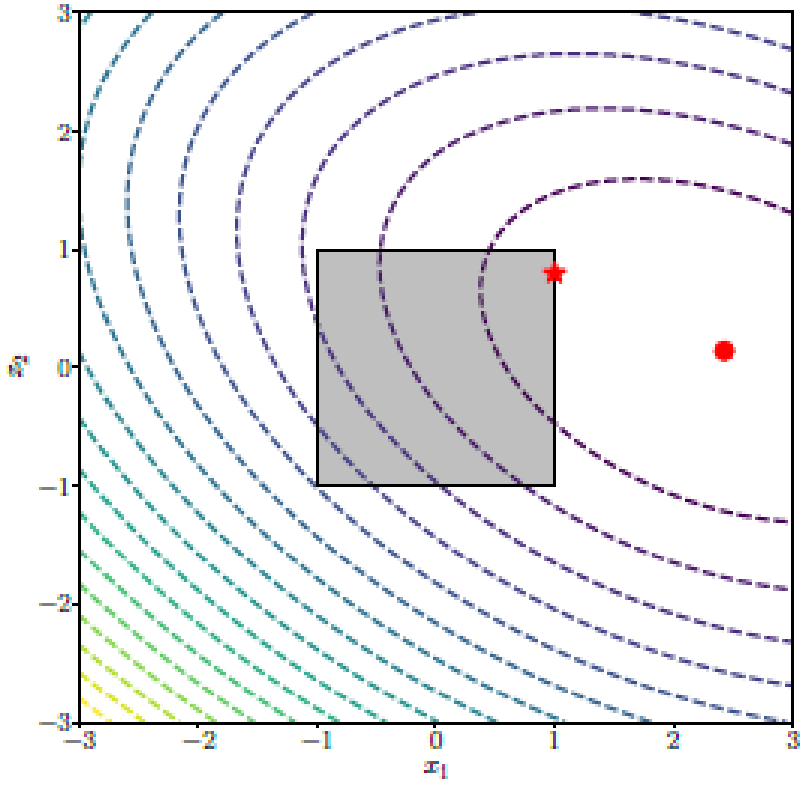
SÜREKLİ ENİYİLEME

(CHAPTER 7)

Matematiksel Eniyileme Problemi

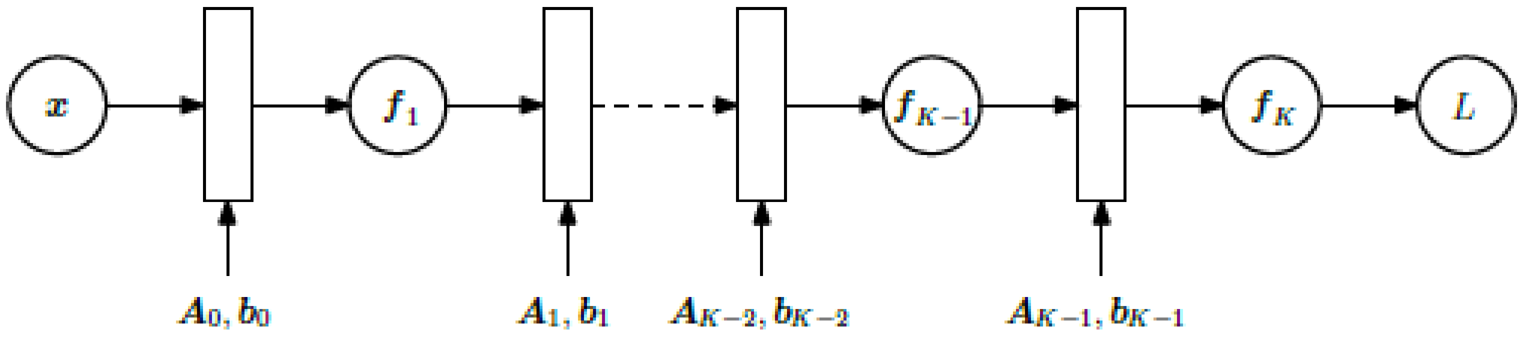


$$
\begin{aligned}
&\min_{\boldsymbol{x}} && f(\boldsymbol{x}) \\
&\text{subject to} && g_i(\boldsymbol{x}) \leqslant 0 \quad \text{for all} \quad i = 1, \ldots, m\,.
\end{aligned}
$$

- Terminoloji
- Bu dersin kapsamı

Yapay Öğrenme için Eniyileme

- Yapay sinir ağı



$$x_{k+1} = f_k(x_k)$$
$$f_k(x_k) = \sigma_k(A_k x_k + b_k)$$

$$y = (f_K \circ f_{K-1} \circ \cdots \circ f_1)(x) = f_K(f_{K-1}(\cdots(f_1(x))\cdots))$$

- Kayıp fonksiyonu $L(\theta)$, $\theta = \{A_0, b_0, \ldots, A_{K-1}, b_{K-1}\}$

Yapay Öğrenme için Eniyileme

- Boyut İndirgeme (Chapter 10)

$$\max_{b_1} b_1^\top S b_1$$
$$\text{subject to } \|b_1\|^2 = 1\,.$$

- Sınıflandırma (Chapter 12)

$$\min_{w,b} \quad \underbrace{\frac{1}{2}\|w\|^2}_{\textit{margin}}$$
$$\text{subject to} \quad \underbrace{y_n(\langle w, x_n\rangle + b) \geqslant 1}_{\textit{data fitting}}\,.$$

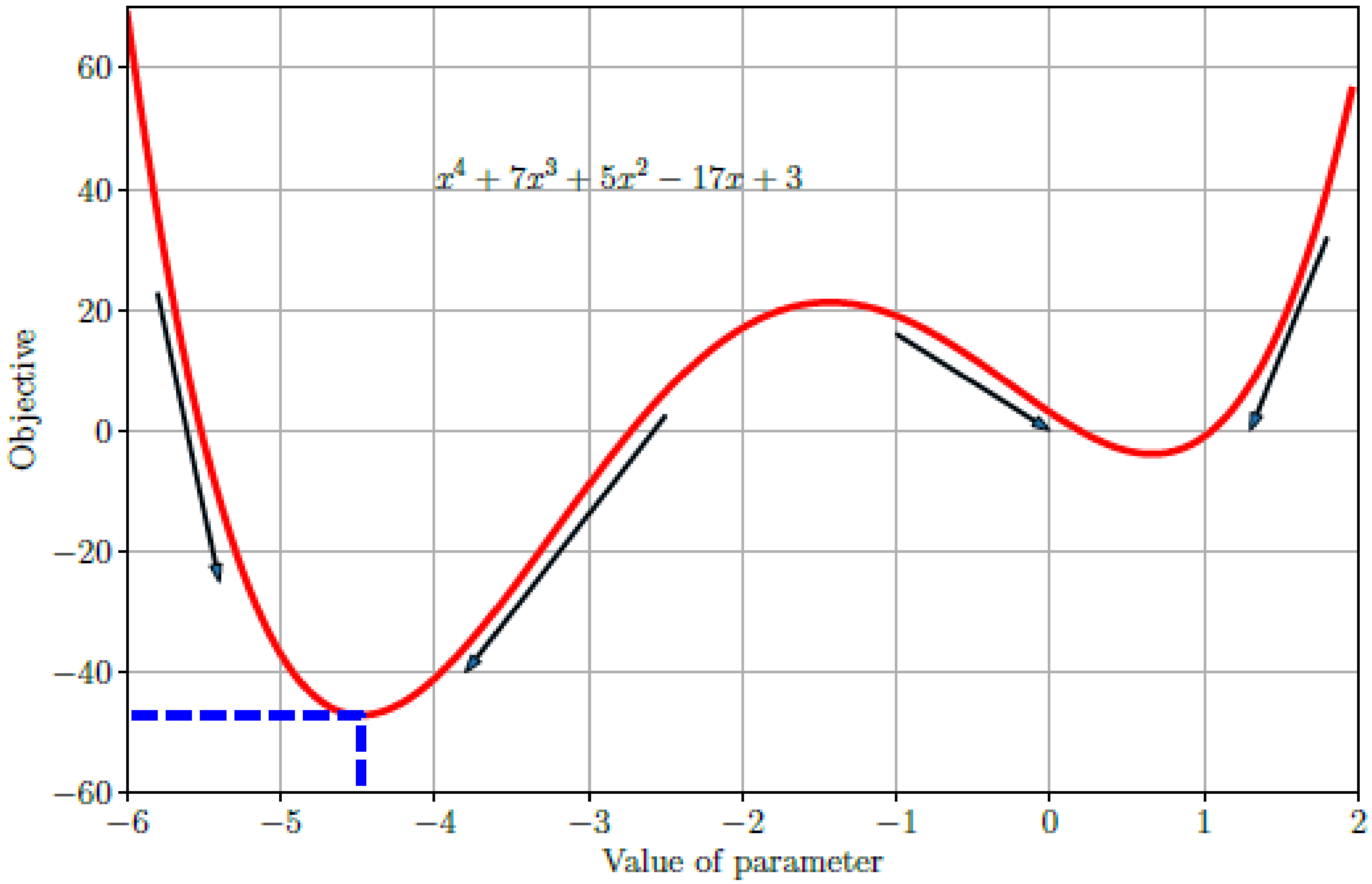
Kritik Konular

Yöntem geliştirirken ya da seçerken

- Sürekli mi?
- Türevlenebilir mi?
- Dışbükey mi?
- Deterministik mi?
- Problemin özel bir yapısı var mı?
- Problem boyutu

$$\begin{array}{ll} \min_{x} & f(x) \\ \text{subject to} & g_i(x) \leqslant 0 \quad \text{for all} \quad i = 1, \ldots, m\,. \end{array}$$

Eniyileme Probleminin Çözümü



- Gerek koşul
- Yeter koşul

- Yerel çözüm
- Global çözüm

Eniyileme Probleminin Çözümü

$$\begin{array}{lll} \min\limits_{x} & f(x) & \\ \text{subject to} & g_i(x) \leqslant 0 \quad \text{for all} & i = 1, \ldots, m\,. \end{array}$$

- Bir çözüm bulmak her zaman mümkün mü?
- Teori: Amaç fonksiyonu alttan sınırlı, olurlu çözümler kümesi boş değil ve kapalıysa

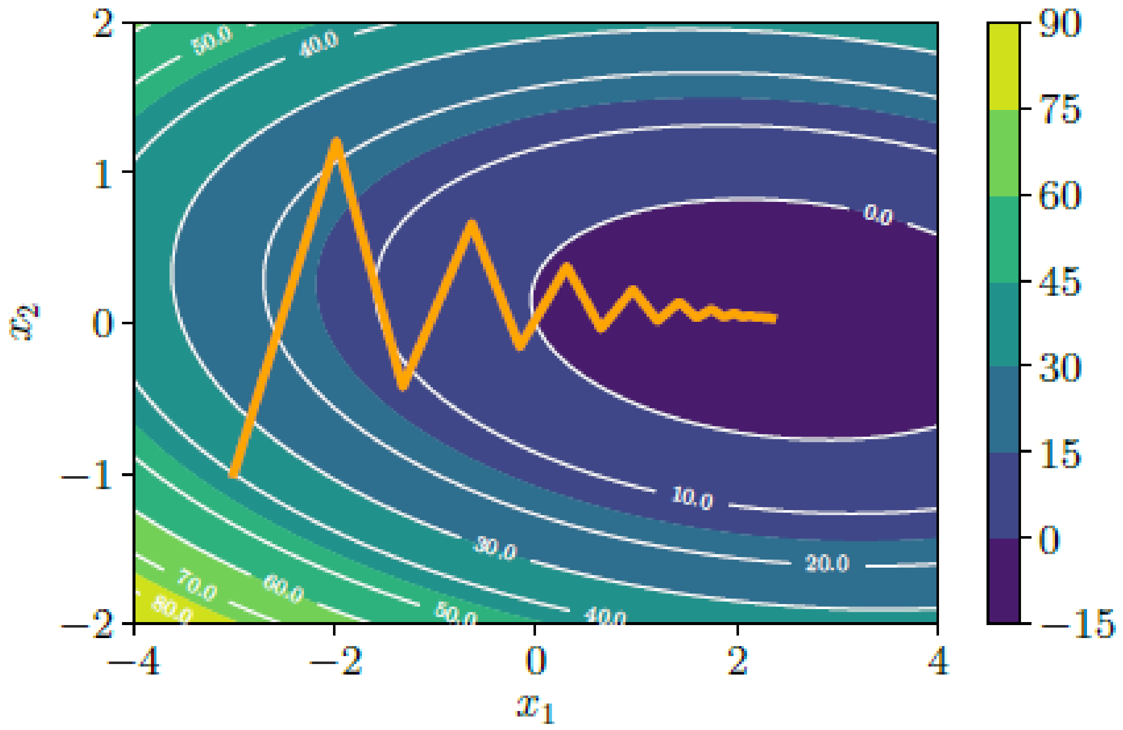
- Pratikte çoğu zaman yaklaşık çözümler hesaplarız

Gradyan Algoritması

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i - \gamma_i((\nabla f)(\boldsymbol{x}_i))^\top$$

- Analitik çözüm nadiren mümkün
- En hızlı düşüş
- Adım boyu

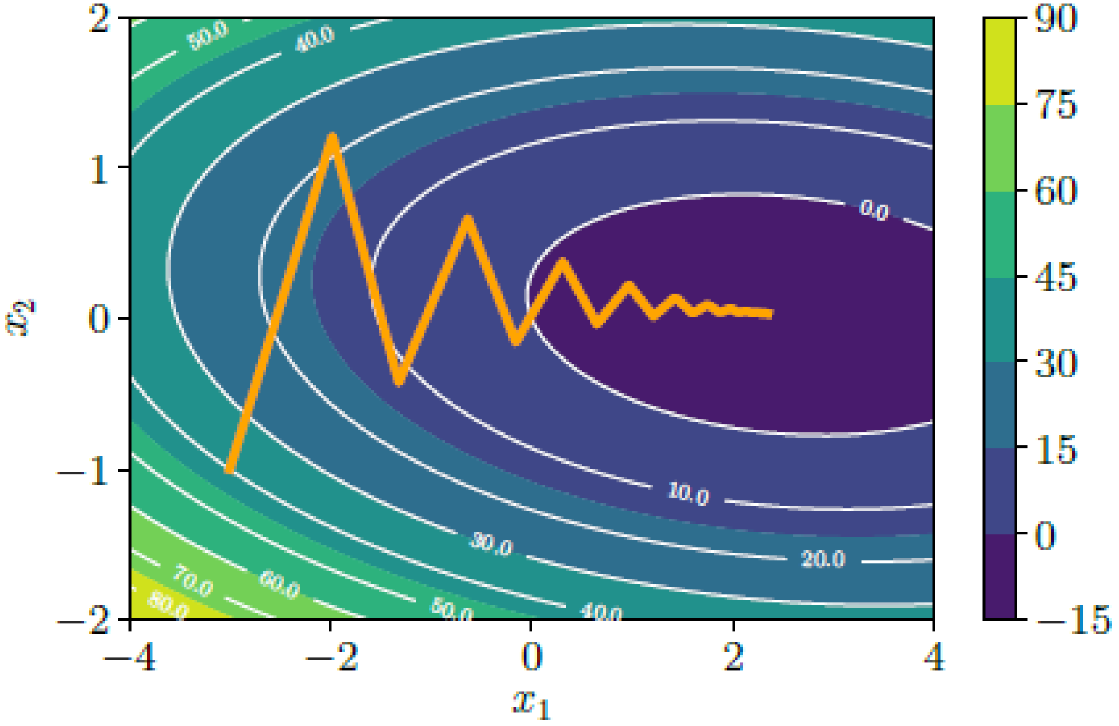


Gradyan Algoritması

$$f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^\top \begin{bmatrix} 2 & 1 \\ 1 & 20 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 5 \\ 3 \end{bmatrix}^\top \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\nabla f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^\top \begin{bmatrix} 2 & 1 \\ 1 & 20 \end{bmatrix} - \begin{bmatrix} 5 \\ 3 \end{bmatrix}^\top$$

$$x_0 = [-3, -1]^\top$$



- Yakınsama hızı
- Hesse matrisinin sağlamlık sayısı

Gradyan Algoritması

- Momentum

$$x_{i+1} = x_i - \gamma_i((\nabla f)(x_i))^\top + \alpha \Delta x_i \tag{7.11}$$

$$\Delta x_i = x_i - x_{i-1} = \alpha \Delta x_{i-1} - \gamma_{i-1}((\nabla f)(x_{i-1}))^\top , \tag{7.12}$$

- Hızlandırma
- Newtonumsu (Değişken-metrik)

Stokastik Gradyan Algoritması

- Sistematik hata içermeyen yaklaşık gradyan

- Ampirik risk (SAA)

$$L(\theta) = \sum_{n=1}^{N} L_n(\theta) \,,$$

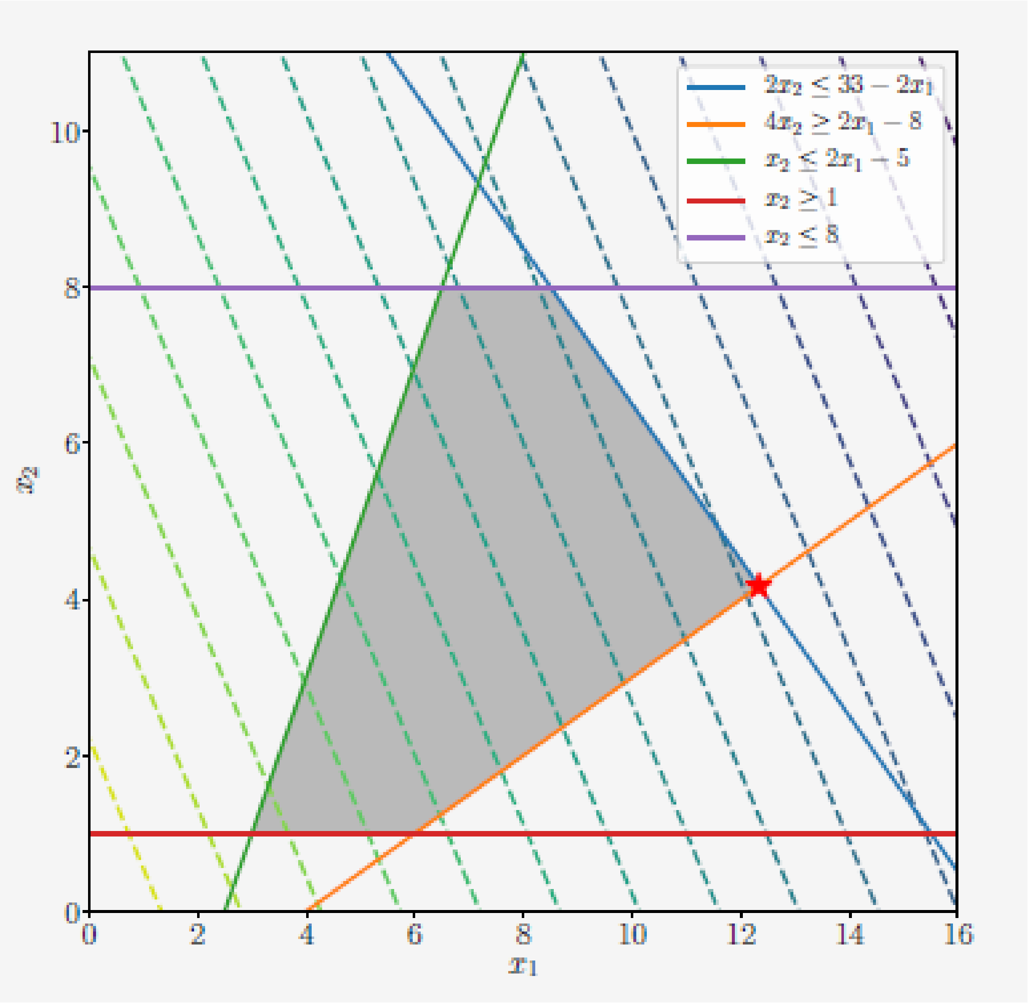
  - En büyük olabilirlik kestirimi

$$L(\theta) = -\sum_{n=1}^{N} \log p(y_n | x_n, \theta) \,,$$

- Doğrudan beklenen kayıp fonksiyonuna uygulama (SA)

- Varyans düşürme, ölçekleme, momentum, ...

Kısıtlı Eniyileme Problemi



$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leqslant 0 \quad \text{for all} \quad i = 1, \ldots, m\,. \end{aligned}$$

- Aktif kısıt
- Olurlu yönler konisi
- Kısıtlı durumda gerek koşul

Lagrange Çarpanları

$$J(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \mathbf{1}(g_i(\boldsymbol{x})), \quad \mathbf{1}(z) = \begin{cases} 0 & \text{if } z \leqslant 0 \\ \infty & \text{otherwise} \end{cases}.$$

$$\begin{aligned} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda}) &= f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i g_i(\boldsymbol{x}) \\ &= f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{x}), \end{aligned}$$

$$J(\boldsymbol{x}) = \max_{\boldsymbol{\lambda} \geqslant \mathbf{0}} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda}).$$

- Amaç fonksiyonunun kısıtlara duyarlılığı
- Aktif olmayan kısıtlara karşılık gelen çarpanlar
- Çarpanların sonlu değer alması

Lagrange Eşiz

$$J(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \mathbf{1}(g_i(\boldsymbol{x}))\,, \quad \mathbf{1}(z) = \begin{cases} 0 & \text{if } z \leqslant 0 \\ \infty & \text{otherwise} \end{cases}.$$

$$\mathfrak{D}(\boldsymbol{\lambda}) = \min_{\boldsymbol{x}\in\mathbb{R}^d} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda}).$$

$$J(\boldsymbol{x}) = \max_{\boldsymbol{\lambda}\geqslant\mathbf{0}} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})\,.$$

$$\begin{aligned} \max_{\boldsymbol{\lambda}\in\mathbb{R}^m} \quad & \mathfrak{D}(\boldsymbol{\lambda}) \\ \text{subject to} \quad & \boldsymbol{\lambda} \geqslant \mathbf{0}\,, \end{aligned}$$

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \max_{\boldsymbol{\lambda}\geqslant\mathbf{0}} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \geqslant \max_{\boldsymbol{\lambda}\geqslant\mathbf{0}} \min_{\boldsymbol{x}\in\mathbb{R}^d} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})\,.$$

- Optimallik açıklığı
- Güçlü düalite özelliği

Dışbükey Eniyileme

- Dışbükey küme

Figure 7.5 Example of a convex set.

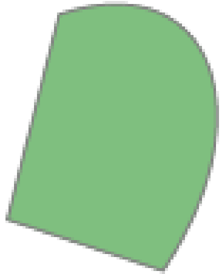
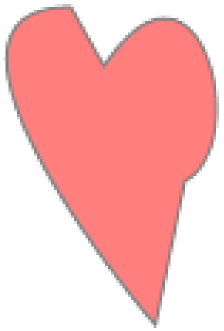


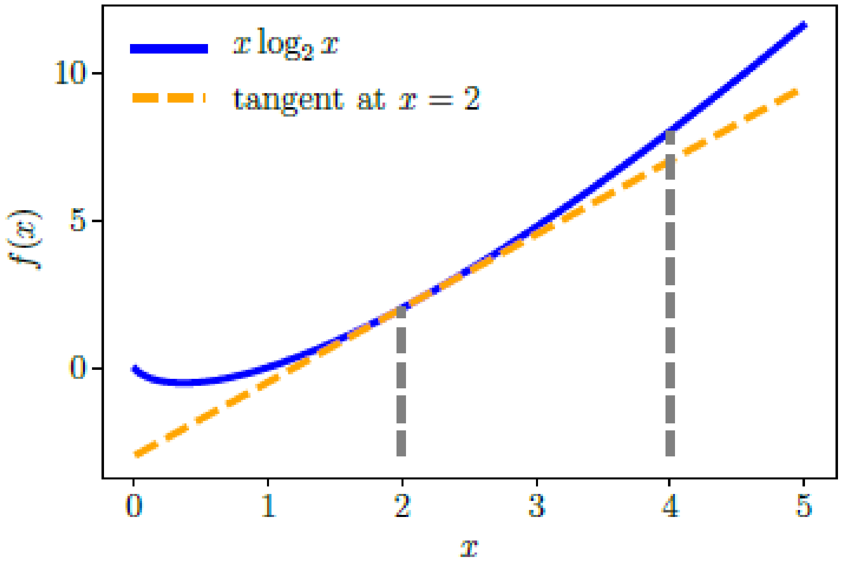
Figure 7.6 Example of a nonconvex set.



- Dışbükey fonksiyon

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leqslant \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}) . \quad 0 \leqslant \theta \leqslant 1$$

$$f(\boldsymbol{y}) \geqslant f(\boldsymbol{x}) + \nabla_{\boldsymbol{x}} f(\boldsymbol{x})^{\top} (\boldsymbol{y} - \boldsymbol{x}) .$$



Doğrusal Program

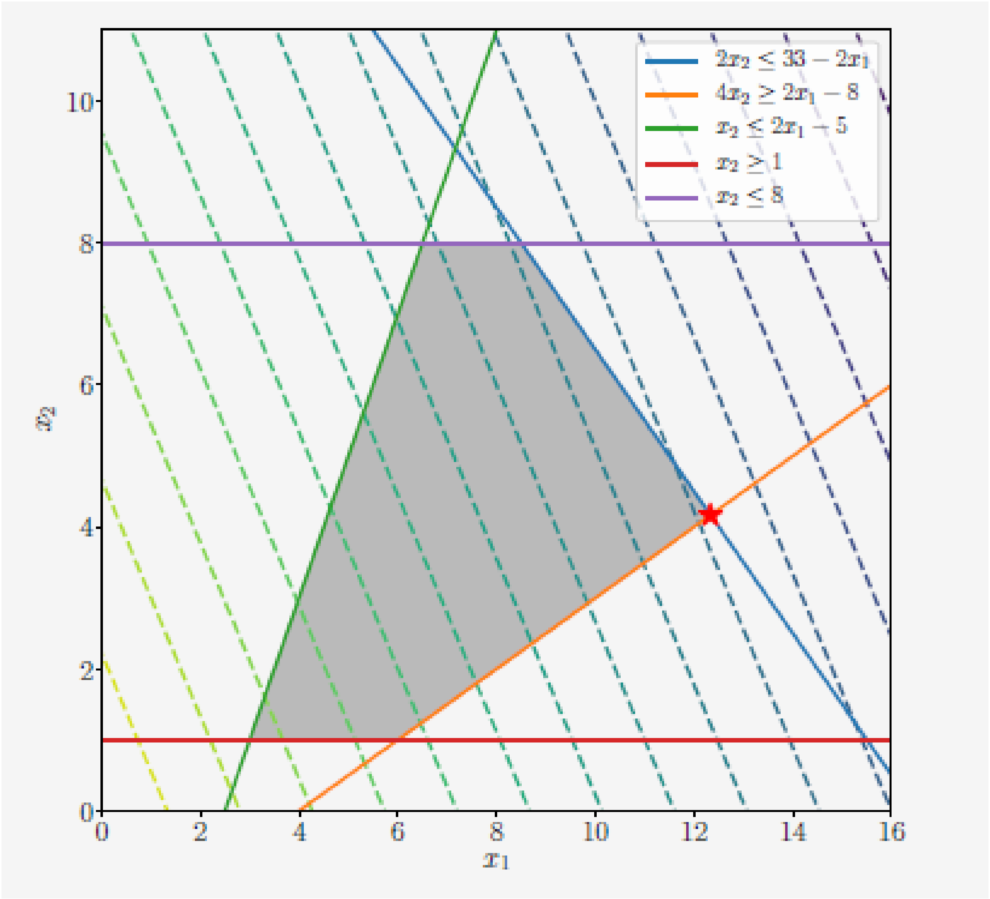
$$\min_{\boldsymbol{x} \in \mathbb{R}^d} \quad \boldsymbol{c}^\top \boldsymbol{x}$$

$$\text{subject to} \quad \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b}\,,$$

$$\max_{\boldsymbol{\lambda} \in \mathbb{R}^m} \quad -\boldsymbol{b}^\top \boldsymbol{\lambda}$$

$$\text{subject to} \quad \boldsymbol{c} + \boldsymbol{A}^\top \boldsymbol{\lambda} = \boldsymbol{0}$$

$$\boldsymbol{\lambda} \geqslant \boldsymbol{0}\,.$$

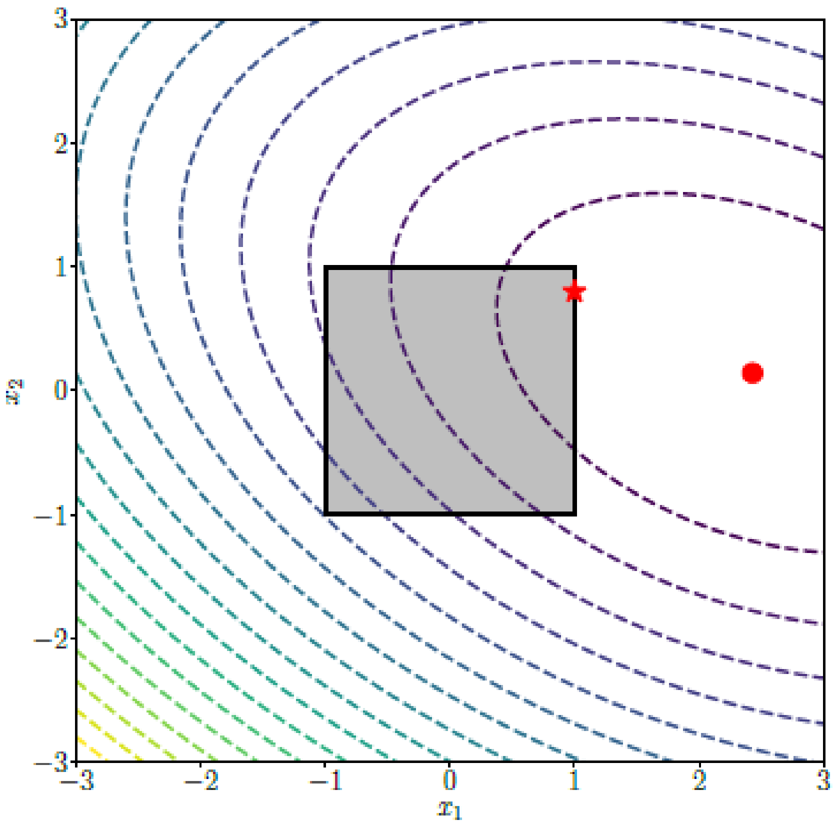


Karesel Program

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \quad \frac{1}{2}\boldsymbol{x}^\top Q\boldsymbol{x} + \boldsymbol{c}^\top\boldsymbol{x}$$
$$\text{subject to} \quad \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b}\,,$$

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^m} \quad -\frac{1}{2}(\boldsymbol{c}+\boldsymbol{A}^\top\boldsymbol{\lambda})^\top Q^{-1}(\boldsymbol{c}+\boldsymbol{A}^\top\boldsymbol{\lambda}) - \boldsymbol{\lambda}^\top\boldsymbol{b}$$
$$\text{subject to} \quad \boldsymbol{\lambda} \geqslant \boldsymbol{0}\,.$$

- SVM



Dışbükey Eşlenik

$$f^*(s) = \sup_{x \in \mathbb{R}^D} \left(\langle s, x \rangle - f(x) \right) .$$

- Dual problem türetimi

$$\min_{x} f(Ax) + g(x) = \max_{u} -f^*(u) - g^*(-A^\top u) .$$

- Örnek

$$f(y) = \frac{\lambda}{2} y^\top K^{-1} y$$