# Convergenta liniara a algoritmilor de ordinul I pentru probleme *ne-tari* convexe

**Ion Necoara**
Fac. de Automatica, UPB

Colaborare cu: **Yu. Nesterov** & **F. Glineur**
UCL, Belgium

2016

# Ion Necoara

- 2000 matematica: Fac. Matematica, UB
- 2002 master in Statistica si Optimizari (Fac. Matematica)
- 2002–2006 doctorat in control, TU Delft
- 2007 –2009 postdoc, KU Leuven
- 2009 cadru didactic Fac. Automatica, UPB (2015 - profesor)
- 2014 - teza abilitare, conducator de doctorat din 2015
- domenii de interes
  - optimizare convexa/ne-convexa/Big Data
  - algoritmi numerici cu garantari matematice a eficientei
  - modelare/control bazat pe optimizare
- web-page: http://acse.pub.ro/person/ion-necoara

# Motivatie

- Problema generala de optimizare:

$$\min_{x \in X \subseteq \mathbb{R}^n} f(x)$$

- Presupuneri standard:
  - $f$ functie convexa si $X$ multime convexa
  - $f$ continuu diferentiabila cu gradient Lipschitz (smooth)
  - ce alte conditii...?

Sub ce conditii/cum putem garanta convergenta liniara pentru metode de tip gradient?

- raspuns partial: N, Clipici, *Parallel random coordinate descent method for composite minimization: onvergence analysis and error bounds*, SIAM J. Optimization, 2016
- astazi un raspuns mai complet: N, Nesterov, Glineur: *Linear convergence of first-order methods for non-strongly convex optimization*, submitted, 2015

# Clasificare metode de optimizare

• Informatia ce indica comportamentul unei functii $f \in \mathbb{R}^n \to \mathbb{R}$ intr-un punct $x \in \mathbb{R}^n$ se poate clasifica:

- Informatie de ordin 0: $f(x)$
- Informatie de ordin 1: $f(x), \nabla f(x)$
- Informatie de ordin 2: $f(x), \nabla f(x), \nabla^2 f(x)$
- ...

• Fie algoritmul iterativ definit de $x_{k+1} = \mathcal{M}(x_k)$; in functie de ordinul informatiei utilizate in expresia lui $\mathcal{M}$:

- Metode de ordin 0: $f(x_k)$
- Metode de ordin 1: $f(x_k), \nabla f(x_k)$
- Metode de ordin 2: $f(x_k), \nabla f(x_k), \nabla^2 f(x_k)$
- ...

# Rate de convergenta in optimizare

Problema generala de optimizare si metoda iterativa:

$$f^* = \min_{x \in X \subseteq \mathbb{R}^n} f(x) \quad \& \quad x_{k+1} = \mathcal{M}(x_k)$$

Dorim caracterizare rata convergenta la solutie (local/global):

$$x_k \to x^* \;\vee\; f(x_k) \to f^* \;\vee\; \|\nabla f(x_k)\| \to 0 \; (X = \mathbb{R}^n)$$

- **Rezultat** classic din analiza: orice functie continua poate fi aproximata cu o functie diferentiabila, arbitrar de bine!
- Deci presupunand doar diferentiabilitate nu va fi suficient pentru a caracteriza rata de convergenta a unei metode
- E nevoie sa impunem o conditie pe *magnitudinea derivatei*

Classic in optimizare, se considera urmatoarele conditii:

- functia obiectiv $f$ are gradient Lipschitz (permite rata locala)
- in plus: $f$ (tare) convexa & $X$ convexa (permit rate globale)

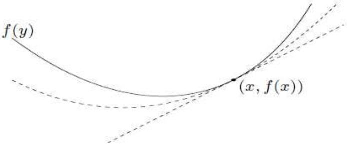
# Convexitate tare - continuitate Lipschitz

- $f$ diferentiabila este convexa daca:

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \forall x, y$$

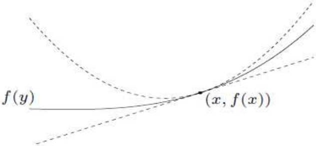
- $f$ diferentiabila este *tare* convexa daca $\exists \sigma > 0$ a.i.:

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{\sigma}{2} \|y - x\|^2 \quad \forall x, y$$




- $f$ are gradient Lipschitz daca $\exists L > 0$ a.i.:

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| \quad \forall x, y$$



# Istoric - Metode de ordinul I

Cea mai *simpla* metoda de ordinul I: Metoda Gradient

$$\text{solutie ec. } F(x) = 0 \Longleftarrow \text{punct fix iter. } x_{k+1} = x_k - \alpha F(x_k)$$

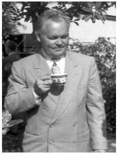
- pasul $\alpha > 0$ constant sau variabil
- iteratia foarte simpla (operatii vectoriale)!
- convergenta rapida/lenta?
- adecvata pentru $x$ de dimensiune mare
  - Prima aparitie in lucrarea [Cau:47] a lui Augustin-Louis Cauchy, 1847
  - Cauchy rezolva un sistem neliniar de 6 ecuatii cu 6 necunoscute, utilizand metoda gradient

[Cau:47] A. Cauchy. *Methode generale pour la resolution des systemes dequations simultanees*, C. R. Acad. Sci. Paris, 25, 1847

# Istoric - Metode de ordinul I

Rata de convergenta slaba a metodei gradient reprezinta motivatia dezvoltarii de alte metode de ordin I cu performante superioare

- Metoda de Gradienti Conjugati - autori independenti Lanczos, Hestenes, Stiefel (1952)
  • QP convex solutia in $n$ iteratii



- Metoda de Gradient Accelerat - dezvoltata de Yurii Nesterov (1983)
  • cu un ordin mai rapida decat gradientul clasic

- MGA neutilizata pentru 2 decade! - acum este una din cele mai folosite metode de optimizare
- Google returneaza aprox. 20 mil. rezultate ($\approx$ 2000 citari)

# Metoda Gradient

$$\boxed{\min_{x \in \mathbb{R}^n} f(x)}$$

- conditii de optimalitate: $\nabla f(x^*) = 0$
- metoda gradient (MG) pentru problema de optimizare:

$$\boxed{x_{k+1} = x_k - \alpha_k \nabla f(x_k)}$$

- Pasul $\alpha_k$ poate fi ales:
  - pas constant, conditii Wolfe, backtracking, pas ideal,...
- Avantaje MG:
  - complexitate de calcul redusa - $\mathcal{O}(n)+$ calcul $\nabla f(x)$
  - nu necesita informatii de ordin II
  - convergenta globala garantata in conditii obisnuite
  - robust la erori de calcul/la gradient inexact [NecNed:13]
- Dezavantaje MG:
  - rata convergenta destul de lenta - subliniara sau cel mult liniara (sub anumite conditii de regularitate)

[NecNed:13] Necoara, Nedelcu, *Rate analysis of inexact dual first order methods: application to dual decomposition*, IEEE T. Automatic Control, 59(5), 2014

# Rate de convergenta pentru MG

Presupunem $f$ are $\nabla f(x)$ Lipschitz, i.e.

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad \forall x, y \in \mathrm{dom} f$$

Metoda gradient (MG) cu pas constant $\alpha = 1/L$

$$\boxed{x_{k+1} = x_k - 1/L \nabla f(x_k)}$$

**Teorema** Sub convexitate si gradient Lipschitz, MG are rata de convergenta *subliniara* $\mathcal{O}(1/k)$:

$$f(x_k) - f^* \leq \frac{L\|x_0 - x^*\|^2}{2k}$$

**Teorema** Daca in plus $f$ tare convexa cu constanta $\sigma$, i.e.

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\sigma}{2}\|x - y\|^2 \quad \forall x, y$$

atunci rata de convergenta liniara (numar conditionare $\mu = \sigma/L$):

$$f(x_k) - f^* \leq \left(\frac{1-\mu}{1+\mu}\right)^k \frac{L\|x_0 - x^*\|^2}{2}$$

# Metoda de Gradient Accelerat (MGA)

Convergenta slaba a MG $\Longrightarrow$ dezvoltarea unor metode cu performante superioare:

- Metoda de Gradient Accelerat (Nesterov 1983) - cu un ordin mai rapida decat gradientul clasic in cazul problemelor convexe
- **Metoda de Gradient Accelerat implica:**

$$\boxed{\text{Pas de gradient: } x_{k+1} = y_k - \frac{1}{L}\nabla f(y_k)}$$

$$\boxed{\text{O combinatie liniara : } y_{k+1} = x_{k+1} + \beta_k(x_{k+1} - x_k)}$$

- Punctele initiale $x_0 = y_0$ si $\beta_k$ ales in mod adecvat, e.g. pentru $f$ tare convexa luam $\beta_k = \frac{\sqrt{L}-\sqrt{\sigma}}{\sqrt{L}+\sqrt{\sigma}}$.
- MGA are performante superioare fata de MG insa complexitatea pe iteratie ramane aceeasi (MGA este optima!).

[Nes:83] Nesterov, *A method for unconstrained convex minimization problem with the rate of convergence* $\mathcal{O}(1/k^2)$, Soviet. Math. Docl., 269, 1983.

# Rate de convergenta pentru MGA

**Metoda de Gradient Accelerat**

$$\boxed{\text{Pas de gradient: } x_{k+1} = y_k - \frac{1}{L}\nabla f(y_k)}$$

$$\boxed{\text{O combinatie liniara : } y_{k+1} = x_{k+1} + \beta_k(x_{k+1} - x_k)}$$

**Teorema** Sub convexitate si gradient Lipschitz, MGA are rata de convergenta *subliniara* $\mathcal{O}(1/k^2)$:

$$f(x_k) - f^* \leq \frac{L\|x_0 - x^*\|^2}{k^2}$$

**Teorema** Daca in plus $f$ tare convexa cu constanta $\sigma$, i.e.

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{\sigma}{2}\|x - y\|^2 \quad \forall x, y$$

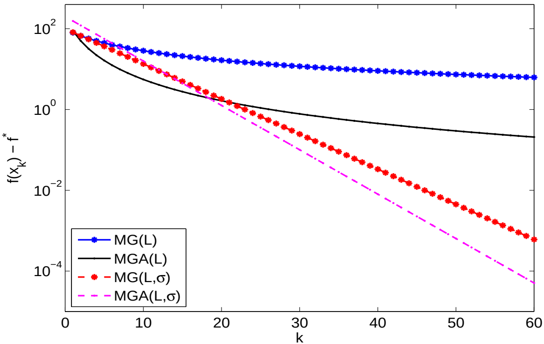
atunci rata de convergenta liniara (numar conditionare $\mu = \sigma/L$):

$$f(x_k) - f^* \leq (1 - \sqrt{\mu})^k \frac{L\|x_0 - x^*\|^2}{2}$$

Observam ca numarul de conditionare este sub radical in MGA!!!

# Convergenta MG versus MGA

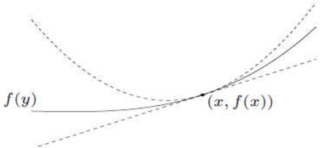
| **Conditii** $f$ | **MG** | **MGA** |
| --- | --- | --- |
| $\nabla f(x)$ Lipschitz$(L)$ | $\mathcal{O}\left(\frac{LR^2}{k}\right)$ | $\mathcal{O}\left(\frac{LR^2}{k^2}\right)$ |
| $\nabla f(x)$ Lipschitz si $f(x)$ tare convexa$(\sigma)$ | $\mathcal{O}\left(\left(\frac{L-\sigma}{L+\sigma}\right)^k\right)$ | $\mathcal{O}\left(\left(1-\sqrt{\frac{\sigma}{L}}\right)^k\right)$ |



# Continuitate Lipschitz - convexitate tare

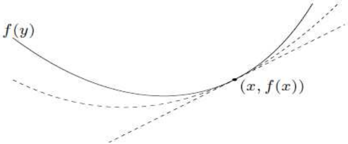
- Daca $f$ are gradient Lipschitz atunci:

$$f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2}\|y - x\|^2$$




- Daca $f$ este tare convexa atunci:

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{\sigma}{2}\|y - x\|^2$$




- Daca $f$ tare convexa si gradient Lipschitz atunci co-coercivitate:

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq \frac{\sigma L}{\sigma + L}\|x - y\|^2 + \frac{1}{\sigma + L}\|\nabla f(y) - \nabla f(x)\|^2$$

# Problema convexitatii tari a lui $f$

- Clasic: pentru a demonstra rata de convergenta liniara in metode de tip gradient este necesar ca $f$ sa fie tare convexa
- In practica: convexitatea tare a lui $f$ nu este des intalnita

**Exemplu 1**:

$$f(x) = x^T Q x \Longrightarrow L = \|Q\| = \lambda_{\max}(Q) \quad \& \quad \sigma = 1/\|Q^{-1}\| = \lambda_{\min}(Q)$$

**Exemplu 2**:

$$f(x) = \log\left(1 + e^{a^T x}\right) \Longrightarrow L = \|\nabla^2 f(x)\| = \|\frac{e^{a^T x}}{(1+e^{a^T x})^2} a a^T\| \quad \& \quad \sigma = 0$$

Insa, multe aplicatii se modeleaza ca $\min\limits_{x \in X} f(x)$, in care:

- funct. obiectiv de forma $f(x) = g(Ax) + c^T x$, cu $g$ tare convexa
- multimea constragerilor $X$ este poliedrala: $X = \{x : \; Cx \leq c\}$

# Aplicatie 1

- Problema celor mai mici patrate liniara (regularizata):

$$\min_{x\in\mathbb{R}^n} \|Ax - b\|^2 \;\; (+\lambda\|x\|_1)$$

- Problema CMMP:

$$f(x) = x^T A^T Ax - 2b^T Ax + b^T b = g(Ax) + \langle q, x\rangle + r$$

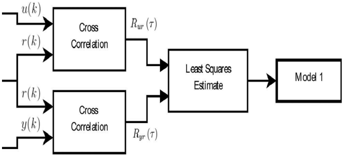
  - observam $g(t) = t^T t \implies g$ tare convexa in $t$ cu $\sigma = 2$
- Problema CMMMP regularizata (lasso in compressed sensing):

$$z = (x, u) \implies f(z) = \|[A \;\; 0]z - b\|^2 + \lambda(0, e)^T z \;\; \& \;\; -u \le x \le u$$

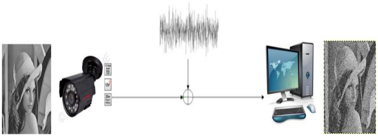
  - observam $g(t) = t^T t \implies g$ tare convexa in $t$ cu $\sigma = 2$
  - constrangeri poliedrale $X = \{z = (x, u) : \; -u \le x \le u\}$

Problema CMMP liniara (regularizata) apare in multe aplicatii:

identificare sisteme




reconstructie semnale



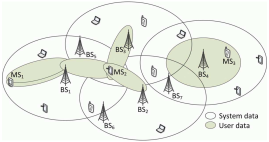
# Aplicatie 2

- Consideram o problema de optimizare cu constrangeri liniare:

$$f^* = \min_{u} \quad g(u)$$
$$\text{s.l.} \quad Cu \leq c$$

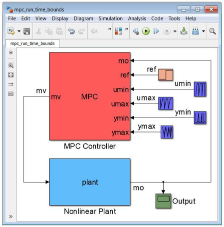
- Probleme practice ce se pot formula astfel enumeram: network utility maximization (NUM), direct current optimal power flow (DC-OPF), model predictive control (MPC)

NUM




DC-OPF

MPC



# Aplicatie 2

- Obtinem functia duala: $f(x) = \min_{u \in \mathbb{R}^m} g(u) + \langle x, Cu - c \rangle$
- Sub dualitate puternica avem problema duala:

$$f^* = \max_{x \in \mathbb{R}^n_+} f(x)$$

- Fiind data $g$ convexa, notam $\tilde{g}(t)$ conjugata ei:

$$\tilde{g}(t) = \max_{u \in \mathbb{R}^m} \langle t, u \rangle - g(u)$$

- **Lema**: daca $g$ are $\nabla g(u)$ Lipschitz cu constanta $L$, atunci $\tilde{g}(t)$ este tare convexa cu $\sigma = \frac{1}{L}$ si reciproc
- Obsv.1: functia duala poate fi scrisa drept

$$f(x) = -\tilde{g}(-C^T x) - c^T x$$

unde $\tilde{g}(t)$ este tare convexa in $t$ cu $\sigma = \frac{1}{L}$
- Obsv.2: multimea constrangerilor $X = \mathbb{R}^n_+$ poliedrala simpla

# Aplicatie 2: MPC

- Exemplu din practica: problema MPC pe orizont $N$

$$V_N^*(x) = \min_{x(t),u(t)} V_N(x(t), u(t)) \quad \left( = \sum_{t=0}^{N-1} \ell(x(t), u(t) + \ell_{\mathrm{f}}(x(N)) \right)$$
$$\text{s.l. } x(t+1) = Ax(t) + Bu(t), \quad x(0) = x$$
$$u(t) \in U, \quad \forall t \geq 0$$

unde $U$ este multime poliedrala.

- Daca luam

$$\mathbf{u} = \left[ u(0)^T, u(1)^T, \ldots, u(N-1)^T \right]^T$$

atunci prin eliminarea starilor, problema de MPC poate fi scrisa drept:

$$V_N^*(x) = \min_{\mathbf{u}} \quad g(\mathbf{u})$$
$$\text{s.l.} \quad C\mathbf{u} \leq c$$

# Metode de tip gradient: abordarea primala vrs. duala

- Reamintim problema de optimizare generala:

$$\min_{u \in U} g(u)$$

- Iteratiile metodei gradient trebuie sa ramana fezabile $\rightarrow$ metoda de gradient proiectat

$$\boxed{u_{k+1} = [u_k - \alpha_k \nabla g(u_k)]_U}$$

- Aavantaj major abordare primala: sub $\nabla g(u)$ Lipschitz si convexitate tare $\Rightarrow$ convergenta liniara pentru MG proiectat
- Dezavantaj major abordare primala: trebuie realizata proiectia $[u_k - \alpha_k \nabla g(u_k)]_U$
- Daca $U$ nu este simpla (e.g. descrisa de $Cu \leq c$, cu $C$ matrice generala), atunci proiectia este foarte dificil de calculat $\rightarrow$ abordarea duala

# Metode de tip gradient: abordarea primala vrs. duala

- Abordarea duala: rezolvam problema duala

$$\max_{x \in X} f(x) \qquad \text{cu} \quad X = \mathbb{R}^n_+$$

- Avantaj major fata de abordarea primala: proiectia nu trebuie realizata decat pentru multiplicatorul Lagrange pentru constrangerile de inegalitate $x \in \mathbb{R}^n_+$ - simpla!
- Dezavantaj major: din literatura existenta prin abordarea duala convergenta NU este liniara
- Rezultate existente pentru algoritmi de tip dual gradient sunt de convergenta subliniara $\left(\mathcal{O}\left(\frac{1}{k^p}\right),\ p = 1, 2\right)$ [NecSuy:08], [NecNed:13] sau covergenta locala liniara [LuoTse:93]

gaura in teoria convergentei metodelor de tip gradient dual!


[NecSuy:08] Necoara, Suykens, *Application of a smoothing technique to decomposition in convex optimization*, IEEE T. Automatic Control, 53(11), 2008

[NecNed:13] Necoara, Nedelcu, *Rate analysis of inexact dual first order methods: application to dual decomposition*, IEEE T. Automatic Control, 59(5), 2014

[LuoTse:93] Luo, Tseng,, *On the convergence rate of dual ascent methods for strictly convex minimization*, Math. Oper. Res., 18, 1993

# Obtinerea convergentei liniare pt. metode de tip gradient

- Intrebari cheie pt. problema convexa smooth $\min_{x \in X} f(x)$:
  - Se poate obtine o rata de convergenta liniara pentru problema $\min_{x \in X} f(x)$ cand $f$ nu este tare convexa?
  - Se poate obtine rata liniara si pe abordarea duala a problemei primale cu constrangeri liniare $\min_{x:\ Cx \leq c} f(x)$?
- Raspuns: DA (in anumite cazuri).
- Intrebare cheie: cum?
- Raspuns: prin conceptul de Relaxarea Convexitatii Tari.
- Relaxarea convexitatii tari: in esenta presupune relaxarea conditiei de convexitate tare (valabila pentru orice $x, y \in \mathbb{R}^n$) la o conditie valabila pentru $x$ sau $y$ bine fixat:

$$\boxed{f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{\sigma}{2}\|y - x\|^2 \quad \forall x, y \in \mathbb{R}^n}$$

- Aspect esential: nu este necesar ca $f$ sa fie tare convexa (TC) pentru a avea o proprietate de tip TC numai in $x$ sau $y$.

# Resultate existente - marginirea erorii in optimizare

Primele contributii aduse de Tseng (1992)
Cadrul general: pentru problema convexa

$$X^* = \arg\min_{x \in X} f(x)$$

$\bar{x} = [x]_{X^*}$ & $\nabla^+ f(x) = x - [x - \nabla f(x)]_X$

**Lema**: Daca $f$ tare convexa si $\nabla f(x)$ Lipschitz atunci:

$$\|x-y\|^2 \leq \frac{1+L}{\sigma} \|\nabla^+ f(x) - \nabla^+ f(y)\| \|x-y\| \;\; \forall x, y$$

Fixand $y = \bar{x}$ rezulta o relatie de marginire a erorii (ME):

$$ME: \quad \|x - \bar{x}\| \leq \frac{1+L}{\sigma} \|\nabla^+ f(x)\| \;\; \forall x \in X$$

- $f$ are gradient Lipschitz si convexa
- in plus, relatia de marginire a erorii, ME, are loc

$$\Downarrow$$

- clasica metoda gradient converge liniar!

[Tse:92] Luo, Tseng, *On the linear convergence of descent methods for convex essentially smooth minimization*, SIAM J. Contr. Optim., 30, 1992

# Resultate existente - marginirea erorii in optimizare

Din conditiile de optimalitate - multimea solutiilor se scrie ca:

$$X^* = \{x \in \mathbb{R}^n : x = [x - \nabla f(x)]_X\}$$

- Intuitiv, pentru un algoritm iterativ, cantitatea:

  $$\|x_k - [x_k - \nabla f(x_k)]_X\|$$

  reprezinta o masura buna a optimalitatii punctului $x_k$.
- Contributia lui Tseng la problema $\min_{x \in X} f(x)$:
  - pentru probleme unde $f(x) = g(Ax)$, iar $g(t)$ tare convexa
  - multimea constrangerilor $X$ poliedrala
- utilizand conditiile de optimalitate si concepte din analiza variationala (Robinson'78), a demonstrat ca $\exists \kappa > 0$ a.i. notand $\bar{x} = [x]_{X^*}$ & $\nabla^+ f(x) = x - [x - \nabla f(x)]_X$ avem :

  $$\|x - \bar{x}\| \leq \kappa \|\nabla^+ f(x)\| \quad \forall x \in \mathcal{N}(X^*)$$

- Neajunsuri ale contributiilor lui Tseng: proprietatea de error bound este asigurata doar local, i.e. pentru $x$ suficient de aproape de $X^*$ (intr-o vecintata a lui $X^*$)

# Resultate existente - marginirea erorii in optimizare

**Teorema** Sub proprietatea de marginirea erorii locala, Tseng a demonstrat convergenta liniara *locala* a gradientului proiectat.

**Neajunsuri si probleme deschise**:

1. Desi convergenta liniara locala este o contributie majora, dorim convergenta globala (ca si in cazul tare convex)
2. Problema deschisa: met. gradient accelerat poate converge linear pt. probleme care relaxeaza convexitatea tare?
3. De ce este mai dificil de analizat MGA fata de MG?
   - MG desi nu utilizeaza informatie legata de convexitate tare in iteratie ($\sigma$), totusi converge liniar pt fuctii smooth tari convexe!
   - Pe de alta parte, MGA utilizeaza informatie legata de convexitate tare in iteratie $\beta = \frac{\sqrt{L}-\sqrt{\sigma}}{\sqrt{L}+\sqrt{\sigma}}$. Deci daca aceasta conditie TC nu mai este valabila ($\sigma = 0$) ce putem face?

# Ideea de baza a lui Tseng?

Functia $f$ are proprietatea de convexitatea tare $(\sigma)$, atunci:

$$(TC): \; f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{\sigma}{2}\|y - x\|^2 \quad \forall x, y$$

Reamintim notatiile: $\bar{x} = [x]_{X^*}$ & $\nabla^+ f(x) = x - [x - \nabla f(x)]_X$

**Lema**: Daca $f$ tare convexa $(\sigma)$ si $\nabla f(x)$ Lipschitz $(L)$, atunci:

$$\|x - y\|^2 \leq \frac{1+L}{\sigma}\|\nabla^+ f(x) - \nabla^+ f(y)\|\|x - y\| \;\; \forall x, y$$

Fixand $y = \bar{x}$ rezulta o relatie de marginire a erorii (ME):

$$ME: \quad \|x - \bar{x}\| \leq \frac{1+L}{\sigma}\|\nabla^+ f(x)\| \;\; \forall x \in X$$

**Obs.**: Marginirea erorii (ME) NU implica convexitatea tare (TC)!!!

Vom proceda intr-o maniera similara!

# Ideea de baza a noastra

Functia $f$ cu gradient Lipschitz $(L)$ si tare convexa $(\sigma)$, atunci:

$$(TC): \; f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{\sigma}{2}\|y-x\|^2 \quad \forall x, y$$

Fie $X^* = \arg\min_{x \in X} f(x)$ si notam $\bar{x} = [x]_{X^*}$. Atunci:

**Lema**: Avem urmatoarele relaxari ale convexitatii tari (TC):

1. Luand $y = \bar{x}$ obt. prima cond. de relaxare a convexitatii tari:
$$\textbf{TC1}: \; f^* \geq f(x) + \nabla f(x)^T (\bar{x} - x) + \frac{\sigma}{2}\|x - \bar{x}\|^2 \quad \forall x,$$
2. Luand $y = x$ & $x = \bar{x}$ obt. a doua conditie de relaxare a TC:
$$\textbf{TC2}: \; f(x) \geq f^* + \nabla f(\bar{x})^T (x - \bar{x}) + \frac{\sigma}{2}\|x - \bar{x}\|^2 \quad \forall x,$$
3. Luand $y = x$ & $x = \bar{x}$ si tinand seama ca $\nabla f(\bar{x})^T (x - \bar{x}) \geq 0$ obtinem a treia conditie de relaxare a convexitatii tari:
$$\textbf{TC3}: \; f(x) \geq f^* + \frac{\sigma}{2}\|x - \bar{x}\|^2 \quad \forall x,$$

**Teorema**: Avem $TC \Rightarrow TC1 \Rightarrow TC2 \Rightarrow TC3 \;\Leftrightarrow\; ME$ (implicatii stricte!)

# Relaxarea convexitatii tari 1

Luand $y = \bar{x}$ obtinem prima conditie de relaxare a convexitatii tari:

$$\textbf{TC1}: \ f^* \geq f(x) + \nabla f(x)^T(\bar{x} - x) + \frac{\sigma_1}{2}\|x - \bar{x}\|^2 \quad \forall x$$

**Teorema**. Sub prima conditie de relaxare a convexitatii tari avem (numar conditionare $\mu_1 = \sigma_1/L$):

- Metoda gradient converge liniar:

$$f(x_k) - f^* \leq \left(\frac{1-\mu_1}{1+\mu_1}\right)^k \frac{L\|x_0 - x^*\|^2}{2}$$

- Metoda gradient accelerat converge liniar cu $\beta_k = \frac{\sqrt{L}-\sqrt{\sigma_1}}{\sqrt{L}+\sqrt{\sigma_1}}$

$$f(x_k) - f^* \leq (1 - \sqrt{\mu_1})^k (f(x_0) - f^*)$$

**Teorema**. Problema smooth convexa:

$$\min_{x:\ Cx \leq d} g(Ax)$$

unde $g$ este tare convexa si cu gradiet Lipschitz, iar matricea $A$ oarecare, satisface TC1 global!

# Relaxarea convexitatii tari 3

Luand $y = \bar{x}$ obtinem prima conditie de relaxare a convexitatii tari:

$$\textbf{TC3}: \; f(x) \geq f^* + \frac{\sigma_3}{2}\|x - \bar{x}\|^2 \quad \forall x$$

**Teorema**. Sub a treia conditie de relaxare a convexitatii tari avem (numar conditionare $\mu_3 = \sigma_3/L$):

- Metoda gradient converge liniar:

$$f(x_k) - f^* \leq \left(\frac{1}{1+\mu 3}\right)^k \frac{L\|x_0 - x^*\|^2}{2}$$

- Metoda gradient accelerat *restartata* la fiecare $K_e = 2e/\sqrt{\mu_3}$ iteratii, converge liniar:

$$f(x_k) - f^* \leq \left(1 - \frac{\sqrt{\mu_3}}{e}\right)^k (f(x_0) - f^*).$$

**Teorema**. Problema smooth convexa:

$$\min_{x:\, Cx \leq d} g(Ax) + c^T x$$

unde $g$ este tare convexa si cu gradiet Lipschitz, iar matricea $A$ oarecare, satisface TC3 pe orice multime subnivel $\{x : \; f(x) \leq M\}$.

# Relaxarea convexitatii tari 4: "Marginirea Erorii Generalizata"

- Extindem clasa de probleme $\min_{x \in X} f(x)$ ce satisfac proprietatea de error bound la proprietate de error bound generalizat:

$$\boxed{\text{(GEB):} \quad \|x - \bar{x}\| \leq \left(\kappa_1 + \kappa_2 \|x - \bar{x}\|^2\right) \|\nabla^+ f(x)\| \quad \forall x \in X}$$

  - unde $\bar{x} = [x]_{X^*}$ si $\nabla^+ f(x) = x - [x - \nabla f(x)]_X$
  - unde $\kappa_1 > 0$ si $\kappa_2 \geq 0$

- Observatie: proprietatea de error bound introdusa de Tseng este inclusa in categoria (GEB):

$$\kappa_1 = \kappa, \ \kappa_2 = 0 \rightarrow \|x - \bar{x}\| \leq \kappa \|\nabla^+ f(x)\|$$

- Contributia noastra:
  (i) proprietatea de marginirea erorii generalizata este globala pentru o clase de probleme de optimizare mai larga
  (ii) include toate celelalte clase definite pana in prezent

# Marginirea erorii generalizata global in optimizare: caz 1

- Consideram problema generala de tip composite:

$$\min_{x \in \mathbb{R}^n} F(x) \quad (= f(x) + \Psi(x)),$$

in care presupunem $f(x) = g(Ax)$, cu $A$ matrice oarecare, si urmatoarele conditii:
  - $g(t)$ tare convexa in $t$ si cu gradient Lipschitz
  - $\Psi(x)$ functie poliedrala (supragraficul - multime poliedrala)
  - $\Psi(x)$ este mariginita superior (i.e. $\Psi(x) \leq \bar{\Psi} \; \forall x \in dom\Psi$)
  - $\Psi(x)$ este Lipschitz continua cu constanta $L_\Psi$
- Atunci $F$ va satisface proprietatea (GEB):

$$\boxed{\|x - \bar{x}\| \leq \left(\kappa_1 + \kappa_2 \|x - \bar{x}\|^2\right) \|\nabla^+ F(x)\| \quad \forall x \in X}$$

unde $\nabla^+ F(x) = x - \text{prox}_\Psi (x - \nabla f(x))$

Cerintele nu sunt restrictive:

(i) $\psi(x) = 1_X(x)$, unde $X$ poliedru (nu neaparat marginit)

(ii) $\psi(x) = \|x\|_p + 1_X(X)$, unde $X$ politop si $p = 1, \infty$

# Marginirea erorii generalizata global in optimizare: caz 2

Consideram urmatoarea problema primala constransa liniar

$$\min_{u \in \mathbb{R}^m} \{g(u) : \; Au \leq 0\}$$

unde $g$ tare convexa si gradient Lipschitz $\Rightarrow$ gradientul proiectat pe primala converge liniar!

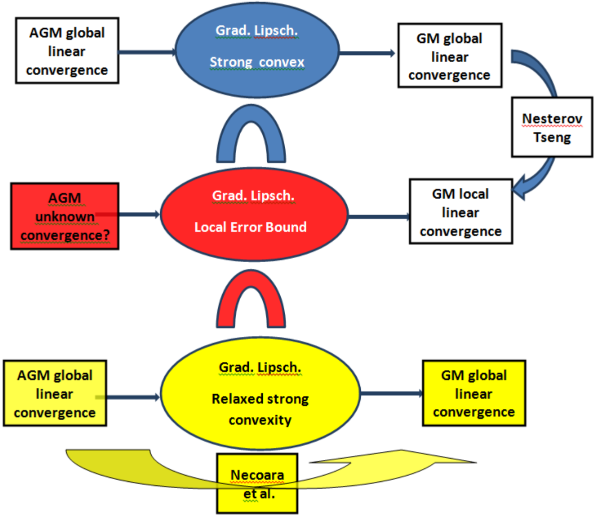
- In multe aplicatii insa, proiectia pe multimea fezabila $U = \{u : Au \leq 0\}$ este prea complexa $\Rightarrow$ abordam problema duala in loc de problema primala:

$$\max_x -\tilde{g}(-A^T x) - \Psi(x),$$

  unde $\tilde{g}$ este conjugata convexa a lui $g$ iar $\Psi(x) = \mathbf{I}_{\mathbb{R}^n_+}(x)$.
- Daca $g(x)$ tare convexa si cu gradient Lipschitz, atunci problema duala va satisface (GEB) global

Theorema: sub (GEB) gradientul clasic converge liniar!

# Recap: raspunsuri la cateva probleme deschise




- Conditiile de relaxare a convexitatii tari contin clase cunoscute de probleme + altele noi cu multe implicatii in aplicatii reale $\Rightarrow$ convergenta liniara pt. GM/AGM

# Aplicatia 1: rezolvarea sistemelor liniare

Matrice simetrica pozitiv semidefinita $Q \succeq 0$ si sistemul:

$$Qx + q = 0 \quad \Leftrightarrow \quad \min_{x \in \mathbb{R}^n} f(x) \quad =: \frac{1}{2} x^T Q x + q^T x$$

Daca consideram descompunerea Cholesky $Q = L_Q^T L_Q$, atunci:

$$f(x) = \frac{1}{2} \|L_Q x\|^2 + (L_Q^T L_Q x_s)^T x = g(L_Q x)$$

unde $g(z) = \frac{1}{2} \|z\|^2 + (L_Q x_s)^T z$ functie tare convexa!

**Teorema**. Functia $f$ satisface TC1 cu $\sigma_1 = \lambda_{\min}(Q)$ (cea mai mica valoare proprie nenula a lui $Q$).

- Teoria dezvoltata permite rezolvarea sistemului cu acuratete $\epsilon$ in $\sqrt{\mathrm{cond}(Q)} \log \frac{1}{\epsilon}$ iteratii de metoda gradient accelerat.
- Gradientul conjugat are aceeasi complexitate dar pentru matrici $Q \succ 0$! Pentru $Q \succeq 0$ exista rezultate???
- Alte rezultate din literatura au complexitate $\mathrm{cond}(Q) \log \frac{1}{\epsilon}$.
- Extensie imediata la cazul general $Ax = b$ cu $A \in \mathbb{R}^{m \times n}$ utilizand $f(x) = \|Ax - b\|^2$.

# Aplicatia 2: Programare liniara

Fie problema de programare liniara:

$$(LP): \quad min_u \; c^T u \quad \text{s.t.}: \quad Eu = b, \; u \geq 0.$$

Considerand formularea primala-duala obtinem sistemul liniar:

$$\begin{cases} Ax = d \\ x \in \mathbf{K} = \mathbb{R}^n_+ \times \mathbb{R}^m \times \mathbb{R}^n_+, \end{cases}$$

unde $x = \begin{bmatrix} u \\ v \\ s \end{bmatrix}, \quad A = \begin{bmatrix} 0 & E^T & I_n \\ E & 0 & 0 \\ c^T & -b^T & 0 \end{bmatrix}, \quad d = \begin{bmatrix} c \\ b \\ 0 \end{bmatrix}.$

LP echivalenta cu urmatoarea problema QP cu structura:

$$\min_{x \in \mathbf{K}} \|Ax - d\|^2.$$

**Teorema**. Functia $f$ satisface TC1 si deci teoria dezvoltata permite rezolvarea problemelor de programare liniara in $\log \frac{1}{\epsilon}$ iteratii de metoda gradient accelerat (operatii matrice-vector).

# Aplicatia 3: optimizare in sisteme retelizate

- In multe aplicatii, e.g. controlul sistemelor retelizate, data ranking sau procesare imagine, avem un sistem format din mai multe entitati interconectate
- Interconexiunea dintre sisteme se denota printr-un graf bipartit:

  $G = ([N] \times [\bar{N}], E)$, unde $[N] = \{1, \ldots, N\}$, $[\bar{N}] = \{1, \ldots, \bar{N}\}$ iar $E \in \{0, 1\}^{N \times \bar{N}}$ este o matrice de incidenta
- Introducem conceptul de sisteme vecine:

$$\mathcal{N}_j = \{i \in [N] : \; E_{ij} = 1\} \quad \forall j \in [\bar{N}]$$

$$\bar{\mathcal{N}}_i = \{j \in [\bar{N}] : \; E_{ij} = 1\} \quad \forall i \in [N].$$

- Multimile $\mathcal{N}_j$ si $\bar{\mathcal{N}}_i$, reprezinta e.g. multimea surselor ce impart link-ul $j \in [\bar{N}]$, respectiv link-urile utilizate de sursele $i \in [N]$

# Aplicatia 3: optimizare in sisteme retelizate

- Sub un sistem retelizat, multe probleme din practica pot fi formulate drept probleme composite partial separabile:

$$\boxed{F^* = \min_{x \in \mathbb{R}^n} F(x) \quad \left(= \sum_{j=1}^{\bar{N}} f_j(x_{\mathcal{N}_j}) + \sum_{i=1}^{N} \Psi_i(x_i)\right)}$$

- In mod obisnuit se considera componenta smooth a functiei $F(x)$ are proprietate de gradient Lipschitz pe coordonate:

$$\|\nabla_i f(x + U_i y_i) - \nabla_i f(x)\| \leq L_i \|y_i\| \quad \forall x \in \mathbb{R}^n,\ y_i \in \mathbb{R}^{n_i}$$

- Noi consideram insa ca functiile individuale $f_j(x)$ au gradient Lipschitz:

$$\|\nabla f_j(x_{\mathcal{N}_j}) - \nabla f_j(y_{\mathcal{N}_j})\| \leq L_{\mathcal{N}_j} \|x_{\mathcal{N}_j} - y_{\mathcal{N}_j j}\| \quad \forall x_{\mathcal{N}_j}, y_{\mathcal{N}_j} \in \mathbb{R}^{n_{\mathcal{N}_j}}.$$

# Aplicatia 3: problema Lasso

- Problema Lasso constransa (utilizata in reconstructie de semnale, etc):

$$\min_{x_i \in X_i \subseteq \mathbb{R}^{n_i}} F(x) \quad \left( = \frac{1}{2}\|Ax - b\|^2 + \sum_{i=1}^{N} \lambda_i \|x_i\|_1 \right),$$

  Aceasta problema poate fi scrisa sub forma composite separabila:
  - $\Psi(x) = \sum_{i=1}^{N} [\lambda_i \|x_i\|_1 + \mathbf{I}_{X_i}(x_i)]$ separabila
  - $f_j(x_{\mathcal{N}_j}) = \frac{1}{2}(a_{\mathcal{N}_j}^T x_{\mathcal{N}_j} - b_j)^2$, unde $a_{\mathcal{N}_j} \to$ reprezinta vectorul format din componentele nenule ale liniei $j$ a lui $A$, corespunzand unei multimi $\mathcal{N}_j$.
- Theorema: Problema Lasso satisface (GEB) daca $X_i$ politop si deci o putem rezolva cu complexitate liniara prin metoda gradient (accelerat).

# Aplicatia 3: duala

- Unele probleme din practica (NUM, DC-OPF) pot fi formulate:

$$f^* = \min_{u: Au \leq b} \sum_{j=1}^{\bar{N}} g_j(u_j)$$

unde $g_j$ sunt tari convexe cu constantele $\sigma_j$. Drept rezultat, conjugatele convexe $\tilde{g}_j$ au gradient Lipschitz cu $L_j = \frac{1}{\sigma_j}$.

- Duala problemei originale are forma:

$$f^* = \max_{x \in \mathbb{R}^n} \sum_{j=1}^{\bar{N}} -\tilde{g}_j(-A_j^T x) - \langle x, b \rangle - \Psi(x),$$

unde $\Psi(x) = \mathbf{I}_{\mathbb{R}^n_+}(x)$.

- Daca luam $f_j(x_{\mathcal{N}_j}) = -\tilde{g}_j\left(-A_{\mathcal{N}_j}^T x_{\mathcal{N}_j}\right) - \langle x_{\mathcal{N}_j}, \bar{b}_{\mathcal{N}_j} \rangle$ atunci problema dual satisface (GEB) si poate fi rezolvata cu complexitate liniara.

# Concluzii finale

- Relaxam conditia de convexitate tare; relaxare ce conduce la probleme de optimizare des intalnite in practica (sisteme liniare, lasso, duala, optimizare in retele): $g(Ax)$
- Derivam clase mai generale de probleme de optimizare pentru care metode de tip gradient converg liniar
- Aratam ca si gradientul accelerat converge liniar pe aceste clase de probleme

Multumesc colaboratorilor:

Yu. Nesterov F. Glineur D. Clipici

Mai multe detalii pe pagina mea web:

http://acse.pub.ro/person/ion-necoara

Intrebari?