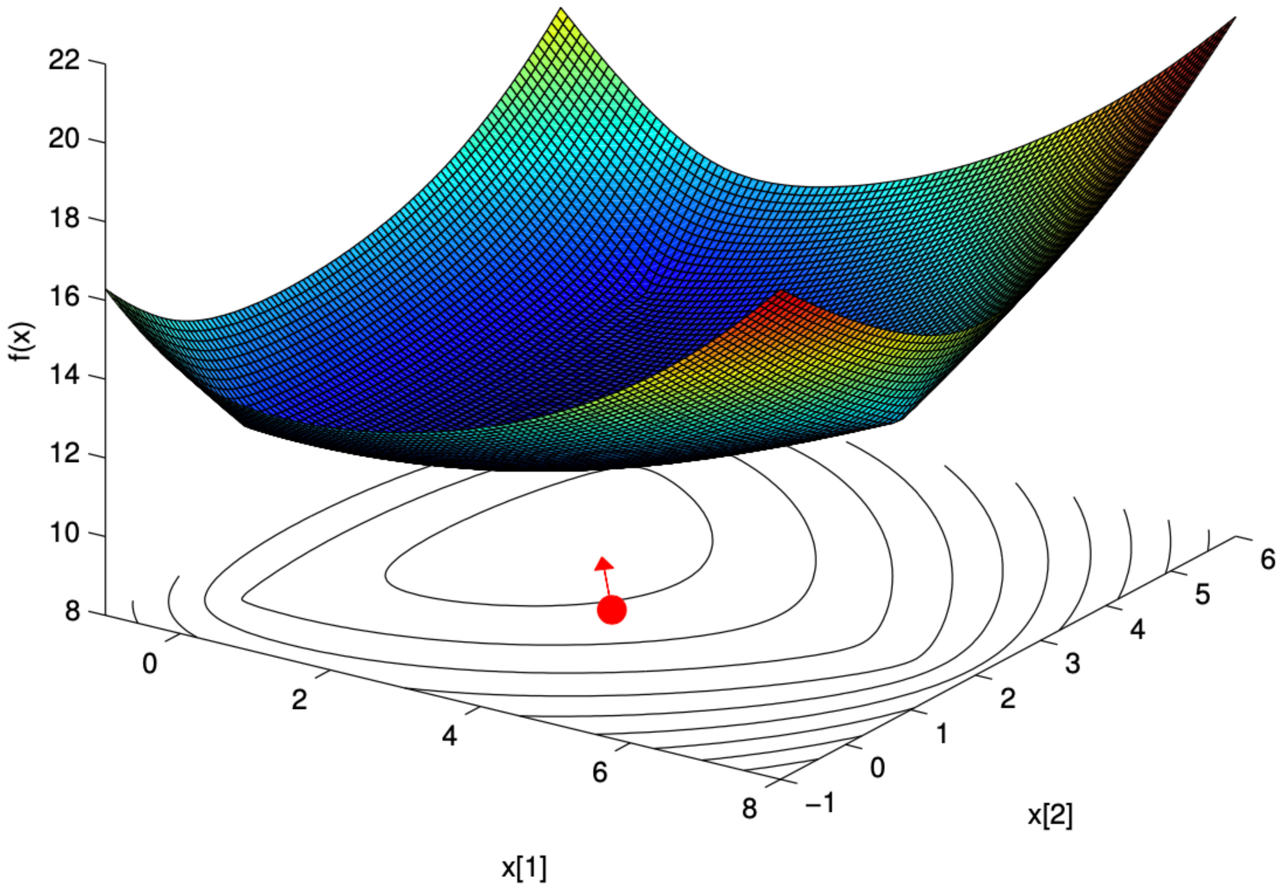
Gradyan İnişi Yöntemi (Çok Değişken)



Gradyan İnişi Yöntemi (Çok Değişken)



Gradyan İnişi Yöntemi (Çok Değişken)



Uygulama: Doğrusal Bağlanım

$$y_i \approx \underbrace{\hat{\beta}_0 + \hat{\beta}_1 x_i}_{\hat{y}_i}, \quad i = 1, \ldots, n$$

$$e_i = y_i - \hat{y}_i$$ *i*. kalıntı (*i*th residual)

Kalıntı Kareler Toplamı (Residual Sum of Squares)

$$\mathrm{KKT} = e_1^2 + e_2^2 + \cdots + e_n^2$$

(Amaç: Bu toplamı minimize eden $\hat{\beta}_0$ *ve* $\hat{\beta}_1$ *değerlerini bulalım)*

En Küçük Kareler (EKK) Yöntemi (Least Squares Method)

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \text{KKT} = \min_{\hat{\beta}_0, \hat{\beta}_1} (y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1)^2 + (y_2 - \hat{\beta}_0 - \hat{\beta}_1 x_2)^2 + \ldots + (y_n - \hat{\beta}_0 - \hat{\beta}_1 x_n)^2$$

Minimize edilmek istenen fonksiyon basit (doğrusal) olduğundan analitik bir çözüm bulabildik:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Burada $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ ve $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$

Şimdi analitik çözüm yerine **gradian iniş yöntemi ile** minimizasyon problemini çözelim?

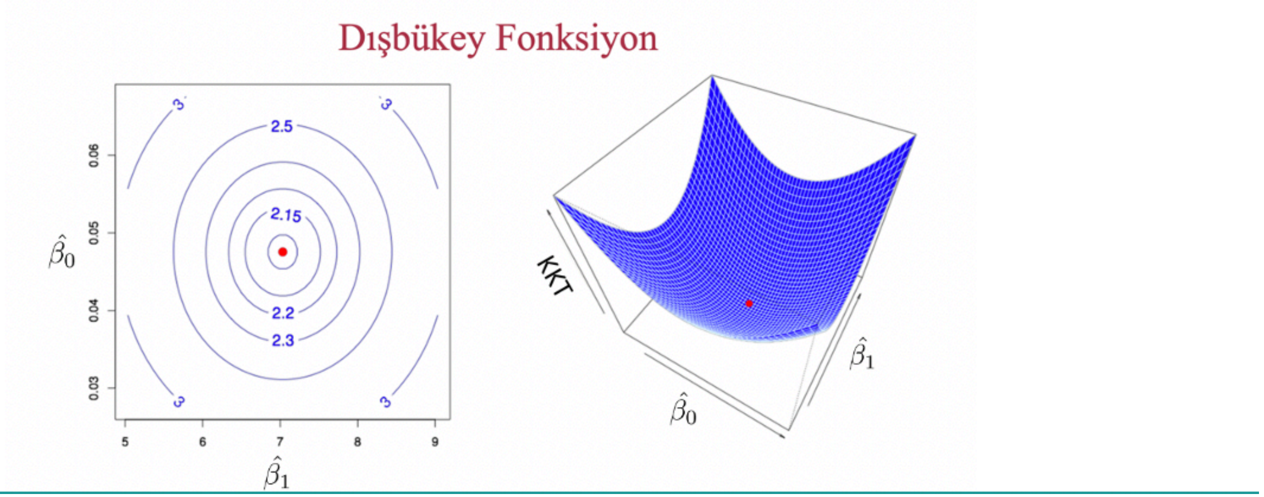
Gradyan İniş İle Çözüm

Problem:
$$\min_{(\hat{\beta}_0,\hat{\beta}_1)\in\mathbb{R}^2} f(\hat{\beta}_0,\hat{\beta}_1)$$

En Küçük Kareler (EKK) Yöntemi (Least Squares Method)

$$\min_{\hat{\beta}_0,\hat{\beta}_1} \text{KKT} = (y_1-\hat{\beta}_0-\hat{\beta}_1x_1)^2+(y_2-\hat{\beta}_0-\hat{\beta}_1x_2)^2+\ldots+(y_n-\hat{\beta}_0-\hat{\beta}_1x_n)^2$$

Çözüm:



Problem: $$\min_{(\hat\beta_0,\hat\beta_1)\in\mathbb{R}^2} f(\hat\beta_0,\hat\beta_1)$$

$$f(\hat\beta_0,\hat\beta_1) = \sum_{i=1}^{n}(y_i - \hat\beta_0 - \hat\beta_1 x_i)^2$$

$$\nabla f(\hat\beta_0,\hat\beta_1) = \begin{bmatrix} \dfrac{\partial f}{\partial\hat\beta_0} \\ \dfrac{\partial f}{\partial\hat\beta_1} \end{bmatrix} = \begin{bmatrix} \displaystyle\sum_{i=1}^{n} \frac{\partial}{\partial\hat\beta_0}(y_i - \hat\beta_0 - \hat\beta_1 x_i)^2 \\ \displaystyle\sum_{i=1}^{n} \frac{\partial}{\partial\hat\beta_1}(y_i - \hat\beta_0 - \hat\beta_1 x_i)^2 \end{bmatrix} = \begin{bmatrix} \displaystyle\sum_{i=1}^{n} -2(y_i - \hat\beta_0 - \hat\beta_1 x_i) \\ \displaystyle\sum_{i=1}^{n} -2x_i(y_i - \hat\beta_0 - \hat\beta_1 x_i) \end{bmatrix}$$

$$= \begin{bmatrix} \displaystyle -2\sum_{i=1}^{n} y_i + 2n\hat\beta_0 + 2\hat\beta_1\sum_{i=1}^{n} x_i \\ \displaystyle -2\sum_{i=1}^{n} x_i y_i + 2\hat\beta_0\sum_{i=1}^{n} x_i + 2\hat\beta_1\sum_{i=1}^{n} {x_i}^2 \end{bmatrix}$$