PENNON

A Generalized Augmented Lagrangian Method for Nonconvex NLP and SDP

Michal Kočvara

Institute of Information Theory and Automation

Academy of Sciences of the Czech Republic

and

Czech Technical University

kocvara@utia.cas.cz

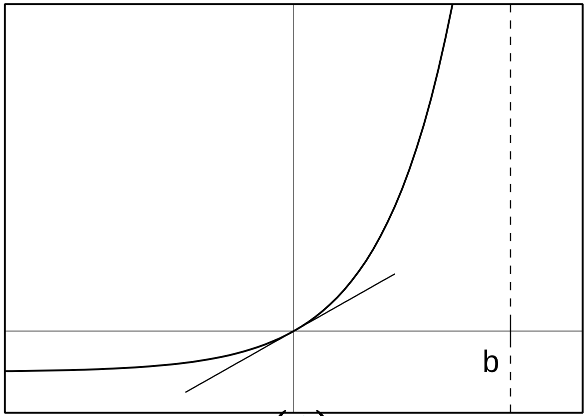
http://www.utia.cas.cz/kocvara

PBM Method for convex NLP

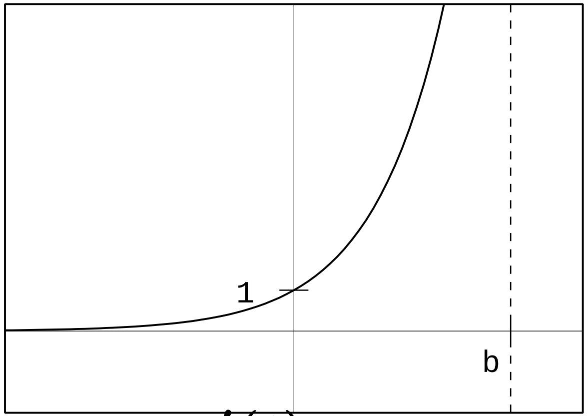
R. Polyak '87
Ben-Tal, Zibulevsky '92, '97
Breitfeld, Shanno '94

Combination of:
(exterior) **P**enalty meth., (interior) **B**arrier meth., Method of **M**ultipliers

$$(CP) \qquad \min_{x\in\mathbb{R}^n} \{f(x) : g_i(x) \leq 0\,, \quad i = 1,\ldots,m\}$$



$\varphi(t)$



$\varphi'(t)$

PBM in semidefinite programming

Problem: $$\min_{x\in\mathbb{R}^n} \{f(x) : \mathcal{A}(x) \preccurlyeq 0\}$$

where

1. $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is C^2
2. $\mathcal{A} : \mathbb{R}^n \longrightarrow \mathbb{S}_d$ is generally nonconvex matrix opretaror

PBM in semidefinite programming

Problem: $$\min_{x\in\mathbb{R}^n} \{f(x) : \mathcal{A}(x) \preccurlyeq 0\}$$

where

1. $f : \mathbb{R}^n \to \mathbb{R}$ is C^2
2. $\mathcal{A} : \mathbb{R}^n \to \mathbb{S}_d$ is generally nonconvex matrix opretaror

Question: How can the matrix constraint

$$\mathcal{A}(x) \preccurlyeq 0$$

be treated by Penalty-Barrier approach ?

Idea: Find an *augmented Lagrangian* as follows:

$$F(x, U, p) = f(x) + \langle U, \Phi_p(\mathcal{A}(x))\rangle_{\mathbb{S}_d}$$

Construction of the penalty function Φ_p

Given:
scalar valued penalty function φ
matrix $A = S^\top \Lambda S$, where Λ = $\mathbf{diag}\ (\lambda_1, \lambda_2, \ldots, \lambda_d)^\top$

Define

$$A \overset{\Phi_p}{\longmapsto} S^T \begin{pmatrix} p\varphi\left(\frac{\lambda_1}{p}\right) & 0 & \ldots & 0 \\ 0 & p\varphi\left(\frac{\lambda_2}{p}\right) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \ldots & 0 & p\varphi\left(\frac{\lambda_d}{p}\right) \end{pmatrix} S$$

$\longrightarrow$ any positive eigenvalue of A is “penalized” by φ

PBM algorithm for semidefinite problems

We have

$$\mathcal{A}(x) \preccurlyeq 0 \Longleftrightarrow \Phi_p(\mathcal{A}(x)) \preccurlyeq 0$$

and the corresponding *augmented Lagrangian*:

$$F(x, U, p) := f(x) + \langle U, \Phi_p(\mathcal{A}(x)) \rangle_{\mathbb{S}_d}$$

PBM algorithm:

(i)	Find x^{k+1} satisfying $\|\nabla_x F(x, U^k, p^k)\| \leq \varepsilon^k$
(ii)	$U^{k+1} = D_{\mathcal{A}} \Phi_p(\mathcal{A}(x); U^k)$
(iii)	$p^{k+1} < p^k$

PENNON for nonconvex problems

Idea: replace Newton by Levenberg-Marquardt

Given x, compute the gradient g and Hessian H at x.
Compute the minimal eigenvalue $\lambda_{\min}$ of H.
If $\lambda_{\min} < 10^{-3}$, set

$$\widehat{H}(\alpha) = H + (-\lambda_{\min} + \alpha)I$$

Compute the search direction

$$d(\alpha) = -\widehat{H}(\alpha)^{-1} g$$

Line-search in direction $d(\alpha)$. Step-length s.
Set $\quad x_{\text{new}} = x + s d(\alpha)$

PBM for nonconvex problems

$$\widehat{H}(\alpha) = H + (-\lambda_{\min} + \alpha)I$$

$$d(\alpha) = -\widehat{H}(\alpha)^{-1}g$$

$$x_{\text{new}} = x + sd(\alpha)$$

Simple version:

$\alpha \in [-\lambda_{\min}, -2\lambda_{\min}]$; no convergence proof, works very well in praxis

Sophisticated version:

Full Trust-Region method; convergence proof, often slower in praxis

Computing β

Choose initial $\widehat{\beta} > 0$.
Perform Cholesky factorization of $H + \widehat{\beta} I$.
If it fails, go to Step (i); otherwise go to Step (iii).

(i) Set $\widehat{\beta} \leftarrow 2\widehat{\beta}$.

(ii) Perform Cholesky factorization of $H + \widehat{\beta} I$.
If it fails, go to Step (i); otherwise stop and return $\beta = \widehat{\beta}$.

(iii) Set $\widehat{\beta} \leftarrow \widehat{\beta}/2$.

(iv) Perform Cholesky factorization of $H + \widehat{\beta} I$.
If it fails stop and return $\beta = 2\widehat{\beta}$; otherwise go to Step (iii).

On output, $\beta \in [-\lambda_{\min}, -2\lambda_{\min}]$

PENNON for NSDP: theory

Based on Breitfeld-Shanno, 1993; generalized by M. Stingl, 2003–2004

Assume:

1. $f, \mathcal{A} \in C^2$
2. $x \in \Omega$ nonempty, bounded
3. Constraint Qualification

Then $\exists$ an index set $\mathcal{K}$ so that:

- $x_k \longrightarrow \hat{x}, \ k \in \mathcal{K}$
- $U_k \longrightarrow \widehat{U}, \ k \in \mathcal{K}$
- $(\hat{x}, \widehat{U})$ satisfies first-order optimality conditions

The reciprocal barrier function in SDP

Find a penalty function φ which allows "direct" computation of Φ, its gradient and Hessian:

$$\Phi(A) = (A - I)^{-1} - I \qquad (\varphi := \frac{1}{t-1} - 1)$$

Then

$$\frac{\partial}{\partial x_i}\Phi(\mathcal{A}(x)) = -(A - I)^{-1}\frac{\partial}{\partial x_i}A(A - I)^{-1}$$

The reciprocal barrier function in SDP

Find a penalty function φ which allows "direct" computation of Φ, its gradient and Hessian:

$$\Phi(A) = (A - I)^{-1} - I \qquad (\varphi := \frac{1}{t-1} - 1)$$

Then

$$\frac{\partial^2}{\partial x_i \partial x_j}\Phi(\mathcal{A}(x)) =$$

$$(\mathcal{A}(x) - I)^{-1}\frac{\partial \mathcal{A}(x)}{\partial x_i}(\mathcal{A}(x) - I)^{-1}\frac{\partial \mathcal{A}(x)}{\partial x_j}(\mathcal{A}(x) - I)^{-1}$$

$$+ (\mathcal{A}(x) - I)^{-1}\frac{\partial^2 \mathcal{A}(x)}{\partial x_i \partial x_j}(\mathcal{A}(x) - I)^{-1}$$

$$+ (\mathcal{A}(x) - I)^{-1}\frac{\partial \mathcal{A}(x)}{\partial x_j}(\mathcal{A}(x) - I)^{-1}\frac{\partial \mathcal{A}(x)}{\partial x_i}(\mathcal{A}(x) - I)^{-1}$$

Construction of the penalty function

Complexity of Hessian assembling - linear SDP:

- $O(d^3n + d^2n^2)$ for dense matrices
- $O(n^2K^2)$ for sparse matrices
  ($K \ldots$ max. number of nonzeros in $A_i,\ i = 1, \ldots, n$)
- Compare to $O(d^4 + d^3n + d^2n^2)$ in the general case

$$\min_{x \in \mathbb{R}^n} \{b^T x : \mathcal{A}(x) \preccurlyeq 0\} \qquad \mathcal{A} : \mathbb{R}^n \longrightarrow \mathbb{S}_d$$

PENNON

BMI problems

The same technique as for nonconvex NLP can be used for nonconvex SDP problems, in particular for optimization problems with bilinear matrix inequalities:

$$\min_{x \in \mathbb{R}^n} \quad \langle c, x \rangle$$

s.t.

$$x_i \in [a, b] \qquad i = 1, \ldots, n$$

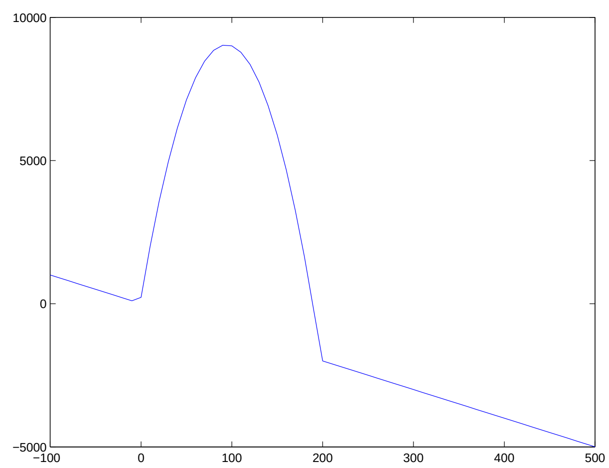
$$A_0 + \sum_{i=1}^{n} x_i A_i + \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j K_{ij} \preccurlyeq 0$$

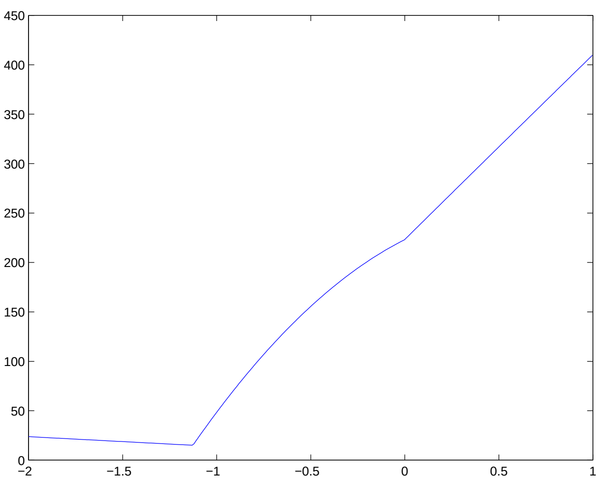
Feasibility BMI problem:

$$\min_{x\in\mathbb{R}^n,\lambda} \quad \lambda$$

$$\text{s.t.} \quad x_i \in [a,b] \qquad i = 1,\ldots,n$$

$$A_0 + \sum_{i=1}^{n} x_i A_i + \sum_{i=1}^{n}\sum_{j=i+1}^{n} x_i x_j K_{ij} - \lambda I \preccurlyeq 0$$





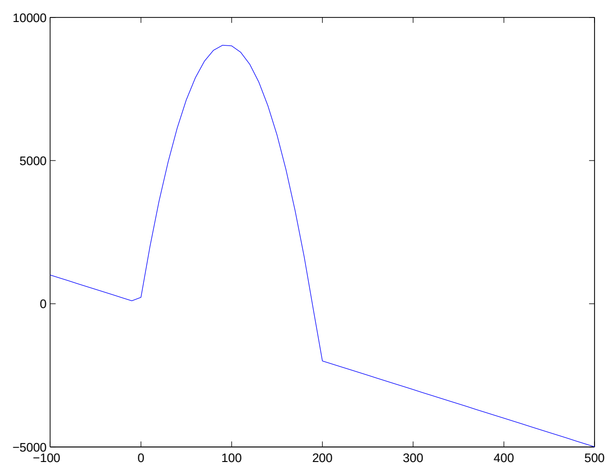
BMI problems

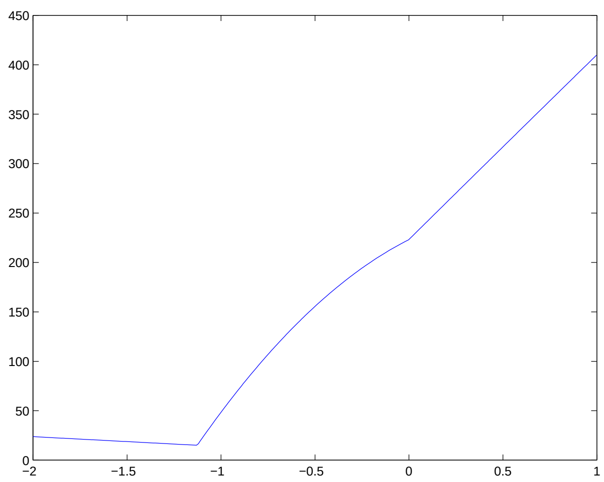
Feasibility BMI problem:

$$\min_{x\in\mathbb{R}^n,\lambda} \quad \lambda + \rho\|x\|^2$$

$$\text{s.t.} \quad x_i \in [a,b] \qquad i = 1,\ldots,n$$

$$A_0 + \sum_{i=1}^{n} x_i A_i + \sum_{i=1}^{n}\sum_{j=i+1}^{n} x_i x_j K_{ij} - \lambda I \preccurlyeq 0$$





GEVP optimization problems

Special kind of BMI: generalized eigenvalue problem.
Find, w.r.t. variable x, the maximal eigenvalue of

$$\mathcal{A}(x)\omega = \lambda\mathcal{B}(x)\omega\,.$$

GEVP optimization problem:

$$\min_{x\in\mathbb{R}^n,\lambda} \quad \lambda$$
$$\text{s.t.} \quad \mathcal{A}(x) - \lambda\mathcal{B}(x) \preccurlyeq 0$$

This is a quasiconvex problem: there exists a unique global minimum

Quasiconvex functions

Definition: A function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is called quasiconvex if its domain and all level sets

$$\{x \in \mathrm{dom} f \mid f(x) \leq \alpha\}$$

are convex.