(NLP): programare neliniara constransa

Problema de optimizare neliniara supusa la constrangeri:

$$
\begin{aligned}
(NLP): \quad & \min_{x \in \mathbb{R}^n} f(x) \\
& \text{s.l.}: \quad g_1(x) \leq 0, \ldots, g_m(x) \leq 0 \\
& \qquad\quad h_1(x) = 0, \ldots, h_p(x) = 0
\end{aligned}
$$

unde functiile f, g_i si h_j are continue si diferentiabile

$$
\begin{aligned}
& g(x) = [g_1(x), \ldots, g_m(x)]^T, && h(x) = [h_1(x), \ldots, h_p(x)]^T, \\
& g: \mathbb{R}^n \to \mathbb{R}^m, && h: \mathbb{R}^n \to \mathbb{R}^p.
\end{aligned}
$$

$$
\begin{aligned}
(NLP): \quad & \min_{x \in \mathbb{R}^n} \quad f(x) \\
& \text{s.l. } g(x) \leq 0, \quad h(x) = 0.
\end{aligned}
$$

$$
(NLP): \qquad f^* = \min_{x \in X} f(x)
$$

- functia obiectiv f
- multimea fezabila $X = \{x : g_i(x) \leq 0, \ h_j(x) = 0 \ \forall i, j\}$

(CP): programare convexa constransa

Problema de optimizare convexa supusa la constrangeri:

$$(CP): \quad \min_{x \in \mathbb{R}^n} f(x)$$
$$\text{s.l.}: \quad g_1(x) \leq 0, \ldots, g_m(x) \leq 0$$
$$a_1^T x = b_1, \ldots, a_p^T x = b_p$$

unde functiile f si g_i are continue si convexe (e.g. Hesianele sunt pozitiv semidefinite), constrangerile de egalitate sunt liniare.

$$g(x) = [g_1(x), \ldots, g_m(x)]^T, \qquad h(x) = Ax - b,$$
$$g: \mathbb{R}^n \to \mathbb{R}^m, \qquad A = [a_1 \cdots a_p]^T \in \mathbb{R}^{p \times n}.$$

$$(CP): \quad \min_{x \in \mathbb{R}^n} \quad f(x)$$
$$\text{s.l. } g(x) \leq 0, \quad Ax = b.$$

$$(CP): \qquad f^* = \min_{x \in X} f(x)$$

- functia obiectiv f convexa
- multimea fezabila convexa $X = \{x : g_i(x) \leq 0 \ \forall i, \ Ax = b\}$

Exemplu 1: proiectia Euclidiana

$$\min_{x \in X} \|x - x^0\|_2^2 \quad \text{sau} \quad \min_{x \in \cap_{i=1}^m X_i} \|x - x^0\|_2^2$$

Proiectia Euclidiana a punctului x^0 pe multimea X: $[x^0]_X$
determinarea celui mai "apropiat" punct (in distanta Euclidiana) din multimea X de punctul x^0

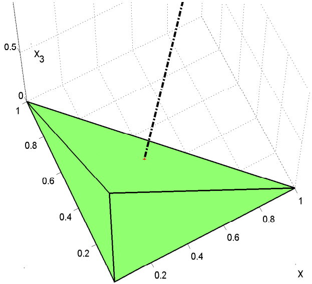
- functie obiectiv patratica convexa (Hesiana este I_n)
- proiectia exista si este unica in cazul X multime convexa (caz particular de (CP))
- aplicatii practice: imagistica medicala, tomografie computerizata, microscopie electronica
- solutie analitica pentru multimi "simple" (semispatiu, hyperplan, bila, box);

Exemplu: $\min_{x \in \mathbb{R}^n_+} \|x - x^0\|_2^2$, i.e. $X = \mathbb{R}^n_+ : \ x \geq 0$

$$([x^0]_{\mathbb{R}^n_+})_i = 0 \ \text{ daca } \ x_i^0 < 0 \quad \bigvee \quad = x_i^0 \ \text{ daca } \ x_i^0 \geq 0 \quad \forall i$$

Proiectia Euclidiana - exemple

$$\min_{x\in\mathbb{R}^3} \quad \|x - [1\ 1\ 1]^T\|_2^2$$
$$\text{s.l. } x_1 + x_2 + x_3 \le 1, x \ge 0.$$

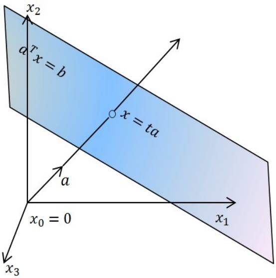


$$\min_{x\in\mathbb{R}^3} \quad \|x\|_2^2$$
$$\text{s.l. } a_1x_1 + a_2x_2 + a_3x_3 = b$$

$$\Downarrow$$

$$\mathcal{H} = \{x : \ a^T x = b\}, \quad [0]_H = ta$$

$$t = \frac{b}{\|a\|^2} \Rightarrow [0]_{\mathcal{H}} = \frac{b}{\|a\|^2} a$$

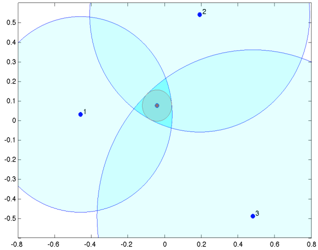


Exemplu 2: problema localizarii

Se cunosc locatiile a m senzori $s_i \in \mathbb{R}^3$ si distantele R_i de la senzori la obiectul necunoscut $x \in \mathbb{R}^3$

$$\min_{x \in \mathbb{R}^3, R > 0} \quad R$$
$$\text{s.l. } R + \|s_i - x\| \leq R_i \quad \forall i = 1, \ldots, m.$$

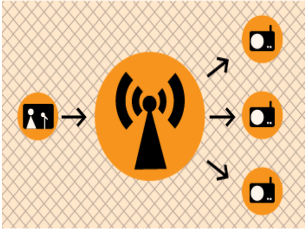
- se determina coordonatele: centrul $x \in \mathbb{R}^3$ si raza $R > 0$
- functie obiectiv liniara
- constrangeri patratice convexe
- caz particular de (CP)

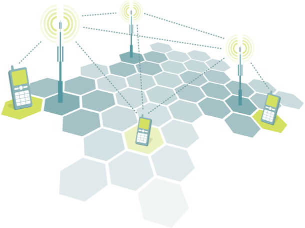


Exemplu 3: controlul puterii in sisteme de comunicatie

Dispunem de n transmitatori, cu nivelurile de putere $P_1, P_2, \ldots, P_n$, si de n receptori.

- fiecare receptor i receptioneaza semnalul transmitatorului i
- puterea receptata de receptorul i de la transmitatorul j (puterea semnalului de interferenta) este data de $G_{ij}P_j$, unde G_{ij} reprezinta amplificarea pe linia de comunicatie (i, j)
- *Puterea semnalului* primit de receptorul i de la transmitatorul i este $G_{ii}P_i$





Exemplu 3: controlul puterii in sisteme de comunicatie

- Zgomotul din retea este dat de σ_i.
- Raportul semnal - interferenta si zgomot este definit de

$$S_i = \frac{G_{ii} P_i}{\sigma_i + \sum_{j \neq i} G_{ij} P_j} = \frac{\text{semnal util}}{\text{zgomot} + \text{interferenta}}.$$

Constrangeri ce apar natural in medii de retea:
$P_{\min} \leq P_i \leq P_{\max}$ si $S_i \geq S_{\min}$.

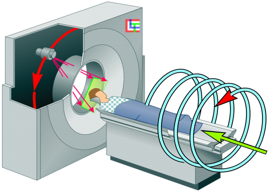
Problema de programare convexa

$$\begin{aligned} \min_{P \in \mathbb{R}^n} \ & P_1 + \cdots + P_n \\ \text{s.l.} \ & P_{\min} \leq P_i \leq P_{\max} \\ & G_{ii} P_i / (\sigma_i + \textstyle\sum_{j \neq i} G_{ij} P_j) \geq S_{\min} \end{aligned}$$

Exemplu 4: reconstructie tomografica

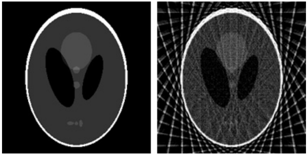
Tomografie computerizata = tehnica noninvaziva ce foloseste raze X (sau alte tipuri de radiatii) pentru a produce imagini 2D/3D ale interiorului obiectului scanat.
Procedura de functionare consta in:

1. Se achizitioneaza o serie de proiectii, din diferite unghiuri, ale obiectului scanat;



www.merckmanuals.com

2. Prin intermediul proiectiilor obtinute, se reconstruieste interiorul obiectului cu ajutorul unui algoritm iterativ;



www.mathworks.com

In majoritatea cazurilor, radiatiile folosite sunt daunatoare; de aceea se urmareste achizitionarea unui numar minim de proiectii.

Exemplu 4: reconstructie tomografica

Formularea problemei:

• Fie $x \in \mathbb{R}^n$ imaginea interiorului de reconstruit.
• Pentru reconstructie, dispunem de diferite masuratori liniare (proiectii) ale imaginii x: $\quad b_i = A_i x, \quad i = 1, \cdots, m$.
• Notam vectorul proiectiilor $b \in \mathbb{R}^m$ si $A = [A_1^T \cdots A_m^T]^T \in \mathbb{R}^{m \times n}$ matricea de achizitie.
• Imaginea interiorului reprezinta solutia sistemului liniar (subdeterminat deoarece sunt mai putine masuratori m decat dimensiunea imaginii n): $Ax = b$.
• Reformulare in termeni de problema CMMP:

$$\min_{x \in \mathbb{R}^n:\ Ax=b} \|x\|_\alpha$$

unde de obicei se alege $\alpha = 2$ sau $\alpha = 0$ sau $\alpha = 1$. Alegem $\alpha = 0 \ \vee\ 1$ pentru a induce o reprezentare rara a imaginii (vectorul solutie). Se doreste o reprezentare rara a imaginii, deoarece aceasta permite: compresie usoara; algoritmi rapizi pt. procesare; memorie de stocare mica; eliminarea usoara a zgomotului;...

Exemplu 5: matrix completion

- se da o matrice X cu elemente lipsa (e.g. o imagine)
- se presupune ca matricea X este de rang scazut
- scopul este sa se gaseasca elementele lipsa din X
- pentru a impune rang mic asupra unei matrici se foloseste **nuclear norm** $\|\cdot\|_*$: daca A are descompunerea DVS $A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$, atunci $\|A\|_* = \sum_{i=1}^{r} \sigma_i$
- problema se pune astfel:

$$\min_{X \in \mathbb{R}^{m \times n}} \operatorname{rang}(X) \quad \underbrace{\Longrightarrow}_{\text{relaxare convexa}} \quad \min_{X \in \mathbb{R}^{m \times n}} \|X\|_*$$

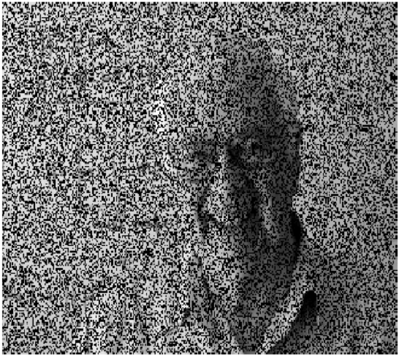
$$\text{s.l.: } X_{ij} = A_{ij} \ \ \forall i,j \in S \qquad\qquad \text{s.l.: } X_{ij} = A_{ij} \quad \forall i,j \in S$$

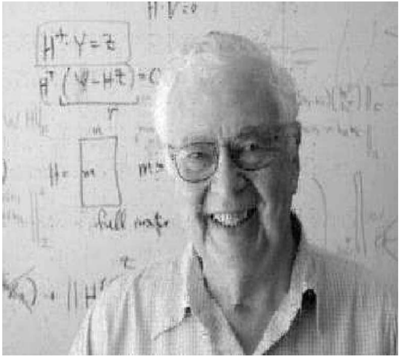
  unde se dau valorile A_{ij} cu $(i,j) \in S$ o submultime a elementelor matricei cautate
- relaxarea convexa se poate scrie ca un SDP:

$$\min_{X, W_1, W_2} \ \operatorname{tr}(W_1) + \operatorname{tr}(W_2)$$

$$\text{s.l.: } X_{ij} = A_{ij} \ \ \forall i,j \in S, \quad \begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \succeq 0$$

Exemplu 5: matrix completion cu aplicare in recuperarea de imagine





imaginea data (40% elemente lipsa) si imaginea recuperata

Exemplu 5: raritate in solutie pt. cazul vector

Definim o masura a raritatii $\|x\|_0 = \sharp \{i : \ x_i \neq 0\}$ si $\|x\|_1 = \sum_i |x_i|$ si apoi problemele de optimizare asociata sistemului $Ax = b$:

$$\min_{x \in \mathbb{R}^n} \|x\|_0 \ \text{ s.l. } Ax = b \qquad \underbrace{\Longrightarrow}_{\text{relaxare convexa}} \qquad \min_{x \in \mathbb{R}^n} \|x\|_1 \ \text{ s.l. } Ax = b$$

• **Mutual coherence**: $\mu(A) = \max\limits_{1 \leq k,j \leq m,\ k \neq j} \frac{|a_k^T a_j|}{\|a_k\|_2 \|a_j\|_2}$

Pentru o matrice A cu rang intreg pe linii (i.e. $m < n$), daca exista o solutie x^* satisfacand relatia:

$$\|x^*\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu(A)}\right)$$

atunci x^* este solutie pentru amandoua problemele de optimizare!

• **Restricted Isometry Property (RIP)**: matricea $A \in \mathbb{R}^{m \times n}$ are RIP(δ, k) daca orice submatrice A_I (obtinute din combinarea a cel mult k coloane din A) are toate valorile singulare marginite intre $1 - \delta$ si $1 + \delta$. Pentru matricea A cu RIP$(0.41, 2k)$ avem ca cele doua probleme de mai sus au aceeasi multime de solutii pe multimea vectorilor de raritate k (Candes & Tao 2005).

Exemplu 5: raritate in solutie pt. cazul matrice

Definim o masura a raritatii $\|X\|_0 = \text{rang}(X)$ si $\|X\|_* = \sum_i \sigma_i$ si problemele de optimizare asociate functiei liniare $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$:

$$\min_{X\in\mathbb{R}^{m\times n}} \|X\|_0 \text{ s.l. } \mathcal{A}(X) = b \quad \underbrace{\Longrightarrow}_{\text{relaxare convexa}} \quad \min_{X\in\mathbb{R}^{m\times n}} \|X\|_* \text{ s.l. } \mathcal{A}(X) = b$$

• **Restricted Isometry Property** de ordin r: cea mai mica constanta $\delta_r(\mathcal{A})$ a.i. pt. orice matrice X de rang cel mult r avem:

$$1 - \delta_r(\mathcal{A}) \leq \frac{\|\mathcal{A}(X)\|}{\|X\|_F} \leq 1 + \delta_r(\mathcal{A})$$

• Presupunem ca $\delta_{2r} < 1$ pentru un intreg $r \geq 1$. Atunci unica solutie a problemei originale neconvexe este matricea de rang cel mult r satisfacand $\mathcal{A}(X) = b$.

• Presupunem ca $r \geq 1$ satisface $\delta_{5r} < 0.1$, atunci cele doua probleme de mai sus (cea originala si relaxarea ei convexa) au aceeasi solutie optima (Recht & Parrilo 2010).

Functia Lagrange

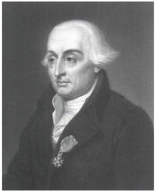
- Reamintim problema NLP generala:

$$f^* = \left\{ \min_{x \in \mathbb{R}^n} \; f(x) : \;\; g(x) \leq 0, \; h(x) = 0 \right\}$$

- Reamintim: $g(x) : \mathbb{R}^n \to \mathbb{R}^m$, $h(x) : \mathbb{R}^n \to \mathbb{R}^p$
- Definim functia Lagrange $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \bar{\mathbb{R}}$:

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \lambda^T g(x) + \mu^T h(x)$$

Numita si Lagrangianul, functia Lagrange a fost introdusa pentru prima data de Lagrange pentru probleme cu constrangeri de egalitate. Variabilele: $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$ se numesc multiplicatori Lagrange (variabile duale).



Functia Lagrange

- Regula generala: impunem ca multiplicatorii lagrange pentru constrangeri de inegalitate sa fie nenegativi, i.e. $\lambda \geq 0$
- Pentru orice variabila primala $\tilde{x}$ fezabila pentru problema de optimizare (NLP) (i.e. $g(\tilde{x}) \leq 0$ si $h(\tilde{x}) = 0$) si pentru orice variabila duala fezabila $(\tilde{\lambda}, \tilde{\mu})$ (i.e. $\tilde{\lambda} \in \mathbb{R}^m_+$ si $\tilde{\mu} \in \mathbb{R}^p$) avem:

$$\mathcal{L}(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \leq f(\tilde{x})$$

  anume functia Lagrange este marginita superior de functia obiectiv a problemei primale.
- Demonstratie: inegalitatea este evidenta deoarece $\tilde{\lambda} \geq 0$, $g(\tilde{x}) \leq 0$ iar $h(\tilde{x}) = 0$

Functia duala

- Functia duala $q : \mathbb{R}^m \times \mathbb{R}^p \to \bar{\mathbb{R}}$ este infimul neconstrans al Lagrangianului in variabila primala x, pentru multiplicatorii λ si μ fixati, i.e:

$$q(\lambda, \mu) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu).$$

- Daca $q(\lambda, \mu)$ este nemarginita inferior, i.e. $q(\lambda, \mu) = -\infty$, atunci spunem ca perechea (λ, μ) este dual infezabila.
- **Lema:** Pentru orice pereche duala $(\tilde{\lambda}, \tilde{\mu})$ fezabila, i.e. $\tilde{\lambda} \in \mathbb{R}^m_+$ si $\tilde{\mu} \in \mathbb{R}^p$, avem

$$q(\tilde{\lambda}, \tilde{\mu}) \leq f^*$$

- **Demonstratie:** pentru orice punct fezabil $\tilde{x}$ avem

$$q(\tilde{\lambda}, \tilde{\mu}) \leq \mathcal{L}(\tilde{x}, \tilde{\lambda}, \tilde{\mu}) \leq f(\tilde{x}) \quad \forall \tilde{x} \in X,\ \tilde{\lambda} \in \mathbb{R}^m_+,\ \tilde{\mu} \in \mathbb{R}^p.$$

atunci pentru $\tilde{x} = x^*$ rezulta $q(\tilde{\lambda}, \tilde{\mu}) \leq f^*$

Functia duala

- **Lema:** Functia duala $q : \mathbb{R}^m \times \mathbb{R}^p \to \bar{\mathbb{R}}$ este intotdeauna functie concava.
- **Demonstratie:** observam ca Lagrangianul $\mathcal{L}(x, \cdot, \cdot)$ este functie afina in (λ, μ) pentru x fixat. Fie $\alpha \in [0,\ 1]$, atunci pentru (λ_1, μ_1) și (λ_2, μ_2) avem:

$$
\begin{aligned}
& q(\alpha\lambda_1 + (1-\alpha)\lambda_2, \alpha_1\mu_1 + (1-\alpha)\mu_2) \\
& \quad = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \alpha\lambda_1 + (1-\alpha)\lambda_2, \alpha_1\mu_1 + (1-\alpha)\mu_2) \\
& \quad = \inf_{x \in \mathbb{R}^n} \alpha\mathcal{L}(x, \lambda_1, \mu_1) + (1-\alpha)\mathcal{L}(x, \lambda_2, \mu_2) \\
& \quad \geq \alpha \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda_1, \mu_1) + (1-\alpha) \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda_2, \mu_2) \\
& \quad = \alpha q(\lambda_1, \mu_1) + (1-\alpha) q(\lambda_2, \mu_2).
\end{aligned}
$$

Problema duala

- Problema duala este definita ca fiind problema de maximizare concava a functiei duale:

$$q^* = \max_{\lambda \geq 0,\ \mu \in \mathbb{R}^p} q(\lambda, \mu),$$

unde notam cu q^* valoarea optima duala.

- **Observatie**: din moment ce q este intotdeauna concava ($-q$ convexa), problema duala este intotdeauna problema convexa:

$$q^* = \max_{\lambda \geq 0,\ \mu \in \mathbb{R}^p} q(\lambda, \mu) = -\min_{\lambda \geq 0,\ \mu \in \mathbb{R}^p} -q(\lambda, \mu)$$

chiar daca problema primala (UNLP) nu este convexa!

- **Observatie**: multimea fezabila duala $\Omega = \mathbb{R}^m_+ \times \mathbb{R}^p$ este o multime simpla (proiectia pe aceasta multime se poate realiza usor)

Exemplu 1 de problema duala

- Fie urmatoarea problema patratica convexa:

$$\min_{x\in\mathbb{R}^n} \quad 1/2x^T x$$
$$\text{s.l:} \quad Ax = b, \quad \text{unde } A \in \mathbb{R}^{p\times n}$$

- **Observam**:
  - $f(x) = 1/2x^T x$ patratica convexa ($Q = I_n \succ 0$) si $g(x) = 0$ (fara inegalitati)
  - $h(x) = Ax - b$, $h : \mathbb{R}^n \to \mathbb{R}^p$, deci $\mu \in \mathbb{R}^p$
  - Lagrangianul $\mathcal{L}(x,\mu) = f(x) + \mu^T h(x)$ este:

$$\mathcal{L}(x,\mu) = 1/2x^T x + \mu^T(Ax - b) \Rightarrow q(\mu) = \min_{x\in\mathbb{R}^n} 1/2x^T x + \mu^T(Ax - b)$$

$$x(\mu) : \quad \nabla_x \mathcal{L}(x,\mu) = 0 \Rightarrow x(\mu) + A^T\mu = 0 \Rightarrow x(\mu) = -A^T\mu$$

$$q(\mu) = \mathcal{L}(x(\mu),\mu) = 1/2\mu^T AA^T\mu - \mu^T AA^T\mu - b^T\mu$$
$$= -1/2\mu^T AA^T\mu - b^T\mu$$

- Astfel, problema duala este de fapt (QP) fara constrangeri, dar Hesiana este pozitiv semidefinita!

$$q^* = \max_{\mu\in\mathbb{R}^p} -1/2\mu^T AA^T\mu - b^T\mu$$

Exemplu 2 de problema duala

- Fie urmatoarea problema CMMP cu restrictii patratice:

$$\min_{x \in \mathbb{R}^n} \quad 1/2\|Ax - b\|^2$$
$$\text{s.l:} \quad 1/2\|x\|^2 \leq c$$

- **Observam**:
  - $f(x) = 1/2\|Ax - b\|^2$ patratica convexa ($Q = A^T A \succeq 0$) si $h(x) = 0$ (fara egalitati)
  - $g(x) = 1/2\|x\|^2 - c$ patratica convexa, deci $\lambda \in \mathbb{R}_+$
  - Lagrangianul $\mathcal{L}(x, \lambda) = f(x) + \lambda g(x)$ este:

$$\mathcal{L}(x, \lambda) = 1/2\|Ax - b\|^2 + 1/2\lambda(x^T x - 2c) \Rightarrow q(\lambda) = \min_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda)$$

$$x(\lambda): \quad \nabla_x \mathcal{L}(x, \lambda) = 0 \Rightarrow A^T A x(\lambda) - A^T b + \lambda I_n x(\lambda) = 0$$

$$\Rightarrow x(\lambda) = (A^T A + \lambda I_n)^{-1} A^T b$$

$$q(\lambda) = \mathcal{L}(x(\lambda), \lambda) = -1/2 b^T A (A^T A + \lambda I_n)^{-1} A^T b - \lambda c + 1/2\|b\|^2$$

- Astfel, problema duala devine (este concava!)

$$q^* = \max_{\lambda \in \mathbb{R}_+} -1/2 b^T A (A^T A + \lambda I_n)^{-1} A^T b - \lambda c + 1/2\|b\|^2$$

Exemplu 3 de problema duala

- Fie urmatoarea problema de optimizare NLP:

$$\min_{x \in \mathbb{R}^2} \quad x_1^4 + x_2^4 - 4x_1x_2$$
$$\text{s.l:} \quad x \geq 0, x_1 + x_2 = 5$$

- **Observam**:
  - $f(x) = x_1^4 + x_2^4 - 4x_1x_2$, $f : \mathbb{R}^2 \to \mathbb{R}$, $n = 2$
  - $g(x) = -x$, $g : \mathbb{R}^2 \to \mathbb{R}^2$, deci $m = 2$ si $\lambda \in \mathbb{R}^2$
  - $h(x) = x_1 + x_2 - 5$, $h : \mathbb{R}^2 \to \mathbb{R}$, deci $p = 1$ si $\mu \in \mathbb{R}$
  - Lagrangianul $\mathcal{L}(x, \lambda, \mu) = f(x) + \lambda^T g(x) + \mu^T h(x)$ este:

$$\mathcal{L}(x, \lambda, \mu) = x_1^4 + x_2^4 - 4x_1x_2 - \lambda_1 x_1 - \lambda_2 x_2 + \mu(x_1 + x_2 - 5)$$

- Astfel, problema duala este de fapt

$$q^* = \max_{\lambda \in \mathbb{R}^2_+,\, \mu \in \mathbb{R}} \left(\min_{x \in \mathbb{R}^2} x_1^4 + x_2^4 - 4x_1x_2 - \lambda_1 x_1 - \lambda_2 x_2 + \mu(x_1 + x_2 - 5) \right)$$

Dualitate slaba

- Din moment ce functia duala este marginita superior de functia obiectiv, rezulta urmatoarea teorema:
- **Teorema (dualitate slaba)**: urmatoarea inegalitate are loc pentru orice problema de optimizare (NLP):

$$q^* \leq f^*$$

- **Observatie:** daca avem x^* fezabil pentru primala si (λ^*, μ^*) fezabil pentru duala astfel incat $q(\lambda^*, \mu^*) = f^*$ atunci x^* este punct de minim global pentru primala iar (λ^*, μ^*) este punct de maxim global pentru duala
- **Observatie:** daca $f^* = -\infty$ atunci $q(\lambda, \mu) = \infty, \forall\ (\lambda, \mu) \in \Omega$. De asemenea, daca q^* atunci problema primala este infezabila

Dualitate puternica

- Diferenta $f^* - q^*$ se numeste duality gap, iar din dualitate slaba este nenegativa.
- In anumite cazuri (e.g. optimizare convexa in care multimea fezabila indeplineste anumite criterii), putem obtine o versiune mai puternica a dualitatii.
- **Conditia Slater**: presupunem ca problema (NLP) este convexa (i.e. f si $g_1, \cdots, g_m$ sunt convexe, iar $h_1, \ldots, h_p$ sunt afine) si exista $\bar{x} \in \mathbb{R}^n$ fezabil astfel incat $g(\bar{x}) < 0$ și $h(\bar{x}) = 0$
- **Teorema (dualitate puternica)**: Daca problema (NLP) primala satisface conditia Slater, atunci valorile optime pentru problemele primale si duale sunt egale, anume:

$$q^* = f^*$$

Mai mult $(\lambda^*)^T g(x^*) = 0$ (complementaritatea), unde x^* este punct de minim global pentru problema primala si (λ^*, μ^*) este punct de maxim global pentru problema duala.

Dualitate: interpretarea min-max

- problema originala (primala) este echivalenta cu

$$\min_{x\in\mathbb{R}^n}\left(\max_{\lambda\geq 0,\mu}\mathcal{L}(x,\lambda,\mu)\right)$$

- problema duala este interschimbarea intre min si max

$$\max_{\lambda\geq 0,\mu}\left(\min_{x\in\mathbb{R}^n}\mathcal{L}(x,\lambda,\mu)\right)$$

- dualitatea slaba are loc intotdeauna deoarece

$$\max\min\mathcal{L}(\cdot)\leq\min\max\mathcal{L}(\cdot)$$

- dualitatea puternica are loc cand

$$\min\max\mathcal{L}(\cdot)\leq\max\min\mathcal{L}(\cdot)\quad\Leftrightarrow\quad\min\max\mathcal{L}(\cdot)=\max\min\mathcal{L}(\cdot)$$

Intuitie

Putem interpreta dualitatea ca o aproximare liniara:

- Definim $1_{-}(y) = \infty$ daca $y > 0$ si 0 altfel (functia indicator al lui $\mathbb{R}_{+}$).
- Definim $1_{0}(y) = \infty$ daca $y \neq 0$ si 0 altfel (functia indicator al lui $\{0\}$).
- Rescriem problema originala (NLP) ca:

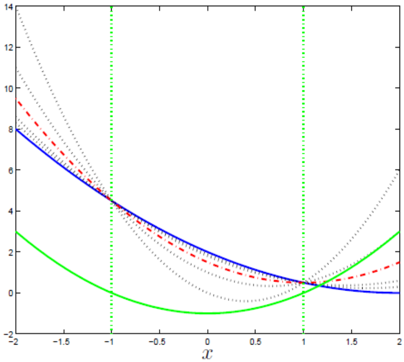
$$\min_{x \in \mathbb{R}^n} f(x) + \sum_i 1_{-}(g_i(x)) + \sum_j 1_0(h_j(x))$$

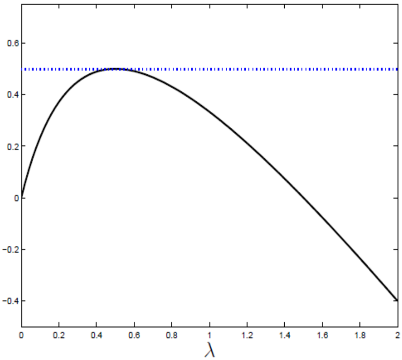
- pentru a obtine dualitate inlocuim $1_{-}(f_i(x))$ cu $\lambda_i f_i(x)$ (o masura de disconfort cand $\lambda_i \geq 0$ si $g_i(x) > 0$).
- De asemenea $\mu_j h_j(x)$ este o margina inferioara pe $1_0(h_j(x))$

$$\min_{x \in \mathbb{R}^n} f(x) + \sum_i \lambda_i g_i(x) + \sum_j \mu_j h_j(x)$$

Interpretare geometrica

$$\min_x (x - c - 1)^2 \qquad \text{s.l.} \quad x^2 \leq c$$



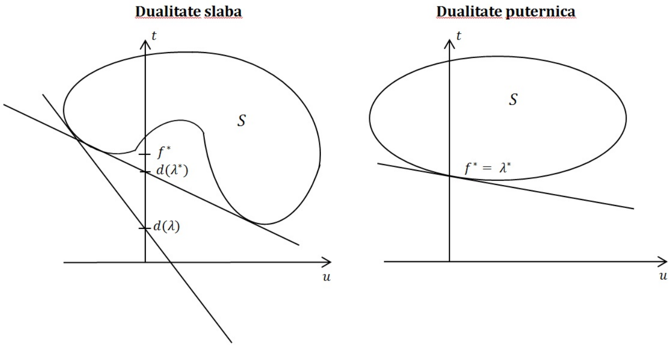


Functia originala (albastru), constrangeri (verde), $\mathcal{L}(x, \lambda)$ pentru diferiti λ (punctat)

Functia duala $q(\lambda)$ si valoarea optima primala f^* (punctat)

Dualitate slaba/puternica, interpretare geometrica

- Pentru a vizualiza grafic diferenta dintre dualitate slaba/puternica, consideram un caz particular de problema (NLP): $\min_{x \in \mathbb{R}^n} \{f(x) : \; g(x) \leq 0\}$, avand o singura constrangere de inegalitate
- Definim multimea $S = \{(u, t) : \; \exists x \in \mathbb{R}^n, f(x) = t, g(x) = u\}$.
- Interpretarea grafica:



Duala unei probleme QP strict convexa

- Fie o problema QP strict convexa ($Q \succ 0$):

$$
\begin{aligned}
f^* = \min_{x \in \mathbb{R}^n} \ & \frac{1}{2} x^T Q x + q^T x \\
\text{s.l.:} \quad & Cx - d \leq 0, \quad Ax - b = 0.
\end{aligned}
$$

  unde multimea fezabila este presupusa a fi nevida.

- Lagrangianul este dat de urmatoare expresie:

$$
\begin{aligned}
\mathcal{L}(x, \lambda, \mu) &= \frac{1}{2} x^T Q x + q^T x + \lambda^T (Cx - d) + \mu^T (Ax - b) \\
&= -\lambda^T d - \mu^T b + \frac{1}{2} x^T Q x + \left(q + C^T \lambda + A^T \mu\right)^T x.
\end{aligned}
$$

- Din moment ce $q(\lambda, \mu) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu)$, avem:

$$
q(\lambda, \mu) = -\lambda^T d - \mu^T b + \inf_{x \in \mathbb{R}^n} \left(\frac{1}{2} x^T Q x + \left(q + C^T \lambda + A^T \mu\right)^T x \right)
$$

Duala unei probleme QP strict convexa

- Observam ca problema:

$$\inf_{x \in \mathbb{R}^n} t(x) \quad \left(= \frac{1}{2} x^T Q x + \left(q + C^T \lambda + A^T \mu \right)^T x \right)$$

  este de fapt o problema QP strict convexa.

- Notam solutia acestei probleme drept $x(\lambda, \mu)$. Din moment ce problema este strict convexa, putem determina $x(\lambda, \mu)$ din conditiile de optimalitate $\nabla t(x(\lambda, \mu)) = 0$, i.e:

$$Qx(\lambda, \mu) + q + C^T \lambda + A^T \mu = 0 \overset{Q \succ 0}{\rightarrow} x(\lambda, \mu) = -Q^{-1}(q + C^T \lambda + A^T \mu)$$

- Inlocuind $x(\lambda, \mu)$ inapoi in expresia dualei, rezulta

$$q(\lambda, \mu) = -\lambda^T d - \mu^T b - \frac{1}{2} \left(q + C^T \lambda + A^T \mu \right)^T Q^{-1} \left(q + C^T \lambda + A^T \mu \right)$$

Duala unei probleme QP strict convexa

- Astfel, problema de optimizare duala este data de expresia:

$$q^* = \max_{\lambda \geq 0,\ \mu \in \mathbb{R}^p} -\frac{1}{2} \begin{bmatrix} \lambda \\ \mu \end{bmatrix}^T \begin{bmatrix} C \\ A \end{bmatrix} Q^{-1} \begin{bmatrix} C \\ A \end{bmatrix}^T \begin{bmatrix} \lambda \\ \mu \end{bmatrix}$$
$$- \begin{bmatrix} d + CQ^{-1}q \\ b + AQ^{-1}q \end{bmatrix}^T \begin{bmatrix} \lambda \\ \mu \end{bmatrix} - \frac{1}{2} q^T Q^{-1} q.$$

- Observatii
  - functia duala este patratica in (λ, μ) si concava
  - problema duala este un QP convex (Hessiana nu mai este pozitiv definita)
  - formularea duala a unei probleme QP are constrangeri mult mai simple: $\lambda \geq 0$ si $\mu \in \mathbb{R}^p$
  - ultimul termen in functia duala este o constanta, insa trebuie pastrat pentru a mentine dualitatea puternica $q^* = f^*$

Exemplu de problema QP strict convexa si duala sa

- Fie urmatoarea problema de optimizare NLP:

$$\min_{x \in \mathbb{R}^2} \quad x_1^2 + x_2^2 + x_1 x_2 + x_1 - x_2$$
$$\text{s.l:} \quad x_1 \geq 0, x_1 - x_2 \geq 1, x_1 + x_2 = 3$$

- Daca exprimam detaliat functie patratica $f(x) = \frac{1}{2} x^T Q x + q^T x$, $f : \mathbb{R}^2 \to \mathbb{R}$, avem:

$$f(x) = \frac{Q_{11}}{2} x_1^2 + \frac{Q_{22}}{2} x_2^2 + \frac{Q_{12} + Q_{21}}{2} x_1 x_2 + q_1 x_1 + q_2 x_2$$

- Astfel, pentru problema noastra (dorim Q simetrica, $Q_{12} = Q_{21}$), rezulta:

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \; q = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Exemplu de problema QP strict convexa si duala sa

- Observam: $\Lambda(Q) = \{1, 3\}$, $Q \succ 0$, $\det(Q) = 3$ si

$$Q^{-1} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

- Observam:
  - $g : \mathbb{R}^2 \to \mathbb{R}^2$, $g(x) = \begin{bmatrix} g_1(x) \\ g_2(x) \end{bmatrix} = \begin{bmatrix} -x_1 \\ x_2 - x_1 + 1 \end{bmatrix}$
  - $g(x) = Cx - d \Rightarrow C \in \mathbb{R}^{2\times 2},\ d \in \mathbb{R}^2$

$$C = \begin{bmatrix} -1 & 0 \\ -1 & 1 \end{bmatrix},\ d = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

  - $h : \mathbb{R}^2 \to \mathbb{R}$, $h(x) = x_1 + x_2 - 3$
  - $h(x) = Ax - b \Rightarrow A \in \mathbb{R}^{1\times 2},\ b \in \mathbb{R}$, $A = \begin{bmatrix} 1 & 1 \end{bmatrix},\ b = 3$

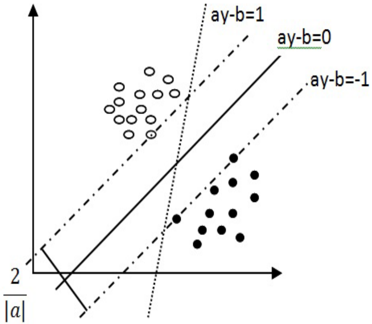
Clasificare liniara binara

Problema clasificarii binare: se urmareste separarea unui set de obiecte in doua clase.

Modelare matematica: sa se determine hiperplanul(a) optim ce separa un set de vectori (obiectele) x_i in doua clase:

(*i*) clasa vectorilor x_i pt. care $a^T x_i - b \leq 1$;

(*ii*) clasa vectorilor x_j pt. care $a^T x_j - b \geq -1$.



Clasificare liniara binara - e-mail filtering

E-mail filtering - determinarea unui estimator ce separa un set dat de email-uri in doua clase: spam e-mail si wanted e-mail.

- dispunem de un dictionar de cuvinte $D = \{c_1, c_2, \ldots, c_p\}$, unde c_i reprezinta cuvantul i.
- antrenam estimatorul printr-un set de e-mail-uri cunoscute
- pentru fiecare e-mail cunoscut i asociem vectorul $x_i = [n_1, n_2, \ldots, n_p]^T$, unde n_j reprezinta numarul de aparitii ale cuvantului c_j in e-mail-ul i.
- pentru fiecare e-mail cunoscut i asociem eticheta $y_i = \{-1, 1\}$, unde $y_i = -1$ daca e-mail-ul cunoscut i este spam, altfel $y_i = 1$.

$$
\begin{aligned}
\min_{w \in \mathbb{R}^n, b \in \mathbb{R}, \xi \geq 0} \quad & \frac{1}{2}\|w\|_2^2 + C\Sigma_{i=1}^m \xi_i \\
\text{s.l.} \quad & y_1(w^T x_1 - b) \leq 1 - \xi_1, \\
& \ldots \\
& y_m(w^T x_m - b) \leq 1 - \xi_m.
\end{aligned}
$$

Programare liniara (LP) si dualitate

- Reamintim ca o problema de programare liniara (LP) are urmatoarea forma:

$$
\begin{aligned}
f^* = \min_{x \in \mathbb{R}^n} \; & c^T x \\
\text{s.l.:} \quad & Cx - d \leq 0, \quad Ax - b = 0.
\end{aligned}
$$

- In programarea liniara putem cauta solutia problemei printre punctele de extrem ale poliedrului ce defineste multimea fezabila:

$$X = \{x \in \mathbb{R}^n : \; Cx - d \leq 0, \quad Ax - b = 0\}.$$

- **Teorema:** daca multimea fezabila lui X este un politop, atunci exista un punct de minim al problemei (LP) intr-unul din varfurile politopului.

Programare liniara (LP) si dualitate

- Demonstratie: o multime X politop se poate scrie ca acoperirea convexa generata de varfurile sale:

$$X = \text{Conv}(\{v_1, \dots, v_q\}).$$

Deoarece $x^* \in X$, putem scrie $x^* = \sum_{i=1}^{q} \alpha_i v_i$, ude $\alpha_i \geq 0$ si $\sum_{i=1}^{q} \alpha_i = 1$. Este clar ca $c^T v_i \geq f^*$, orice v_i fiind fezabil. Notăm cu $\mathcal{I}$ multimea de indecsi: $\mathcal{I} = \{i : \ \alpha_i > 0\}$. Daca exista $i_0 \in \mathcal{I}$ astfel incat $c^T v_{i_0} > f^*$, atunci:

$$f^* = c^T x^* = \alpha_{i_0} c^T v_{i_0} + \sum_{i \in \mathcal{I} \setminus \{i_0\}} \alpha_i c^T v_i > \sum_{i \in \mathcal{I}} \alpha_i f^* = f^*$$

si deci obtinem o contradictie. Aceasta implica ca orice varf pentru care $\alpha_i > 0$ este un punct de minim.

Programare liniara (LP) si dualitate

- Exprimam acum duala unui (LP). Lagrangianul asociata unui (LP) general este dat de:

$$
\begin{aligned}
\mathcal{L}(x,\lambda,\mu) &= c^T x + \lambda^T (Cx - d) + \mu^T (Ax - b) \\
&= -\lambda^T d - \mu^T b + \left(c + C^T\lambda + A^T\mu\right)^T x.
\end{aligned}
$$

- Functia duala corespunzatoare este:

$$
\begin{aligned}
q(\lambda,\mu) &= -\lambda^T d - \mu^T b + \inf_{x\in\mathbb{R}^n} \left(c + C^T\lambda + A^T\mu\right)^T x \\
&= -\lambda^T d - \mu^T b + \begin{cases} 0 & \text{dacă } \; c + C^T\lambda + A^T\mu = 0 \\ -\infty & \text{altfel.} \end{cases}
\end{aligned}
$$

- Observatie: problema din membrul drept de mai sus este un (LP) nemarginit, a carui minim este $-\infty$ daca $c + C^T\lambda + A^T\mu \neq 0$

Programare liniara (LP) si dualitate

- Astfel, putem exprima problema duala drept:

$$q^* = \max_{\lambda \in \mathbb{R}^m,\ \mu \in \mathbb{R}^p} \begin{bmatrix} -d \\ -b \end{bmatrix}^T \begin{bmatrix} \lambda \\ \mu \end{bmatrix}$$

$$\text{s.l.:}\ \ \lambda \geq 0, \quad c + C^T \lambda + A^T \mu \ = \ 0.$$

- Observatie: un LP poate fi scris in forma standard:

$$\min_{x \in \mathbb{R}^n} \{c^T x : \ \ Ax = b,\ x \geq 0\},$$

prin folosirea de variabile suplimentare (numite si variabile artificiale)

- Duala sa rezulta: $\max_{\mu \in \mathbb{R}^n} \{b^T \mu : \ A^T \mu \leq c\}$
- **Teorema de dualitate pentru (LP):** daca una dintre problemele (LP), primala sau duala, are solutie optima atunci si cealalta problema are solutie optima si valorile optime corespunzatoare sunt egale. Mai mult, daca una dintre probleme, primala sau duala, are functie obiectiv nemarginita, atunci cealalta problema nu are puncte fezabile.

Exemplu de (LP) din practica (problema vinarului)

- Un vinar trebuie sa decida cate sticle de vin rosu si cate de vin alb sa produca. El poate vinde o sticla de vin rosu pentru \$12 iar una de vin alb pentru \$7.
- Obiectivul lui este desigur de a-si maximiza venitul:

  $$F(x_1, x_2) = 12x_1 + 7x_2$$

  unde x_1 reprezinta numarul de sticle de vin rosu ce vor fi produse, iar x_2 cele de vin alb.
- Vinul trebuie desigur invechit (4 ani pentru rosu si 3 ani pentru alb), inainte de a fi vandut, insa spatiul de depozitare este limitat la 1000 ani-invechire-sticla pentru o productie. Astfel, la problema noastra adaugam constrangerea:

  $$4x_1 + 3x_2 \leq 1000$$

Exemplu de (LP) din practica (problema vinarului)

- Cantitatea de struguri ce poate fi procesat pentru a produce vinul este limitata, e.g. la $16.000kg$, si este necesar de a procesa 3 kg pentru a produce pe cel rosu ,si 2 kg pentru al produce pe cel alb. Avem astfel urmatoarea constrangere:

$$3x_1 + 2x_2 \leq 16000$$

- O constrangere desigur evidenta este ca nu putem produce cantitati negative, i.e. $x_1 \geq 0, x_2 \geq 0$. Astfel, problema vinarului poate fi formulata drept o problema (LP) cu constrangeri de inegalitate:

$$\min_{x \in \mathbb{R}^n} \{c^T x : \quad Cx \leq d\},$$

unde

$$c = \begin{bmatrix} 12 \\ 7 \end{bmatrix}, C = \begin{bmatrix} 4 & 3 \\ 3 & 2 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}, d = \begin{bmatrix} 1000 \\ 16000 \\ 0 \\ 0 \end{bmatrix}$$